```
race:  1=white, 2=black, 3=Asian
gender:  1=male, 0=female
```

Mean systolic blood pressures (using **adjmean**) are calculated for six categories: white males, white females, black males, black females, Asian males, and Asian females. Had the outcome been a dichotomous variable for hypertension (1=hypertensive, 0=normal), the probability of being hypertensive (using **adjprop**) would be calculated for each of the six categories.

An overall test of differences is printed, and additionally, a test for interaction. The names for the dummy variables for *xvar1* are **X12**, **X13**, **X14**, etc., where the first number ("1") specifies dummy variables representing *xvar1*, and the second number designates which dummy variable. Dummy variables for *xvar2* are **X22**, **X23**, **X24**, etc. Interaction terms are created by multiplying combinations for the *xvar1* and *xvar2* dummy variables in the model. An F test for a linear model or the likelihood ratio test for a logistic model is printed to test an interaction effect.

adjust(*covlist***)** specifies an optional list of covariates. If this option is not specified, unadjusted estimates are reported. Values for the covariates default to their means. Alternatively, some (or all) covariates can be set to any value for prediction after estimation of the model. For example, the following statement estimates the systolic blood pressure means for each race category adjusted for age, gender, and cholesterol level:

```
. adjmean sbp, by(race) adjust(age gender chl)
```

If we wanted to calculate these means for someone who is 50 years old, male, with a cholesterol level of 250, we would write

```
. adjmean sbp, by(race) adjust(age=50 gender=1 chl=250)
```

model displays the output from **regress** or **logistic**. If this option is not requested, the model is not shown.

level(#) specifies the confidence level in percent for the confidence intervals. The default is 95%.

graph displays points for each adjusted mean or probability (y-axis) by *xvar1* (x-axis). If only *xvar1* is requested, confidence intervals are shown also. If *xvar2* is specified, the points are separated by categories of *xvar2*, with *xvar1* still defining the x-axis. This sounds more confusing than it is—see examples which follow.

bar displays a bar graph rather than point estimates when specified with graph.

graph_options are any of the options allowed with **graph**, **twoway**.

Examples

To illustrate how these programs work, let's take a look at a study of the cost of treating patients with acute low back pain (Carey et al. 1995). All patients ($n = 1406$) had low back pain (LBP) at entry into the study, and were followed for 6 months.

```
Contains data from LBP.DTA
 obs:         1,406                      Low Back Pain: Outcomes of Care
 vars:           10                      20 Jan 1998 13:28
-------------------------------------------------------------------------------
  1. id          long   %9.0g             Patient ID
  2. provtype    byte   %9.0g    praclbl  Provider Type
  3. chiromd     byte   %9.0g    chirlbl  Chiropractor vs. MD
  4. age         byte   %8.0g             Patient Age in Years
  5. gender      byte   %8.0g    genlbl   Patient Gender
  6. cost        int    %10.0g            Total Cost of LBP Treatment
  7. status4w    byte   %9.0g    func4    Low Back Pain at 4 Weeks
  8. severity    byte   %8.0g             Severity of LBP
  9. severe3     byte   %9.0g    sevlbl   Severity of LBP
 10. sciatica    byte   %8.0g    yesno    Sciatica Present
-------------------------------------------------------------------------------
```

Example 1

Does the cost of care for an episode of low back pain differ by the type of provider seen? Provider types are 1 for primary care physician, 2 for chiropractor, 3 for orthopedist, and 4 for HMO physician. To answer this, we would look at the mean cost for each provider type (**provtype**).

```
. adjmean cost, by(provtype)

*Unadjusted* Means and 95% Confidence Intervals
  Outcome:     Total Cost of LBP Treatment -- cost
  Nominal X:   Provider Type -- provtype
  Covariates:  (none)
```

```
 provtype      numobs       mean         se      lower      upper
    1:PC          538   398.0911    26.1085   346.9193   449.2628
  2:Chiro         514   626.3891   26.71108   574.0363   678.7419
  3:Ortho         162   627.5185   47.57903   534.2653   720.7717
    4:HMO         192   340.5104   43.70413   254.8519   426.1689
  Test for difference of 4 means:
    F(3, 1402) =   19.38
    Prob > F   <   0.0001
```

