```
* Probit and Logit Models in Stata
* Copyright 2013 by Ani Katchova

clear all
set more off

use C:\Econometrics\Data\probit_insurance

global ylist ins
global xlist retire age hstatusg hhincome educyear married hisp

describe $ylist $xlist
summarize $ylist $xlist

tabulate $ylist

* Regression
reg $ylist $xlist

* Probit model
probit $ylist $xlist

* Logit model
logit $ylist $xlist


* Marginal effects (at the mean and average marginal effect)
quietly reg $ylist $xlist
margins, dydx(*) atmeans
margins, dydx(*)

quietly logit $ylist $xlist
margins, dydx(*) atmeans
margins, dydx(*)

quietly probit $ylist $xlist
margins, dydx(*) atmeans
margins, dydx(*)


*Logistic model gives odds ratio
logistic $ylist $xlist


* Predicted probabilities
quietly logit $ylist $xlist
predict plogit, pr

quietly probit $ylist $xlist
predict pprobit, pr

quietly regress $ylist $xlist
predict pols, xb

summarize $ylist plogit pprobit pols
```

```
* Percent correctly predicted values
quietly logit $ylist $xlist
estat classification

quietly probit $ylist $xlist
estat classification
```

```
. * Probit and Logit Models in Stata
. * Copyright 2013 by Ani Katchova
.
. clear all

. set more off

.
. use C:\Econometrics\Data\probit_insurance

.
. global ylist ins

. global xlist retire age hstatusg hhincome educyear married hisp

.
. describe $ylist $xlist

              storage  display     value
variable name   type   format      label      variable label
------------------------------------------------------------------------------
------------------
ins             float  %9.0g
retire          double %12.0g
age             double %12.0g
hstatusg        float  %9.0g
hhincome        float  %9.0g
educyear        double %12.0g
married         double %12.0g
hisp            double %12.0g

. summarize $ylist $xlist

    Variable |       Obs        Mean    Std. Dev.       Min        Max
-------------+--------------------------------------------------------
         ins |      3206    .3870867    .4871597          0          1
      retire |      3206    .6247661    .4842588          0          1
         age |      3206    66.91391    3.675794         52         86
    hstatusg |      3206    .7046163    .4562862          0          1
    hhincome |      3206    45.26391    64.33936          0   1312.124
-------------+--------------------------------------------------------
    educyear |      3206    11.89863    3.304611          0         17
     married |      3206    .7330006     .442461          0          1
        hisp |      3206    .0726762    .2596448          0          1

.
. tabulate $ylist

        ins |      Freq.     Percent        Cum.
------------+-----------------------------------
          0 |      1,965       61.29       61.29
          1 |      1,241       38.71      100.00
------------+-----------------------------------
      Total |      3,206      100.00
```

```
.
. * Regression
. reg $ylist $xlist

      Source |       SS       df       MS              Number of obs =    3206
-------------+------------------------------           F(  7,  3198) =   41.14
       Model |  62.8403396     7  8.97719137           Prob > F      =  0.0000
    Residual |   697.78505  3198    .2181942           R-squared     =  0.0826
-------------+------------------------------           Adj R-squared =  0.0806
       Total |   760.62539  3205  .237324615           Root MSE      =  .46711

------------------------------------------------------------------------------
         ins |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
      retire |   .0408508   .0182197     2.24   0.025     .0051273    .0765743
         age |  -.0028955   .0024189    -1.20   0.231    -.0076383    .0018473
    hstatusg |   .0655583   .0194531     3.37   0.001     .0274166    .1037001
    hhincome |   .0004921   .0001375     3.58   0.000     .0002225    .0007617
    educyear |   .0233686   .0028672     8.15   0.000      .017747    .0289903
     married |   .1234699   .0193618     6.38   0.000     .0855071    .1614326
        hisp |  -.1210059    .033666    -3.59   0.000     -.187015   -.0549969
       _cons |   .1270857   .1605628     0.79   0.429    -.1877308    .4419021
------------------------------------------------------------------------------

.
