Lecture 6: Bivariate Regression, Part 2

Dr. Julie VanDusky-Allen[1]

[1]Boise State University

Slides prepared for the Political Science MA Program

Objectives

- Know how to run an ordinary least squares regression in STATA
- Know how to write a regression equation from regression output
- Know how to interpret the significance test for a coefficient in STATA
- Know how to interpret a significant coefficient for a continuous variable
- Know how to interpret a significant coefficient for a dummy variable

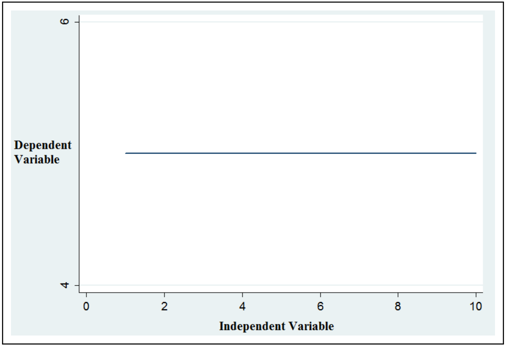
Objectives (cont.)

- Know how to interpret the significance test for a constant in STATA
- Know how to interpret a constant
- Know how to calculate predictions by hand and in STATA
- Know how to graph estimates in STATA
- Know how to calculate and interpret an R-square and adjusted R-square

Review: Bivariate Regression

- Independent Variable $\rightarrow$ Dependent Variable
- Run a regression
- $\widehat{\text{Dependent Variable}} = \hat{b}$*Independent Variable + $\hat{a}$
- $\hat{b}$ tells us the effect of the independent variable on the dependent variable

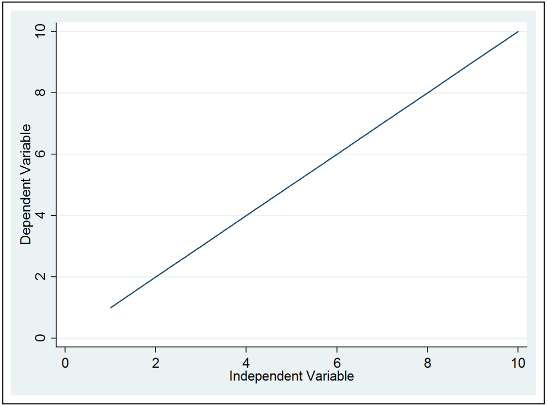
No Relationship



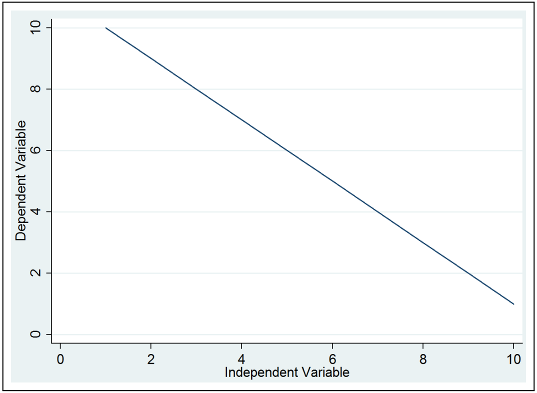
Is $\hat{b}$ Different from Zero?

- But we do not expect the coefficient to be 0. Instead, we expect it to either be greater than 0 (or positive) or less than 0 (negative).

