# Quantitative Methods in Political Science

## Recitation

Mai Nguyen

New York University

October 7, 2013

# Review from Last Week's Lab Session

- Correlation Analysis

# Review from Last Week's Lab Session

- Correlation Analysis
- Scatterplots

# Review from Last Week's Lab Session

- Correlation Analysis
- Scatterplots
- Creating and recoding variables

# Review from Last Week's Lab Session

- Correlation Analysis
- Scatterplots
- Creating and recoding variables
  - *generate* **newvariablename=something**

# Review from Last Week's Lab Session

- Correlation Analysis
- Scatterplots
- Creating and recoding variables
  - *generate* **newvariablename=something**
  - *recode* **variablename something**

# Review from Last Week's Lab Session

- Correlation Analysis
- Scatterplots
- Creating and recoding variables
  - *generate* **newvariablename=something**
  - *recode* **variablename something**
  - Addressing problems with missing data

# Review from Last Week's Lab Session

- Correlation Analysis
- Scatterplots
- Creating and recoding variables
  - *generate* **newvariablename=something**
  - *recode* **variablename something**
  - Addressing problems with missing data
    - Recall that missing data is represented by "." in Stata

# This Week...

- Review of regression

# This Week...

- Review of regression
- Bivariate regressions in Stata

# This Week...

- Review of regression
- Bivariate regressions in Stata
- Interpretation of regression output

# This Week...

- Review of regression
- Bivariate regressions in Stata
- Interpretation of regression output
- More work with scatterplots

# Getting Started

- Open up the Przeworski dataset

# Getting Started

- Open up the Przeworski dataset
- If you didn't save your dataset from last week, make sure you recode the necessary variables:

# Getting Started

- Open up the Przeworski dataset
- If you didn't save your dataset from last week, make sure you recode the necessary variables:
  - *recode gdppc min/0=.*

# Getting Started

- Open up the Przeworski dataset
- If you didn't save your dataset from last week, make sure you recode the necessary variables:
  - *recode gdppc min/0=.*
  - *recode investment -999=.*

# Getting Started

- Open up the Przeworski dataset
- If you didn't save your dataset from last week, make sure you recode the necessary variables:
  - *recode gdppc min/0=.*
  - *recode investment -999=.*
  - What does this do again?

# Regression

- What is a regression?

# Regression

- What is a regression?

$$\hat{y} = a + bx \tag{1}$$

# Regression

- What is a regression?

$$\hat{y} = a + bx \tag{1}$$

  - $\hat{y}$ is the predicted value of y

# Regression

- What is a regression?

$$\hat{y} = a + bx \tag{1}$$

  - $\hat{y}$ is the predicted value of y
  - $a$ is the intercept

# Regression

- What is a regression?

$$\hat{y} = a + bx \tag{1}$$

  - $\hat{y}$ is the predicted value of y
  - $a$ is the intercept
  - $b$ is the slope

# Regression

Bivariate regressions in Stata:

- The basic command for running a regression in Stata is *regress*

# Regression

Bivariate regressions in Stata:

- The basic command for running a regression in Stata is *regress*
  - Type *regress* **yvariablename xvariablename**

# Regression

Bivariate regressions in Stata:

- The basic command for running a regression in Stata is *regress*
  - Type *regress* **yvariablename xvariablename**
  - Example: *regress gdppc agehinst*

# Regression

Bivariate regressions in Stata:

- The basic command for running a regression in Stata is *regress*
  - Type *regress* **yvariablename xvariablename**
  - Example: *regress gdppc agehinst*
  - As always you can shorten the command in Stata and use *reg* (the command then would be *reg gdppc agehinst*)

# Regression

Bivariate regressions in Stata:

- The basic command for running a regression in Stata is *regress*
  - Type *regress* **yvariablename xvariablename**
  - Example: *regress gdppc agehinst*
  - As always you can shorten the command in Stata and use *reg* (the command then would be *reg gdppc agehinst*)
- How do we interpret the results? Let's take a look at the output.

