Stata commands to use

Before following the tips in this document, please import your excel document into Stata. Below examples are based off Research Office Biostatistician data.

Tips

Syntax is the grammar of Stata commands. Stata is a command-based language. Most Stata commands are verbs. They tell Stata to do something: summarize, tabulate, regress, etc.

Basic Stata command syntax:

The name of the command followed by a list of variables and instructions. For the specific syntax for a particular command click help. For example, to do a regression, regress is the name of the command. The first variable listed (`logTimeClose`) is the outcome variable the others (`CHRONICINFECTION Sex BCereus BMI`) are predictor variables. The i. and ## are specific instructions on how to work with the variables. After a comma, Stata allows you to use "options" which modify how the command is used. Vce (robust) tells Stata to calculate the robust standard error. Help will list all the available options for a command.

```
. regress logTimeClose i.CHRONICINFECTION##i.Sex i.BCereus BMI, vce(robust)
```

Some commands can be combined using brackets. For example, a scatter graph and a line fit graph are combined on a single two-way graph.

```
. twoway (scatter logTimeClose Age, sort) (lfit logTimeClose Age, sort)
```

You can access the Help Tab from the command line. For example, to identify the collinearity diagnostics command collin…

```
. help collin
```

If you do not know the specific Stata command, use:

```
. search collinearity
```

In Stata, you can keep a log file of all commands and results (Graphical output is exempt). You can also copy and paste these to a work file. If you want to provide annotation, put a * before your comment as shown below:

```
. * Here's how to use help from the command line
. help collin
```

If the comment is long, use /*…*/. Example below:

```
/* Here I will recode BMI into categories using the
ABS categories. Personally, I don't think we should use categories */
```

Using /* … */ comments can be appended directly to commands. The recode command is used to create categories. 9/18 means values 9 to 18, 1 is the numerical value of the category, and Underweight is the label for the category. If the label has a space in it you have to put it in quotations. Once the categories are defined the option is to generate a new variable called `BMI_Cat`.

```
recode BMI (9/18 = 1 Underweight) (19/24 = 2 Healthy) (25/29 = 3 Overweight) (30/39 = 4
Obese) (40/max = 5 "Extremely Obese"), gen(BMI_Cat) /* ABS categories */
```

A command can be conditional by using if along with other logical operators. For example, if you wanted to do a regression of BMI as the outcome and education as the predictor but only for single parent households with more than one child.

```
regress BMI education if parents==1 & nchildren > 1
```

Stata's Logical and Relational Operators

`& and`

`| or`

```
! not
~ not
> greater than
< less than
>= greater or equal
<= less than or equal
== equal
!= not equal
```

If you want to repeat a command for each subgroup of a variable use `by`. It usually comes before the command separated by a colon: For example, to create a table of statistics for `TIMETOCLOSURE` for each category of `BCereus` (presence or absence of infection with this bacteria).

```
. by BCereus, sort: tabstat TIMETOCLOSURE,  stat(mean sd p25 p50 p75)
```

Note: `by` usually only works if `BCereus` is sorted, hence using the option `sort`. You can do both together with `bysort`.

```
. bysort BCereus: tabstat TIMETOCLOSURE,  stat(mean sd p25 p50 p75)
```

`by` can also be used as an option with some commands using the syntax `by(varname)`,For example:

```
. tabstat TIMETOCLOSURE, by(BCereus) stat(mean sd p25 p50 p75)
```

This gives a neater output and does not require the subgrouping variable to be sorted.

In addition to the resources listed in the “Introduction to Using Stata 15 at Gold Coast Health”, http://wlm.userweb.mwn.de/Stata/wstatbas.htm is an informative website to follow.

Summary Statistics and Graphics

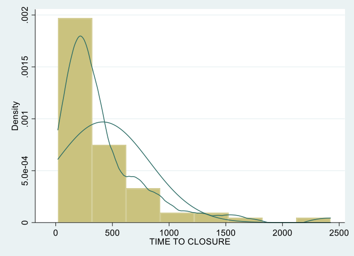
To start understanding your data, run some basic summary statistics and graphs to visualize the data.

```
. summarize TIMETOCLOSURE, detail

TIME TO CLOSURE
-------------------------------------------------------------
Percentiles      Smallest
1%          19             19
5%        60.5             34
10%         72             52       Obs                  71
25%        180           60.5       Sum of Wgt.          71

50%        288                      Mean           414.9155
Largest       Std. Dev.      411.4733
75%        560           1224
90%      843.7           1446       Variance       169310.2
95%       1224           1629       Skewness       2.437968
99%       2424           2424       Kurtosis        10.5143
```

```
. histogram TIMETOCLOSURE
(bin=8, start=19, width=300.625)
```