. * Probit model
. probit $ylist $xlist

Iteration 0:   log likelihood = -2139.7712
Iteration 1:   log likelihood = -1994.4552
Iteration 2:   log likelihood =  -1993.624
Iteration 3:   log likelihood = -1993.6237
Iteration 4:   log likelihood = -1993.6237

Probit regression                                 Number of obs   =       3206
                                                  LR chi2(7)      =     292.30
                                                  Prob > chi2     =     0.0000
Log likelihood = -1993.6237                       Pseudo R2       =     0.0683

------------------------------------------------------------------------------
         ins |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
      retire |   .1183567   .0512678     2.31   0.021     .0178736    .2188397
         age |  -.0088696    .006899    -1.29   0.199    -.0223914    .0046521
    hstatusg |   .1977357   .0554868     3.56   0.000     .0889835    .3064878
    hhincome |    .001233   .0003866     3.19   0.001     .0004754    .0019907
    educyear |   .0707477   .0084782     8.34   0.000     .0541308    .0873647
     married |    .362329   .0560031     6.47   0.000      .252565    .4720931
        hisp |  -.4731099   .1104393    -4.28   0.000     -.689567   -.2566529
       _cons |  -1.069319   .4580794    -2.33   0.020    -1.967139   -.1715002
------------------------------------------------------------------------------

.
. * Logit model
. logit $ylist $xlist
```

```
Iteration 0:   log likelihood = -2139.7712
Iteration 1:   log likelihood = -1996.7434
Iteration 2:   log likelihood = -1994.8864
Iteration 3:   log likelihood = -1994.8784
Iteration 4:   log likelihood = -1994.8784

Logistic regression                                     Number of obs   =       3206
                                                        LR chi2(7)      =     289.79
                                                        Prob > chi2     =     0.0000
Log likelihood = -1994.8784                             Pseudo R2       =     0.0677

------------------------------------------------------------------------------
         ins |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
      retire |   .1969297   .0842067     2.34   0.019     .0318875    .3619718
         age |  -.0145955   .0112871    -1.29   0.196    -.0367178    .0075267
    hstatusg |   .3122654   .0916739     3.41   0.001     .1325878     .491943
    hhincome |   .0023036    .000762     3.02   0.003       .00081    .0037972
    educyear |   .1142626   .0142012     8.05   0.000     .0864288    .1420963
     married |    .578636   .0933198     6.20   0.000     .3957327    .7615394
        hisp |  -.8103059   .1957522    -4.14   0.000    -1.193973   -.4266387
       _cons |  -1.715578   .7486219    -2.29   0.022     -3.18285   -.2483064
------------------------------------------------------------------------------

.
.
. * Marginal effects (at the mean and average marginal effect)
. quietly reg $ylist $xlist

. margins, dydx(*) atmeans

Conditional marginal effects                            Number of obs   =       3206
Model VCE    : OLS

Expression   : Linear prediction, predict()
dy/dx w.r.t. : retire age hstatusg hhincome educyear married hisp
at           : retire          =    .6247661 (mean)
               age             =    66.91391 (mean)
               hstatusg        =    .7046163 (mean)
               hhincome        =    45.26391 (mean)
               educyear        =    11.89863 (mean)
               married         =    .7330006 (mean)
               hisp            =    .0726762 (mean)

------------------------------------------------------------------------------
             |            Delta-method
             |      dy/dx   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
      retire |   .0408508   .0182197     2.24   0.025     .0051408    .0765608
         age |  -.0028955   .0024189    -1.20   0.231    -.0076366    .0018455
    hstatusg |   .0655583   .0194531     3.37   0.001      .027431    .1036856
    hhincome |   .0004921   .0001375     3.58   0.000     .0002226    .0007616
    educyear |   .0233686   .0028672     8.15   0.000     .0177491    .0289882
     married |   .1234699   .0193618     6.38   0.000     .0855215    .1614183
```

```
        hisp |  -.1210059    .033666    -3.59   0.000      -.18699   -.0550218
------------------------------------------------------------------------------

. margins, dydx(*)

Average marginal effects                        Number of obs   =       3206
Model VCE    : OLS

Expression   : Linear prediction, predict()
dy/dx w.r.t. : retire age hstatusg hhincome educyear married hisp

------------------------------------------------------------------------------
             |            Delta-method
             |      dy/dx   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
      retire |   .0408508   .0182197     2.24   0.025     .0051408    .0765608
         age |  -.0028955   .0024189    -1.20   0.231    -.0076366    .0018455
    hstatusg |   .0655583   .0194531     3.37   0.001      .027431    .1036856
    hhincome |   .0004921   .0001375     3.58   0.000     .0002226    .0007616
    educyear |   .0233686   .0028672     8.15   0.000     .0177491    .0289882
     married |   .1234699   .0193618     6.38   0.000     .0855215    .1614183
        hisp |  -.1210059    .033666    -3.59   0.000      -.18699   -.0550218
------------------------------------------------------------------------------

.
