回帰分析I

14. 相互作用項（または交差項）

イントロダクション

- 前回取り上げたモデルの推定結果の一部です。

```
. reg lwage female belavg abvavg

      Source |       SS       df       MS              Number of obs =    1260
-------------+------------------------------           F(  3,  1256) =  105.08
       Model |  89.2739026     3  29.7579675           Prob > F      =  0.0000
    Residual |   355.70607  1256   .28320547           R-squared     =  0.2006
-------------+------------------------------           Adj R-squared =  0.1987
       Total |  444.979972  1259  .353439215           Root MSE      =  .53217

------------------------------------------------------------------------------
       lwage |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
      female |  -.5426538   .0315669   -17.19   0.000    -.6045836   -.4807241
      belavg |  -.1773534   .0471458    -3.76   0.000    -.2698466   -.0848601
      abvavg |  -.0165335   .0336826    -0.49   0.624    -.0826139    .0495469
       _cons |   1.873418   .0222717    84.12   0.000     1.829724    1.917112
------------------------------------------------------------------------------
```

- $ABVAVG$の係数はゼロという帰無仮説は10%水準で棄却できません。
- なので、これからする議論を簡単にするために、以下ではモデルを少しだけ単純化します。

$$\ln(WAGE) = \beta_0 + \beta_1 FEMALE + \beta_2 BELAVG + U$$

- β_1は「$BELAVG$が一定の下で、女性であることの賃金（対数値）への効果」を測ります。
  - $BELAVG$が一定の下っていうのは、「外見を一定とすると、、、」または「外見が賃金に与える影響をコントロールしたとき、、、」っていうことですね。
- β_2は「$FEMALE$が一定の下で、外見が平均に満たないことの賃金（対数値）への効果」を測ります。
  - $FEMALE$が一定の下っていうのは、「性別を一定とすると、、、」または「性別が賃金に与える影響をコントロールしたとき、、、」っていうことですね。

$$\ln(WAGE) = \beta_0 + \beta_1 FEMALE + \beta_2 BELAVG + U$$

```
. reg lwage female belavg

      Source |       SS           df       MS      Number of obs   =     1,260
-------------+----------------------------------   F(2, 1257)      =    157.59
       Model |  89.2056659         2  44.6028329   Prob > F        =    0.0000
    Residual |  355.774306     1,257  .283034452   R-squared       =    0.2005
-------------+----------------------------------   Adj R-squared   =    0.1992
       Total |  444.979972     1,259  .353439215   Root MSE        =    .53201

------------------------------------------------------------------------------
       lwage |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
      female |  -.5434273   .0315181   -17.24   0.000    -.6052611   -.4815935
      belavg |  -.1715922   .0456477    -3.76   0.000    -.2611464   -.0820381
       _cons |   1.867951   .0192814    96.88   0.000     1.830124    1.905778
------------------------------------------------------------------------------
```

- 「係数はゼロ」という帰無仮説は有意水準1%で棄却できます。
- 係数の推定値は、「（他の条件を一定にした時）、外見が平均に満たない人は、そうでない人に比べて、賃金が17%低くなる」ということを示しています。

$$\ln(WAGE) = \beta_0 + \beta_1 FEMALE + \beta_2 BELAVG + U$$

- モデルを見たら、いろいろ考えてみる（妄想する？）ことが大切です。
  - 欠落変数があるのではないか、、、、（上のモデルではいっぱいありそうですね）
  - 関数形はこれでいいのか、、、（上のモデルなら左辺に対数を取る？取らない？）
- ここでは、これまでとは少し違うことを考えてみましょう。

$$\ln(WAGE) = \beta_0 + \beta_1 FEMALE + \beta_2 BELAVG + U$$

- この定式化では、

「男性であれ、女性であれ、外見が平均未満であることの賃金への効果はβ_2で一定」

ですよね。

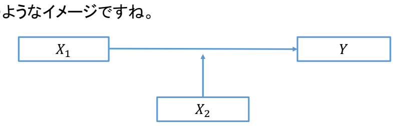
- でも、次のような可能性、あるかもしれません。
- 労働市場において、女性の外見は男性の外見よりも重要？
- もしそうであるなら、「外見が平均未満であることの賃金への負の効果」は女性の方が男性に比べて大きいかもしれません。
  - 外見が賃金へ与える効果は性別に依存しているかもしれない、ということです。
- しかし、今の定式化は、この可能性をはじめから排除してしまっていますよね。

