Lab7: Binary Outcomes Models

Introduction to Econometrics,Spring 2023

Jiayi Cheng

Nanjing University

21/04/2023

Section 1

Binary Outcomes Models

Subsection 1

Review the Theroy

- 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

Nonlinear probability model

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.

Binary Outcomes Models

- Logit Model
  - Review

Nonlinear probability model

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.

- 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)

- 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
------------------------------------------------------------------------------
```

- 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
```

	dy/dx	Delta-method 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
------------------------------------------------------------------------------
```

- 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)
```

work	Coef.	Robust 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

Binary Outcomes Models

- 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
------------------------------------------------------------------------------
```

- 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

------------------------------------------------------------------------------
             |            Delta-method
             |      dy/dx   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%
--------------------------------------------------
```

- 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.

Section 2

Oaxaca-Blinder Decomposition in Stata

Subsection 1

Review and Introduction

Oaxaca-Blinder Decomposition in Stata

- Review the Theory

Oaxaca-Blinder Decomposition: difference in mean

- The difference in mean of Y_i of group A and B is

$$\bar{Y}_A - \bar{Y}_B = \hat{\beta}_A \bar{X}'_A - \hat{\beta}_B \bar{X}'_B$$

- A small trick: plus and minus a term $\hat{\beta}_B \bar{X}'_A$,then

$$\begin{aligned}\bar{Y}_A - \bar{Y}_B &= \hat{\beta}_A \bar{X}'_A - \hat{\beta}_B \bar{X}'_B \\ &= \hat{\beta}_A \bar{X}'_A - \hat{\beta}_B \bar{X}'_A + \hat{\beta}_B \bar{X}'_A - \hat{\beta}_B \bar{X}'_B \\ &= (\hat{\beta}_A - \hat{\beta}_B)\bar{X}'_A + \hat{\beta}_B(\bar{X}'_A - \bar{X}'_B)\end{aligned}$$

- Then the second term is **characteristics effect** which describes how much the difference of outcome, Y, in mean is due to differences in the levels of explanatory variables(characteristics).
- the first term is **coefficients effect** which describes how much the difference of outcome, Y, in mean is due to differences in the magnitude of regression coefficients.

Oaxaca-Blinder Decomposition in Stata

- Commands-*oaxaca*-

```
help oaxaca
findit oaxaca
ssc install oaxaca
```

Oaxaca-Blinder Decomposition in Stata

- Syntax

```
oaxaca depvar [indepvars] [if] [in] [weight] , by(groupvar) [options]
```

Oaxaca-Blinder Decomposition in Stata

- Options
  - by(groupvar)：指定分组变量
  - swap：交换组别
  - threefold：three-fold decomposition; the default.
  - weight(#) / pooled / omega：two-fold decomposition; 设置参照系数 (nondiscriminatory coefficients).

    weight(#)—# 是 Group1 相对于 Group2 的权重，如 weight(1)，将 Group1 系数设置为参照系数.

    omega—using pooled model excluding groupvar.

    pooled—using pooled model including groupvar.
  - relax—do no stop on dropped coefficients/zero variances.
  - noisily—oaxaca first estimates two group-specific regression models and then performs the decomposition. noisily causes the group models' results to be displayed.

Subsection 2

An Example : Twofold decomposition

Oaxaca-Blinder Decomposition in Stata

- An Example
  - The standard application of the Blinder–Oaxaca Decomposition is to divide **Male-female average wage gap** into two parts:
    Explained Part: due to differences in the **levels** of explanatory variables such as schooling years,experience,tenure,industry,occupation,etc.
    Unexplained Part: due to differences in the **coefficients** to explanatory variables
    such as returns to schooling years, experience and tenure and premium in industry and occupation, etc.
  - An example using data from the Swiss Labor Market Survey 1998 (Jann 2003)
  - Reference: Jann B. The Blinder–Oaxaca decomposition for linear regression models[J]. The Stata Journal, 2008, 8(4): 453-479.