. quietly logit $ylist $xlist

. margins, dydx(*) atmeans

Conditional marginal effects                    Number of obs   =       3206
Model VCE    : OIM

Expression   : Pr(ins), predict()
dy/dx w.r.t. : retire age hstatusg hhincome educyear married hisp
at           : retire          =    .6247661 (mean)
               age             =    66.91391 (mean)
               hstatusg        =    .7046163 (mean)
               hhincome        =    45.26391 (mean)
               educyear        =    11.89863 (mean)
               married         =    .7330006 (mean)
               hisp            =    .0726762 (mean)

------------------------------------------------------------------------------
             |            Delta-method
             |      dy/dx   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
      retire |   .0460479   .0196856     2.34   0.019     .0074648     .084631
         age |  -.0034129   .0026389    -1.29   0.196     -.008585    .0017592
    hstatusg |   .0730168    .021412     3.41   0.001     .0310499    .1149836
    hhincome |   .0005386   .0001785     3.02   0.003     .0001888    .0008885
    educyear |   .0267179   .0033025     8.09   0.000     .0202452    .0331907
     married |    .135302   .0217469     6.22   0.000     .0926789    .1779251
        hisp |  -.1894732    .045563    -4.16   0.000    -.2787749   -.1001714
------------------------------------------------------------------------------
```

```
. margins, dydx(*)

Average marginal effects                          Number of obs   =       3206
Model VCE    : OIM

Expression   : Pr(ins), predict()
dy/dx w.r.t. : retire age hstatusg hhincome educyear married hisp

------------------------------------------------------------------------------
             |            Delta-method
             |      dy/dx   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
      retire |   .0427616    .018228     2.35   0.019     .0070354    .0784878
         age |  -.0031693   .0024486    -1.29   0.196    -.0079686      .00163
    hstatusg |   .0678058   .0197778     3.43   0.001     .0290419    .1065696
    hhincome |   .0005002   .0001646     3.04   0.002     .0001777    .0008228
    educyear |   .0248111   .0029705     8.35   0.000     .0189891    .0306332
     married |   .1256459   .0198205     6.34   0.000     .0867985    .1644933
        hisp |   -.175951   .0421962    -4.17   0.000     -.258654   -.0932481
------------------------------------------------------------------------------

.
. quietly probit $ylist $xlist

. margins, dydx(*) atmeans

Conditional marginal effects                      Number of obs   =       3206
Model VCE    : OIM

Expression   : Pr(ins), predict()
dy/dx w.r.t. : retire age hstatusg hhincome educyear married hisp
at           : retire          =    .6247661 (mean)
               age             =    66.91391 (mean)
               hstatusg        =    .7046163 (mean)
               hhincome        =    45.26391 (mean)
               educyear        =    11.89863 (mean)
               married         =    .7330006 (mean)
               hisp            =    .0726762 (mean)

------------------------------------------------------------------------------
             |            Delta-method
             |      dy/dx   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
      retire |   .0448503   .0194248     2.31   0.021     .0067783    .0829222
         age |  -.0033611   .0026141    -1.29   0.199    -.0084846    .0017624
    hstatusg |   .0749303   .0210134     3.57   0.000     .0337448    .1161158
    hhincome |   .0004672   .0001466     3.19   0.001       .00018    .0007545
    educyear |   .0268093   .0032034     8.37   0.000     .0205306    .0330879
     married |   .1373016   .0211866     6.48   0.000     .0957766    .1788267
        hisp |  -.1792811   .0417552    -4.29   0.000    -.2611197   -.0974425
------------------------------------------------------------------------------

. margins, dydx(*)

Average marginal effects                          Number of obs   =       3206
```

```
Model VCE    : OIM

Expression   : Pr(ins), predict()
dy/dx w.r.t. : retire age hstatusg hhincome educyear married hisp

------------------------------------------------------------------------------
             |            Delta-method
             |      dy/dx   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
      retire |   .0419784   .0181404     2.31   0.021     .0064239    .0775329
         age |  -.0031459   .0024451    -1.29   0.198    -.0079382    .0016465
    hstatusg |   .0701324   .0195584     3.59   0.000     .0317987     .108466
    hhincome |   .0004373   .0001366     3.20   0.001     .0001696     .000705
    educyear |   .0250926   .0029084     8.63   0.000     .0193923    .0307929
     married |   .1285099   .0194623     6.60   0.000     .0903646    .1666552
        hisp |  -.1678014   .0388964    -4.31   0.000     -.244037   -.0915657
------------------------------------------------------------------------------

.