- ちょっとお堅く言えば、一般的には、

「説明変数X_1が従属変数Yに与える影響が、別の説明変数X_2の値に依存する」

ようなことはありえますね。

- 次のようなイメージですね。



- 今の例では、Yが賃金（の対数値）、X_1が$BELAVG$、X_2が$FEMALE$と考えればいいですね。

- このような、

「ある説明変数の役割が、別の説明変数とかかわりあっているような状況」

別の言い方をするなら、

「ある説明変数と別の説明変数との間の相互作用的な関係」

のモデル化がこの講義ノートの目的となります。

- 以下の3つのケースを考えていきます。

（1）二つの説明変数がともにダミー変数のとき。
（2）一つの説明変数がダミー変数、もう一方が連続変数のとき。
（3）二つの説明変数ともに連続変数のとき。

二つのダミー変数の相互作用

$$\ln(WAGE) = \beta_0 + \beta_1 FEMALE + \beta_2 BELAVG + U$$

- このモデルを次のように変えることで、「外見が賃金へ与える効果が性別に依存している」可能性を考慮に入れることができます。

$$\ln(WAGE) = \beta_0 + \beta_1 FEMALE + \beta_2 BELAVG + \beta_3 (BELAVG \times FEMALE) + U$$

- この新しい説明変数$BELAVG \times FEMALE$は「**相互作用項**（interaction term）」や「**交差項**（cross term）などと呼ばれます。
- 相互作用項の働きを理解するために条件付期待値を考えてみましょう。
- ここの目的は相互作用項の働きの理解なので、$E(U|FEMALE, BELAVG) = 0$ とします。

9

$$\ln(WAGE) = \beta_0 + \beta_1 FEMALE + \beta_2 BELAVG + \beta_3 (BELAVG \times FEMALE) + U$$

- 男性、外見平均以上

$E(\ln(WAGE)|\, FEMALE = 0, BELAVG = 0) = \boxed{\beta_0}$

男性ならば、
外見が平均未満の効果はβ_2

- 男性、外見平均未満

$E(\ln(WAGE)|\, FEMALE = 0, BELAVG = 1) = \boxed{\beta_0} + \beta_2$

10

$$\ln(WAGE) = \beta_0 + \beta_1 FEMALE + \beta_2 BELAVG + \beta_3 (BELAVG \times FEMALE) + U$$

- 女性、外見平均以上

$E(\ln(WAGE)|\, FEMALE = 1, BELAVG = 0) = \boxed{\beta_0 + \beta_1}$

女性ならば、
外見が平均未満の効果は
$\beta_2 + \beta_3$

- 女性、外見平均未満

$E(\ln(WAGE)|\, FEMALE = 1, BELAVG = 1) = \boxed{\beta_0 + \beta_1} + \beta_2 + \beta_3$

11

$$\ln(WAGE) = \beta_0 + \beta_1 FEMALE + \beta_2 BELAVG + \beta_3 (BELAVG \times FEMALE) + U$$

- 男性、外見平均以上

$E(\ln(WAGE)|\, FEMALE = 0, BELAVG = 0) = \boxed{\beta_0}$

男性ならば、
外見が平均未満の効果は
β_2

- 男性、外見平均未満

$E(\ln(WAGE)|\, FEMALE = 0, BELAVG = 1) = \boxed{\beta_0} + \beta_2$

- 女性、外見平均以上

$E(\ln(WAGE)|\, FEMALE = 1, BELAVG = 0) = \boxed{\beta_0 + \beta_1}$

女性ならば、
外見が平均未満の効果は
$\beta_2 + \beta_3$

- 女性、外見平均未満

$E(\ln(WAGE)|\, FEMALE = 1, BELAVG = 1) = \boxed{\beta_0 + \beta_1} + \beta_2 + \beta_3$

相互作用項にかかる係数β_3は、
外見が平均未満であることの効果の男女差
を捉えるもの。

12

まとめ

（1） $\ln(WAGE) = \beta_0 + \beta_1 FEMALE + \beta_2 BELAVG + U$

相互作用項のないモデルでは、外見（$BELAVG$）が賃金へ与える効果は、性別にかかわらずβ_2で一定

（2） $\ln(WAGE) = \beta_0 + \beta_1 FEMALE + \beta_2 BELAVG + \beta_3(BELAVG \times FEMALE) + U$

相互作用項のあるモデルでは、外見（$BELAVG$）が賃金へ与える効果は性別に依存する。

- 男性ならβ_2
- 女性なら$\beta_2 + \beta_3$
- 効果の男女差はβ_3

- β_3が統計的に有意にゼロと異なるなら、相互作用があるってことでモデル（2）が妥当。
  - このときモデル（1）は妥当ではない。
- β_3が統計的に有意にゼロと違わないなら、相互作用はない（厳密には、相互作用があるという十分な証拠はない）ってこと。
  - もしそうなら、モデル（1）で十分。
- つまり、モデル（1）でいいかどうかは、モデル（2）を推定してみなければ分からない。
- よって、説明変数間に相互作用的な関係がある可能性があるなら、相互作用項を入れたモデルを推定してみることは重要です。