Oaxaca-Blinder Decomposition in Stata

- Twofold decomposition
  - Different weight
  - Remember: Weight $\beta^* = W \times \beta_0 + (I - W)\beta_1$

Oaxaca-Blinder Decomposition in Stata

- Twofold decomposition
  - weight=1: $\beta^* = \beta_0$: Wage of men is viewed as non-discriminatory.

```
. oaxaca lnwage educ exper tenure, by(female) weight(1)

Blinder-Oaxaca decomposition                    Number of obs     =      1,434
                                                  Model           =     linear
Group 1: female = 0                               N of obs 1      =        751
Group 2: female = 1                               N of obs 2      =        683

------------------------------------------------------------------------------
      lnwage |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
overall      |
     group_1 |   3.440222   .0174874   196.73   0.000     3.405947    3.474497
     group_2 |   3.266761   .0218522   149.49   0.000     3.223932    3.309591
  difference |   .1734607    .027988     6.20   0.000     .1186052    .2283163
   explained |   .0908978   .0136196     6.67   0.000     .0642038    .1175917
 unexplained |    .082563   .0255804     3.23   0.001     .0324263    .1326996
-------------+----------------------------------------------------------------
explained    |
        educ |    .047771   .0110011     4.34   0.000     .0262092    .0693328
       exper |   .0190709   .0060299     3.16   0.002     .0072526    .0308892
      tenure |   .0240559   .0064492     3.73   0.000     .0114156    .0366961
-------------+----------------------------------------------------------------
unexplained  |
        educ |  -.0640559   .1193498    -0.54   0.591    -.2979772    .1698654
       exper |  -.0397237     .04058    -0.98   0.328    -.1192591    .0398117
      tenure |   .0420986   .0270201     1.56   0.119    -.0108598    .0950571
       _cons |   .1442439   .1340957     1.08   0.282    -.1185788    .4070667
------------------------------------------------------------------------------

. est store oa1
```

Oaxaca-Blinder Decomposition in Stata

- Threefold decomposition
  - **The decomposition output reports in the first panel** :

    3.44—the mean of log wages (lnwage) for men

    3.27—the mean of log wages (lnwage) for women

    0.17—wage gap

    explained—the mean increase in women's wages if they had the same characteristics as men

    unexplained—the change in women's wages when applying the men's coefficients to the women's characteristics

Oaxaca-Blinder Decomposition in Stata

- Twofold decomposition
  - weight=0: $\beta^* = \beta_1$: Wage of women is viewed as non-discriminatory.

```
. oaxaca lnwage educ exper tenure, by(female) weight(0)

Blinder-Oaxaca decomposition                    Number of obs     =      1,434
                                                  Model           =     linear
Group 1: female = 0                               N of obs 1      =        751
Group 2: female = 1                               N of obs 2      =        683

------------------------------------------------------------------------------
      lnwage |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
overall      |
     group_1 |   3.440222   .0174874   196.73   0.000     3.405947    3.474497
     group_2 |   3.266761   .0218522   149.49   0.000     3.223932    3.309591
  difference |   .1734607    .027988     6.20   0.000     .1186052    .2283163
   explained |   .0852798    .015693     5.43   0.000     .0545222    .1160375
 unexplained |   .0881809    .026692     3.30   0.001     .0358655    .1404963
-------------+----------------------------------------------------------------
explained    |
        educ |   .0510912    .012239     4.17   0.000     .0271031    .0750792
       exper |   .0254173   .0088089     2.89   0.004     .0081522    .0426824
      tenure |   .0087714   .0086201     1.02   0.309    -.0081238    .0256665
-------------+----------------------------------------------------------------
unexplained  |
        educ |  -.0673761   .1255358    -0.54   0.591    -.3134218    .1786696
       exper |  -.0460701   .0470666    -0.98   0.328    -.1383189    .0461788
      tenure |   .0573831   .0368225     1.56   0.119    -.0147877     .129554
       _cons |   .1442439   .1340957     1.08   0.282    -.1185788    .4070667
------------------------------------------------------------------------------

. est store oa2
```

