Lab7: Regresstion Tables and Binary Outcomes Models

Introduction to Econometrics,Fall 2020

Yi Wang

Nanjing University

26/11/2020

Section 1

Making Regression Tables in Stata : estout package

Subsection 1

Description

Making Regression Tables in Stata : estout package

- Description
  - The estout package provides tools for making regression tables in Stata. The package currently contains the following commands.
    - ★ esttab : A command for **publication-style regression tables** that display nicely in Stata's results window, and can be exported to various formats such as **CSV**, **RTF**, **HTML**, **LaTeX**.
    - ★ estout : A **generic program for making a table** from one or more sets of estimation results. estout is **the engine behind esttab**.
    - ★ eststo : A utility command to **store estimation results** for later tabulation.
    - ★ estadd : A utility command to **add additional results** to an existing estimation set.
    - ★ estpost : A utility command to **post results** from various **non-eclass commands** as estimation results (so that they can be tabulated).

Making Regression Tables in Stata : estout package

- Installation

```
ssc install estout, replace
help esttab
```

- Reference
  Making regression tables in Stata

Subsection 2

-esttab-

Making Regression Tables in Stata : estout package

- -esttab-
  - *-esttab-* is a wrapper for *-estout-*

```
esttab [ namelist ] [ using filename ] [, options estout_options ]
```

- The procedure :
  - Store a number of models
  - Apply esttab to these stored estimation sets to compose a regression
  - Table produces a fully formatted right away

Making Regression Tables in Stata : estout package

- -esttab-
  - Example

```
. sysuse auto, clear
(1978 Automobile Data)
. qui eststo: reg price weight mpg
. qui eststo: reg price weight mpg foreign
. esttab        //Models stored are automatically picked up by esttab
----------------------------------------------------------------------------
                      (1)             (2)             (3)             (4)
                    price           price           price           price
----------------------------------------------------------------------------
weight              1.747**         3.465***        1.747**         3.465***
                   (2.72)          (5.49)          (2.72)          (5.49)
mpg                -49.51           21.85          -49.51           21.85
                  (-0.57)          (0.29)         (-0.57)          (0.29)
foreign                            3673.1***                       3673.1***
                                   (5.37)                          (5.37)
_cons              1946.1         -5853.7          1946.1         -5853.7
                   (0.54)         (-1.73)          (0.54)         (-1.73)
----------------------------------------------------------------------------
N                      74              74              74              74
----------------------------------------------------------------------------
t statistics in parentheses
* p<0.05, ** p<0.01, *** p<0.001
. eststo clear    //removes the models from memory

* eststo : Store the regression models.
```

Making Regression Tables in Stata : estout package

- -esttab-
  - Or −>

```
. sysuse auto, clear
(1978 Automobile Data)
. qui reg price weight mpg
. est store model1
. qui reg price weight mpg foreign
. est store model2
. esttab model1 model2
```

	(1) price	(2) price
weight	1.747**	3.465***
	(2.72)	(5.49)
mpg	-49.51	21.85
	(-0.57)	(0.29)
foreign		3673.1***
		(5.37)
_cons	1946.1	-5853.7
	(0.54)	(-1.73)
N	74	74

```
t statistics in parentheses
* p<0.05, ** p<0.01, *** p<0.001
. est clear
```

Making Regression Tables in Stata : estout package

- -esttab-
  - Standard errors, p-values, and summary statistics

```
* Default : (t-statistics) and the number of observations in the table footer


. sysuse auto, clear
(1978 Automobile Data)
. qui eststo: reg price weight mpg
. qui eststo: reg price weight mpg foreign
. esttab, se ar2 //replace by standard errors and add the adjusted R-squared
--------------------------------------------
                      (1)             (2)
                    price           price
--------------------------------------------
weight              1.747**         3.465***
                  (0.641)         (0.631)
mpg                -49.51           21.85
                  (86.16)         (74.22)
foreign                            3673.1***
                                  (684.0)
_cons              1946.1         -5853.7
                 (3597.0)        (3377.0)
--------------------------------------------
N                      74              74
adj. R-sq           0.273           0.478
--------------------------------------------
Standard errors in parentheses
* p<0.05, ** p<0.01, *** p<0.001
* t-statistics can also be replaced by : p, ci, aux
```

Making Regression Tables in Stata : estout package

- -esttab-
  - Standard errors, p-values, and summary statistics

```
* Further summary statistics options : pr2, bic, scalars()

. esttab, p scalars(F df_m df_r)
--------------------------------------------
                      (1)             (2)   
                    price           price   
--------------------------------------------
weight              1.747**         3.465***
                  (0.008)         (0.000)   
mpg                -49.51           21.85   
                  (0.567)         (0.769)   
foreign                            3673.1***
                                  (0.000)   
_cons              1946.1         -5853.7   
                  (0.590)         (0.087)   
--------------------------------------------
N                      74              74   
F                   14.74           23.29   
df_m                    2               3   
df_r                   71              70   
--------------------------------------------
p-values in parentheses
* p<0.05, ** p<0.01, *** p<0.001
```

Making Regression Tables in Stata : estout package

- -esttab-
  - Numerical formats

```
* Default :
* t-statistics are printed using 2 decimal places.
* R-squared measures are printed using 3 decimal places.
* point estimates, standard errors using an adaptive display format.(a3)

. esttab, b(a6) p(4) r2(4) nostar
----------------------------------------------------
                      (1)             (2)
                    price           price
----------------------------------------------------
weight           1.746559        3.464706
                 (0.0081)        (0.0000)
mpg             -49.51222        21.85360
                 (0.5673)        (0.7693)
foreign                          3673.060
                                 (0.0000)
_cons            1946.069       -5853.696
                 (0.5902)        (0.0874)
----------------------------------------------------
N                      74              74
R-sq               0.2934          0.4996
----------------------------------------------------
p-values in parentheses

* increase precision for the point estimates
* display p-values and the R-squared using 2 decimal places
```

Making Regression Tables in Stata : estout package

- -esttab-
  - Labels, titles, and notes

```
. esttab, label                                   ///
      title(This is a regression table)           ///
      nonumbers mtitles("Model A" "Model B")      ///
      addnote("Source: auto.dta")
This is a regression table
----------------------------------------------------
                          Model A         Model B
----------------------------------------------------
Weight (lbs.)              1.747**         3.465***
                          (2.72)          (5.49)
Mileage (mpg)             -49.51           21.85
                         (-0.57)          (0.29)
Car type                                  3673.1***
                                          (5.37)
Constant                  1946.1         -5853.7
                          (0.54)         (-1.73)
----------------------------------------------------
Observations                  74              74
----------------------------------------------------
t statistics in parentheses
Source: auto.dta
* p<0.05, ** p<0.01, *** p<0.001

. eststo clear
```

