.
. * 2. Estudiamos la distribución de la vble endógena Status
. tab status

    Status | Freq.   Percent   Cum.
------------+-----------------------------------
        0 | 259   31.24   31.24
        1 | 570   68.76   100.00
------------+-----------------------------------
     Total | 829   100.00

.
. * 3. Estimamos vía logit el modelo original
. logit status gender education workexp programmingexp

Iteration 0:   log likelihood = -514.83131
Iteration 1:   log likelihood = -507.91144
Iteration 2:   log likelihood = -507.88834
Iteration 3:   log likelihood = -507.88834

Logistic regression   Number of obs =   829
                                                LR chi2(4)   = 13.89
                                                Prob > chi2   = 0.0077
Log likelihood = -507.88834   Pseudo R2   = 0.0135

----
       status | Coefficient Std. err.   z   P>|z| [95% conf. interval]
--------------+---
       gender | .1591773   .1775537   0.90   0.370   -.1888216   .5071762
    education | .5065154   .1909185   2.65   0.008   .132322   .8807088
      workexp | .2750172   .1619233   1.70   0.089   -.0423467   .5923811
programmingexp | -.2159291 .1653185   -1.31   0.192   -.5399475   .1080892
        _cons | .2260515   .2446994   0.92   0.356   -.2535506   .7056535

.
. * 4. Calculamos los marginales
. margins

Predictive margins   Number of obs = 829
Model VCE: OIM

Expression: Pr(status), predict()

             | Delta-method
             | Margin   std. err.   z   P>|z| [95% conf. interval]

```
-------------+-----------------------------------------------------------
      _cons |   .6875754   .0159602    43.08   0.000     .656294    .7188568
-----------------------------------------------------------------------
```

.
. * 5. Calculamos la matriz de aciertos y errores
. predict status_pr
(option pr assumed; Pr(status))

. estat class

Logistic model for status

```
              -------- True --------
Classified |         D        ~D  |     Total
-----------+-------------------------+-----------
    +    |      570        259  |      829
    -    |       0          0  |       0
-----------+-------------------------+-----------
  Total  |      570        259  |      829
```

Classified + if predicted Pr(D) >= .5
True D defined as status != 0
```
--------------------------------------------------
Sensitivity                    Pr( +| D)  100.00%
Specificity                    Pr( -|~D)    0.00%
Positive predictive value      Pr( D| +)   68.76%
Negative predictive value      Pr(~D| -)      .%
--------------------------------------------------
False + rate for true ~D       Pr( +|~D)  100.00%
False - rate for true D        Pr( -| D)    0.00%
False + rate for classified +  Pr(~D| +)   31.24%
False - rate for classified -  Pr( D| -)      .%
--------------------------------------------------
Correctly classified                       68.76%
--------------------------------------------------
```

.
. * 6. Realizamos un análisis univariante básico de la columna de predicciones
. sum status_pr

```
   Variable |      Obs       Mean    Std. dev.      Min        Max
-------------+---------------------------------------------------------
  status_pr |      829   .6875754    .0604164   .5025306   .7625591
```

.
. * 7. Realizamos un análisis univariante básico de la columna Status
. sum status

```
   Variable |      Obs       Mean    Std. dev.      Min        Max
-------------+---------------------------------------------------------
     status |      829   .6875754    .4637617        0          1
```

.
. * Como resultado vemos que la media es la misma, por lo que aunque status_pr no

. * esté normalizada y Status sí, tienen un resultado muy parecido.
.
. * 8. Calculamos la capacidad predictiva con un límite de 0.5
. estat class, cutoff(0.5)

Logistic model for status

```
            -------- True --------
Classified |      D         ~D  |     Total
-----------+------------------------+-----------
    +    |    570          259 |     829
    -    |     0            0  |      0
-----------+------------------------+-----------
  Total  |    570          259 |     829
```

Classified + if predicted Pr(D) >= .5
True D defined as status != 0

Sensitivity   Pr(+| D) 100.00%
Specificity   Pr(-|~D)   0.00%
Positive predictive value   Pr(D| +)   68.76%
Negative predictive value   Pr(~D| -)   .%

False + rate for true ~D   Pr(+|~D) 100.00%
False - rate for true D   Pr(-| D)   0.00%
False + rate for classified + Pr(~D| +)   31.24%
False - rate for classified - Pr(D| -)   .%

Correctly classified   68.76%

.
. * 9. Comprobamos que añadir matching no tiene sentido porque imita el comportamiento
. *   de Status.
. logit status gender education workexp programmingexp matching

note: matching != 0 predicts success perfectly;
    matching omitted and 487 obs not used.

note: education != 0 predicts failure perfectly;
    education omitted and 201 obs not used.

Iteration 0:   log likelihood = -95.505679
Iteration 1:   log likelihood = -94.646437
Iteration 2:   log likelihood = -94.645706
Iteration 3:   log likelihood = -94.645706

Logistic regression   Number of obs =   141
                                       LR chi2(3)   =   1.72
                                       Prob > chi2   = 0.6325
Log likelihood = -94.645706   Pseudo R2   = 0.0090

----
     status | Coefficient Std. err.   z   P>|z| [95% conf. interval]
------------+--

```
        gender |  -.3968064   .4666304    -0.85  0.395   -1.311385    .5177723
     education |        0  (omitted)
       workexp |  .3998252   .3822363     1.05  0.296   -.3493442    1.148995
programmingexp |  .0800877  .4181156     0.19  0.848   -.7394038    .8995792
      matching |        0  (omitted)
         _cons |   .493503  .4871577     1.01  0.311   -.4613086    1.448315
-------------------------------------------------------------------------------
```

.
. * 10. Comprobamos el comportamiento entre matching y status. Los resultados son muy buenos.
. tab matching status

```
           |       Status
  Matching |        0         1 |    Total
-----------+----------------------+----------
         0 |      259        83 |     342
         1 |        0       487 |     487
-----------+----------------------+----------
     Total |      259       570 |     829
```

.
. * 11. Se calculan los marginales de todas las vbles (comprobamos p-value y signo del coeficiente)
. margins, dydx(*)

Average marginal effects   Number of obs = 141
Model VCE: OIM

Expression: Pr(status), predict()
dy/dx wrt: gender education workexp programmingexp matching

```
-------------------------------------------------------------------------------
               |             Delta-method
               |     dy/dx   std. err.     z   P>|z|    [95% conf. interval]
---------------+---------------------------------------------------------------
        gender |  -.0949222   .1105392    -0.86  0.390   -.3115749    .1217306
     education |        0  (omitted)
       workexp |  .0956443   .0900663     1.06  0.288   -.0808825    .2721711
programmingexp |  .0191582   .0999705     0.19  0.848   -.1767803    .2150967
      matching |        0  (omitted)
-------------------------------------------------------------------------------
```

.
end of do-file