Add in the options normal and kdensity superimpose a normal curve and an empirical estimate of the frequency distribution.

```
. histogram TIMETOCLOSURE, normal kdensity
(bin=8, start=19, width=300.625)
```



The distribution above does not look normal (no left tail and skewed to the right). To do a formal test of normality

```
. swilk TIMETOCLOSURE

Shapiro-Wilk W test for normal data

    Variable |        Obs       W           V         z       Prob>z
-------------+------------------------------------------------------
TIMETOCLOS~E |         71     0.74797     15.692     5.992    0.00000
```

As expected, the probability that the underlying data from which this sample was taken is normal and is tiny. The data is significantly different from being normally distributed.

<u>Generating variables and cleaning and organising your data</u>
[Data tab, Create or Change Data].

```
. gen AgeDec = Age/10
```

Once a variable is generated it can be modified with the replace command.

```
. replace AgeDec = 8 if AgeDec>8
```

The BMI variable, in this example was in red because it was text due to an “NA” appearing in the data. First, we remove the NA by replacing it with nothing.

```
. replace BMI = "" in 27
(1 real change made)
```

There are two ways to convert tie text variable to a numerical variable. First, generate a new variable that is numerical using the real command.

```
. gen BMIr =real(BMI)
(1 missing value generated)
```

Second, convert the original text variable to a numerical variable using destring.

```
. destring BMI, replace
BMI: all characters numeric; replaced as double
(1 missing value generated)
```

The other text variable is Sex. There are two ways to convert this to numeric.

First, use encode. [Data tab, Create or Change Data, Other Variable Transformation Commands].

```
. encode Sex, gen(Sex12)
```

New variables are added to the bottom of the variable list but can be moved to a more convenient location with the order command.

```
. order Sex12, after( Sex)
```

Second, replace the values and convert to numeric.

```
. replace Sex="0" if Sex=="F"
(43 real changes made)

. replace Sex="1" if Sex=="M"
(28 real changes made)

. destring Sex, replace
Sex: all characters numeric; replaced as byte
```

Values in numeric variables can be labelled.

```
. label define Gender 0 "F" 1 "M"
. label values (Sex) Gender
```

NOTE: Using encode automatically labelled the new variable Sex12 with F and M. However, it encoded the original Fs and Ms with 1s and 2s (not, as I had hoped, with 0s and **1s).**

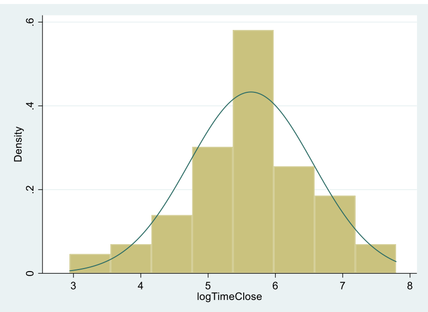
Basic data analysis

Association between two categorical variables.

```
. tabulate CHRONICINFECTION Bonehealing, chi2 exact

Enumerating sample-space combinations:
stage 5:  enumerations = 1
stage 4:  enumerations = 2
stage 3:  enumerations = 5
stage 2:  enumerations = 16
stage 1:  enumerations = 0

  CHRONIC |                    Bone healing
INFECTION |         1          2          3          4          5 |     Total
----------+-------------------------------------------------------+----------
        0 |        38          2          5          3          1 |        49
        1 |        14          3          2          3          0 |        22
----------+-------------------------------------------------------+----------
    Total |        52          5          7          6          1 |        71

          Pearson chi2(4) =   3.8521   Pr = 0.426
           Fisher's exact =                 0.402

. tabulate Bonehealing CHRONICINFECTION, chi2 exact
```

Log transformation:

```
. gen logTimeClose =ln( TIMETOCLOSURE)

. histogram logTimeClose, normal
(bin=8, start=2.9444389, width=.60609192)
```




```
. swilk logTimeClose

                   Shapiro-Wilk W test for normal data
```

```
    Variable |        Obs       W           V         z       Prob>z
-------------+------------------------------------------------------
logTimeClose |         71    0.98790      0.753    -0.616    0.73112
```