Oaxaca-Blinder Decomposition in Stata

- Twofold decomposition
  - weight=0.5: Reimers(1983)

```
. oaxaca lnwage educ exper tenure, by(female) weight(0.5)

Blinder-Oaxaca decomposition                     Number of obs     =      1,434
                                                   Model           =     linear
Group 1: female = 0                                N of obs 1      =        751
Group 2: female = 1                                N of obs 2      =        683

------------------------------------------------------------------------------
      lnwage |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
overall      |
     group_1 |   3.440222   .0174874   196.73   0.000     3.405947    3.474497
     group_2 |   3.266761   .0218522   149.49   0.000     3.223932    3.309591
  difference |   .1734607    .027988     6.20   0.000     .1186052    .2283163
   explained |   .0880888   .0136315     6.46   0.000     .0613715    .1148061
 unexplained |   .0853719   .0255607     3.34   0.001      .035274    .1354699
-------------+----------------------------------------------------------------
explained    |
        educ |   .0494311   .0112121     4.41   0.000     .0274558    .0714064
       exper |   .0222441   .0067646     3.29   0.001     .0089858    .0355024
      tenure |   .0164136   .0056741     2.89   0.004     .0052925    .0275347
-------------+----------------------------------------------------------------
unexplained  |
        educ |   -.065716   .1224423    -0.54   0.591    -.3056985    .1742665
       exper |  -.0428969   .0438153    -0.98   0.328    -.1287734    .0429796
      tenure |   .0497409   .0318942     1.56   0.119    -.0127706    .1122523
       _cons |   .1442439   .1340957     1.08   0.282    -.1185788    .4070667
------------------------------------------------------------------------------

. est store oa3
```

Oaxaca-Blinder Decomposition in Stata

- Twofold decomposition
  - weight the coefficients by the group sizes: Cotton(1988)

```
. sum female
. local p_1=1-r(mean)

. oaxaca lnwage educ exper tenure, by(female) weight(`p_1´)

Blinder-Oaxaca decomposition                    Number of obs     =      1,434
                                                  Model           =     linear
Group 1: female = 0                               N of obs 1      =        751
Group 2: female = 1                               N of obs 2      =        683
------------------------------------------------------------------------------
      lnwage |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
overall      |
     group_1 |   3.440222   .0174874   196.73   0.000     3.405947    3.474497
     group_2 |   3.266761   .0218522   149.49   0.000     3.223932    3.309591
  difference |   .1734607    .027988     6.20   0.000     .1186052    .2283163
   explained |   .0878688   .0137253     6.40   0.000     .0609678    .1147698
 unexplained |    .085592   .0256087     3.34   0.001     .0353997    .1357842
-------------+----------------------------------------------------------------
explained    |
        educ |   .0495611   .0112649     4.40   0.000     .0274824    .0716398
       exper |   .0224927   .0068879     3.27   0.001     .0089927    .0359927
      tenure |    .015815   .0057996     2.73   0.006     .0044481     .027182
-------------+----------------------------------------------------------------
unexplained  |
        educ |   -.065846   .1226846    -0.54   0.591    -.3063033    .1746113
       exper |  -.0431454   .0440695    -0.98   0.328      -.12952    .0432291
      tenure |   .0503395   .0322786     1.56   0.119    -.0129253    .1136043
       _cons |   .1442439   .1340957     1.08   0.282    -.1185788    .4070667
------------------------------------------------------------------------------

. est store oa4
```

