Quantitative Methods in Political Science

Recitation

Mai Nguyen

New York University

October 16, 2013

Review from Last Week's Lab Session

- Regression

Review from Last Week's Lab Session

- Regression
- Bivariate Regressions in Stata using the *regress* command

Review from Last Week's Lab Session

- Regression
- Bivariate Regressions in Stata using the *regress* command
- Interpretation of regression output

Review from Last Week's Lab Session

- Regression
- Bivariate Regressions in Stata using the *regress* command
- Interpretation of regression output
    - Slope (coefficient)

Review from Last Week's Lab Session

- Regression
- Bivariate Regressions in Stata using the *regress* command
- Interpretation of regression output
  - Slope (coefficient)
  - Intercept (constant)

Review from Last Week's Lab Session

- Regression
- Bivariate Regressions in Stata using the *regress* command
- Interpretation of regression output
  - Slope (coefficient)
  - Intercept (constant)
  - R-squared

Review from Last Week's Lab Session

- Regression
- Bivariate Regressions in Stata using the *regress* command
- Interpretation of regression output
  - Slope (coefficient)
  - Intercept (constant)
  - R-squared
  - Number of observations

Review from Last Week's Lab Session

- Regression
- Bivariate Regressions in Stata using the *regress* command
- Interpretation of regression output
  - Slope (coefficient)
  - Intercept (constant)
  - R-squared
  - Number of observations
- Scatter plots with a fitted line using *lfit*

This Week...

- More on bivariate regressions

This Week...

- More on bivariate regressions
- Pairwise correlations

This Week...

- More on bivariate regressions
- Pairwise correlations
- Depicting correlations in a plot

This Week...

- More on bivariate regressions
- Pairwise correlations
- Depicting correlations in a plot
- Multiple regression

Quick Recap of Bivariate Regressions

As you recall for bivariate regressions in Stata we use the *regress* command:

- *regress* **yvariablename xvariablename**

Quick Recap of Bivariate Regressions

As you recall for bivariate regressions in Stata we use the *regress* command:

- *regress* **yvariablename xvariablename**
- Let's try *regress gdppc openk*

Quick Recap of Bivariate Regressions

As you recall for bivariate regressions in Stata we use the *regress* command:

- *regress* **yvariablename xvariablename**
- Let's try *regress gdppc openk*

```
. regress gdppc openk

      Source |       SS       df       MS              Number of obs =     132
-------------+------------------------------           F(  1,   130) =    2.63
       Model |  94269760.9     1  94269760.9           Prob > F      =  0.1070
    Residual |  4.6531e+09   130  35793061.5           R-squared     =  0.0199
-------------+------------------------------           Adj R-squared =  0.0123
       Total |  4.7474e+09   131  36239448.5           Root MSE      =  5982.7

------------------------------------------------------------------------------
       gdppc |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
       openk |   21.60243   13.31116     1.62   0.107    -4.732103    47.93696
       _cons |   4129.315   1090.924     3.79   0.000     1971.052    6287.579
------------------------------------------------------------------------------
```

Quick Recap of Bivariate Regressions

- Recall that from our output we can create our regression line equation and interpret each part:

Quick Recap of Bivariate Regressions

- Recall that from our output we can create our regression line equation and interpret each part:

  $\hat{y} = a + bx$

Quick Recap of Bivariate Regressions

- Recall that from our output we can create our regression line equation and interpret each part:

$$\hat{y} = a + bx$$
$$\hat{y} = 4129 + 22x$$

Quick Recap of Bivariate Regressions

- Recall that from our output we can create our regression line equation and interpret each part:

$$\hat{y} = a + bx$$
$$\hat{y} = 4129 + 22x$$
$$\widehat{IncomePerCapita} = 4129 + 22 * OpennessToTrade$$

Quick Recap of Bivariate Regressions

- Recall that from our output we can create our regression line equation and interpret each part:

$$\hat{y} = a + bx$$
$$\hat{y} = 4129 + 22x$$
$$\widehat{IncomePerCapita} = 4129 + 22 * OpennessToTrade$$

- Anyone want to try interpreting each part of this?