Compare time to closure by Chronic infection. Non-parametric ranksum (Man-Whitney U).

```
. ranksum TIMETOCLOSURE, by( CHRONICINFECTION)
```

```
Two-sample Wilcoxon rank-sum (Mann-Whitney) test

CHRONICINF~N |      obs    rank sum    expected
-------------+---------------------------------
           0 |       49      1708.5        1764
           1 |       22       847.5         792
-------------+---------------------------------
    combined |       71        2556        2556

unadjusted variance     6468.00
adjustment for ties       -3.69
                     ----------
adjusted variance       6464.31

Ho: TIMETO~E(CHRONI~N==0) = TIMETO~E(CHRONI~N==1)
             z =  -0.690
    Prob > |z| =   0.4900
```

<u>t-test of log transformed data:</u>

```
. ttest logTimeClose, by( CHRONICINFECTION)
```

```
Two-sample t test with equal variances
------------------------------------------------------------------------------
   Group |     Obs        Mean    Std. Err.   Std. Dev.   [95% Conf. Interval]
---------+--------------------------------------------------------------------
       0 |      49    5.563024    .1269281    .8884966    5.307818     5.81823
       1 |      22    5.802506    .2113022    .9910952    5.363079    6.241933
---------+--------------------------------------------------------------------
combined |      71     5.63723    .1093155    .9211087    5.419207    5.855253
---------+--------------------------------------------------------------------
    diff |           -.2394812    .2363458               -.7109781    .2320157
------------------------------------------------------------------------------
    diff = mean(0) - mean(1)                                      t =  -1.0133
Ho: diff = 0                                     degrees of freedom =       69

    Ha: diff < 0                 Ha: diff != 0                 Ha: diff > 0
 Pr(T < t) = 0.1572         Pr(|T| > |t|) = 0.3145          Pr(T > t) = 0.8428
```

The tabstat command can be used to compare summary statistics across categories of a categorical variable. For example below, time to closure by infection with B. Cereus.

```
. tabstat TIMETOCLOSURE, by( BCereus) stat(mean sd p25 p50 p75)
```

```
Summary for variables: TIMETOCLOSURE
     by categories of: BCereus (B.Cereus)

 BCereus |      mean        sd       p25       p50       p75
---------+--------------------------------------------------
       0 |  280.7451  196.3073       144       221       384
       1 |    757.05  592.3052       264       720      1068
---------+--------------------------------------------------
   Total |  414.9155  411.4733       180       288       560
------------------------------------------------------------
```

<u>To estimate statistics like means or proportions, use:</u>

[Statistics / Summaries, tables, and tests / Summary and descriptive statistics / Proportions]

```
. proportion Sex12
```

```
Proportion estimation                Number of obs    =         71

--------------------------------------------------------------
             |                                   Logit
             | Proportion   Std. Err.     [95% Conf. Interval]
```

```
-------------+------------------------------------------------
Sex12        |
           F |   .6056338   .0579997       .4861713     .713679
           M |   .3943662   .0579997        .286321    .5138287
--------------------------------------------------------------
```

Then test against a specified value, example below.

```
Statistics / Summaries, tables, and tests / Classical tests of hypotheses / Proportion
test
```

```
. prtest Sex12 == 0.5
Sex12 is not a 0/1 variable
r(450);
```

NOTE: There is an Error message because this test was expecting the variable to consist of 0s or 1s and as noted when Sex12 was encoded. Redoing it with the Sex variable will work.

```
. prtest Sex == 0.5
```

```
One-sample test of proportion                  Number of obs      =        71

------------------------------------------------------------------------------
    Variable |       Mean   Std. Err.                     [95% Conf. Interval]
-------------+----------------------------------------------------------------
         Sex |   .3943662   .0579997                      .2806889    .5080435
------------------------------------------------------------------------------
    p = proportion(Sex)                                           z =  -1.7802
Ho: p = 0.5

     Ha: p < 0.5                 Ha: p != 0.5                 Ha: p > 0.5
 Pr(Z < z) = 0.0375         Pr(|Z| > |z|) = 0.0750          Pr(Z > z) = 0.9625
```

<u>To graph your data, use:</u>

```
Graphics / Twoway graph (scatter, line, etc)
```

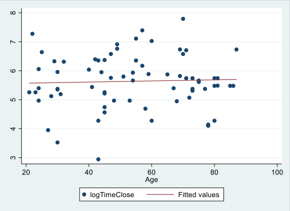
Two-way is the most used graphics command. Use the help tab to find out more about two-way.

