```
       ___  ___  ___  ___  ___   (R)
      /__    /  ___/  /  ___/
     ___/   /  /__/   /  /__/    14.2    Copyright 1985-2015 StataCorp LLC
       Statistics/Data Analysis          StataCorp
                                         4905 Lakeway Drive
          MP – Parallel Edition          College Station, Texas 77845 USA
                                         800-STATA-PC      http://www.stata.com
                                         979-696-4600      stata@stata.com
                                         979-696-4601 (fax)


     Single-user 8-core Stata perpetual license:
            Serial number:  10699393
              Licensed to:  Prashant Yadav
                            Student


     Notes:
          1.  Unicode is supported; see help unicode advice.
          2.  More than 2 billion observations are allowed; see help_obs_advice.
          3.  Maximum number of variables is set to 5000; see help set maxvar.
```

1 . sysuse nlsw88, clear
  (NLSW, 1988 extract)

2 .
3 . *Description and summarize data

4 . describe

```
Contains data from /Applications/Stata/ado/base/n/nlsw88.dta
  obs:         2,246                        NLSW, 1988 extract
 vars:            17                        1 May 2014 22:52
 size:        60,642                        (_dta has notes)
```

variable name	storage type	display format	value label	variable label
idcode	int	%8.0g		NLS id
age	byte	%8.0g		age in current year
race	byte	%8.0g	racelbl	race
married	byte	%8.0g	marlbl	married
never_married	byte	%8.0g		never married
grade	byte	%8.0g		current grade completed
collgrad	byte	%16.0g	gradlbl	college graduate
south	byte	%8.0g		lives in south
smsa	byte	%9.0g	smsalbl	lives in SMSA
c_city	byte	%8.0g		lives in central city
industry	byte	%23.0g	indlbl	industry
occupation	byte	%22.0g	occlbl	occupation
union	byte	%8.0g	unionlbl	union worker
wage	float	%9.0g		hourly wage
hours	byte	%8.0g		usual hours worked
ttl_exp	float	%9.0g		total work experience

tenure	float	%9.0g	**job tenure (years)**

Sorted by: **idcode**

5 . summarize

Variable	Obs	Mean	Std. Dev.	Min	Max
idcode	2,246	2612.654	1480.864	1	5159
age	2,246	39.15316	3.060002	34	46
race	2,246	1.282725	.4754413	1	3
married	2,246	.6420303	.4795099	0	1
never_marr~d	2,246	.1041852	.3055687	0	1
grade	2,244	13.09893	2.521246	0	18
collgrad	2,246	.2368655	.4252538	0	1
south	2,246	.4194123	.4935728	0	1
smsa	2,246	.7039181	.4566292	0	1
c_city	2,246	.2916296	.4546139	0	1
industry	2,232	8.189516	3.010875	1	12
occupation	2,237	4.642825	3.408897	1	13
union	1,878	.2454739	.4304825	0	1
wage	2,246	7.766949	5.755523	1.004952	40.74659
hours	2,242	37.21811	10.50914	1	80
ttl_exp	2,246	12.53498	4.610208	.1153846	28.88461
tenure	2,231	5.97785	5.510331	0	25.91667

6 .
7 . *Explore the union variable

8 . tab union

union worker	Freq.	Percent	Cum.
nonunion	1,417	75.45	75.45
union	461	24.55	100.00
Total	1,878	100.00	

9 .
10 . *Plot the union variable against age

11 . tw (sc union age)

12 . tw (sc union age) (lpoly union age)

13 .
14 . *Logit regression model – simple

15 . logit union age

```
Iteration 0:   log likelihood = -1046.6242
Iteration 1:   log likelihood = -1046.5069
Iteration 2:   log likelihood = -1046.5069
```

```
Logistic regression                          Number of obs    =        1,878
                                             LR chi2(1)       =         0.23
                                             Prob > chi2      =       0.6282
Log likelihood = -1046.5069                  Pseudo R2        =       0.0001
```

union	Coef.	Std. Err.	z	P>|z|	[95% Conf. Interval]	
age	.0085532	.0176572	0.48	0.628	-.0260542	.0431606
_cons	-1.458568	.6953749	-2.10	0.036	-2.821478	-.0956578

16 . predict phat
(option **pr** assumed; Pr(union))

17 . tw (sc union age) (sc phat age)

18 . tw (fn y = logistic(.0085532*x+ -1.458568), range(-1000 1000))