# Regression

```
. regress gdppc agehinst
```

| Source | SS | df | MS |
|---|---|---|---|
| Model | 2.7428e+09 | 1 | 2.7428e+09 |
| Residual | 2.4133e+09 | 152 | 15876745.7 |
| Total | 5.1561e+09 | 153 | 33699753.5 |

| | |
|---|---|
| Number of obs | 154 |
| F( 1, 152) | 172.76 |
| Prob > F | 0.0000 |
| R-squared | 0.5320 |
| Adj R-squared | 0.5289 |
| Root MSE | 3984.6 |

| gdppc | Coef. | Std. Err. | t | P>\|t\| | [95% Conf. | Interval] |
|---|---|---|---|---|---|---|
| agehinst | 162.9329 | 12.39631 | 13.14 | 0.000 | 138.4415 | 187.4242 |
| _cons | 1543.919 | 433.1102 | 3.56 | 0.000 | 688.2255 | 2399.612 |

# Regression

From the output we can create our regression line:

- The slope is given by the coefficient for *agehinst*: 162.9329 or we'll say 163

# Regression

From the output we can create our regression line:

- The slope is given by the coefficient for *agehinst*: 162.9329 or we'll say 163
- The intercept is given by the constant: 1543.919 or we'll say 1544

# Regression

From the output we can create our regression line:

- The slope is given by the coefficient for *agehinst*: 162.9329 or we'll say 163
- The intercept is given by the constant: 1543.919 or we'll say 1544
- We can plug this into our general regression line formula

$$\hat{y} = 1544 + 163x$$

$$\widehat{IncomePerCapita} = 1544 + 163 * AgeCurrentRegime$$

# Regression

We can interpret using the templates from class:

- Slope:

# Regression

We can interpret using the templates from class:

- Slope:
  - For each [unit] increase in [x], y is expected to [increase/decrease] on average by [the slope].

# Regression

We can interpret using the templates from class:

- Slope:
  - For each [unit] increase in [x], y is expected to [increase/decrease] on average by [the slope].
- Intercept:

# Regression

We can interpret using the templates from class:

- Slope:
  - For each [unit] increase in [x], y is expected to [increase/decrease] on average by [the slope].
- Intercept:
  - When [x] is 0, [y] is expected to equal the [y-intercept].

# Regression

We can interpret using the templates from class:

- Slope:
  - For each [unit] increase in [x], y is expected to [increase/decrease] on average by [the slope].
- Intercept:
  - When [x] is 0, [y] is expected to equal the [y-intercept].
- R-squared (proportion of variability of dependent variable that is accounted for by the independent variable (or the model):

# Regression

We can interpret using the templates from class:

- Slope:
  - For each [unit] increase in [x], y is expected to [increase/decrease] on average by [the slope].
- Intercept:
  - When [x] is 0, [y] is expected to equal the [y-intercept].
- R-squared (proportion of variability of dependent variable that is accounted for by the independent variable (or the model):
  - So [XXX] % of the variability in y is explained by x (or by the model).

# Regression

Your turn. How does investment affect income per capita?

- Answer the research question by running a bivariate regression

# Regression

Your turn. How does investment affect income per capita?

- Answer the research question by running a bivariate regression
- Construct the regression equation line

# Regression

Your turn. How does investment affect income per capita?

- Answer the research question by running a bivariate regression
- Construct the regression equation line
- Interpret the slope

# Regression

Your turn. How does investment affect income per capita?

- Answer the research question by running a bivariate regression
- Construct the regression equation line
- Interpret the slope
- Interpret the intercept

# Regression

Your turn. How does investment affect income per capita?

- Answer the research question by running a bivariate regression
- Construct the regression equation line
- Interpret the slope
- Interpret the intercept
- Interpret the R-squared

# Regression

```
. regress gdppc investment
```

| Source | SS | df | MS |
|---|---|---|---|
| Model | 1.7454e+09 | 1 | 1.7454e+09 |
| Residual | 2.9898e+09 | 129 | 23176820.5 |
| Total | 4.7352e+09 | 130 | 36424558.6 |

| | |
|---|---|
| Number of obs | 131 |
| F( 1, 129) | 75.31 |
| Prob > F | 0.0000 |
| R-squared | 0.3686 |
| Adj R-squared | 0.3637 |
| Root MSE | 4814.2 |

| gdppc | Coef. | Std. Err. | t | P>\|t\| | [95% Conf. | Interval] |
|---|---|---|---|---|---|---|
| investment | 424.7486 | 48.94559 | 8.68 | 0.000 | 327.9086 | 521.5887 |
| _cons | -613.219 | 841.499 | -0.73 | 0.467 | -2278.145 | 1051.707 |

# More on scatterplots

Recall last week we made scatterplots using the *scatter* command.