It is important to include the sort option

```
. twoway (scatter logTimeClose Age, sort)
```

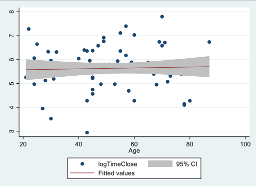
You can combine graphs on the same graph, For example, adding a linear best fit line (`lfit`).

```
. twoway (scatter logTimeClose Age, sort) (lfit logTimeClose Age, sort)
```



With confidence intervals

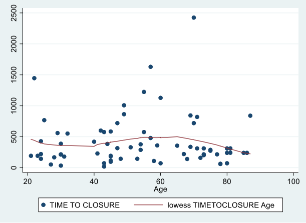
```
. twoway (scatter logTimeClose Age, sort) (lfitci logTimeClose Age, sort)
```




```
. twoway (scatter TIMETOCLOSURE Age, sort) (lfitci TIMETOCLOSURE Age, sort)
```

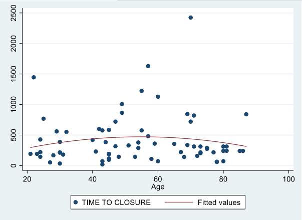
<u>Lowess smothing</u>

```
. twoway (scatter TIMETOCLOSURE Age, sort) (lowess TIMETOCLOSURE Age, sort)
```



<u>Quadratic</u>

```
. twoway (scatter TIMETOCLOSURE Age, sort) (qfit TIMETOCLOSURE Age, sort)
```



Linear regression:

```
Statistics / Linear models and related / linear regression
```

In terms of the syntax, write the command followed by the dependent (outcome) variable and then the list of independent (predictor) variables. Following a comma, add in options.

NOTE: the results below for `CHRONICINFECTION` are the same as for the t-test, P=0.314, so chronic infection does not significantly affect log time to closure.
NOTE: The i. before `CHRONICINFECTION`. This tells Stata to treat the variable as a set of indicator variables (a 0 or 1 for each category of a categorical variable, except the base category to which the other categories are compared). If you don't use the i. the variable is treated like a scale variable assuming equal distances between each category. If there are only two categories in the categorical variable there would only be one indicator variable so the analyses would be equivalent. It is best to routinely use the i. for any categorical variable.

```
. regress logTimeClose i.CHRONICINFECTION
```

```
      Source |       SS           df       MS      Number of obs   =        71
-------------+----------------------------------   F(1, 69)        =      1.03
       Model |  .870769496         1  .870769496   Prob > F        =    0.3145
    Residual |  58.5201186        69  .848117661   R-squared       =    0.0147
-------------+----------------------------------   Adj R-squared   =    0.0004
       Total |  59.3908881        70  .848441259   Root MSE        =    .92093
```

```
------------------------------------------------------------------------------------
      logTimeClose |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------------+----------------------------------------------------------------
1.CHRONICINFECTION |   .2394812   .2363458     1.01   0.314    -.2320157    .7109781
             _cons |   5.563024   .1315619    42.28   0.000     5.300566    5.825483
------------------------------------------------------------------------------------
```

Predictor variables are added below.
NOTE: Bone healing is presented, with coefficients for each category, these being relative to the base category which is category 1.

```
. regress logTimeClose i.CHRONICINFECTION Age BMI i.Sex i.Bonehealing i.BCereus
```

```
      Source |       SS           df       MS      Number of obs   =        70
-------------+----------------------------------   F(9, 60)        =      3.04
       Model |  18.5791296         9  2.06434773   Prob > F        =    0.0046
    Residual |  40.7421598        60  .679035996   R-squared       =    0.3132
-------------+----------------------------------   Adj R-squared   =    0.2102
```

```
       Total |  59.3212893        69  .859728831   Root MSE        =    .82404

------------------------------------------------------------------------------------
      logTimeClose |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------------+----------------------------------------------------------------
1.CHRONICINFECTION |   .0823017   .2224124     0.37   0.713    -.3625892    .5271927
               Age |  -.0006995   .0055025    -0.13   0.899    -.0117062    .0103071
               BMI |   .0227162   .0172758     1.31   0.194    -.0118405     .057273
                   |
               Sex |
                M  |  -.1788183   .2194999    -0.81   0.418    -.6178835    .2602468
                   |
       Bonehealing |
                2  |   .5973152   .4096912     1.46   0.150    -.2221892     1.41682
                3  |   .2816708   .3360304     0.84   0.405    -.3904901    .9538317
                4  |   .2717252    .367179     0.74   0.462    -.4627422    1.006193
                5  |    -.16918   .8444604    -0.20   0.842    -1.858352    1.519992
                   |
         1.BCereus |   .9843949   .2231356     4.41   0.000     .5380572    1.430732
             _cons |   4.685803   .5934943     7.90   0.000     3.498637    5.872968
------------------------------------------------------------------------------------
```