19 .
20 . *Logit regression model - multiple

21 . logit union age wage married collgrad

```
Iteration 0:   log likelihood = -1046.6242
Iteration 1:   log likelihood = -1022.7905
Iteration 2:   log likelihood = -1022.4618
Iteration 3:   log likelihood = -1022.4618
```

```
Logistic regression                          Number of obs    =        1,878
                                             LR chi2(4)       =        48.32
                                             Prob > chi2      =       0.0000
Log likelihood = -1022.4618                  Pseudo R2        =       0.0231
```

union	Coef.	Std. Err.	z	P>|z|	[95% Conf. Interval]	
age	.0079164	.0179577	0.44	0.659	-.0272801	.0431129
wage	.0670861	.0132001	5.08	0.000	.0412143	.0929579
married	-.2175451	.1124157	-1.94	0.053	-.4378758	.0027856
collgrad	.2598352	.1288056	2.02	0.044	.0073809	.5122894
_cons	-1.893258	.7195491	-2.63	0.009	-3.303548	-.4829673

22 .
23 . *Marginals effects

24 . margins, dydx(_all)

```
Average marginal effects                      Number of obs    =      1,878
Model VCE      : OIM

Expression     : Pr(union), predict()
dy/dx w.r.t. : age wage married collgrad
```

	dy/dx	Delta-method Std. Err.	z	P>|z|	[95% Conf. Interval]	
age	.0014259	.0032342	0.44	0.659	-.004913	.0077648
wage	.0120839	.0023262	5.19	0.000	.0075246	.0166432
married	-.0391854	.0201932	-1.94	0.052	-.0787634	.0003925
collgrad	.0468029	.0231306	2.02	0.043	.0014678	.0921381

25 . margins, dydx(married) at(age==40 collgrad == 1 wage == 30)

```
Average marginal effects                      Number of obs    =      1,878
Model VCE      : OIM

Expression     : Pr(union), predict()
dy/dx w.r.t. : married
at             : age            =             40
                 wage           =             30
                 collgrad       =              1
```

	dy/dx	Delta-method Std. Err.	z	P>|z|	[95% Conf. Interval]	
married	-.0503114	.0263575	-1.91	0.056	-.1019712	.0013484

26 .
27 . *Classification table

28 . estat classification

Logistic model for union

Classified	True D	~D	Total
+	6	14	20
-	455	1403	1858
Total	461	1417	1878

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

```
Sensitivity                    Pr( +| D)      1.30%
Specificity                    Pr( -|~D)     99.01%
Positive predictive value      Pr( D| +)     30.00%
Negative predictive value      Pr(~D| -)     75.51%
─────────────────────────────────────────────────────
False + rate for true ~D       Pr( +|~D)      0.99%
False - rate for true D        Pr( -| D)     98.70%
False + rate for classified +  Pr(~D| +)     70.00%
False - rate for classified -  Pr( D| -)     24.49%
─────────────────────────────────────────────────────
Correctly classified                         75.03%
─────────────────────────────────────────────────────
```

29 .
30 . *Table of regressions

31 . reg union age wage married collgrad, robust

```
Linear regression                     Number of obs   =      1,878
                                      F(4, 1873)      =      12.13
                                      Prob > F        =     0.0000
                                      R-squared       =     0.0272
                                      Root MSE        =    .42504
```

union	Coef.	Robust Std. Err.	t	P>|t|	[95% Conf. Interval]	
age	.0014404	.0032183	0.45	0.655	-.0048714	.0077521
wage	.0135656	.0026387	5.14	0.000	.0083904	.0187407
married	-.0403997	.0209978	-1.92	0.055	-.0815813	.0007818
collgrad	.0494278	.0262915	1.88	0.060	-.002136	.1009916
_cons	.1004228	.1292727	0.78	0.437	-.1531108	.3539564

32 . estimates store a1

33 . logit union age wage married collgrad

```
Iteration 0:   log likelihood = -1046.6242
Iteration 1:   log likelihood = -1022.7905
Iteration 2:   log likelihood = -1022.4618
Iteration 3:   log likelihood = -1022.4618

Logistic regression                   Number of obs   =      1,878
                                      LR chi2(4)      =      48.32
                                      Prob > chi2     =     0.0000
Log likelihood = -1022.4618           Pseudo R2       =     0.0231
```

union	Coef.	Std. Err.	z	P>|z|	[95% Conf. Interval]	
age	.0079164	.0179577	0.44	0.659	-.0272801	.0431129

wage	.0670861	.0132001	5.08	0.000	.0412143	.0929579
married	-.2175451	.1124157	-1.94	0.053	-.4378758	.0027856
collgrad	.2598352	.1288056	2.02	0.044	.0073809	.5122894
_cons	-1.893258	.7195491	-2.63	0.009	-3.303548	-.4829673