Oaxaca-Blinder Decomposition in Stata

- Twofold decomposition
  - Omega: Neumark (1988), Oaxaca and Ransom (1994)

```
. oaxaca lnwage educ exper tenure, by(female) omega

Blinder-Oaxaca decomposition                     Number of obs     =      1,434
                                                   Model           =     linear
Group 1: female = 0                                N of obs 1      =        751
Group 2: female = 1                                N of obs 2      =        683

------------------------------------------------------------------------------
             |               Robust
      lnwage |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
overall      |
     group_1 |   3.440222   .0174586   197.05   0.000     3.406004     3.47444
     group_2 |   3.266761   .0218042   149.82   0.000     3.224026    3.309497
  difference |   .1734607   .0279325     6.21   0.000      .118714    .2282075
   explained |   .0925597   .0140752     6.58   0.000     .0649728    .1201466
 unexplained |    .080901   .0243589     3.32   0.001     .0331585    .1286435
-------------+----------------------------------------------------------------
explained    |
        educ |   .0506419   .0115857     4.37   0.000     .0279344    .0733495
       exper |    .021852   .0064912     3.37   0.001     .0091295    .0345744
      tenure |   .0200658   .0054145     3.71   0.000     .0094536     .030678
-------------+----------------------------------------------------------------
unexplained  |
        educ |  -.0669269   .1395872    -0.48   0.632    -.3405128     .206659
       exper |  -.0425047   .0412027    -1.03   0.302    -.1232605     .038251
      tenure |   .0460887   .0271554     1.70   0.090    -.0071349    .0993122
       _cons |   .1442439   .1624352     0.89   0.375    -.1741233    .4626112
------------------------------------------------------------------------------

. est store oa5
```

Oaxaca-Blinder Decomposition in Stata

- Twofold decomposition
  - Pooled: Jann(2008) and Fortin(2011)

```
. oaxaca lnwage educ exper tenure, by(female) pooled

Blinder-Oaxaca decomposition                      Number of obs     =      1,434
                                                    Model           =     linear
Group 1: female = 0                                 N of obs 1      =        751
Group 2: female = 1                                 N of obs 2      =        683

------------------------------------------------------------------------------
             |               Robust
      lnwage |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
overall      |
     group_1 |   3.440222   .0174586   197.05   0.000     3.406004     3.47444
     group_2 |   3.266761   .0218042   149.82   0.000     3.224026    3.309497
  difference |   .1734607   .0279325     6.21   0.000      .118714    .2282075
   explained |    .089347   .0137531     6.50   0.000     .0623915    .1163026
 unexplained |   .0841137    .025333     3.32   0.001      .034462    .1337654
-------------+----------------------------------------------------------------
explained    |
        educ |   .0493404   .0113168     4.36   0.000     .0271599     .071521
       exper |   .0215214   .0064081     3.36   0.001     .0089617    .0340811
      tenure |   .0184852   .0051833     3.57   0.000     .0083262    .0286443
-------------+----------------------------------------------------------------
unexplained  |
        educ |  -.0656254    .139432    -0.47   0.638    -.3389072    .2076564
       exper |  -.0421741   .0411638    -1.02   0.306    -.1228537    .0385055
      tenure |   .0476693   .0271699     1.75   0.079    -.0055828    .1009213
       _cons |   .1442439   .1624352     0.89   0.375    -.1741233    .4626112
------------------------------------------------------------------------------

. est store oa6
```

Oaxaca-Blinder Decomposition in Stata

- Twofold decomposition
  - Oaxaca-Blinder Table:

```
. esttab oa1 oa2 oa5 oa6,                          ///
        b(%6.3f) se(%6.3f)                          ///
        star(* 0.1 ** 0.05 *** 0.01) replace       ///
        mtitle(male female omega pooled)           ///
        obslast nogaps compress

. esttab oa1 oa2 oa5 oa6 using OB.csv,             ///
        b(%6.3f) se(%6.3f) r2(%6.3f)               ///
        star(* 0.1 ** 0.05 *** 0.01) replace       ///
        mtitle(male female omega pooled)           ///
        obslast nogaps compress
```