Making Regression Tables in Stata : estout package

- -esttab-
  - Labels, titles, and notes

```
* About factor variables and interactions

. sysuse auto, clear
(1978 Automobile Data)
. qui eststo: reg price mpg i.foreign
. qui eststo: reg price c.mpg##i.foreign
. esttab, varwidth(25) label nobaselevels interaction(" X ")
------------------------------------------------------------
                                   (1)             (2)
                                 Price           Price
------------------------------------------------------------
Mileage (mpg)                  -294.2***       -329.3***
                               (-5.28)         (-4.39)
Foreign                         1767.3*         -13.59
                                (2.52)         (-0.01)
Foreign X Mileage (mpg)                          78.89
                                                (0.70)
Constant                       11905.4***      12600.5***
                               (10.28)          (8.25)
------------------------------------------------------------
Observations                        74              74
------------------------------------------------------------
t statistics in parentheses
* p<0.05, ** p<0.01, *** p<0.001

. eststo clear
```

Making Regression Tables in Stata : estout package

- -esttab-
  - Compressed table

```
. sysuse auto, clear
(1978 Automobile Data)
. qui eststo: reg price weight
. qui eststo: reg price weight mpg
. qui eststo: reg price weight mpg foreign
. qui eststo: reg price weight mpg foreign displacement
. esttab, compress
--------------------------------------------------------------
                  (1)          (2)          (3)          (4)
                price        price        price        price
--------------------------------------------------------------
weight          2.044***     1.747**      3.465***     2.458**
               (5.42)       (2.72)       (5.49)       (2.82)
mpg                         -49.51        21.85        19.08
                           (-0.57)       (0.29)       (0.26)
foreign                                  3673.1***    3930.2***
                                         (5.37)       (5.67)
displace~t                                             10.22

                                                      (1.65)
_cons          -6.707       1946.1      -5853.7      -4846.8
              (-0.01)       (0.54)      (-1.73)      (-1.43)
--------------------------------------------------------------
N                  74           74           74           74
--------------------------------------------------------------
t statistics in parentheses
* p<0.05, ** p<0.01, *** p<0.001
* reduces horizontal spacing to fit more models on screen without line breaking

. eststo clear
```

Making Regression Tables in Stata : estout package

- -esttab-
  - Significance stars

```
. sysuse auto, clear
(1978 Automobile Data)
. qui eststo: reg price weight mpg
. qui eststo: reg price weight mpg foreign
. esttab, star(+ 0.10 * 0.05)
----------------------------------------------
                      (1)             (2)
                    price           price
----------------------------------------------
weight              1.747*          3.465*
                   (2.72)          (5.49)
mpg                -49.51           21.85
                  (-0.57)          (0.29)
foreign                            3673.1*
                                   (5.37)
_cons              1946.1         -5853.7+
                   (0.54)         (-1.73)
----------------------------------------------
N                      74              74
----------------------------------------------
t statistics in parentheses
+ p<0.10, * p<0.05
* default symbols and thresholds are:  * for p<.05, ** for p<.01, and *** for p<.001.
```

Making Regression Tables in Stata : estout package

- -esttab-
  - With Excel

```
esttab using example.csv
esttab using example.csv,replace wide plain
```

  - With Word

```
esttab using example.rtf
esttab using example.rtf, append wide label modelwidth(8)

* varwidth() and modelwidth() change the column widths


lab var mpg "The mgp variable has a really long label"
esttab using example.rtf, replace label nogap onecell

* onecell : placed beneath one another in the same table cell
```

Making Regression Tables in Stata : estout package

- -esttab-
  - With LaTeX

```
. esttab using example.tex, label nostar replace page

* page[(packages)] adds opening and closing code to define a whole LaTeX document
```

Making Regression Tables in Stata : estout package

- -esttab-
  - With LaTeX

	(1)	(2)
	Price	Price
Weight (lbs.)	1.747	3.465
	(2.72)	(5.49)
Mileage (mpg)	-49.51	21.85
	(-0.57)	(0.29)
Car type		3673.1
		(5.37)
Constant	1946.1	-5853.7
	(0.54)	(-1.73)
Observations	74	74

t statistics in parentheses

Making Regression Tables in Stata : estout package

- -esttab-
  - With LaTeX

```
. esttab using example.tex, label nostar replace page booktabs

* produces a LaTeX formatted table for use with LaTeX´s booktabs package
```

Making Regression Tables in Stata : estout package

- -esttab-
  - With LaTeX

	(1) Price	(2) Price
Weight (lbs.)	1.747	3.465
	(2.72)	(5.49)
Mileage (mpg)	-49.51	21.85
	(-0.57)	(0.29)
Car type		3673.1
		(5.37)
Constant	1946.1	-5853.7
	(0.54)	(-1.73)
Observations	74	74

t statistics in parentheses

Making Regression Tables in Stata : estout package

- -esttab-
  - With LaTeX

```
. esttab using example.tex, label nostar replace page booktabs ///
    width(0.8\hsize)


* width(\hsize) in LaTeX or width(100%) in HTML to span the whole page
```

Making Regression Tables in Stata : estout package

- -esttab-
  - With LaTeX

	(1) Price	(2) Price
Weight (lbs.)	1.747	3.465
	(2.72)	(5.49)
Mileage (mpg)	-49.51	21.85
	(-0.57)	(0.29)
Car type		3673.1
		(5.37)
Constant	1946.1	-5853.7
	(0.54)	(-1.73)
Observations	74	74

t statistics in parentheses

Making Regression Tables in Stata : estout package

- -esttab-
  - With LaTeX

```
. esttab using example.tex, label nostar replace page booktabs ///
    width(0.8\hsize) alignment(ll)


* specify the alignment of the models´ columns in LaTeX
```

Making Regression Tables in Stata : estout package

- -esttab-
  - With LaTeX

	(1) Price	(2) Price
Weight (lbs.)	1.747	3.465
	(2.72)	(5.49)
Mileage (mpg)	-49.51	21.85
	(-0.57)	(0.29)
Car type		3673.1
		(5.37)
Constant	1946.1	-5853.7
	(0.54)	(-1.73)
Observations	74	74

t statistics in parentheses

Making Regression Tables in Stata : estout package

- -esttab-
  - With LaTeX

```
. esttab using example.tex, label nostar replace page booktabs ///
    width(0.8\hsize) alignment(ll) title(Regression table)
```

Making Regression Tables in Stata : estout package

- -esttab-
  - With LaTeX

表 1: Regression table

	(1) Price	(2) Price
Weight (lbs.)	1.747	3.465
	(2.72)	(5.49)
Mileage (mpg)	-49.51	21.85
	(-0.57)	(0.29)
Car type		3673.1
		(5.37)
Constant	1946.1	-5853.7
	(0.54)	(-1.73)
Observations	74	74

t statistics in parentheses

- -esttab-
  - With LaTeX–分组样本回归mgroup()选项

```
sysuse auto, clear
eststo clear
eststo: qui reg weight mpg
eststo: qui reg weight mpg foreign
eststo: qui reg price weight mpg
eststo: qui reg price weight mpg foreign
esttab using mgroups.tex, replace                       ///
         star(* 0.1 ** 0.05 *** 0.01)                   ///
         compress nogaps                                ///
         title(An Illustration of mgroup() in esttab)   ///
         mgroups("Group A" "Group B",                   ///
             pattern(1 0 1 0) span                      ///
             prefix(\multicolumn{@span}{c}{) suffix(})  ///
             erepeat(\cmidrule(lr){@span}) )            ///
         booktabs page(dcolumn) alignment(D{.}{.}{-1})
```