More on Correlations

- A couple of weeks ago we learned the *correlate* (or *corr*) command in Stata, let's work with that some more:

More on Correlations

- A couple of weeks ago we learned the *correlate* (or *corr*) command in Stata, let's work with that some more:
  - *corr gdppc agehinst openk*

More on Correlations

- A couple of weeks ago we learned the *correlate* (or *corr*) command in Stata, let's work with that some more:
  - *corr gdppc agehinst openk*

```
. corr gdppc agehinst openk
(obs=132)

             |    gdppc agehinst    openk
-------------+---------------------------
       gdppc |   1.0000
    agehinst |   0.7347   1.0000
       openk |   0.1409  -0.0111   1.0000
```

More on Correlations

- A couple of weeks ago we learned the *correlate* (or *corr*) command in Stata, let's work with that some more:
  - *corr gdppc agehinst openk*

```
. corr gdppc agehinst openk
(obs=132)

             |    gdppc agehinst    openk
-------------+---------------------------
       gdppc |   1.0000
    agehinst |   0.7347   1.0000
       openk |   0.1409  -0.0111   1.0000
```

- Now let's try a new command *pwcorr* which gives us pairwise correlations:

More on Correlations

- A couple of weeks ago we learned the *correlate* (or *corr*) command in Stata, let's work with that some more:
  - *corr gdppc agehinst openk*

```
. corr gdppc agehinst openk
(obs=132)

             |    gdppc agehinst    openk
-------------+---------------------------
       gdppc |   1.0000
    agehinst |   0.7347   1.0000
       openk |   0.1409  -0.0111   1.0000
```

- Now let's try a new command *pwcorr* which gives us pairwise correlations:
  - *pwcorr gdppc agehinst openk*

More on Correlations

- A couple of weeks ago we learned the *correlate* (or *corr*) command in Stata, let's work with that some more:
  - *corr gdppc agehinst openk*

```
. corr gdppc agehinst openk
(obs=132)
```

	gdppc	agehinst	openk
gdppc	1.0000		
agehinst	0.7347	1.0000	
openk	0.1409	-0.0111	1.0000

- Now let's try a new command *pwcorr* which gives us pairwise correlations:
  - *pwcorr gdppc agehinst openk*

```
. pwcorr gdppc agehinst openk
```

	gdppc	agehinst	openk
gdppc	1.0000		
agehinst	0.7294	1.0000	
openk	0.1409	0.0061	1.0000

More on Correlations

- A couple of weeks ago we learned the *correlate* (or *corr*) command in Stata, let's work with that some more:
  - *corr gdppc agehinst openk*

```
. corr gdppc agehinst openk
(obs=132)

             |    gdppc agehinst    openk
-------------+---------------------------
       gdppc |   1.0000
    agehinst |   0.7347   1.0000
       openk |   0.1409  -0.0111   1.0000
```

- Now let's try a new command *pwcorr* which gives us pairwise correlations:
  - *pwcorr gdppc agehinst openk*

```
. pwcorr gdppc agehinst openk

             |    gdppc agehinst    openk
-------------+---------------------------
       gdppc |   1.0000
    agehinst |   0.7294   1.0000
       openk |   0.1409   0.0061   1.0000
```

- Notice any difference?