.
. *Logistic model gives odds ratio
. logistic $ylist $xlist

Logistic regression                               Number of obs   =       3206
                                                  LR chi2(7)      =     289.79
                                                  Prob > chi2     =     0.0000
Log likelihood = -1994.8784                       Pseudo R2       =     0.0677

------------------------------------------------------------------------------
         ins | Odds Ratio   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
      retire |   1.217658    .102535     2.34   0.019     1.032401    1.436159
         age |   .9855105   .0111235    -1.29   0.196     .9639481    1.007555
    hstatusg |   1.366517    .125274     3.41   0.001     1.141779    1.635491
    hhincome |   1.002306   .0007638     3.02   0.003      1.00081    1.003804
    educyear |   1.121046   .0159201     8.05   0.000     1.090274    1.152688
     married |   1.783604   .1664455     6.20   0.000     1.485472     2.14157
        hisp |    .444722   .0870553    -4.14   0.000     .3030149    .6526993
       _cons |   .1798597   .1346469    -2.29   0.022     .0414673    .7801209
------------------------------------------------------------------------------

.
.
. * Predicted probabilities
. quietly logit $ylist $xlist

. predict plogit, pr

.
. quietly probit $ylist $xlist

. predict pprobit, pr

.
. quietly regress $ylist $xlist
```

```
. predict pols, xb

.
. summarize $ylist plogit pprobit pols

    Variable |       Obs        Mean    Std. Dev.       Min        Max
-------------+--------------------------------------------------------
         ins |      3206    .3870867    .4871597          0          1
      plogit |      3206    .3870867    .1418287   .0340215   .9649615
     pprobit |      3206    .3861139    .1421416   .0206445   .9647618
        pols |      3206    .3870867    .1400249  -.1557328   1.197223

.
.
. * Percent correctly predicted values
. quietly logit $ylist $xlist

. estat classification

Logistic model for ins

              -------- True --------
Classified |         D            ~D  |      Total
-----------+--------------------------+-----------
     +     |       345           308  |        653
     -     |       896          1657  |       2553
-----------+--------------------------+-----------
   Total   |      1241          1965  |       3206

Classified + if predicted Pr(D) >= .5
True D defined as ins != 0
--------------------------------------------------
Sensitivity                     Pr( +| D)   27.80%
Specificity                     Pr( -|~D)   84.33%
Positive predictive value       Pr( D| +)   52.83%
Negative predictive value       Pr(~D| -)   64.90%
--------------------------------------------------
False + rate for true ~D        Pr( +|~D)   15.67%
False - rate for true D         Pr( -| D)   72.20%
False + rate for classified +   Pr(~D| +)   47.17%
False - rate for classified -   Pr( D| -)   35.10%
--------------------------------------------------
Correctly classified                        62.45%
--------------------------------------------------

.
. quietly probit $ylist $xlist

. estat classification

Probit model for ins

              -------- True --------
Classified |         D            ~D  |      Total
```

```
-----------+-------------------------+-----------
     +     |       335           305  |        640
     -     |       906          1660  |       2566
-----------+-------------------------+-----------
   Total   |      1241          1965  |       3206

Classified + if predicted Pr(D) >= .5
True D defined as ins != 0
--------------------------------------------------
Sensitivity                     Pr( +| D)   26.99%
Specificity                     Pr( -|~D)   84.48%
Positive predictive value       Pr( D| +)   52.34%
Negative predictive value       Pr(~D| -)   64.69%
--------------------------------------------------
False + rate for true ~D        Pr( +|~D)   15.52%
False - rate for true D         Pr( -| D)   73.01%
False + rate for classified +   Pr(~D| +)   47.66%
False - rate for classified -   Pr( D| -)   35.31%
--------------------------------------------------
Correctly classified                        62.23%
--------------------------------------------------
```