Making Regression Tables in Stata : estout package

- -esttab-
  - With LaTeX–分组样本回归mgroup()选项

```
* "Group A" "Group B" : 组别名称
* mgroup()子选项 :
* pattern(1 0 1 0) : 1--该组别的第一个模型, 0--该组别的其他模型
* span : 定组别名能在表格中跨列
* prefix(\multicolumn{@span}{c}{)和suffix()}):组别名在LaTeX代码中跨行
* erepeat(\cmidrule(lr){@span}) : 设定跨行代码,下面加底部表格线
* page(dcolumn) : 添加加载宏包dcolumn
* alignment(D{.}{.}{-1}) : 调整单元格对齐方式,小数点对齐

ssc install texify
texify mgroups.tex
```

Making Regression Tables in Stata : estout package

- -esttab-
  - With LaTeX–分组样本回归mgroup()选项

Table 1: An Illustration of mgroup() in esttab

	Group A		Group B	
	(1) weight	(2) weight	(3) price	(4) price
mpg	-108.4^{***}	-91.22^{***}	-49.51	21.85
	(-11.60)	(-10.34)	(-0.57)	(0.29)
foreign		-550.1^{***}		3673.1^{***}
		(-4.96)		(5.37)
weight			1.747^{***}	3.465^{***}
			(2.72)	(5.49)
_cons	5328.8^{***}	5125.7^{***}	1946.1	-5853.7^{*}
	(25.85)	(27.93)	(0.54)	(-1.73)
N	74	74	74	74

t statistics in parentheses

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Making Regression Tables in Stata : estout package

- -esttab-
  - Non-standard contents
    - ⋆ Sometimes it is necessary to include parameter statistics in a table
    - ⋆ main() option : replacing the point-estimates
    - ⋆ aux() option : replacing the t-statistics

Making Regression Tables in Stata : estout package

- -esttab-
  - Non-standard contents
    - For example, to include VIF(variance inflation factors) instead of t-statistics after reg

```
. sysuse auto, clear
. eststo clear
. reg price weight mpg foreign
. estadd vif

    Variable |       VIF       1/VIF
-------------+----------------------
      weight |      3.86    0.258809
         mpg |      2.96    0.337297
     foreign |      1.59    0.627761
-------------+----------------------
    Mean VIF |      2.81
added matrix:
                e(vif) :  1 x 4
```

- -esttab-
  - Non-standard contents

```
. esttab, aux(vif 2) wide nopar
----------------------------------------
                      (1)
                    price
----------------------------------------
weight              3.465***        3.86
mpg                 21.85           2.96
foreign            3673.1***        1.59
_cons             -5853.7
----------------------------------------
N                      74
----------------------------------------
vif in second column
* p<0.05, ** p<0.01, *** p<0.001
```

Making Regression Tables in Stata : estout package

- -esttab-
  - Non-standard contents
    - More than two kinds of parameter statistics
    - Switch to **estout syntax** and make use of the cells() option
    - **cells()** disables b(), beta(), main(), t(), abs, se(), p(), ci(). . .

Making Regression Tables in Stata : estout package

- -esttab-
  - Non-standard contents
    - For example,to print point estimates, t-statistics, and variance inflation factors in one table

```
. esttab, cells("b(fmt(a3) star) vif(fmt(2))" t(par fmt(2)))
----------------------------------------------
                      (1)
                    price
                      b/t              vif
----------------------------------------------
weight              3.465***          3.86
                   (5.49)
mpg                 21.85             2.96
                   (0.29)
foreign            3673.1***          1.59
                   (5.37)
_cons             -5853.7
                  (-1.73)
----------------------------------------------
N                      74
----------------------------------------------
```

Making Regression Tables in Stata : estout package

- -esttab-
  - Non-standard contents
    - complicated summary statistics section in the table footer
    - r2, ar2, pr2, aic, bic, scalars()...
    - estout's stats() option, equivalently

Making Regression Tables in Stata : estout package

- -esttab-
  - Non-standard contents
  - Use esttab to assemble a basic table and then **hand-edit and re-run the estout call**

Making Regression Tables in Stata : estout package

- -esttab-
  - Non-standard contents

```
. sysuse auto, clear
. eststo: qui reg price weight mpg
. eststo: qui reg price weight mpg foreign

. esttab, noisily notype
estout ,
 cells(b(fmt(a3) star) t(fmt(2) par("{ralign @modelwidth:{txt:(}" "{txt:)}}")))
 stats(N, fmt(%18.0g) labels(`"N"'))
 starlevels(* 0.05 ** 0.01 *** 0.001)
 varwidth(12)
 modelwidth(12)
 abbrev
 delimiter(" ")
 smcltags
 prehead(`"{hline @width}"')
 posthead("{hline @width}")
 prefoot("{hline @width}")
 postfoot(`"{hline @width}"' `"t statistics in parentheses"' `"@starlegend"')
 varlabels(, end("" "") nolast)
 mlabels(, depvar)
 numbers
 collabels(none)
 eqlabels(, begin("{hline @width}" "") nofirst)
 interaction(" # ")
 notype
 level(95)
 style(esttab)
```

Making Regression Tables in Stata : estout package

- -esttab-
  - Non-standard contents

```
return list
dis r(cmdline)
```

Subsection 3

-estout-

Making Regression Tables in Stata : estout package

- -estout-
  - The full syntax of estout is rather complex
  - The most important options:

```
estout [ namelist ] [ using filename ] [, cells(array)
                      stats(scalarlist) style(style) options ]
```

  - cells() and stats() options : determine the primary contents of the table
  - style() option : determines the basic formatting of the table.

Making Regression Tables in Stata : estout package

- -estout-
  - Choosing an output format

```
. sysuse auto, clear
. eststo: qui reg price weight mpg
. eststo: qui reg price weight mpg foreign
. estout, style(tex)
            &        est1&        est2&        est3&        est4\\
            &           b&           b&           b&           b\\
weight      &    1.746559&    3.464706&    1.746559&    3.464706\\
mpg         &   -49.51222&     21.8536&   -49.51222&     21.8536\\
foreign     &            &     3673.06&            &     3673.06\\
_cons       &    1946.069&   -5853.696&    1946.069&   -5853.696\\
```

Making Regression Tables in Stata : estout package

- -estout-
  - The cells option
  - The stats option

```
. eststo clear
. estout, stats(r2 bic N)
------------------------------------
                                .
                                b
------------------------------------
weight                   3.464706
mpg                       21.8536
foreign                   3673.06
_cons                   -5853.696
------------------------------------
r2                       .4995594
bic                      1357.414
N                              74
------------------------------------
```