More on Correlations

We can depict these pairwise correlations in one plot:

- Type *graph matrix* **yvariablename xvariablename1 xvariablename2**

More on Correlations

We can depict these pairwise correlations in one plot:

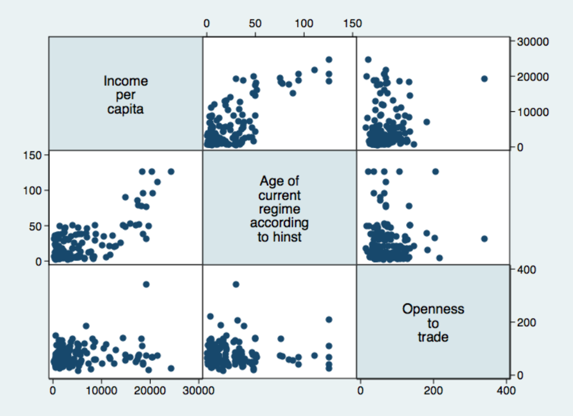
- Type *graph matrix* **yvariablename xvariablename1 xvariablename2**
- Try *graph matrix gdppc agehinst openk*

More on Correlations

We can depict these pairwise correlations in one plot:

- Type *graph matrix* **yvariablename xvariablename1 xvariablename2**
- Try *graph matrix gdppc agehinst openk*
- The plot itself may be a little confusing so let's go over it...

More on Correlations



More on Correlations

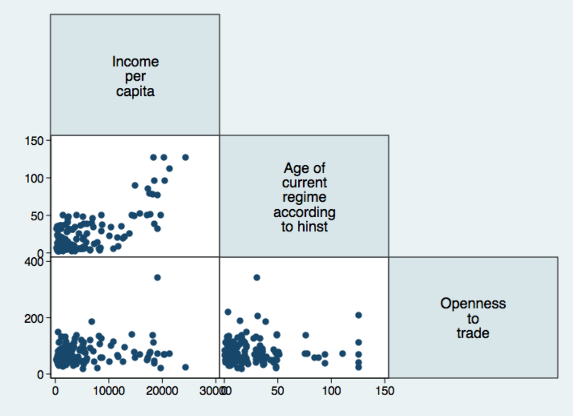
You also have the option of just showing half of the plot:

- Type: *graph matrix gdppc agehinst openk, half*

More on Correlations

You also have the option of just showing half of the plot:

- Type: *graph matrix gdppc agehinst openk, half*



Multiple Regression

Now the moment everyone has been waiting for...multiple regressions!

- Before I show you the mechanics of how to run multiple regressions, think about why we would even want to in the first place. Why would it be important to analyze multiple variables?

Multiple Regression

Now the moment everyone has been waiting for...multiple regressions!

- Before I show you the mechanics of how to run multiple regressions, think about why we would even want to in the first place. Why would it be important to analyze multiple variables?
- The format for multiple regression in Stata is identical to the format for bivariate regressions

Multiple Regression

Now the moment everyone has been waiting for...multiple regressions!

- Before I show you the mechanics of how to run multiple regressions, think about why we would even want to in the first place. Why would it be important to analyze multiple variables?
- The format for multiple regression in Stata is identical to the format for bivariate regressions
- Type *regress* **yvariablename xvariablename1 xvariablename2** and so on

Multiple Regression

Now the moment everyone has been waiting for...multiple regressions!

- Before I show you the mechanics of how to run multiple regressions, think about why we would even want to in the first place. Why would it be important to analyze multiple variables?
- The format for multiple regression in Stata is identical to the format for bivariate regressions
- Type *regress* **yvariablename xvariablename1 xvariablename2** and so on
- Try *regress gdppc agehinst openk*

Multiple Regression

Now the moment everyone has been waiting for...multiple regressions!

- Before I show you the mechanics of how to run multiple regressions, think about why we would even want to in the first place. Why would it be important to analyze multiple variables?
- The format for multiple regression in Stata is identical to the format for bivariate regressions
- Type *regress* **yvariablename xvariablename1 xvariablename2** and so on
- Try *regress gdppc agehinst openk*
- As you can see, the output also looks similar to the bivariate regression output. Let's go over it again.