実際に推定してみる

- それでは、外見の効果には男女差があるのかどうか調べるために

$$\ln(WAGE) = \beta_0 + \beta_1 FEMALE + \beta_2 BELAVG + \beta_3(BELAVG \times FEMALE) + U$$

を推定してみましょう。

- まず相互作用項$BELAVG \times FEMALE$を作ります（generateコマンド）。
  - 名前は適当につければいいです。ここでは$BELAVG_FEM$とします。
- あとはこれまで通りに推定すればよいです。


```
. generate belavg_fem = belavg*female

. reg lwage female belavg belavg_fem

      Source |       SS       df       MS              Number of obs =    1260
-------------+------------------------------           F(  3,  1256) =  105.03
       Model |  89.2420975     3  29.7473658           Prob > F      =  0.0000
    Residual |  355.737875  1256  .283230792           R-squared     =  0.2006
-------------+------------------------------           Adj R-squared =  0.1986
       Total |  444.979972  1259  .353439215           Root MSE      =  .53219

------------------------------------------------------------------------------
       lwage |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
      female |  -.5477646   .0337688   -16.22   0.000     -.614014   -.4815153
      belavg |  -.1842937   .0577873    -3.19   0.001     -.297664   -.0709234
  belavg_fem |    .033818    .094293     0.36   0.720    -.1511711    .2188072
       _cons |   1.869431   .0197244    94.78   0.000     1.830735    1.908127
------------------------------------------------------------------------------
```

- 相互作用項の係数の推定値の符号は正です。
- これは、外見が平均未満であることが賃金に与える負の影響は、女性の方が男性よりも「小さい」ことを示しています。具体的には、、、
  - 男性だと18%程度賃金が低くなる（-0.184）。
  - 女性だと15%程度賃金が低くなる（$-0.184 + 0.034$）。
- 直感とは違いますね、、、ただし、相互作用項の係数がゼロであるという帰無仮説は、10%の有意水準で棄却することはできません。
- 従って、外見が平均未満であることが賃金に与える影響は性別に依存する、という仮説に対する十分な統計的なサポートはない、ということになります。

相互作用項：Stataでの扱い方

- Stataは相互作用項の取り扱いに長けています。
- さっきは、まずgenerateコマンドを使って相互作用項を作り、そしてその変数を使ってモデルを推定という手順で行きました。
- しかし、 Stataでは、相互作用項をわざわざ作ることなく、相互作用項入りのモデルを推定することができます。
- 今のモデルであれば以下のようにすれば同じモデルを推定できます。

```
. reg lwage i.female i.belavg i.belavg#i.female

      Source |       SS       df       MS              Number of obs =    1260
-------------+------------------------------           F(  3,  1256) =  105.03
       Model |  89.2420975     3  29.7473658           Prob > F      =  0.0000
    Residual |  355.737875  1256  .283230792           R-squared     =  0.2006
-------------+------------------------------           Adj R-squared =  0.1986
       Total |  444.979972  1259  .353439215           Root MSE      =  .53219

------------------------------------------------------------------------------
       lwage |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
    1.female |  -.5477646   .0337688   -16.22   0.000     -.614014   -.4815153
    1.belavg |  -.1842937   .0577873    -3.19   0.001     -.297664   -.0709234
             |
     belavg#|
      female |
        1 1  |    .033818    .094293     0.36   0.720    -.1511711    .2188072
             |
       _cons |   1.869431   .0197244    94.78   0.000     1.830735    1.908127
------------------------------------------------------------------------------
```

17

reg lwage i.female i.belavg i.belavg#i.female

- ポイントとなるのは、i. と # の二つです。
- i.　は序数的な変数に使いましたね（講義ノート12参照）。
  - ➢ Stataに「これは序数的な変数です。この変数よりダミー変数を作成して下さい」とお願いするものでした。
  - ➢ ダミー変数は0−1ですから序数的な変数の一種です。
  - ➢ i.femaleで、「femaleは序数的な変数です。この変数よりダミー変数を作成して下さい」とお願いします（もともとダミー変数ですが）。
  - ➢ i.belavgで、「belavgは序数的な変数です。この変数よりダミー変数を作成して下さい」とお願いします（もともとダミー変数ですが）。
- #　は、i.belavg#i.female　というように使うことで、Stataに「i.belavgとi.femaleの相互作用項をモデルの中に入れてくれ」とお願いできます。