Making Regression Tables in Stata : estout package

- -estout-
  - Using labels

```
. sysuse auto, clear
. eststo clear
. eststo, title("Model 1"): qui reg price weight mpg
. eststo, title("Model 2"): qui reg price weight mpg foreign
. label variable foreign "Car type (1=foreign)"
. estout, cells("b(star label(Coef.)) se(label(Std. err.))")  ///
     stats(r2 N, labels(R-squared "N. of cases"))            ///
     label legend varlabels(_cons Constant)
------------------------------------------------------------------------------
                         Model 1                        Model 2
                           Coef.       Std. err.          Coef.       Std. err.
------------------------------------------------------------------------------
Weight (lbs.)           1.746559**      .6413538       3.464706***     .630749
Mileage (mpg)          -49.51222        86.15604        21.8536       74.22114
Car type (1=foreign)                                    3673.06***    683.9783
Constant                1946.069         3597.05      -5853.696       3376.987
------------------------------------------------------------------------------
R-squared               .2933891                       .4995594
N. of cases                   74                             74
------------------------------------------------------------------------------
* p<0.05, ** p<0.01, *** p<0.001
```

Subsection 4

-eststo-

Making Regression Tables in Stata : estout package

- -eststo-
  - Stores a copy of the active estimation results for later tabulation.
  - Analogous to official Stata's **estimates store**.
  - Does not require the user to specify a name for the stored estimates.

```
sysuse auto, clear
reg price weight mpg
eststo
reg price weight mpg foreign
eststo
esttab
```

  - As a prefix

```
sysuse auto, clear
eststo: qui reg price weight mpg
eststo: qui reg price weight mpg foreign
esttab
```

Making Regression Tables in Stata : estout package

- -eststo-
  - Using by

```
. sysuse auto, clear
. eststo clear
. by foreign : eststo: qui reg price weight mpg
------------------------------------------------------------------------------
-> Domestic
(est1 stored)
------------------------------------------------------------------------------
-> Foreign
(est2 stored)
. esttab, label nodepvar nonumber
----------------------------------------------------
                          Domestic         Foreign
----------------------------------------------------
Weight (lbs.)               4.415***        5.156***
                           (4.66)          (5.85)
Mileage (mpg)               237.7          -19.78
                           (1.71)         (-0.34)
Constant                 -13285.4*        -5065.8
                          (-2.32)         (-1.58)
----------------------------------------------------
Observations                   52              22
----------------------------------------------------
t statistics in parentheses
* p<0.05, ** p<0.01, *** p<0.001
```

Making Regression Tables in Stata : estout package

- -eststo-
  - Adding additional statistics

```
. sysuse auto, clear
. eststo clear
. qui reg price weight mpg
. test weight = mpg
 ( 1)  weight - mpg = 0
       F(  1,    71) =    0.36
            Prob > F =    0.5514
. eststo, addscalars(p_diff r(p))
(e(p_diff) = .55138216 added)
(est1 stored)
. esttab, scalars(p_diff) obslast
----------------------------
                      (1)
                    price
----------------------------
weight              1.747**
                   (2.72)
mpg                -49.51
                  (-0.57)
_cons              1946.1
                   (0.54)
----------------------------
p_diff              0.551
N                      74
----------------------------
t statistics in parentheses
* p<0.05, ** p<0.01, *** p<0.001
```

Subsection 5

-estadd-

Making Regression Tables in Stata : estout package

- -estadd-
  - Add results to (stored) estimates
  - Results that are included in the e()-returns for the models can be tabulated by estout or esttab
  - 举例：向内存中添加两个统计量：
    - 一个是文字类型的返回值Industry，采用暂元(local)存储
    - 一个是数值类型的返回值Mean_Wage，采用单值(scalar)来存储

Making Regression Tables in Stata : estout package

- -estadd-

```
. eststo clear
. sysuse nlsw88.dta, clear
(NLSW, 1988 extract)
. qui reg wage ttl_exp married
. estadd  local Industry "Yes"
added macro:
           e(Industry) : "Yes"
```

```
. qui sum wage
. estadd  scalar Mean_Wage = r(mean)
added scalar:
          e(Mean_Wage) =  7.766949


. qui ereturn list
```

Making Regression Tables in Stata : estout package

- -estadd-

```
sysuse nlsw88.dta, clear

global xx "ttl_exp married south hours tenure age i.industry"

qui reg wage $xx if race==1
estadd local Industry   "Yes"
estadd local Occupation "No"
est store m1

qui reg wage $xx if race==2
estadd local Industry   "Yes"
estadd local Occupation "No"
est store m2

qui reg wage $xx i.occupation if race==1
estadd local Industry   "Yes"
estadd local Occupation "Yes"
est store m3

qui reg wage $xx i.occupation if race==2
estadd local Industry   "Yes"
estadd local Occupation "Yes"
est store m4
```

Making Regression Tables in Stata : estout package

- -estadd-

```
local m "m1 m2 m3 m4"
esttab `m´, mtitle(White Black White Black) b(%6.3f) nogap compress  ///
            star(* 0.1 ** 0.05 *** 0.01)                              ///
            drop(*.industry *.occupation)                             ///
            ar2 scalar(N Industry Occupation)
--------------------------------------------------------------
                  (1)          (2)          (3)          (4)
                White        Black        White        Black
--------------------------------------------------------------
ttl_exp         0.251***     0.271***     0.176***     0.193***
               (6.49)       (4.82)       (4.61)       (3.61)
married        -0.737**      0.082       -0.657**      0.099
              (-2.31)       (0.21)      (-2.12)       (0.27)
south          -0.814***    -2.038***    -0.758***    -1.791***
              (-2.72)      (-4.92)      (-2.61)      (-4.56)
hours           0.051***     0.036        0.021        0.007
               (3.81)       (1.35)       (1.56)       (0.25)
tenure          0.025       -0.003        0.039       -0.014
               (0.76)      (-0.08)       (1.25)      (-0.34)
age            -0.073       -0.216***    -0.058       -0.148**
              (-1.58)      (-3.12)      (-1.31)      (-2.23)
_cons           5.576**      9.490**      9.607***     8.420**
               (2.22)       (2.53)       (3.82)       (2.39)
--------------------------------------------------------------
N                1615          572         1612          571
adj. R-sq       0.112        0.166        0.176        0.281
Industry          Yes          Yes          Yes          Yes
Occupation         No           No          Yes          Yes
--------------------------------------------------------------
t statistics in parentheses
* p<0.1, ** p<0.05, *** p<0.01
```

Subsection 6

-estpost-

Making Regression Tables in Stata : estout package

- -estpost-
  - esttab and estout tabulate the e()-returns of a command, but not all commands return their results in e()
  - Posting results from **non-eclass** commands
  - Review

- 结果导出-esttab-

```
sysuse auto,clear

local var price wei len mpg
qui estpost ttest `var´, by(foreign)
esttab using ttable2.rtf, cell("mu_1(fmt(2)) mu_2(fmt(2)) b(star fmt(2)) t(fmt(2))")  ///
        starlevels(* 0.10 ** 0.05 *** 0.01) replace noobs compress                  ///
        title(esttab_Table: T_test)
```