Oaxaca-Blinder Decomposition in Stata

- Twofold decomposition
  - Oaxaca-Blinder Table:

	(1) male	(2) female	(3) omega	(4) pooled
overall				
group_1	3.440***	3.440***	3.440***	3.440***
	(0.017)	(0.017)	(0.017)	(0.017)
group_2	3.267***	3.267***	3.267***	3.267***
	(0.022)	(0.022)	(0.022)	(0.022)
difference	0.173***	0.173***	0.173***	0.173***
	(0.028)	(0.028)	(0.028)	(0.028)
explained	0.091***	0.085***	0.093***	0.089***
	(0.014)	(0.016)	(0.014)	(0.014)
unexplai~d	0.083***	0.088***	0.081***	0.084***
	(0.026)	(0.027)	(0.024)	(0.025)
explained				
educ	0.048***	0.051***	0.051***	0.049***
	(0.011)	(0.012)	(0.012)	(0.011)
exper	0.019***	0.025***	0.022***	0.022***
	(0.006)	(0.009)	(0.006)	(0.006)
tenure	0.024***	0.009	0.020***	0.018***
	(0.006)	(0.009)	(0.005)	(0.005)
unexplai~d				
educ	-0.064	-0.067	-0.067	-0.066
	(0.119)	(0.126)	(0.140)	(0.139)
exper	-0.040	-0.046	-0.043	-0.042
	(0.041)	(0.047)	(0.041)	(0.041)
tenure	0.042	0.057	0.046*	0.048*
	(0.027)	(0.037)	(0.027)	(0.027)
_cons	0.144	0.144	0.144	0.144
	(0.134)	(0.134)	(0.162)	(0.162)
N	1434	1434	1434	1434

Standard errors in parentheses
* p<0.1, ** p<0.05, *** p<0.01

Subsection 3

An Example : Robust and Bootstrap S.E.

- Standard error: Robust and Bootstrap S.E.
  - bootstrap 自助法（又称为自举法）：是对原始样本进行“再抽样”的方法
  - 最常见实现自助法的方法：使用可选项 vce(bootstrap)
  - 例如：reg y x1 x2 x3, vce(boot,reps(400))
    reps(400) 表明抽样的样本个数为 400 次
  - 对于一般的统计量使用自助法，可以使用命令 bootstrap,
    如：bootstrap, _b _se, reps(400): reg y x1 x2 x3

Oaxaca-Blinder Decomposition in Stata

- Standard error: Robust and Bootstrap S.E.

```
qui oaxaca lnwage educ exper tenure, by(female) pooled
est store oase1

qui oaxaca lnwage educ exper tenure, by(female) pooled vce(robust)
est store oase2

qui oaxaca lnwage educ exper tenure, by(female) pooled vce(boot,r(
est store oase3

esttab oase1 oase2 oase3                          ///
   using OBSE.csv,                                ///
   b(%6.3f) se(%6.3f)                             ///
   star(* 0.1 ** 0.05 *** 0.01) replace           ///
   obslast nogaps compress
```