To compare to another category, in this example to category 4. Put b(4). rather than i. before Bonehealling. In this example, B.Cereus looks to be the only important influence on log time to closure.

```
.
. regress logTimeClose i.CHRONICINFECTION Age BMI i.Sex b(4).Bonehealing i.BCereus
```

```
      Source |       SS           df       MS      Number of obs   =        70
-------------+----------------------------------   F(9, 60)        =      3.04
       Model |  18.5791296         9  2.06434773   Prob > F        =    0.0046
    Residual |  40.7421598        60  .679035996   R-squared       =    0.3132
-------------+----------------------------------   Adj R-squared   =    0.2102
       Total |  59.3212893        69  .859728831   Root MSE        =    .82404

------------------------------------------------------------------------------------
      logTimeClose |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------------+----------------------------------------------------------------
1.CHRONICINFECTION |   .0823017   .2224124     0.37   0.713    -.3625892    .5271927
               Age |  -.0006995   .0055025    -0.13   0.899    -.0117062    .0103071
               BMI |   .0227162   .0172758     1.31   0.194    -.0118405     .057273
                   |
               Sex |
                M  |  -.1788183   .2194999    -0.81   0.418    -.6178835    .2602468
                   |
       Bonehealing |
                1  |  -.2717252    .367179    -0.74   0.462    -1.006193    .4627422
                2  |     .32559   .5247338     0.62   0.537    -.7240339    1.375214
                3  |   .0099457   .4678493     0.02   0.983    -.9258923    .9457836
                5  |  -.4409052   .9053846    -0.49   0.628    -2.251944    1.370134
                   |
         1.BCereus |   .9843949   .2231356     4.41   0.000     .5380572    1.430732
             _cons |   4.957528    .729117     6.80   0.000     3.499077    6.415979
------------------------------------------------------------------------------------
```

To include an interaction effect use ## to automatically include both main effects and the interaction effect. Using a single # means you must add in the main effects separately. For example:

```
., regress logTimeClose i.CHRONICINFECTION i.Sex i.CHRONICINFECTION#i.Sex i.BCereus BMI
```

```
. regress logTimeClose i.CHRONICINFECTION##i.Sex i.BCereus BMI
```