18


```
. reg lwage i.female i.belavg i.belavg#i.female

      Source |       SS       df       MS              Number of obs =    1260
-------------+------------------------------           F(  3,  1256) =  105.03
       Model |  89.2420975     3  29.7473658           Prob > F      =  0.0000
    Residual |  355.737875  1256  .283230792           R-squared     =  0.2006
-------------+------------------------------           Adj R-squared =  0.1986
       Total |  444.979972  1259  .353439215           Root MSE      =  .53219

------------------------------------------------------------------------------
       lwage |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
    1.female |  -.5477646   .0337688   -16.22   0.000     -.614014   -.4815153
    1.belavg |  -.1842937   .0577873    -3.19   0.001     -.297664   -.0709234
             |
     belavg#|
      female |
        1 1  |    .033818    .094293     0.36   0.720    -.1511711    .2188072
             |
       _cons |   1.869431   .0197244    94.78   0.000     1.830735    1.908127
------------------------------------------------------------------------------
```


19


```
. reg lwage i.female i.belavg i.belavg#i.female

      Source |       SS       df       MS              Number of obs =    1260
-------------+------------------------------           F(  3,  1256) =  105.03
       Model |  89.2420975     3  29.7473658           Prob > F      =  0.0000
    Residual |  355.737875  1256  .283230792           R-squared     =  0.2006
-------------+------------------------------           Adj R-squared =  0.1986
       Total |  444.979972  1259  .353439215           Root MSE      =  .53219

------------------------------------------------------------------------------
       lwage |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
    1.female |  -.5477646   .0337688   -16.22   0.000     -.614014   -.4815153
    1.belavg |  -.1842937   .0577873    -3.19   0.001     -.297664   -.0709234
             |
     belavg#|
      female |
        1 1  |    .033818    .094293     0.36   0.720    -.1511711    .2188072
             |
       _cons |   1.869431   .0197244    94.78   0.000     1.830735    1.908127
------------------------------------------------------------------------------
```


- lincomやtestコマンドを使うときは少し注意が必要です。例えば、

test _b[1.belavg#1.female] = 0

とか

lincom _b[1.belavg] + _b[1.belavg#1.female]

という感じでアウトプットの表示に合わせて下さい（つまり　i. では無くて1. です）。

```
. lincom _b[1.belavg] + _b[1.belavg#1.female]

 ( 1)  1.belavg + 1.belavg#1.female = 0

------------------------------------------------------------------------------
       lwage |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
         (1) |  -.1504757   .0745104    -2.02   0.044    -.2966542   -.0042971
------------------------------------------------------------------------------
```

20

- Stataでの相互作用項の使い方の理解を深めるために、もう一つモデルを推定してみます。
- 新しい変数を作りましょう（やってみて下さい。使うのはgenerateとreplaceです。）

$$LOOKS3 = \begin{cases} 1 & if\ LOOKS \le 2 \\ 2 & if\ LOOKS = 3 \\ 3 & if\ LOOKS \ge 4 \end{cases}$$

- これを使って、、、

```
. reg lwage female i.looks3

      Source |       SS           df       MS      Number of obs   =     1,260
-------------+----------------------------------   F(3, 1256)      =    105.08
       Model |  89.2739026         3  29.7579675   Prob > F        =    0.0000
    Residual |   355.70607     1,256   .28320547   R-squared       =    0.2006
-------------+----------------------------------   Adj R-squared   =    0.1987
       Total |  444.979972     1,259  .353439215   Root MSE        =    .53217

------------------------------------------------------------------------------
       lwage |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
      female |  -.5426538   .0315669   -17.19   0.000    -.6045836   -.4807241
             |
      looks3 |
          2  |   .1773534   .0471458     3.76   0.000     .0848601    .2698466
          3  |   .1608199   .0506616     3.17   0.002     .0614293    .2602105
             |
       _cons |   1.696065   .0444017    38.20   0.000     1.608955    1.783174
------------------------------------------------------------------------------
```

21

```
. reg lwage female i.female#i.looks3
note: 1.female#3.looks3 omitted because of collinearity

      Source |       SS           df       MS      Number of obs   =     1,260
-------------+----------------------------------   F(5, 1254)      =     63.28
       Model |   89.653589         5  17.9307178   Prob > F        =    0.0000
    Residual |  355.326383     1,254  .283354373   R-squared       =    0.2015
-------------+----------------------------------   Adj R-squared   =    0.1983
       Total |  444.979972     1,259  .353439215   Root MSE        =    .53231

------------------------------------------------------------------------------
       lwage |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
      female |  -.3426796   .0701381    -4.89   0.000    -.4802805   -.2050787
             |
female#looks3 |
        0 2  |   .1987381   .0594228     3.34   0.001     .0821591    .3153171
        0 3  |   .1547401    .064321     2.41   0.016     .0285514    .2809287
        1 1  |   -.171267   .0822822    -2.08   0.038    -.3326929   -.0098411
        1 2  |  -.0336409   .0564256    -0.60   0.551    -.1443399    .0770582
        1 3  |          0  (omitted)
             |
       _cons |   1.685137   .0543287    31.02   0.000     1.578552    1.791722
------------------------------------------------------------------------------
```