Multiple Regression

```
. regress gdppc agehinst openk

      Source |       SS       df       MS              Number of obs =     132
-------------+------------------------------           F(  2,   129) =   82.76
       Model |  2.6680e+09     2  1.3340e+09           Prob > F      =  0.0000
    Residual |  2.0793e+09   129  16118865.4           R-squared     =  0.5620
-------------+------------------------------           Adj R-squared =  0.5552
       Total |  4.7474e+09   131  36239448.5           Root MSE      =  4014.8

------------------------------------------------------------------------------
       gdppc |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
    agehinst |   161.8961   12.81207    12.64   0.000     136.5471    187.2451
       openk |   22.85708    8.93327     2.56   0.012     5.182386    40.53177
       _cons |   101.9413   798.4555     0.13   0.899    -1477.822    1681.705
------------------------------------------------------------------------------
```

Multiple Regression

- Just as before, we can use the information from the output to create our regression line

Multiple Regression

- Just as before, we can use the information from the output to create our regression line
- $\hat{y} = 102 + 162x_1 + 23x_2$

Multiple Regression

- Just as before, we can use the information from the output to create our regression line
- $\hat{y} = 102 + 162x_1 + 23x_2$
- $\widehat{IncomePerCapita} = 102 + 162 * RegimeAge + 23 * TradeOpenness$

- **The slope for x_1**: For each [unit] increase in [x_1], [y] is expected to [increase/decrease] on average by [the slope], controlling for [x_2]. This is the effect of [x_1] on [y], holding [x_2] constant.

Multiple Regression

- **The slope for** x_1: For each [unit] increase in [x_1], [y] is expected to [increase/decrease] on average by [the slope], controlling for [x_2]. This is the effect of [x_1] on [y], holding [x_2] constant.
- **The slope for** x_2: For each [unit] increase in [x_2], [y] is expected to [increase/decrease] on average by [the slope], controlling for [x_1]. This is the effect of [x_2] on [y], holding [x_1] constant.

Multiple Regression

- **The slope for x_1**: For each [unit] increase in [x_1], [y] is expected to [increase/decrease] on average by [the slope], controlling for [x_2]. This is the effect of [x_1] on [y], holding [x_2] constant.
- **The slope for x_2**: For each [unit] increase in [x_2], [y] is expected to [increase/decrease] on average by [the slope], controlling for [x_1]. This is the effect of [x_2] on [y], holding [x_1] constant.
- **The y-intercept**: when [x_1 and x_2] are 0, [y] is expected to equal the [y-intercept].

Multiple Regression

- **The slope for** x_1: For each [unit] increase in [x_1], [y] is expected to [increase/decrease] on average by [the slope], controlling for [x_2]. This is the effect of [x_1] on [y], holding [x_2] constant.
- **The slope for** x_2: For each [unit] increase in [x_2], [y] is expected to [increase/decrease] on average by [the slope], controlling for [x_1]. This is the effect of [x_2] on [y], holding [x_1] constant.
- **The y-intercept**: when [x_1 and x_2] are 0, [y] is expected to equal the [y-intercept].
- **(Adjusted) R-squared**: [XXX] % of the variability in [y] is explained by both [x_1 and x_2] (or by the model).

Multiple Regression

Now it's your turn. What if I wanted to know the effect of regime age and investment on growth?

- What code would I use?

Multiple Regression

Now it's your turn. What if I wanted to know the effect of regime age and investment on growth?

- What code would I use?
- What would my regression line equation look like?

Multiple Regression

Now it's your turn. What if I wanted to know the effect of regime age and investment on growth?

- What code would I use?
- What would my regression line equation look like?
- Interpret the intercept.

Multiple Regression

Now it's your turn. What if I wanted to know the effect of regime age and investment on growth?

- What code would I use?
- What would my regression line equation look like?
- Interpret the intercept.
- Interpret the slope for x_1.

Multiple Regression

Now it's your turn. What if I wanted to know the effect of regime age and investment on growth?

- What code would I use?
- What would my regression line equation look like?
- Interpret the intercept.
- Interpret the slope for x_1.
- Interpret the slope for x_2.