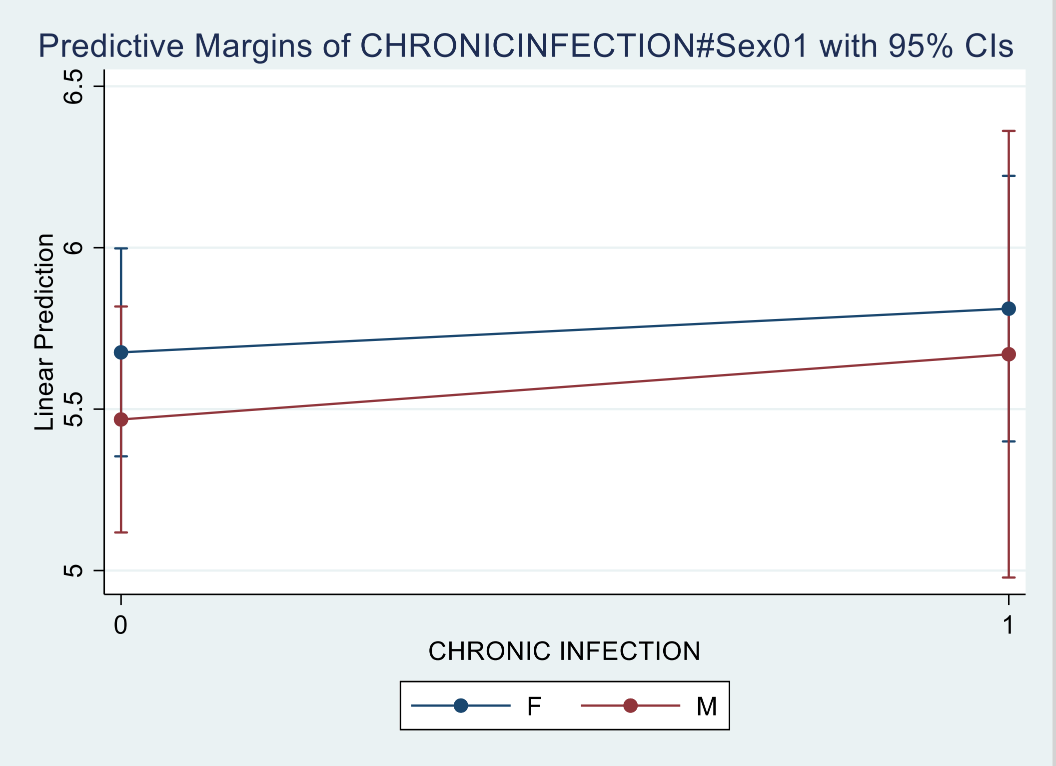
```
      Source |       SS           df       MS      Number of obs   =        70
-------------+----------------------------------   F(5, 64)        =      4.95
       Model |  16.5328238         5  3.30656476   Prob > F        =    0.0007
    Residual |  42.7884655        64  .668569774   R-squared       =    0.2787
-------------+----------------------------------   Adj R-squared   =    0.2223
       Total |  59.3212893        69  .859728831   Root MSE        =    .81766

----------------------------------------------------------------------------------------
          logTimeClose |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-----------------------+----------------------------------------------------------------
    1.CHRONICINFECTION |   .1353685   .2605297     0.52   0.605    -.3850994    .6558365
                       |
                  Sex  |
                    M  |  -.2078547    .239169    -0.87   0.388    -.6856497    .2699404
                       |
CHRONICINFECTION#Sex   |
                  1#M  |   .0666915   .4699383     0.14   0.888    -.8721182    1.005501
                       |
             1.BCereus |   .9863938    .224936     4.39   0.000     .5370325    1.435755
                   BMI |   .0233051   .0167869     1.39   0.170    -.0102307    .0568408
                 _cons |   4.711488     .53271     8.84   0.000     3.647277    5.775698
----------------------------------------------------------------------------------------
```

Margins is a useful post-regression command. It provides the predicted outcome for the model at specified levels of covariates in the model. It is especially useful for understanding the nature of an interaction effect as it gives the predicted outcome at each combination of the interacting variables. For example:

```
. margins CHRONICINFECTION#Sex

Predictive margins                              Number of obs     =         70
Model VCE    : OLS

Expression   : Linear prediction, predict()

-----------------------------------------------------------------------------------------
                     |            Delta-method
                     |      Margin   Std. Err.       t     P>|t|      [95% Conf. Interval]
---------------------+-------------------------------------------------------------------
CHRONICINFECTION#Sex |
                 0#F |    5.675853   .1612051     35.21    0.000      5.353809    5.997897
                 0#M |    5.467998     .17518     31.21    0.000      5.118036    5.817961
                 1#F |    5.811222    .205839     28.23    0.000      5.400011    6.222432
                 1#M |    5.670058   .3462398     16.38    0.000      4.978365    6.361752
-----------------------------------------------------------------------------------------
```

Follow this up with marginsplot to graph the effect. Non-parallel lines suggest a significant interaction.