We see that the "unadjusted" mean costs differ by type of provider seen, with patients of chiropractors and orthopedists incurring the largest costs ($626 and $628), and HMO patients the least ($341). The overall F test shows a significant (p <0.0001) difference for the means. There are a number of other ways we could have gotten these same unadjusted means quite simply in Stata, e.g.,

```
. tabulate provtype, summarize(cost)
. oneway cost provtype, tabulate
```

Example 2

An argument might be made that the cost of treating LBP is higher for chiropractors and orthopedist because their patients have more severe pain. We rerun Example 1, but this time we will adjust for variables such as age, severity (a scale from 0 to 23, where 0 means no impairment and 23 means complete incapacitation), and sciatica (another measure of severity, where 1 means sciatica present, 0 means no sciatica). In addition to requesting the adjusted means, we request the model to be printed and a graph of the results.

```
. adjmean cost, by(provtype) adjust(age severity sciatica) model graph ylabel
xlabel(1,2,3,4)
  Source |       SS       df       MS                  Number of obs =    1401
---------+------------------------------               F(  6,  1394) =   48.04
   Model |  89651393.4     6  14941898.9               Prob > F      =  0.0000
Residual |   433565958  1394  311022.925               R-squared     =  0.1713
---------+------------------------------               Adj R-squared =  0.1678
   Total |   523217351  1400  373726.679               Root MSE      =  557.69
------------------------------------------------------------------------------
    cost |      Coef.   Std. Err.       t     P>|t|       [95% Conf. Interval]
---------+--------------------------------------------------------------------
      X2 |   248.1997   34.51434      7.191   0.000       180.4941    315.9054
      X3 |   200.9192   50.19444      4.003   0.000       102.4544     299.384
      X4 |  -24.06906   47.30386     -0.509   0.611      -116.8635    68.72537
     age |  -3.698749   1.149891     -3.217   0.001      -5.954453   -1.443045
severity |   26.00893   2.204246     11.799   0.000       21.68494    30.33293
sciatica |   238.0384   36.73184      6.480   0.000       165.9827     310.094
   _cons |   201.8666   57.91298      3.486   0.001       88.26064    315.4726
------------------------------------------------------------------------------
*Adjusted* Means and 95% Confidence Intervals
  Outcome:     Total Cost of LBP Treatment -- cost
  Nominal X:   Provider Type -- provtype
  Covariates:  age severity sciatica
 provtype      numobs     adjmean         se      lower      upper
    1:PC          536   386.3189   24.11297   339.0583   433.5794
  2:Chiro         513   634.5186   24.69707   586.1132    682.924
  3:Ortho         161    587.238    44.0196   500.9612   673.5149
    4:HMO         191   362.2498   40.61471   282.6465   441.8532
  Test for difference of 4 means:
    F(3, 1394) =   22.43
    Prob > F   <   0.0001
( See Figure 1 below)
```

The variable, `X2`, is the dummy variable which compares the adjusted mean cost for chiropractors (a `provtype` of 2) versus primary care physicians (a `provtype` of 1), `X3` compares orthopedists (a `provtype` of 3) to primary care, and `X4` compares HMO's (a `provtype` of 4) to primary care. The orthopedists' mean costs decreases from $628 (Example 1 result) to $587 after adjustment for age, severity of low back pain, and sciatica. Primary care physician–adjusted mean costs are lower as well, while chiropractors and HMO–adjusted costs increase. There is still an overall statistical difference in these means. Chiropractors and orthopedists charge more, even after adjusting for age and measures of severity. A graph of the adjusted means and 95% confidence intervals is in Figure 1. Had we preferred a bar graph, we could have typed

```
. adjmean cost, by(provtype) adjust(age severity sciatica) model ylabel graph bar
  (output omitted)
( See Figure 2 below)
```