- 描述性统计表格导出
  - -esttab-

```
estpost summarize price wei len mpg rep78, detail
esttab using Desc4.rtf,                                                         ///
    cells("count mean(fmt(2)) sd(fmt(2)) min(fmt(2)) p50(fmt(2)) max(fmt(2))") ///
    noobs compress replace title(esttab_Table: Descriptive statistics)
```

- -estpost-
  - Review
- 相关系数矩阵导出

```
*<方法一> -logout-
logout, save(Corr1) word replace: pwcorr price wei len mpg rep78, star(.05)


*<方法二> -esttab-
estpost correlate  price wei len mpg rep78, matrix
esttab using Corr2.rtf,                                                ///
    unstack not noobs compress nogaps replace star(* 0.1 ** 0.05 *** 0.01) ///
    b(%8.3f) p(%8.3f) title(esttab_Table: correlation coefficient matrix)


*<方法三> -corr2docx-
corr2docx price wei len mpg rep78 using Corr3.docx,                    ///
        replace spearman(ignore) pearson(pw) star                      ///
        title(corr2docx_Table: correlation coefficient matrix)
```

Section 2

Nonlinear Regression (continued)

Subsection 1

Dummy Variables

- Dummy Variables

```
. use Nations2.dta,clear
(UN Human Development Indicators)
. des
Contains data from Nations2.dta
  obs:           194                         UN Human Development Indicators
 vars:            13                         2 Jul 2012 06:11
-----------------------------------------------------------------------------------
              storage   display    value
variable name   type    format     label      variable label
-----------------------------------------------------------------------------------
country         str21   %21s                  Country
region          byte    %8.0g      region     Region
gdp             float   %9.0g                 Gross domestic product per cap 2005$, 20
school          float   %9.0g                 Mean years schooling (adults) 2005/2010
adfert          float   %8.0g                 Adolescent fertility: births/1000 fem 15-
                                                2010
chldmort        float   %9.0g                 Prob dying before age 5/1000 live births
                                                2005/2009
life            float   %9.0g                 Life expectancy at birth 2005/2010
pop             float   %9.0g                 Population 2005/2010
urban           float   %9.0g                 Percent population urban 2005/2010
femlab          float   %9.0g                 Female/male ratio in labor force 2005/20
literacy        float   %9.0g                 Adult literacy rate 2005/2009
co2             float   %9.0g                 Tons of CO2 emitted per cap 2005/2006
gini            float   %9.0g                 Gini coef income inequality 2005/2009
-----------------------------------------------------------------------------------
Sorted by: region  country
```

Nonlinear Regression (continued)

- Dummy Variables

```
. tab region,gen(reg)  //为每一类创建虚拟变量
     Region |      Freq.     Percent        Cum.
------------+-----------------------------------
     Africa |         52       26.80       26.80
   Americas |         35       18.04       44.85
       Asia |         49       25.26       70.10
     Europe |         43       22.16       92.27
    Oceania |         15        7.73      100.00
------------+-----------------------------------
      Total |        194      100.00
. gen loggdp=ln(gdp)
(15 missing values generated)


. label values reg1 reg1
. label define reg1 0 "others" 1 "Africa" //reg1 = 1 (Africa); = 0 (elsewhere)
```

Nonlinear Regression (continued)

- Dummy Variables

```
. eststo : qui reg life reg1
. eststo : qui reg life loggdp chldmort
. eststo : qui reg life reg1 reg2 reg3 reg4 loggdp chldmort
. eststo : qui reg life reg1 reg2 reg4 loggdp chldmort
. eststo : qui reg life reg1 loggdp chldmort


. esttab using e2.tex, replace nonumbers           ///
      title(Dummy_Var_Example Table)               ///
      mtitles(m1 m2 m3 m4)                         ///
      star( * 0.10 ** 0.05 *** 0.01 ) compress     ///
      b(%6.3f) t(%6.3f) r2(%9.3f)                  ///
      booktabs page
(output written to e2.tex)
```

Nonlinear Regression (continued)

- Dummy Variables

Table 1: Dummy_Var_Example Table

	m1	m2	m3	m4	est5
reg1	-16.721***		-2.151*	-2.927***	-3.144***
	(-15.175)		(-1.775)	(-3.570)	(-3.979)
loggdp		1.525***	1.407***	1.422***	1.483***
		(4.834)	(4.412)	(4.468)	(4.894)
chldmort		-0.146***	-0.127***	-0.127***	-0.128***
		(-19.041)	(-14.605)	(-14.610)	(-14.781)
reg2			1.487	0.692	
			(1.264)	(0.933)	
reg3			0.984		
			(0.871)		
reg4			1.455	0.649	
			(1.213)	(0.852)	
_cons	73.211***	62.286***	62.162***	62.821***	62.657***
	(128.333)	(20.464)	(20.043)	(20.901)	(21.430)
N	194	178	178	178	178
R^2	0.545	0.880	0.891	0.891	0.890

t statistics in parentheses

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Nonlinear Regression (continued)

- Dummy Variables

```
. predict lifehat
. graph twoway scatter life  reg1, msymbol(oh S) jitter(5)           ///
      || lfit life reg1                                               ///
      || , legend(off) xlabel(0 "others" 1 "Africa")  scheme(s2mono)  ///
      xtitle("Global region") ytitle("Life expectancy in years")
. cap graph export nat1.png, replace
```

Nonlinear Regression (continued)

- Dummy Variables
  - Explanation:
    - m1 : The difference between the two means equals –16.72 years.
    - m2 : **loggpd** and **chldmort**(child mortality rate) together explain about 88% of the variance in life expectancy(m2). Including four dummy variables for regions 1–4 raises this only to about 89% (m3).
    - m3 : It is not possible to include all five in one regression because of **multicollinearity**. None of the regional dummy variables have significant effects.
    - m4 : Dropping reg3, the weakest of these predictors.The coefficient on reg1 now appears significant.
    - m5 : A reduced model.
    - Conclusion : The differences in life expectancy among **other regions** of the world are largely accounted for by variations in **wealth** and **child mortality**, but in **Africa** there are circumstances at work (such as **wars**) that further depress life expectancy.

Subsection 2

Interaction Effects

Nonlinear Regression (continued)

- Interaction Effects
  - consider some different variables:
    - per capita carbon dioxide emissions (co2)
    - percent of the population living in urban areas (urban)
    - dummy variable reg4 defined as 1 for European countries and 0 for all others
    - form an interaction term named urb_reg4 by multiplying the dummy variable reg4 times the measurement variable urban

Nonlinear Regression (continued)

- Interaction Effects

```
. label val reg4 reg4
. label define reg4 0 "others" 1 "Europe"
. gen logco2 = log10(co2)
(9 missing values generated)
. label var logco2 "log10(per cap CO2)"


. gen urb_reg4 = urban * reg4
. label variable urb_reg4 "interaction urban*reg4 (Europe)"
```

Nonlinear Regression (continued)