Subsection 4

An Example : Detailed OB decomposition

Oaxaca-Blinder Decomposition in Stata

- Detailed OB decomposition
  - Too long table: detail option

```
. tab isco, nofreq gen(isco)

. oaxaca lnwage educ exper tenure isco2-isco9, by(female) pooled detail
Blinder-Oaxaca decomposition                     Number of obs     =      1,434
                                                   Model           =     linear
Group 1: female = 0                                N of obs 1      =        751
Group 2: female = 1                                N of obs 2      =        683
------------------------------------------------------------------------------
             |               Robust
      lnwage |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
overall      |
     group_1 |   3.440222   .0174589   197.05   0.000     3.406003    3.474441
     group_2 |   3.266761   .0218047   149.82   0.000     3.224025    3.309498
  difference |   .1734607   .0279331     6.21   0.000      .118713    .2282085
   explained |   .0738838    .017772     4.16   0.000     .0390513    .1087163
 unexplained |   .0995769   .0266887     3.73   0.000      .047268    .1518859
-------------+----------------------------------------------------------------
explained    |
        educ |   .0395615   .0097334     4.06   0.000     .0204843    .0586387
       exper |   .0218791   .0064781     3.38   0.001     .0091823    .0345759
      tenure |   .0180525   .0049008     3.68   0.000     .0084471    .0276579
       isco2 |  -.0073025   .0059345    -1.23   0.218    -.0189339    .0043288
       isco3 |   .0079376   .0059685     1.33   0.184    -.0037604    .0196357
       isco4 |   .0148991   .0093095     1.60   0.110    -.0033473    .0331455
       isco5 |   .0342628   .0101794     3.37   0.001     .0143115    .0542142
       isco6 |  -.0067029   .0034618    -1.94   0.053    -.0134879    .0000821
       isco7 |  -.0458051   .0114057    -4.02   0.000    -.0681599   -.0234503
       isco8 |  -.0107627   .0044138    -2.44   0.015    -.0194135   -.0021119
       isco9 |   .0078644   .0039715     1.98   0.048     .0000803    .0156485
-------------+----------------------------------------------------------------
unexplained  |
        educ |  -.1324971   .1788045    -0.74   0.459    -.4829475    .2179533
       exper |  -.0345881   .0400924    -0.86   0.388    -.1131677    .0439914
      tenure |   .0475836   .0262981     1.81   0.070    -.0039597     .099127
       isco2 |  -.0085307   .0228235    -0.37   0.709    -.0532641    .0362026
       isco3 |  -.0461995   .0553921    -0.83   0.404    -.1547659     .062367
       isco4 |  -.0489104   .0325967    -1.50   0.133    -.1127987    .0149779
       isco5 |   -.002608   .0251929    -0.10   0.918    -.0519853    .0467692
```

Oaxaca-Blinder Decomposition in Stata

- Detailed OB decomposition
  - Too long table: detail option
  - detail：不加任何参数可以汇报各因素的单独贡献
  - detail()：在括号中定义因素组，聚合结果。
  - categorical()：识别 dummy-variable sets，变换系数，使类别因素的分解结果不取决于基础组的选择。

Oaxaca-Blinder Decomposition in Stata

- Detailed OB decomposition

```
. oaxaca lnwage educ exper tenure isco2-isco9, by(female) pooled  ///
    detail(exp_ten: exper tenure, isco: isco?) categorical(isco?)
(normalized: isco1 isco2 isco3 isco4 isco5 isco6 isco7 isco8 isco9)
Blinder-Oaxaca decomposition                      Number of obs     =      1,434
                                                    Model           =     linear
Group 1: female = 0                                 N of obs 1      =        751
Group 2: female = 1                                 N of obs 2      =        683
```

lnwage	Coef.	Robust Std. Err.	z	P>|z|	[95% Conf.	Interval]
overall						
group_1	3.440222	.0174589	197.05	0.000	3.406003	3.474441
group_2	3.266761	.0218047	149.82	0.000	3.224025	3.309498
difference	.1734607	.0279331	6.21	0.000	.118713	.2282085
explained	.0738838	.017772	4.16	0.000	.0390513	.1087163
unexplained	.0995769	.0266887	3.73	0.000	.047268	.1518859
explained						
educ	.0395615	.0097334	4.06	0.000	.0204843	.0586387
exp_ten	.0399316	.0089081	4.48	0.000	.022472	.0573911
isco	-.0056093	.012445	-0.45	0.652	-.0300009	.0187824
unexplained						
educ	-.1324971	.1788045	-0.74	0.459	-.4829475	.2179533
exp_ten	.0129955	.0400811	0.32	0.746	-.0655619	.0915529
isco	-.0159367	.0296549	-0.54	0.591	-.0740592	.0421858
_cons	.2350152	.195018	1.21	0.228	-.1472132	.6172435

```
exp_ten: exper tenure
isco: isco1 isco2 isco3 isco4 isco5 isco6 isco7 isco8 isco9
```