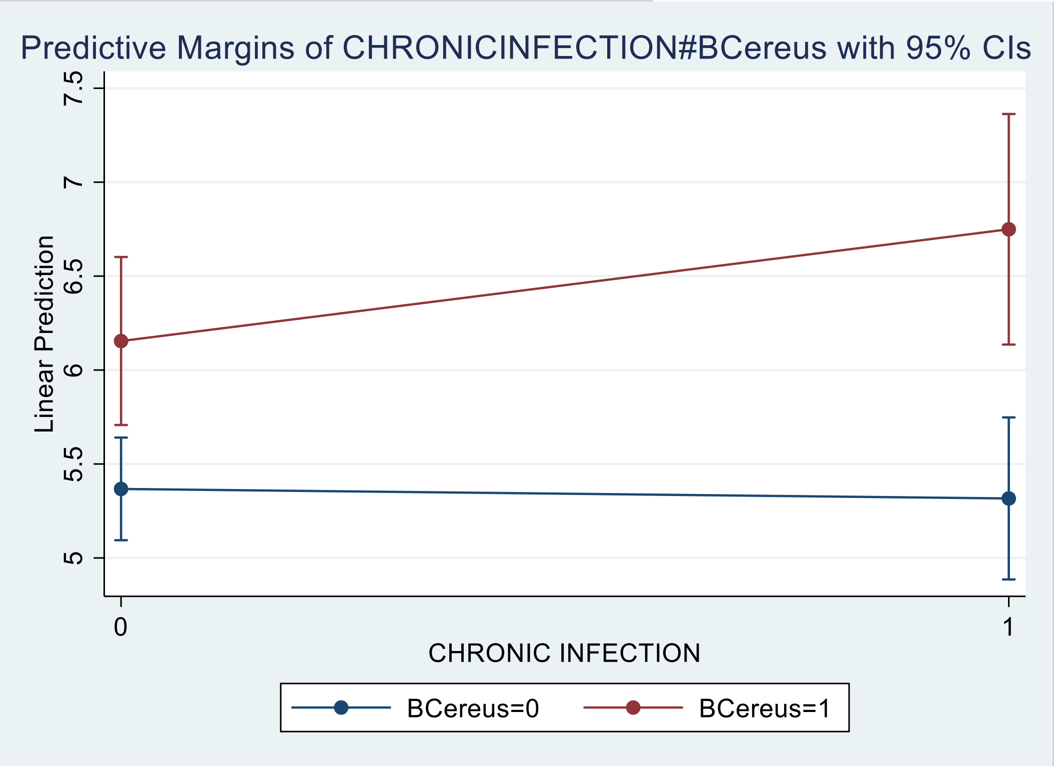
```
. marginsplot
```



In this example, with B.cereus:

```
. regress logTimeClose i.CHRONICINFECTION##i.BCereus BMI i.Sex

      Source |       SS           df       MS      Number of obs   =        70
-------------+----------------------------------   F(5, 64)        =      5.48
       Model |  17.7822956         5  3.55645912   Prob > F        =    0.0003
    Residual |  41.5389937        64  .649046777   R-squared       =    0.2998
-------------+----------------------------------   Adj R-squared   =    0.2451
       Total |  59.3212893        69  .859728831   Root MSE        =    .80563

------------------------------------------------------------------------------------------
            logTimeClose |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------------------+----------------------------------------------------------------
      1.CHRONICINFECTION |  -.0506401    .258335    -0.20   0.845    -.5667235    .4654433
               1.BCereus |    .787167   .2616826     3.01   0.004     .2643959    1.309938
                         |
CHRONICINFECTION#BCereus |
                     1 1 |   .6453348   .4626284     1.39   0.168    -.2788717    1.569541
                         |
                     BMI |   .0212396   .0165988     1.28   0.205    -.0119204    .0543996
```

```
                         |
                     Sex |
                       M |  -.2509926   .2091696    -1.20   0.235    -.6688569    .1668717
                   _cons |   4.845798    .531723     9.11   0.000     3.783559    5.908037
------------------------------------------------------------------------------------------

. margins CHRONICINFECTION#BCereus

Predictive margins                              Number of obs     =         70
Model VCE    : OLS

Expression   : Linear prediction, predict()

------------------------------------------------------------------------------------------
                         |            Delta-method
                         |     Margin   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------------------+----------------------------------------------------------------
CHRONICINFECTION#BCereus |
                     0 0 |   5.367449   .1367317    39.26   0.000     5.094296    5.640602
                     0 1 |   6.154616   .2238037    27.50   0.000     5.707517    6.601715
                     1 0 |   5.316809   .2159119    24.62   0.000     4.885475    5.748143
                     1 1 |   6.749311   .3071336    21.98   0.000     6.135741    7.362881
------------------------------------------------------------------------------------------

. marginsplot

  Variables that uniquely identify margins: CHRONICINFECTION BCereus
```