- 解釈できますか？

22

連続変数とダミー変数の相互作用

- 教育水準と賃金の関係を考えましょう。

$$\ln(WAGE) = \beta_0 + \beta_1 EDUC + U$$

- 男性と女性の定数項が違う可能性を考慮したモデルは、

$$\ln(WAGE) = \beta_0 + \delta_0 FEMALE + \beta_1 EDUC + U$$

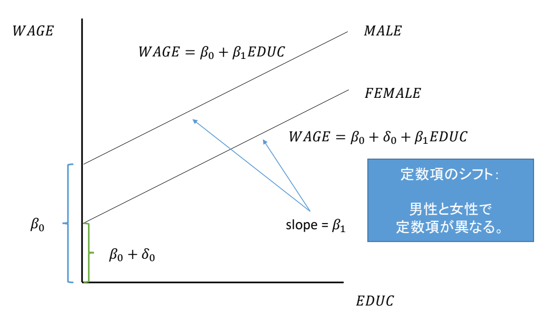
23

- 理解を深めるために図を描いてみましょう。

$E(WAGE|FEMALE = 0, EDUC) = \beta_0 + \beta_1 EDUC$

$E(WAGE|FEMALE = 1, EDUC) = \beta_0 + \delta_0 + \beta_1 EDUC$

- ここでは$\delta_0 < 0$（女性の賃金の方が低い）として図を描いてみます。



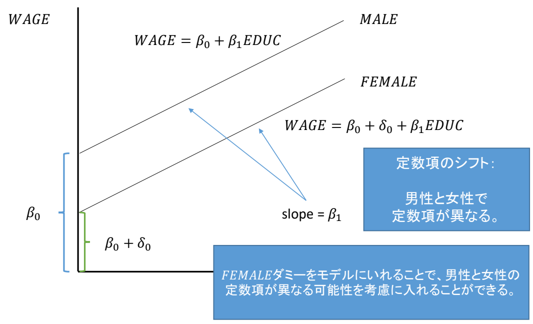
24

- 理解を深めるために図を描いてみましょう。

$E(WAGE|FEMALE = 0, EDUC) = \beta_0 + \beta_1 EDUC$

$E(WAGE|FEMALE = 1, EDUC) = \beta_0 + \delta_0 + \beta_1 EDUC$

- ここでは$\delta_0 < 0$（女性の賃金の方が低い）として図を描い$EDUC$てみます。



連続変数とダミー変数の相互作用

- 教育水準と賃金の関係を考えましょう。

$$\ln(WAGE) = \beta_0 + \beta_1 EDUC + U$$

- 男性と女性の定数項が違う可能性を考慮したモデルは、

$$\ln(WAGE) = \beta_0 + \delta_0 FEMALE + \beta_1 EDUC + U$$

- 他の可能性？
- 教育のリターンが男女で違うかも？
- すなわち、$EDUC$の係数は男女で異なるかも？
- この可能性には$EDUC$と$FEMALE$の相互作用項を入れることで対処できます。

$$\ln(WAGE) = \beta_0 + \delta_0 FEMALE + \beta_1 EDUC + \delta_1 (EDUC \times FEMALE) + U$$

- 条件付期待値を考えてみましょう。
- 男性

$$E(\ln(WAGE)|FEMALE = 0, EDUC) = \beta_0 + \beta_1 EDUC$$

- 女性

$$E(\ln(WAGE)|FEMALE = 0, EDUC) = \beta_0 + \delta_0 + \beta_1 EDUC + \delta_1 EDUC$$

すなわち

$$E(\ln(WAGE)|FEMALE = 0, EDUC) = \beta_0 + \delta_0 + (\beta_1 + \delta_1) EDUC$$

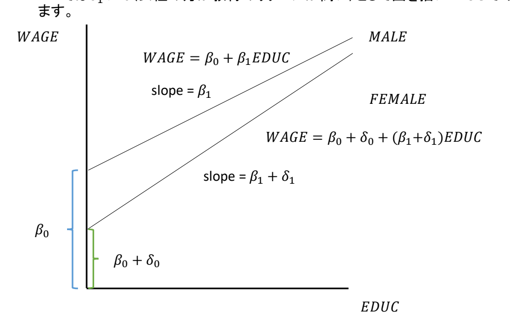
- $EDUC$が（対数）賃金に与える影響は、男性はβ_1、女性は$\beta_1 + \delta_1$、ということです。