34 . margins, dydx(_all) post

Average marginal effects   Number of obs   =   1,878
Model VCE   : **OIM**

Expression   : **Pr(union), predict()**
dy/dx w.r.t. : **age wage married collgrad**

	Delta-method					
	dy/dx	Std. Err.	z	P>|z|	[95% Conf. Interval]	
age	.0014259	.0032342	0.44	0.659	-.004913	.0077648
wage	.0120839	.0023262	5.19	0.000	.0075246	.0166432
married	-.0391854	.0201932	-1.94	0.052	-.0787634	.0003925
collgrad	.0468029	.0231306	2.02	0.043	.0014678	.0921381

35 . estimates store a2

36 . probit union age wage married collgrad

Iteration 0:   log likelihood = **-1046.6242**
Iteration 1:   log likelihood = **-1021.6693**
Iteration 2:   log likelihood = **-1021.6284**
Iteration 3:   log likelihood = **-1021.6284**

Probit regression   Number of obs   =   1,878
                                                LR chi2(**4**)   =   49.99
                                                Prob > chi2   =   0.0000
Log likelihood = **-1021.6284**   Pseudo R2   =   0.0239

union	Coef.	Std. Err.	z	P>|z|	[95% Conf. Interval]	
age	.0042593	.0105376	0.40	0.686	-.016394	.0249125
wage	.0417685	.0079746	5.24	0.000	.0261385	.0573984
married	-.1293822	.0663289	-1.95	0.051	-.2593844	.0006201
collgrad	.1542455	.0766703	2.01	0.044	.0039745	.3045165
_cons	-1.142153	.421949	-2.71	0.007	-1.969158	-.3151487

37 . margins, dydx(_all) post

Average marginal effects   Number of obs   =   1,878
Model VCE   : **OIM**

Expression   : **Pr(union), predict()**
dy/dx w.r.t. : **age wage married collgrad**

	dy/dx	Delta-method Std. Err.	z	P>|z|	[95% Conf. Interval]	
age	.0013045	.0032272	0.40	0.686	−.0050206	.0076296
wage	.0127927	.0023905	5.35	0.000	.0081074	.017478
married	−.0396267	.0202639	−1.96	0.051	−.0793432	.0000899
collgrad	.0472417	.0234193	2.02	0.044	.0013407	.0931428

38 . estimates store a3

39 .
40 . esttab a1 a2 a3 , b(3) se(3) star compress nogap s(N r2) mtitles("LPM" "Logit"

	(1) LPM	(2) Logit	(3) Probit
age	0.001	0.001	0.001
	(0.003)	(0.003)	(0.003)
wage	0.014***	0.012***	0.013***
	(0.003)	(0.002)	(0.002)
married	−0.040	−0.039	−0.040
	(0.021)	(0.020)	(0.020)
collgrad	0.049	0.047*	0.047*
	(0.026)	(0.023)	(0.023)
_cons	0.100		
	(0.129)		
N	1878.000	1878.000	1878.000
r2	0.027		

Standard errors in parentheses
* p<0.05, ** p<0.01, *** p<0.001

41 .
42 . *Latent variable simulation

43 . clear

44 . set obs 1000
number of observations (_N) was 0, now 1,000

45 . set seed 12345

46 . gen x = rnormal()

47 . su

Variable	Obs	Mean	Std. Dev.	Min	Max
x	1,000	.0141333	1.020651	-3.355197	3.653764

48 . kdensity x

49 . gen e = rnormal()

50 . gen ystar = 4*x + 1*e

51 . gen y = (ystar>0)

52 . tab y

y	Freq.	Percent	Cum.
0	490	49.00	49.00
1	510	51.00	100.00
Total	1,000	100.00	

53 . probit y x

```
Iteration 0:   log likelihood = -692.94717
Iteration 1:   log likelihood =  -190.1974
Iteration 2:   log likelihood = -187.16339
Iteration 3:   log likelihood = -187.02261
Iteration 4:   log likelihood = -187.02236
Iteration 5:   log likelihood = -187.02236
```

Probit regression

				Number of obs	=	1,000
				LR chi2(1)	=	1011.85
				Prob > chi2	=	0.0000
Log likelihood = -187.02236				Pseudo R2	=	0.7301

y	Coef.	Std. Err.	z	P>|z|	[95% Conf. Interval]	
x	3.855881	.2775006	13.90	0.000	3.311989	4.399772
_cons	.0329217	.0737001	0.45	0.655	-.1115278	.1773713

Note: 61 failures and 56 successes completely determined.

54 .
55 .
56 .