Oaxaca-Blinder Decomposition in Stata

- Detailed OB decomposition

```
. ereturn display,coeflegend
------------------------------------------------------------------------------
      lnwage |      Coef.  Legend
-------------+----------------------------------------------------------------
overall      |
     group_1 |   3.440222  _b[overall:group_1]
     group_2 |   3.266761  _b[overall:group_2]
  difference |   .1734607  _b[overall:difference]
   explained |   .0738838  _b[overall:explained]
 unexplained |   .0995769  _b[overall:unexplained]
-------------+----------------------------------------------------------------
explained    |
        educ |   .0395615  _b[explained:educ]
     exp_ten |   .0399316  _b[explained:exp_ten]
        isco |  -.0056093  _b[explained:isco]
-------------+----------------------------------------------------------------
unexplained  |
        educ |  -.1324971  _b[unexplained:educ]
     exp_ten |   .0129955  _b[unexplained:exp_ten]
        isco |  -.0159367  _b[unexplained:isco]
       _cons |   .2350152  _b[unexplained:_cons]
------------------------------------------------------------------------------
```

Oaxaca-Blinder Decomposition in Stata

- Detailed OB decomposition

```
nlcom (educ_explained    : _b[explained:educ])           ///
      (expten_explained  : _b[explained:exp_ten])        ///
      (isco_explained    : _b[explained:isco])           ///
      (Total_explained   : _b[overall:explained])        ///
      (educ_unexplained  : _b[unexplained:educ])         ///
      (expten_unexplained: _b[unexplained:exp_ten])      ///
      (isco_unexplained  : _b[unexplained:isco])         ///
      (cons_unexplained  : _b[unexplained:isco])         ///
      (Total_unexplained : _b[overall:unexplained]), post
```

Oaxaca-Blinder Decomposition in Stata

- Detailed OB decomposition

```
educ_expla~d:  _b[explained:educ]
expten_exp~d:  _b[explained:exp_ten]
isco_expla~d:  _b[explained:isco]
Total_expl~d:  _b[overall:explained]
educ_unexp~d:  _b[unexplained:educ]
expten_une~d:  _b[unexplained:exp_ten]
isco_unexp~d:  _b[unexplained:isco]
cons_unexp~d:  _b[unexplained:isco]
Total_unex~d:  _b[overall:unexplained]
```

lnwage	Coef.	Std. Err.	z	P>|z|	[95% Conf.	Interval]
educ_explained	.0395615	.0097334	4.06	0.000	.0204843	.0586387
expten_explained	.0399316	.0089081	4.48	0.000	.022472	.0573911
isco_explained	-.0056093	.012445	-0.45	0.652	-.0300009	.0187824
Total_explained	.0738838	.017772	4.16	0.000	.0390513	.1087163
educ_unexplained	-.1324971	.1788045	-0.74	0.459	-.4829475	.2179533
expten_unexplained	.0129955	.0400811	0.32	0.746	-.0655619	.0915529
isco_unexplained	-.0159367	.0296549	-0.54	0.591	-.0740592	.0421858
cons_unexplained	-.0159367	.0296549	-0.54	0.591	-.0740592	.0421858
Total_unexplained	.0995769	.0266887	3.73	0.000	.047268	.1518859

```
. est store deoa1
```