- 図を描いてみましょう。

$E(WAGE|FEMALE = 0, EDUC) = \beta_0 + \beta_1 EDUC$

$E(WAGE|FEMALE = 1, EDUC) = \beta_0 + \delta_0 + (\beta_1 + \delta_1) EDUC$

- ここでは$\delta_1 > 0$（女性の方が教育のリターンが高い）として図を描い$EDUC$てみます。



- 推定してみましょう。まず相互作用項なしモデルです。

```
. reg lwage female educ

      Source |       SS           df       MS      Number of obs   =     1,260
-------------+----------------------------------   F(2, 1257)      =    225.96
       Model |  117.674658         2  58.8373289   Prob > F        =    0.0000
    Residual |  327.305314     1,257  .260386089   R-squared       =    0.2644
-------------+----------------------------------   Adj R-squared   =    0.2633
       Total |  444.979972     1,259  .353439215   Root MSE        =    .51028

------------------------------------------------------------------------------
       lwage |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
      female |  -.5497287   .0302208   -18.19   0.000    -.6090174     -.49044
        educ |   .0611915   .0054799    11.17   0.000     .0504408    .0719421
       _cons |   1.080245   .0710119    15.21   0.000     .9409298     1.21956
------------------------------------------------------------------------------
```

- 相互作用項ありモデルを推定しましょう。
- まずgenerateコマンドで相互作用項を作ります。
- 変数名はFEM_EDUCとしましょう。


```
. reg lwage female educ fem_educ

      Source |       SS           df       MS      Number of obs   =     1,260
-------------+----------------------------------   F(3, 1256)      =    152.89
       Model |  119.031155         3  39.6770517   Prob > F        =    0.0000
    Residual |  325.948817     1,256  .25951339    R-squared       =    0.2675
-------------+----------------------------------   Adj R-squared   =    0.2657
       Total |  444.979972     1,259  .353439215   Root MSE        =    .50942

------------------------------------------------------------------------------
       lwage |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
      female |  -.8858892   .1500972    -5.90   0.000    -1.180358   -.5914204
        educ |   .0525291   .0066546     7.89   0.000     .0394737    .0655844
    fem_educ |   .0267217   .0116879     2.29   0.022     .0037918    .0496516
       _cons |   1.188924   .0853547    13.93   0.000     1.021471    1.356378
------------------------------------------------------------------------------
```

- 係数は正で5%有意です。
- 推定値は、1年学校に多くいくことによるリターンは、男性がだいたい5.3%、女性はだいたい5.3＋2.7% = 8%であることを示しています。
- 女性の教育のリターンはlincomを使って（標準誤差付きで）計算することができます。

```
. lincom _b[educ] + _b[fem_educ]

 ( 1)  educ + fem_educ = 0

------------------------------------------------------------------------------
       lwage |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
         (1) |   .0792508   .0096084     8.25   0.000     .0604004    .0981011
------------------------------------------------------------------------------
```


$$\ln(WAGE) = \beta_0 + \delta_0 FEMALE + \beta_1 EDUC + \delta_1 (EDUC \times FEMALE) + U$$

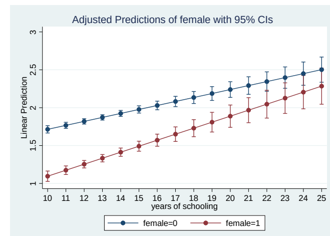
- このモデルについて少し補足します。
- このモデルは定数項が男女で異なる可能性、そして傾きのパラメータが男女で異なる可能性を考慮したモデルです。
- 従って、男女別々に分析しているのとほぼ同じになります。
- 違う点は、男女別々に分析した場合、誤差項の分散が男女で異なる可能性も考慮に入れることができる、ということです。
- その一方、男女で定数項が違う、男女で傾きのパラメータが違う、などの検定をすることはできなくなります。


```
. reg lwage edu if female == 0

      Source |       SS           df       MS      Number of obs   =       824
-------------+----------------------------------   F(1, 822)       =     59.34
       Model |  16.1701882         1  16.1701882   Prob > F        =    0.0000
    Residual |  223.991065       822  .272495213   R-squared       =    0.0673
-------------+----------------------------------   Adj R-squared   =    0.0662
       Total |  240.161253       823  .291811973   Root MSE        =    .52201

------------------------------------------------------------------------------
       lwage |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
        educ |   .0525291    .006819     7.70   0.000     .0391443    .0659138
       _cons |   1.188924   .0874635    13.59   0.000     1.017246    1.360602
------------------------------------------------------------------------------

. reg lwage edu if female == 1

      Source |       SS           df       MS      Number of obs   =       436
-------------+----------------------------------   F(1, 434)       =     75.15
       Model |  17.6547116         1  17.6547116   Prob > F        =    0.0000
    Residual |  101.957752       434  .234925696   R-squared       =    0.1476
-------------+----------------------------------   Adj R-squared   =    0.1456
       Total |  119.612464       435  .274971181   Root MSE        =    .48469

------------------------------------------------------------------------------
       lwage |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
        educ |   .0792508   .0091419     8.67   0.000     .0612828    .0972187
       _cons |   .3030351   .1174712     2.58   0.010      .072152    .5339182
------------------------------------------------------------------------------

. reg lwage i.female c.edu c.edu#i.female

      Source |       SS           df       MS      Number of obs   =     1,260
-------------+----------------------------------   F(3, 1256)      =    152.89
       Model |  119.031155         3  39.6770517   Prob > F        =    0.0000
    Residual |  325.948817     1,256  .25951339    R-squared       =    0.2675
-------------+----------------------------------   Adj R-squared   =    0.2657
       Total |  444.979972     1,259  .353439215   Root MSE        =    .50942

-------------------------------------------------------------------------------
        lwage |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
--------------+----------------------------------------------------------------
     1.female |  -.8858892   .1500972    -5.90   0.000    -1.180358   -.5914204
         educ |   .0525291   .0066546     7.89   0.000     .0394737    .0655844
              |
female#c.educ |
            1 |   .0267217   .0116879     2.29   0.022     .0037918    .0496516
              |
        _cons |   1.188924   .0853547    13.93   0.000     1.021471    1.356378
-------------------------------------------------------------------------------
```