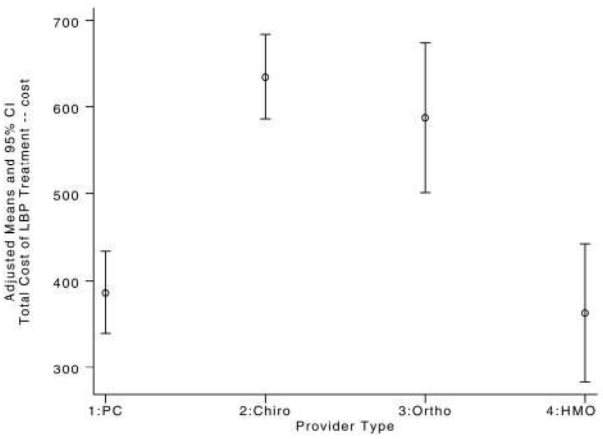

Figure 1. Adjusted means and 95% CI.

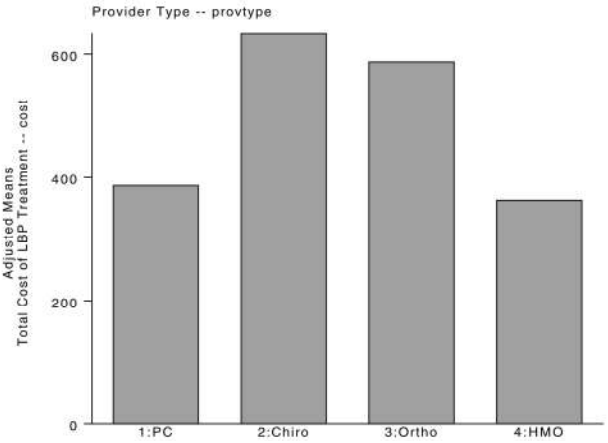

Figure 2. Adjusted means bar graph.

If you would like to test the difference of any pairs of means (not in the model), you can do so with the **test** statement after running **adjmean**. For example, suppose we wanted a test for a comparison between mean costs for chiropractors and orthopedists:

```
. test _b[X2] = _b[X3]
  ( 1)  X2 - X3 = 0.0
     F(  1,  1394) =    0.87
     Prob > F =    0.3499
```

Example 3

Suppose this time we want to compare costs for chiropractors versus any type of physician (**provtype** = 1, 3, and 4 combined), adjusted for age, severity, and sciatica. The variable is called **chiromd** (1=chiropractor, 0=MD). In addition, after fitting the model, let's calculate the predicted mean cost by chiropractor versus MD for someone who is 40 years old, has a low severity score (2), and no sciatica (0). The program does this by estimating the betas for the regression equation, then calculating the adjusted means by substituting specified values for the covariates into the equation.

```
. adjmean cost, by(chiromd) adjust(age=40 severity=2 sciatica=0)

*Adjusted* Means and 95% Confidence Intervals

  Outcome:     Total Cost of LBP Treatment -- cost
  Nominal X:   Chiropractor vs. MD -- chiromd
  Covariates:  age=40 severity=2 sciatica=0

  chiromd      numobs    adjmean         se      lower      upper
     0:MD         888   134.7998    27.6567   80.59362   189.0059
  1:Chiro         513   351.9856   31.48407   290.2779   413.6932

  Test for difference of 2 means:

    F(1, 1396) =   48.28
    Prob > F   <   0.0001
```

Overall, the mean costs are lower for these patients with less severe back pain, though we see that chiropractors have significantly higher costs ($352) than MD's ($135). Let's repeat Example 3 for patients who are the same age (40), but have more severe low back pain, say a high severity score of 20 and the presence of sciatica (1).

```
. adjmean cost, by(chiromd) adjust(age=40 severity=20 sciatica=1)

*Adjusted* Means and 95% Confidence Intervals

  Outcome:        Total Cost of LBP Treatment -- cost
  Nominal X:   Chiropractor vs. MD -- chiromd
  Covariates:     age=40 severity=20 sciatica=1

  chiromd      numobs    adjmean         se      lower      upper
     0:MD         888    847.096   37.21108   774.1641   920.0288
  1:Chiro         513   1064.282   40.35933   985.1794   1143.385

  Test for difference of 2 means:

    F(1, 1396) =   48.28
    Prob > F   <   0.0001
```

Not surprisingly, the cost of care increases substantially when the patients have more severe low back pain. Chiropractors continue to have significantly higher mean costs ($1,064) than MD's ($847).