- Interaction Effects

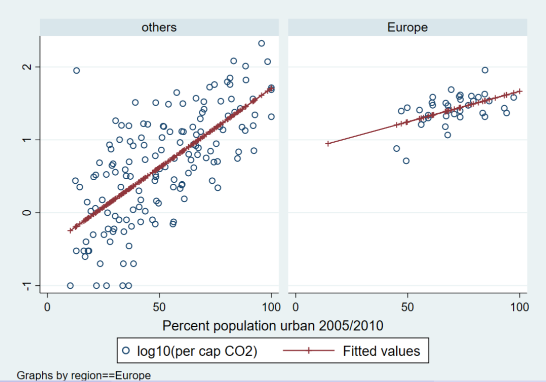
```
. qui reg logco2 urban reg4 urb_reg4
. qui reg logco2 c.urban i.reg4 c.urban#i.reg4
. qui reg logco2 c.urban##i.reg4


 * factor-variable :
 * i. indicator variables
 * c. continuous variables
 * #  an interaction between two variables
 * ## factorial interaction which automatically includes all the lower-level
interactions involving those variables
```

Nonlinear Regression (continued)

- Interaction Effects

```
. predict co2hat
(option xb assumed; fitted values)
. graph twoway scatter logco2 urban, msymbol(Oh)    ///
      || connect co2hat urban, msymbol(+)           ///
      || , by(reg4)
. cap graph export nat2.png, replace
```



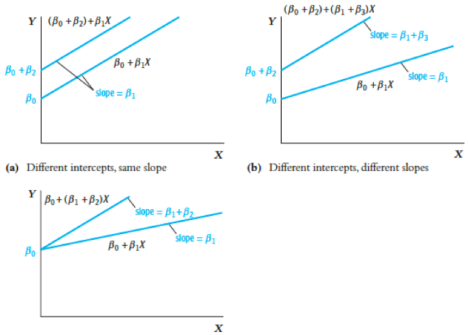
Nonlinear Regression (continued)

- Interaction Effects
  - Explanation:

Interactions Between Independent Variables

Interactions: a Continuous and a Binary Variable

FIGURE 8.8 Regression Functions Using Binary and Continuous Variables



(a) Different intercepts, same slope

(b) Different intercepts, different slopes

(c) Same intercept, different slopes

Interactions of binary variables and continuous variables can produce three different population regression functions: (a) $\beta_0 + \beta_1 X + \beta_2 D$ allows for different intercepts but has the same slope, (b) $\beta_0 + \beta_1 X + \beta_2 D + \beta_3(X \times D)$ allows

Nonlinear Regression (continued)

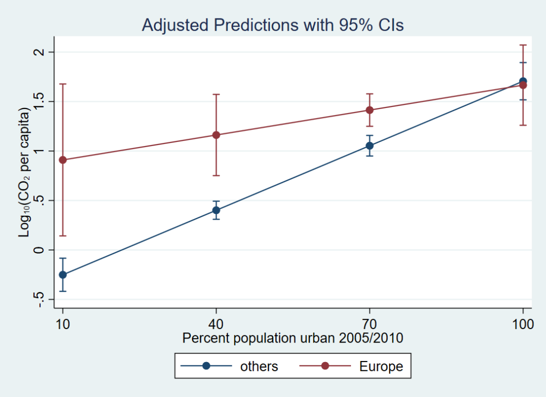
- Interaction Effects
  - Explanation:
    - ★ The line in the left-hand (reg4 $= 0$) panel has a slope of .0217 and y-intercept $-.4682$.
    - ★ The line in the right panel (reg4 $= 1$) has a less-steep slope (.0084) and a higher y-intercept (.826).
    - ★ No European countries exhibit the low-urbanization, low-CO2 profile seen in other parts of the world.
    - ★ Even European nations with middling urbanization have relatively high CO2 emissions.

Nonlinear Regression (continued)

- Interaction Effects

```
. qui margins, at(urban = (10(30)100) reg4 = (0 1)) vsquish
. marginsplot, ytitle("Log{subscript:10}(CO{subscript:2} per capita)") xlabel(10(30)10(

. cap graph export nat3.png, replace
```



Nonlinear Regression (continued)

- Interaction Effects
  - Interactions Between Two Continuous Variables
  - **centering** makes their main effects easier to interpret
  - Interacting variables have been centered, can be interpreted as the effect of each variable when the other is at its mean.

```
. summarize urban loggdp
    Variable |        Obs        Mean    Std. Dev.       Min        Max
-------------+---------------------------------------------------------
       urban |        194    55.43488     23.4391      10.25        100
      loggdp |        179    8.693936    1.297024   5.634075   11.22399
```

Nonlinear Regression (continued)

- Interaction Effects

```
. sum urban
    Variable |        Obs        Mean    Std. Dev.       Min        Max
-------------+---------------------------------------------------------
       urban |        194    55.43488     23.4391      10.25        100
. gen urban0 = urban - r(mean)


. sum loggdp
    Variable |        Obs        Mean    Std. Dev.       Min        Max
-------------+---------------------------------------------------------
      loggdp |        179    8.693936    1.297024   5.634075   11.22399
. gen loggdp0 = loggdp - r(mean)
(15 missing values generated)


. gen urb_gdp = urban0 * loggdp0
(15 missing values generated)
```

Nonlinear Regression (continued)

- Interaction Effects

```
. reg logco2 c.loggdp0 c.urban0 c.loggdp0#c.urban0

      Source |       SS           df       MS      Number of obs   =       175
-------------+----------------------------------   F(3, 171)       =    371.66
       Model |  83.4990751         3  27.833025    Prob > F        =    0.0000
    Residual |  12.8060512       171  .074889188   R-squared       =    0.8670
-------------+----------------------------------   Adj R-squared   =    0.8647
       Total |  96.3051263       174  .553477737   Root MSE        =    .27366

-----------------------------------------------------------------------------------
           logco2 |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
------------------+----------------------------------------------------------------
          loggdp0 |   .4848767   .0242429    20.00   0.000     .4370228    .5327306
           urban0 |   .0025141     .00137     1.84   0.068    -.0001903    .0052185
                  |
c.loggdp0#c.urban0 |  -.0035963   .0007565    -4.75   0.000    -.0050895   -.0021031
                  |
            _cons |   .8903587   .0267759    33.25   0.000     .8375049    .9432124
-----------------------------------------------------------------------------------
```

Nonlinear Regression (continued)

- Interaction Effects
  - Explanation :

Interactions Between Independent Variables

Interactions Between Two Continuous Variables

- Thus the effect on Y of a change in X_1, holding X_2 constant, is

$$\frac{\Delta Y}{\Delta X_1} = \beta_1 + \beta_3 X_2$$

- A similar calculation shows that the effect on Y of a change ΔX_1 in X_2, holding X_1 constant, is

$$\frac{\Delta Y}{\Delta X_2} = \beta_2 + \beta_3 X_1$$

- That is, if X_1 changes by ΔX_1 and X_2 changes by ΔX_2, then the expected change in Y

$$\Delta Y = (\beta_1 + \beta_3 X_2)\Delta X_1 + (\beta_2 + \beta_3 X_1)\Delta X_2 + \beta_3 \Delta X_1 \Delta X_2$$

Nonlinear Regression (continued)