相互作用項：Stataでの扱い方

- Stataでは、相互作用項をあらかじめ作ることなしに、相互作用項を含むモデルを推定することができます。
- 連続変数の場合は　i. ではなく　c.をつけます。
- reg lwage i.female c.edu c.edu#i.female
- こういう具合です。
- このようにStataに推定させると、相互作用項を作らないでいいだけではなく、marginsというコマンドを使っていろいろなことができます。
- 例えば、教育年数が10年、11年、、、、25年のときの（対数）賃金の予測値を、<u>男女別</u>に計算したいとします。

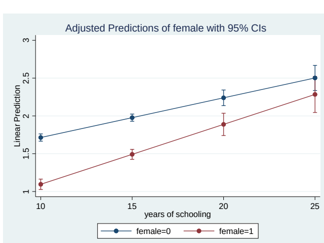
- margins female, at(educ=(10(1)25))
- この後に marginsplot　というコマンド打つと、次のような図を描くことができます。



- 5年ごとでいいなら、

margins female, at(educ=(10(5)25))

- その後に、marginsplotすると、



連続変数と連続変数の相互作用項

- ここでは次のモデルを考えましょう。

$$\ln(WAGE) = \beta_0 + \beta_1 EDUC + \beta_2 EXPER + U$$

- $FEMALE$や$EDUC \times FEMALE$などはモデルから落としました。
  - これは話を簡単にするためです。
  - 実際に分析する際は、落とさなくて構いません。
- $EXPER$はこれまでの勤務年数です。
- このモデルでは、勤務年数が1年増えることの賃金への影響は、教育年数に依存して変わることはありません。

- しかし現実には、この二つの変数には相互にかかわりがあり、勤務年数の賃金への効果は教育年数の値に依存するかもしれません。

- この相互依存は、モデルに$EDUC$と$EXPER$の相互作用項を追加することでモデル化できます。

- すなわち

$$\ln(WAGE) = \beta_0 + \beta_1 EDUC + \beta_2 EXPER + \beta_3 (EDUC \times EXPER) + U$$

ですね。

- これにより、$EXPER$が賃金に与える影響が、$EDUC$の水準によって変化する可能性を考慮することができます。

- 条件付き期待値を見てみましょう。

37

$$E(ln(WAGE)|EDUC, EXPER) = \beta_0 + \beta_1 EDUC + \beta_2 EXPER + \beta_3 (EDUC \times EXPER)$$

- $EDUC$を一定にしたときの、$EXPER$の一年の変化が（対数）賃金の期待値に与える影響は

$$\frac{\partial\, E(ln(WAGE)|EDUC, EXPER)}{\partial EXPER} = \beta_2 + \beta_3 EDUC$$

となり、$EDUC$の水準に依存にします。

- もし$\beta_3 > 0$なら、教育を受けた年数が多いほど、勤務年数が賃金へ及ぼす効果は大きくなります。

- 同様に、$EXPER$一定のもと、$EDUC$の一年の変化が（対数）賃金の期待値に与える影響は、

$$\frac{\partial\, E(ln(WAGE)|EDUC, EXPER)}{\partial EDUC} = \beta_1 + \beta_3 EXPER$$