Example 4

This example estimates the mean costs by categories of two nominal variables: `chiromd` (1=chiropractor, 0=MD) and `sciatica` (1=sciatica present, 0=no sciatica), adjusted for age and severity. The model will be printed, and a graph is requested, defaulting to points which will be connected.

```
. adjmean cost, by(chiromd sciatica) adjust(age severity) model graph ylabel xlabel(0,1) c(ll)
  Source |       SS       df       MS                  Number of obs =    1401
---------+------------------------------               F(  5,  1395) =   53.30
   Model |  83919668.0     5  16783933.6               Prob > F      =  0.0000
Residual |   439297683  1395  314908.733               R-squared     =  0.1604
---------+------------------------------               Adj R-squared =  0.1574
   Total |   523217351  1400  373726.679               Root MSE      =  561.17

------------------------------------------------------------------------------
    cost |      Coef.   Std. Err.       t     P>|t|       [95% Conf. Interval]
---------+--------------------------------------------------------------------
     X12 |   215.2707   35.58975      6.049   0.000       145.4555     285.086
     X22 |     236.85   47.37293      5.000   0.000       143.9201    329.7798
     I22 |    8.35065   74.12591      0.113   0.910      -137.0596    153.7609
     age |  -3.800583   1.151852     -3.300   0.001      -6.060132   -1.541035
severity |   26.23992   2.218988     11.825   0.000       21.88701    30.59283
   _cons |   234.9213   54.74507      4.291   0.000       127.5297    342.3128
------------------------------------------------------------------------------

*Adjusted* Means and 95% Confidence Intervals

  Outcome:      Total Cost of LBP Treatment -- cost
  Nominal X1:   Chiropractor vs. MD -- chiromd
  Nominal X2:   Sciatica Present -- sciatica
  Interaction:  chiromd * sciatica
  Covariates:   age severity

  chiromd  sciatica      numobs    adjmean          se      lower      upper
     0:MD      0:No         705   365.1132    21.21326   323.5359   406.6904
     0:MD     1:Yes         183   601.9631     42.1138   519.4216   684.5047
  1:Chiro      0:No         387   580.3839     28.6184   524.2928    636.475
  1:Chiro     1:Yes         126   825.5845    50.12164    727.348   923.8211

  Test for difference of 4 means:

    F(3, 1395) =   31.83
    Prob > F   <   0.0001

  Test for interaction of chiromd * sciatica:

    F(1, 1395) =    0.01
    Prob > F   =    0.9103
```

(*See Figure 3 below*)

The dummy variable for `chiromd` is `X12` and the dummy variable for `sciatica` is `X22`, with an interaction variable (`I22`) created between them. (Note: This would have been simpler to read if the variables had retained their original names, but often there are more than two categories per variable, and naming gets complicated. Most of the time, we're not interested in seeing the model, anyway. The default summary statistics usually give all the information we want.)

The highest mean cost is for patients with sciatica seeing chiropractors ($826), but costs are higher also for sciatica patients seeing physicians ($602). There is no evidence of interaction ($p = 0.9013$) which we can see from the regression table for `I22`, or from the F statistic at the bottom of the output. The graph (see Figure 3) confirms this. Although it may not be appropriate to connect the points of a nominal variable, it's done here to demonstrate that the `connect` option can work, as well as to show that two lines couldn't be much more parallel (a clear indication of no interaction).