- Interaction Effects
  - Explanation :
    - ★ predicted logco2 rises by 0.48 with each 1-unit increase in loggdp, when urban is at its mean
    - ★ predicted logco2 rises by only a small amount, .0025, with each 1-unit increase in urban when loggdp is at its mean.
    - ★ each 1-unit increase in urbanization, the effect of loggdp on logco2 becomes weaker, decreasing by $-.004$.
    - ★ 二氧化碳排放量随着财富的增加而增加，但在城市化程度较高的国家，二氧化碳排放量的增加幅度较小。

Section 3

Binary Outcomes Models

Subsection 1

Review the Theroy

Binary Outcomes Models

- The Linear Probability Model(LPM)
  - Review

The Linear Probability Model

- The conditional expectation equals the probability that $Y_i = 1$ conditional on $X_{1i}, ..., X_{ki}$

$$E[Y|X_{1i}, ..., X_{ki}] = Pr(Y = 1|X_{1i}, ..., X_{ki})$$
$$= \beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + ... + \beta_k X_{ki}$$

- Now a **Linear Probability Model** can be defined as following

$$Pr(Y = 1|X_{1i}, ..., X_{ki}) = \beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + ... + \beta_k X_{ki}$$

- The **population coefficient** β_j

$$\frac{\partial Pr(Y_i = 1|X_{1i}, ..., X_{ki})}{\partial X_j} = \beta_j$$

- β_j can be explained as *the change in the probability that* $Y = 1$ *associated with a unit change in* X_j

Binary Outcomes Models

- The Linear Probability Model(LPM)
  - Review

The Linear Probability Model(LPM)

Warp up

- Disadvantages of the linear probability model
  - Predicted probability can be above 1 or below 0!(it doesn't make sense)
  - Error terms are heteroskedastic.
- Advantages of the linear probability model:
  - Easy to estimate and inference
  - Coefficient estimates are easy to interpret
  - very useful in some circumstances: Special IV

Binary Outcomes Models

- Probit Model
  - Review

Probit Model

- Probit regression models the probability that $Y = 1$

$$Pr(Y_i = 1|X_1, ...X_k) = \Phi(\beta_0 + \beta_1 X_{1,i} + \beta_2 X_{2,i} + ... + \beta_k X_{k,i})$$

- where $\Phi(Z)$ is the **standard normal** c.d.f, then we have

$$0 \leq \Phi(Z) \leq 1$$

- Then it make sure that the **predicted probabilities** of the probit model are between 0 and 1.

- Logit Model
  - Review

Logit Model

- Using the cumulative standard **logistic** distribution function

$$Pr(Y_i = 1|Z) = \frac{1}{1 + e^{-Z}}$$

- Similar to probit model $Z = \beta_0 + \beta_1 X_{1,i} + \beta_2 X_{2,i} + ... + \beta_k X_{k,i}$
- Since $F(z) = Pr(Z \leq z)$ we have that the predicted probabilities of the logit model are between 0 and 1.

Subsection 2

In practice

Binary Outcomes Models

- In practice
  - Syntax

```
probit y x1 x2 x3,r
logit y x1 x2 x3,or vce(cluster clustvar)

* or--report odds ratios（不显示系数）
* vce()--clustvar为聚类变量的聚类稳健标准误
```

  - 估计完成后进行预测，并计算准确预测的百分比：

```
predict yhat  //计算发生概率的预测值
estat clas    //计算预测准确的百分比
```

- In practice
  - Marginal Effects

Nonlinear probability model

Effect of a change in X: When X is continous

- **Marginal Effects** for X_j
- $$\frac{\partial Pr(Y=1|X_1,...X_k)}{\partial X_j} = \phi(\beta_0+\beta_1 X_{1,i}+\beta_2 X_{2,i}+...+\beta_k X_{k,i})\times\beta_j$$
- Where $\phi()$ is the p.d.f of standard normal.
- Hence, the effect of a change in X depends on the starting value of X like other nonlinear functions.

Binary Outcomes Models

- In practice
  - Marginal Effects

Nonlinear probability model

Effect of a change in X:

- For nonlinear models, the ME varies with the point of evaluation
  - *Marginal Effect at a Representative Value* (MER):ME at $X = X^*$ (at representative values of the regressors)
  - *Marginal Effect at Mean* (MEM): ME at $X = \bar{X}$(at the sample mean of the regressors)
  - *Average Marginal Effect* (AME): average of ME at each $X = X_i$ (at sample values and then average)

Binary Outcomes Models

- In practice
  - Marginal Effects

```
margins,dydx(*)              //计算所有解释变量的平均边际效应
margins,dydx(*) atmeans      //计算所有解释变量样本均值处边际效应
margins,dysx(*) at(x1=0)     //计算所有解释变量在x1=0处边际效应
margins,dydx(x1)             //计算解释变量x1的平均边际效应
margins,eyex(*)              //计算平均弹性
margins,eydx(*)              //计算平均半弹性，x变化1单位使y变化百分之几
margins,dyex(*)              //计算平均半弹性，x变化1%使y变化几个单位
```

Subsection 3

Example

- Example

```
. use womenwk,clear

. reg work age married children education,r

Linear regression                               Number of obs     =      2,000
                                                F(4, 1995)        =     192.58
                                                Prob > F          =     0.0000
                                                R-squared         =     0.2026
                                                Root MSE          =     .41992

------------------------------------------------------------------------------
             |               Robust
        work |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
         age |   .0102552   .0012236     8.38   0.000     .0078556    .0126548
     married |   .1111116   .0226719     4.90   0.000     .0666485    .1555748
    children |   .1153084   .0056978    20.24   0.000     .1041342    .1264827
   education |   .0186011   .0033006     5.64   0.000     .0121282     .025074
       _cons |  -.2073227   .0534581    -3.88   0.000    -.3121622   -.1024832
------------------------------------------------------------------------------
```

- Example

```
. logit work age married children education,nolog    //不显示MLE数值计算的迭代过程
Logistic regression                                 Number of obs     =      2,000
                                                    LR chi2(4)        =     476.62
                                                    Prob > chi2       =     0.0000
Log likelihood = -1027.9144                         Pseudo R2         =     0.1882

------------------------------------------------------------------------------
        work |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
         age |   .0579303    .007221     8.02   0.000     .0437773    .0720833
     married |   .7417775   .1264705     5.87   0.000     .4938998    .9896552
    children |   .7644882   .0515289    14.84   0.000     .6634935     .865483
   education |   .0982513   .0186522     5.27   0.000     .0616936     .134809
       _cons |  -4.159247   .3320401   -12.53   0.000    -4.810034   -3.508461
------------------------------------------------------------------------------
```

Binary Outcomes Models

- Example

```
. logit work age married children education,r or nolog
//为了便于解释回归结果，让Stata 汇报odds ratios（不显示系数）

Logistic regression                             Number of obs     =      2,000
                                                Wald chi2(4)      =     344.54
                                                Prob > chi2       =     0.0000
Log pseudolikelihood = -1027.9144               Pseudo R2         =     0.1882

------------------------------------------------------------------------------
             |               Robust
        work | Odds Ratio   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
         age |   1.059641   .0076352     8.04   0.000     1.044782    1.074712
     married |   2.099664   .2671174     5.83   0.000     1.636292    2.694257
    children |   2.147895   .1068758    15.36   0.000     1.948312    2.367922
   education |    1.10324   .0209737     5.17   0.000     1.062889    1.145123
       _cons |   .0156193   .0051137   -12.70   0.000     .0082221    .0296718
------------------------------------------------------------------------------
Note: _cons estimates baseline odds.

 * LR 统计量为476.62，对应P 值为0.00，故整个方程所有系数（除常数项）联合显著性很高
```