38

- これら二つの効果をあわせて考えると、相互作用項にかかる係数β_3は、$EDUC$と$EXPER$がともに変化するときの影響に相当します。

- それは、「$EDUC$の単独の効果」と「$EXPER$の単独の効果」の<u>合計を超える部分</u>になります。

- すなわち$EDUC$が一年変化し<u>かつ</u>$EXPER$が一年変化したとき、予測される$ln(WAGE)$の変化は、

$$(\beta_1 + \beta_3 EXPER) + (\beta_2 + \beta_3 EDUC) + \beta_3$$

になるということです。

39

- 推定してみましょう。

- 一つの方法は、generateコマンドを使って$EDUC \times EXPER$という変数を作り推定します。

- $EDUC_EXPER$という変数名にします。

```
. generate educ_exper = educ*exper

. reg lwage educ exper educ_exper

      Source |       SS           df       MS      Number of obs   =     1,260
-------------+----------------------------------   F(3, 1256)      =    106.21
       Model |  90.039888         3  30.013296   Prob > F        =    0.0000
    Residual |  354.940084     1,256  .282595609   R-squared       =    0.2023
-------------+----------------------------------   Adj R-squared   =    0.2004
       Total |  444.979972     1,259  .353439215   Root MSE        =     .5316

------------------------------------------------------------------------------
       lwage |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
        educ |   .0707851   .0117582     6.02   0.000     .0477172     .093853
       exper |   .0155045   .0058541     2.65   0.008     .0040197    .0269894
  educ_exper |   .0002343   .0004735     0.49   0.621    -.0006946    .0011631
       _cons |   .4349911   .1511004     2.88   0.004     .1385541    .7314282
------------------------------------------------------------------------------
```

- $EDUC$と$EXPER$の相互作用はなさそうですね。

40

- もう一つの方法は、c. を使う方法です。」

```
. reg lwage c.educ c.exper c.educ#c.exper

      Source |       SS           df       MS      Number of obs   =     1,260
-------------+----------------------------------   F(3, 1256)      =    106.21
       Model |  90.039888         3  30.013296   Prob > F        =    0.0000
    Residual |  354.940084     1,256  .282595609   R-squared       =    0.2023
-------------+----------------------------------   Adj R-squared   =    0.2004
       Total |  444.979972     1,259  .353439215   Root MSE        =     .5316

------------------------------------------------------------------------------
       lwage |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
        educ |   .0707851   .0117582     6.02   0.000     .0477172     .093853
       exper |   .0155045   .0058541     2.65   0.008     .0040197    .0269894
             |
c.educ#c.exper |   .0002343   .0004735     0.49   0.621    -.0006946    .0011631
             |
       _cons |   .4349911   .1511004     2.88   0.004     .1385541    .7314282
------------------------------------------------------------------------------
```

- こちらの方法を使った場合、marginsコマンドが使えるので、いろいろなことができます。

- 例えば、

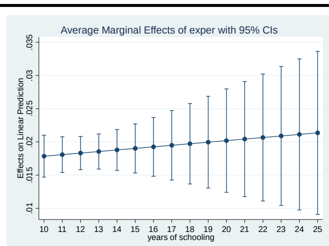
$$\frac{\partial\, E(ln(WAGE)|EDUC, EXPER)}{\partial EXPER} = \beta_2 + \beta_3 EDUC$$

を計算できます。

- 教育年数が10年、11年、、、、25年のときの$EXPER$の効果を計算してみましょう。

- margins, dydx(exper) at(educ=(10(1)25))

- これをmarginsplotで図にすることができます。



- 信頼区間を考慮に入れると、$EXPER$の効果が$EDUC$に依存しているとは言い難いですね。

- これは「相互作用項の係数がゼロ」という帰無仮説を棄却できなかったことと整合的ですね。

- 次に、

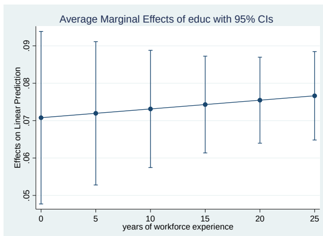
$$\frac{\partial\, E(ln(WAGE)|EDUC, EXPER)}{\partial EDUC} = \beta_1 + \beta_3 EXPER$$

を計算してみましょう。

- 勤務年数が0年、5年、10年、、、25年のときの$EDUC$の効果を計算してみましょう。

- margins, dydx(educ) at(exper=(0(5)25))

- これもmarginsplotで図にすることができます。



- 信頼区間を考慮に入れると、$EDUC$の効果が$EXPER$に依存しているとは言い難いですね。

- これは「相互作用項の係数がゼロ」という帰無仮説を棄却できなかったことと整合的ですね。

45