Had we preferred a bar graph, we could have typed

```
. adjmean cost, by(chiromd sciatica) adjust(age severity) model graph bar ylabel
```

(*See Figure 4 below*)

(*Graphs on next page*)

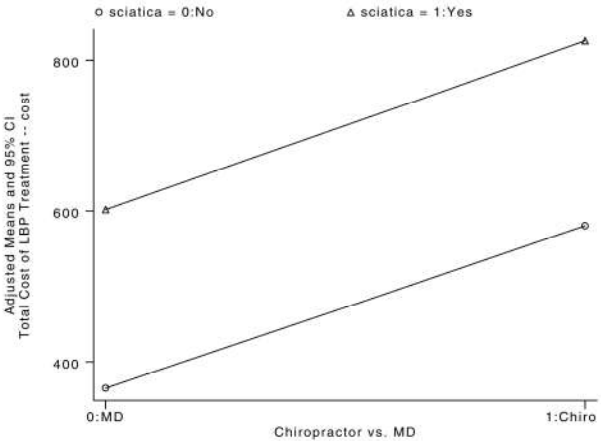

Figure 3. Mean cost of `chirmod` & `sciatica` adjusted for age & severity.

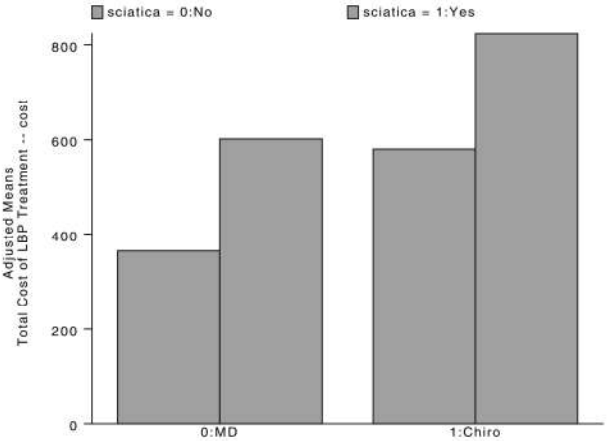

Figure 4. Bar graph version of Figure 3.

Example 5

Let's look at an example where there is significant interaction. When we examine the adjusted mean costs for categories of `chiromd` by gender (1=male, 0=female), we get the following results (model not printed):

```
. adjmean cost, by(chiromd gender) adjust(age severity sciatica) graph bar ylabel
  (output omitted)
*Adjusted* Means and 95% Confidence Intervals
  Outcome:      Total Cost of LBP Treatment -- cost
  Nominal X1:   Chiropractor vs. MD -- chiromd
  Nominal X2:   Patient Gender -- gender
  Interaction:  chiromd * gender
  Covariates:   age severity sciatica
  chiromd    gender      numobs    adjmean         se      lower      upper
     0:MD  0:Female         182   304.1542   41.50648    222.803   385.5054
     0:MD    1:Male         705   446.4079   20.99842   405.2517   487.5641
  1:Chiro  0:Female          60   881.6437   72.07227   740.3847   1022.903
  1:Chiro    1:Male         453   601.9128   26.28687   550.3915   653.4341
  Test for difference of 4 means:
    F(3, 1393) =   23.94
    Prob > F   <   0.0001
  Test for interaction of chiromd * gender:
    F(1, 1393) =   22.19
    Prob > F   <   0.0001
```

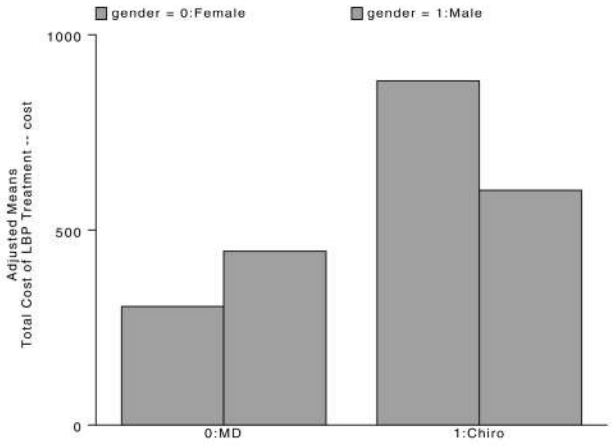

Figure 5. Mean costs by `chiromd` and `gender`.