- Wouldn't it be nice if we could visualize our regression line and add it to a scatterplot?

# More on scatterplots

Recall last week we made scatterplots using the *scatter* command.

- Wouldn't it be nice if we could visualize our regression line and add it to a scatterplot?
- We can do just that:

# More on scatterplots

Recall last week we made scatterplots using the *scatter* command.

- Wouldn't it be nice if we could visualize our regression line and add it to a scatterplot?
- We can do just that:
  - Type: *twoway* (*scatter* **yvariablename xvariablename**) (*lfit* **yvariablename xvariablename**)

# More on scatterplots

Recall last week we made scatterplots using the *scatter* command.

- Wouldn't it be nice if we could visualize our regression line and add it to a scatterplot?
- We can do just that:
  - Type: *twoway* (*scatter* **yvariablename xvariablename**) (*lfit* **yvariablename xvariablename**)
  - Example: *twoway (scatter gdppc agehinst) (lfit gdppc agehinst)*

# More on scatterplots

Recall last week we made scatterplots using the *scatter* command.

- Wouldn't it be nice if we could visualize our regression line and add it to a scatterplot?
- We can do just that:
  - Type: *twoway* (*scatter* **yvariablename xvariablename**) (*lfit* **yvariablename xvariablename**)
  - Example: *twoway (scatter gdppc agehinst) (lfit gdppc agehinst)*
- Additionally, we can add options to make a more complete graph:

# More on scatterplots

Recall last week we made scatterplots using the *scatter* command.

- Wouldn't it be nice if we could visualize our regression line and add it to a scatterplot?
- We can do just that:
  - Type: *twoway* (*scatter* **yvariablename xvariablename**) (*lfit* **yvariablename xvariablename**)
  - Example: *twoway (scatter gdppc agehinst) (lfit gdppc agehinst)*
- Additionally, we can add options to make a more complete graph:
  - *twoway* (*scatter* **yvariablename xvariablename**) (*lfit* **yvariablename xvariablename**), *title( ) ytitle( ) xtitle( )*

# More on scatterplots

Recall last week we made scatterplots using the *scatter* command.

- Wouldn't it be nice if we could visualize our regression line and add it to a scatterplot?
- We can do just that:
  - Type: *twoway* (*scatter* **yvariablename xvariablename**) (*lfit* **yvariablename xvariablename**)
  - Example: *twoway (scatter gdppc agehinst) (lfit gdppc agehinst)*
- Additionally, we can add options to make a more complete graph:
  - *twoway* (*scatter* **yvariablename xvariablename**) (*lfit* **yvariablename xvariablename**), *title( ) ytitle( ) xtitle( )*
  - Example: *twoway (scatter gdppc agehinst) (lfit gdppc agehinst), title(Effect of Regime Age on Income per Capita) ytitle(Income per Capita) xtitle(Age of Current Regime)*

# More on scatterplots

Recall last week we made scatterplots using the *scatter* command.

- Wouldn't it be nice if we could visualize our regression line and add it to a scatterplot?
- We can do just that:
  - Type: *twoway* (*scatter* **yvariablename xvariablename**) (*lfit* **yvariablename xvariablename**)
  - Example: *twoway (scatter gdppc agehinst) (lfit gdppc agehinst)*
- Additionally, we can add options to make a more complete graph:
  - *twoway* (*scatter* **yvariablename xvariablename**) (*lfit* **yvariablename xvariablename**), *title( ) ytitle( ) xtitle( )*
  - Example: *twoway (scatter gdppc agehinst) (lfit gdppc agehinst), title(Effect of Regime Age on Income per Capita) ytitle(Income per Capita) xtitle(Age of Current Regime)*
- As always you can edit and save graphs in the "Graph Editor" window

# More on scatterplots