- Example

```
. margins,dydx(*)          //平均边际效应

Average marginal effects                        Number of obs     =      2,000
Model VCE    : Robust

Expression   : Pr(work), predict()
dy/dx w.r.t. : age married children education

------------------------------------------------------------------------------
             |            Delta-method
             |      dy/dx   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
         age |   .0099674   .0011752     8.48   0.000      .007664    .0122708
     married |    .127629   .0213105     5.99   0.000     .0858612    .1693967
    children |   .1315365   .0068734    19.14   0.000     .1180649    .1450082
   education |   .0169049   .0031833     5.31   0.000     .0106658    .0231441
------------------------------------------------------------------------------
```

- Example

```
. margins,dydx(*) atmeans  //均值处边际效应
Conditional marginal effects                    Number of obs     =      2,000
Model VCE    : Robust

Expression   : Pr(work), predict()
dy/dx w.r.t. : age married children education
at           : age             =      36.208 (mean)
               married         =       .6705 (mean)
               children        =      1.6445 (mean)
               education       =      13.084 (mean)

------------------------------------------------------------------------------
             |            Delta-method
             |      dy/dx   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
         age |   .0115031   .0014318     8.03   0.000     .0086968    .0143094
     married |   .1472934    .024954     5.90   0.000     .0983844    .1962024
    children |    .151803   .0091262    16.63   0.000      .133916    .1696901
   education |   .0195096   .0037642     5.18   0.000     .0121318    .0268874
------------------------------------------------------------------------------
```

- Example

```
. margins,dydx(age) at(age=30)  //age=30处边际效应
Average marginal effects                        Number of obs     =      2,000
Model VCE    : Robust

Expression   : Pr(work), predict()
dy/dx w.r.t. : age
at           : age             =          30

------------------------------------------------------------------------------
             |            Delta-method
             |      dy/dx   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
         age |    .011179   .0014784     7.56   0.000     .0082814    .0140765
------------------------------------------------------------------------------
```

Binary Outcomes Models

- Example

```
. estat clas  //计算logit模型准确预测的比率
Logistic model for work

              ———————— True ————————
Classified |         D            ~D  |      Total
———————————+——————————————————————————+———————————
     +     |      1177           361  |       1538
     -     |       166           296  |        462
———————————+——————————————————————————+———————————
   Total   |      1343           657  |       2000
Classified + if predicted Pr(D) >= .5
True D defined as work != 0
——————————————————————————————————————————————————
Sensitivity                     Pr( +| D)   87.64%
Specificity                     Pr( -|~D)   45.05%
Positive predictive value       Pr( D| +)   76.53%
Negative predictive value       Pr(~D| -)   64.07%
——————————————————————————————————————————————————
False + rate for true ~D        Pr( +|~D)   54.95%
False - rate for true D         Pr( -| D)   12.36%
False + rate for classified +   Pr(~D| +)   23.47%
False - rate for classified -   Pr( D| -)   35.93%
——————————————————————————————————————————————————
Correctly classified                        73.65%
——————————————————————————————————————————————————
```

- Example

```
. logit work age married children education,nolog vce(cluster age)
Logistic regression                             Number of obs     =      2,000
                                                Wald chi2(4)      =     576.81
                                                Prob > chi2       =     0.0000
Log pseudolikelihood = -1027.9144               Pseudo R2         =     0.1882
                                    (Std. Err. adjusted for 40 clusters in age)
------------------------------------------------------------------------------
             |               Robust
        work |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
         age |   .0579303   .0055907    10.36   0.000     .0469728    .0688879
     married |   .7417775   .1084937     6.84   0.000     .5291337    .9544213
    children |   .7644882   .0540759    14.14   0.000     .6585014     .870475
   education |   .0982513   .0148423     6.62   0.000     .0691609    .1273416
       _cons |  -4.159247   .2494119   -16.68   0.000    -4.648086   -3.670409
------------------------------------------------------------------------------
```

* 假设年龄相同的个体存在组内相关，故使用age为聚类变量来计算聚类稳健标准误

- Example

```
. probit work age married children education,nolog
Probit regression                               Number of obs     =      2,000
                                                LR chi2(4)        =     478.32
                                                Prob > chi2       =     0.0000
Log likelihood = -1027.0616                     Pseudo R2         =     0.1889
------------------------------------------------------------------------------
        work |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
         age |   .0347211   .0042293     8.21   0.000     .0264318    .0430105
     married |   .4308575    .074208     5.81   0.000     .2854125    .5763025
    children |   .4473249   .0287417    15.56   0.000     .3909922    .5036576
   education |   .0583645   .0109742     5.32   0.000     .0368555    .0798735
       _cons |  -2.467365   .1925635   -12.81   0.000    -2.844782   -2.089948
------------------------------------------------------------------------------
```

Binary Outcomes Models

- Example

```
. margins,dydx(*)
Average marginal effects                        Number of obs     =      2,000
Model VCE    : OIM

Expression   : Pr(work), predict()
dy/dx w.r.t. : age married children education
```

	dy/dx	Delta-method Std. Err.	z	P>|z|	[95% Conf.	Interval]
age	.0100768	.0011647	8.65	0.000	.0077941	.0123595
married	.1250441	.0210541	5.94	0.000	.0837788	.1663094
children	.1298233	.0068418	18.98	0.000	.1164137	.1432329
education	.0169386	.0031183	5.43	0.000	.0108269	.0230504

- Example

```
. estat clas
Probit model for work

              -------- True --------
Classified |         D            ~D  |      Total
-----------+--------------------------+-----------
     +     |      1177           361  |       1538
     -     |       166           296  |        462
-----------+--------------------------+-----------
   Total   |      1343           657  |       2000
Classified + if predicted Pr(D) >= .5
True D defined as work != 0
--------------------------------------------------
Sensitivity                     Pr( +| D)   87.64%
Specificity                     Pr( -|~D)   45.05%
Positive predictive value       Pr( D| +)   76.53%
Negative predictive value       Pr(~D| -)   64.07%
--------------------------------------------------
False + rate for true ~D        Pr( +|~D)   54.95%
False - rate for true D         Pr( -| D)   12.36%
False + rate for classified +   Pr(~D| +)   23.47%
False - rate for classified -   Pr( D| -)   35.93%
--------------------------------------------------
Correctly classified                        73.65%
--------------------------------------------------
```

Binary Outcomes Models

- Example
  - Logit 模型的边际效应，准R2 以及正确预测比率与Probit模型几乎完全相同，故可视为基本等价.
  - 两者的估计系数虽有差距，但估计系数没有可比性.