Within chiropractors, mean costs are higher for females ($882) than males ($602), whereas within physicians, the relationship is reversed—females: $304, males: $446. The test for interaction is highly significant (p <0.0001), and the bar graph (see Figure 5) illustrates the relationship visually.

Example 6

If we repeat the last example using our original four categories for provider (`provtype`), we find that females have higher mean costs than males if they visit chiropractors, but have about the same or lower costs than males for primary care, orthopedic, or HMO physicians. This is easiest to see in Figure 6.

```
. adjmean cost, by(provtype gender) adjust(age severity sciatica) graph bar ylabel
  (output omitted)
*Adjusted* Means and 95% Confidence Intervals
  Outcome:      Total Cost of LBP Treatment -- cost
  Nominal X1:   Provider Type -- provtype
  Nominal X2:   Patient Gender -- gender
  Interaction:  provtype * gender
  Covariates:   age severity sciatica
 provtype    gender      numobs    adjmean         se      lower      upper
    1:PC  0:Female          89   268.8453   58.88354   153.4357   384.2549
    1:PC    1:Male         447   409.9181   26.25605   358.4572    461.379
  2:Chiro  0:Female          60    882.098   71.71287   741.5433   1022.653
  2:Chiro    1:Male         453   601.6091   26.15592   550.3444   652.8737
  3:Ortho  0:Female          25   569.0792   110.9323    351.656   786.5024
  3:Ortho    1:Male         136   590.6166   47.57316   497.3749   683.8583
    4:HMO  0:Female          68   254.8119   67.46792   122.5772   387.0466
    4:HMO    1:Male         122   419.2326   50.38327   320.4832    517.982
  Test for difference of 8 means:
    F(7, 1389) =   12.94
    Prob > F   <   0.0001
  Test for interaction of provtype * gender:
    F(3, 1389) =    7.34
    Prob > F   <   0.0001
```

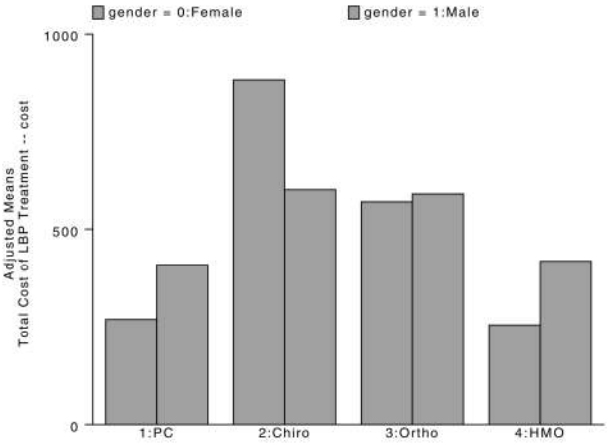

Figure 6. Mean costs adjusted by provider type and gender.

Example 7

In the previous examples, we looked at adjusted means (costs), so let's look at a couple of examples where the outcome is dichotomous rather than continuous, and we want to estimate adjusted probabilities using logistic regression (`adjprop`). Another outcome we considered in this study was how quickly patients got over their back pain. Of course, each provider type was sure their patients would get better quicker. We created a variable, `status4w`, which measured whether patients still had back pain at four weeks (`status4w`: 1=not better, 0=better). So, our logistic regression model predicts the probability of still having low back pain after four weeks. Since patients with less pain at baseline may have improved more quickly, we will adjust for `age`, `severity`, and `sciatica`. The following statement includes a request to display the logistic regression model results and a bar graph by provider type:

```
. adjprop status4w, by(provtype) adj(age severity sciatica) model graph bar ylabel
Logit Estimates                                   Number of obs =   1401
                                                  chi2(6)       = 170.75
                                                  Prob > chi2   = 0.0000
Log Likelihood = -582.29855                       Pseudo R2     = 0.1279
```