Multiple Regression

Now it's your turn. What if I wanted to know the effect of regime age and investment on growth?

- What code would I use?
- What would my regression line equation look like?
- Interpret the intercept.
- Interpret the slope for x_1.
- Interpret the slope for x_2.
- Interpret the R-squared.

Multiple Regression

```
. regress growth agehinst investment

      Source |       SS       df       MS              Number of obs =     132
-------------+------------------------------           F(  2,   129) =    1.05
       Model |  85.0463551     2  42.5231776           Prob > F      =  0.3531
    Residual |  5226.84203   129  40.5181553           R-squared     =  0.0160
-------------+------------------------------           Adj R-squared =  0.0008
       Total |  5311.88839   131  40.5487663           Root MSE      =  6.3654

------------------------------------------------------------------------------
      growth |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
    agehinst |  -.0070879   .0214204    -0.33   0.741    -.0494686    .0352929
  investment |   .0980832   .0680047     1.44   0.152    -.0364658    .2326321
       _cons |    1.19037   1.130256     1.05   0.294    -1.045868    3.426609
------------------------------------------------------------------------------
```

Multiple Regression

- Code: *regress growth agehinst investment*

Multiple Regression

- Code: *regress growth agehinst investment*
- Regression line: $\hat{y} = 1.190 - 0.007x_1 + 0.098x_2$

Multiple Regression

- Code: *regress growth agehinst investment*
- Regression line: $\hat{y} = 1.190 - 0.007x_1 + 0.098x_2$
- Or, $\widehat{Growth} = 1.190 - 0.007 * RegimeAge + 0.098 Investment$

Multiple Regression

- Code: *regress growth agehinst investment*
- Regression line: $\hat{y} = 1.190 - 0.007x_1 + 0.098x_2$
- Or, $\widehat{Growth} = 1.190 - 0.007 * RegimeAge + 0.098 Investment$
- Intercept: When regime age and investment are 0, growth is expected to equal 1.190.

Multiple Regression

- Code: *regress growth agehinst investment*
- Regression line: $\hat{y} = 1.190 - 0.007x_1 + 0.098x_2$
- Or, $\widehat{Growth} = 1.190 - 0.007 * RegimeAge + 0.098 Investment$
- Intercept: When regime age and investment are 0, growth is expected to equal 1.190.
- Slope of x_1: For each year increase in regime age, growth is expected to decrease on average by 0.007, controlling for investment. This is the effect of regime age on growth, holding investment constant.

Multiple Regression

- Code: *regress growth agehinst investment*
- Regression line: $\hat{y} = 1.190 - 0.007x_1 + 0.098x_2$
- Or, $\widehat{Growth} = 1.190 - 0.007 * RegimeAge + 0.098 Investment$
- Intercept: When regime age and investment are 0, growth is expected to equal 1.190.
- Slope of x_1: For each year increase in regime age, growth is expected to decrease on average by 0.007, controlling for investment. This is the effect of regime age on growth, holding investment constant.
- Slope of x_2: For each percent increase in investment, growth is expected to increase on average by 0.098, controlling for regime age. This is the effect of investment on growth, holding regime age constant.

Multiple Regression

- Code: *regress growth agehinst investment*
- Regression line: $\hat{y} = 1.190 - 0.007x_1 + 0.098x_2$
- Or, $\widehat{Growth} = 1.190 - 0.007 * RegimeAge + 0.098 Investment$
- Intercept: When regime age and investment are 0, growth is expected to equal 1.190.
- Slope of x_1: For each year increase in regime age, growth is expected to decrease on average by 0.007, controlling for investment. This is the effect of regime age on growth, holding investment constant.
- Slope of x_2: For each percent increase in investment, growth is expected to increase on average by 0.098, controlling for regime age. This is the effect of investment on growth, holding regime age constant.
- R-squared: 1.6% of the variability in growth is explained by both regime age and investment.