<u>Logistic regression can be performed with the logistic command, as below:</u>

```
. logistic GoodResponse i.Sex BMI i.CHRONICINFECTION i.BCereus

Logistic regression                             Number of obs     =         70
                                                LR chi2(4)        =      44.67
                                                Prob > chi2       =     0.0000
Log likelihood = -22.667119                     Pseudo R2         =     0.4963

------------------------------------------------------------------------------------
      GoodResponse | Odds Ratio   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------------+----------------------------------------------------------------
               Sex |
                 M |   1.887282   1.708759     0.70   0.483     .3200021    11.13066
               BMI |   .8892833    .058282    -1.79   0.073     .7820849    1.011175
1.CHRONICINFECTION |   1.136505   .9736462     0.15   0.881     .2120044    6.092528
         1.BCereus |    .009316   .0093258    -4.67   0.000     .0013096    .0662717
             _cons |   220.1099   488.9732     2.43   0.015     2.829428    17123.02
------------------------------------------------------------------------------------
Note: _cons estimates baseline odds.
```

Logistic regression is one of many generalized linear models and you can get the same result from

```
Statistics / Generalized linear models / Generalized linear model (GLM)
```

Then choosing the binomial family and the logit link.

```
. glm GoodResponse i.Sex BMI i.CHRONICINFECTION i.BCereus, family(binomial 1) link(logit)
```

```
Iteration 0:   log likelihood = -23.418096
Iteration 1:   log likelihood =  -22.67379
Iteration 2:   log likelihood =  -22.66712
Iteration 3:   log likelihood = -22.667119

Generalized linear models                          No. of obs      =         70
Optimization     : ML                              Residual df     =         65
                                                   Scale parameter =          1
Deviance         =  45.33423773                    (1/df) Deviance =   .6974498
Pearson          =   57.9795399                    (1/df) Pearson  =   .8919929

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

                                                   AIC             =   .7904891
Log likelihood   = -22.66711886                    BIC             =   -230.818

------------------------------------------------------------------------------------
                   |                 OIM
      GoodResponse |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------------+----------------------------------------------------------------
               Sex |
                 M |   .6351374   .9054071     0.70   0.483    -1.139428    2.409703
               BMI |  -.1173394   .0655382    -1.79   0.073    -.2457919    .0111131
1.CHRONICINFECTION |   .1279574   .8567023     0.15   0.881    -1.551148    1.807063
         1.BCereus |  -4.676023   1.001055    -4.67   0.000    -6.638055   -2.713991
             _cons |   5.394126   2.221496     2.43   0.015     1.040074    9.748178
------------------------------------------------------------------------------------
```

NOTE: These results are not exactly the same as when the logistic command was used. The results above report the coefficients (βs) of each predictor variable (the weightings of each predictor in the linear part of the model). The output following the logistic command gave the odds ratio. The odds ratio is the exponentiated form of the coefficient, e^{β}, which can be requested from the glm command by adding the eform option.

```
. glm GoodResponse i.Sex BMI i.CHRONICINFECTION i.BCereus, family(binomial 1) link(logit)
eform
```

```
Iteration 0:   log likelihood = -23.418096
Iteration 1:   log likelihood =  -22.67379
Iteration 2:   log likelihood =  -22.66712
Iteration 3:   log likelihood = -22.667119

Generalized linear models                          No. of obs      =         70
Optimization     : ML                              Residual df     =         65
                                                   Scale parameter =          1
Deviance         =  45.33423773                    (1/df) Deviance =   .6974498
Pearson          =   57.9795399                    (1/df) Pearson  =   .8919929

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

                                                   AIC             =   .7904891
Log likelihood   = -22.66711886                    BIC             =   -230.818

------------------------------------------------------------------------------------
                   |                 OIM
      GoodResponse | Odds Ratio   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------------+----------------------------------------------------------------
               Sex |
                 M |   1.887281   1.708758     0.70   0.483      .320002    11.13065
               BMI |   .8892833    .058282    -1.79   0.073     .7820849    1.011175
1.CHRONICINFECTION |   1.136505   .9736461     0.15   0.881     .2120044    6.092527
         1.BCereus |    .009316   .0093258    -4.67   0.000     .0013096    .0662718
             _cons |   220.1097   488.9727     2.43   0.015     2.829426       17123
------------------------------------------------------------------------------------
Note: _cons estimates baseline odds.
```

```
. set seed 22209

. gen GoodResponse = runiform()

. gen Bernouli = rbinomial(1, 0.5)

. replace GoodResponse = 1 if GoodResponse>0.3
(48 real changes made)

. replace GoodResponse =0 if GoodResponse !=1
(23 real changes made)
```