Oaxaca-Blinder Decomposition in Stata

- Detailed OB decomposition

```
. oaxaca lnwage educ exper tenure isco2-isco9, by(female) pooled                    ///
     detail(exp_ten: exper tenure, isco: isco?) categorical(isco?)

. nlcom (educ_explained    : _b[explained:educ]      / _b[overall:difference]*100)  ///
        (expten_explained  : _b[explained:exp_ten]   / _b[overall:difference]*100)  ///
        (isco_explained    : _b[explained:isco]      / _b[overall:difference]*100)  ///
        (Total_explained   : _b[overall:explained]   / _b[overall:difference]*100)  ///
        (educ_unexplained  : _b[unexplained:educ]    / _b[overall:difference]*100)  ///
        (expten_unexplained: _b[unexplained:exp_ten] / _b[overall:difference]*100)  ///
        (isco_unexplained  : _b[unexplained:isco]    / _b[overall:difference]*100)  ///
        (cons_unexplained  : _b[unexplained:_cons]   / _b[overall:difference]*100)  ///
        (Total_unexplained : _b[overall:unexplained] / _b[overall:difference]*100) , post
```

Oaxaca-Blinder Decomposition in Stata

- Detailed OB decomposition

```
educ_expla~d:  _b[explained:educ]      / _b[overall:difference]*100
expten_exp~d:  _b[explained:exp_ten]   / _b[overall:difference]*100
isco_expla~d:  _b[explained:isco]      / _b[overall:difference]*100
Total_expl~d:  _b[overall:explained]   / _b[overall:difference]*100
educ_unexp~d:  _b[unexplained:educ]    / _b[overall:difference]*100
expten_une~d:  _b[unexplained:exp_ten] / _b[overall:difference]*100
isco_unexp~d:  _b[unexplained:isco]    / _b[overall:difference]*100
cons_unexp~d:  _b[unexplained:_cons]   / _b[overall:difference]*100
Total_unex~d:  _b[overall:unexplained] / _b[overall:difference]*100
```

lnwage	Coef.	Std. Err.	z	P>|z|	[95% Conf.	Interval]
educ_explained	22.80716	5.60657	4.07	0.000	11.81849	33.79584
expten_explained	23.02053	5.683341	4.05	0.000	11.88139	34.15967
isco_explained	-3.233732	7.269927	-0.44	0.656	-17.48253	11.01506
Total_explained	42.59396	9.883593	4.31	0.000	23.22247	61.96545
educ_unexplained	-76.38447	102.0626	-0.75	0.454	-276.4234	123.6545
expten_unexplained	7.491894	23.41997	0.32	0.749	-38.41041	53.3942
isco_unexplained	-9.187481	17.16109	-0.54	0.592	-42.8226	24.44763
cons_unexplained	135.4861	108.2158	1.25	0.211	-76.61298	347.5852
Total_unexplained	57.40604	9.883593	5.81	0.000	38.03455	76.77753

```
. est store deoa2
```

Oaxaca-Blinder Decomposition in Stata

- Detailed OB decomposition

```
esttab deoa1 deoa2,                    ///
     b(%6.3f) t(%6.3f)                 ///
     nogaps compress  obslast          ///
     star(* 0.1 ** 0.05 *** 0.01)      ///
     mtitle(coef percent(%))           ///
```

Oaxaca-Blinder Decomposition in Stata

- Detailed OB decomposition

	(1) coef	(2) percent~)
educ_exp~d	0.040***	22.807***
	(4.064)	(4.068)
expten_e~d	0.040***	23.021***
	(4.483)	(4.051)
isco_exp~d	-0.006	-3.234
	(-0.451)	(-0.445)
Total_ex~d	0.074***	42.594***
	(4.157)	(4.310)
educ_une~d	-0.132	-76.384
	(-0.741)	(-0.748)
expten_u~d	0.013	7.492
	(0.324)	(0.320)
isco_une~d	-0.016	-9.187
	(-0.537)	(-0.535)
cons_une~d	-0.016	135.486
	(-0.537)	(1.252)
Total_un~d	0.100***	57.406***
	(3.731)	(5.808)
N	1434	1434

t statistics in parentheses
* p<0.1, ** p<0.05, *** p<0.01