Positive Linear Relationship



Negative Linear Relationship



Running a Regression in STATA

- Hypothesis: As the strength of an upper chamber increases, parties have more control over legislative candidate selection.
- Type “reg depvarname indvarname” into the command box and press enter
- The output should appear in the output window

Reading a Regression Output in STATA

- Focus on the bottom table
- The first column lists the independent variable and the constant ($\hat{a}$)
- The second column lists $\hat{b}$ for the independent variable and $\hat{a}$
- $\hat{b}$ = 2.058028 and $\hat{a}$ = 3.180699

Writing a Regression Output from STATA

- $\hat{y} = \hat{b}x + \hat{a} + \hat{e}$
- $\widehat{\text{Candidate Selection}} = \hat{b}$*Strength of Upper Chamber+ $\hat{a}$
- $\hat{b} = 2.058028$ and $\hat{a} = 3.180699$
- $\widehat{\text{Candidate Selection}}$ = 2.058028*Strength of Upper Chamber+ 3.180699

Example

- As the strength of upper chamber changes, party control over candidate selection stays the same.
  - $H_0 : \hat{b} = 0$
- As the strength of upper chamber increases, party control over candidate selection increases. (Positive relationship)
  - $H_1 : \hat{b} > 0$

How to Interpret Significance Tests for Coefficients in STATA

- Check the sign of $\hat{b}$ first
- If the sign of $\hat{b}$ is not in expected direction, $\hat{b}$ is not significant and is not distinguishable from 0
  - Example: You expected the independent variable to have a positive effect on the dependent variable, but $\hat{b}$ had a negative sign
  - You do not have support for your hypothesis
- If your $\hat{b}$ is in the right direction, move on to the next steps

How to Interpret Significance Tests for Coefficients in STATA (cont.)

- The P>t column tells you the results of the significance tests for ($\hat{b}$)
  - If:
  - $P < 0.002$, the coefficient is significant and different from 0 at the 0.001 level
  - $P < 0.02$, the coefficient is significant and different from 0 at the 0.01 level
  - $P < 0.10$, the coefficient is significant and different from 0 at the 0.05 level
  - $P > 0.10$, the coefficient is not significant and is indistinguishable from 0 (no support)

How to Interpret Significance Tests for Coefficients in STATA (cont.)

- Note: Anytime you see a P value of 0.000, it is less than 0.002, so you have significance at the 0.001 level

How to Interpret Significance Tests for Coefficients in STATA (cont.)

- $\hat{b}$ is positive, as expected
- The P value is less than 0.002, so $\hat{b}$ is significant at the 0.001 level

IF Significant Coefficient

- Results suggest there is a statistical relationship
- Indicate whether $\hat{b}$ is positive or negative
- Indicate what level it is significant at (0.001, 0.01, or 0.05 level)
- If positive $\hat{b}$
  - As the independent variable increases, the dependent variable increases
- If negative $\hat{b}$
  - As the independent variable increases, the dependent variable decreases

IF Insignificant Coefficient (cont.)

- Interpretation: The results of the data analysis suggest that there is a statistical relationship between Strength of Upper Chamber and Party Control Over Candidate Selection. The coefficient was positive and significant at the 0.001 level. As Strength of Upper Chamber increases, Party Control Over Candidate Selection increases.

IF Insignificant Coefficient (cont.)

- Either $\hat{b}$ is in the wrong direction or P $>$ 0.10
- Results do not provide support for your hypothesis
- Cannot make any conclusions

IF Significant Coefficient and Continuous

- If you find that the coefficient for your independent variable is significant, then you can make a specific prediction about how your independent variable should affect your dependent variable
- Continuous variable (0, 1, 2, 3, ...)

IF Significant Coefficient and Continuous (cont.)

- As the independent variable increases by 1 unit, the dependent variable increases/decreases by the value of the $\hat{b}$.
    - Increases if $\hat{b}$ is positive
    - Decreases if $\hat{b}$ is negative

IF Significant Coefficient and Continuous (cont.)

- Example
  - $\widehat{\text{Candidate Selection}}$ = 2.058028*Strength of Upper Chamber+ 3.180699
  - As the strength of upper chamber increases by 1, party control over candidate selection increases by 2.058028.

IF Significant Coefficient and Continuous (cont.)

- Example (cont.)
  - $\widehat{\text{Candidate Selection}}$ = 2.058028*Strength of Upper Chamber+ 3.180699
  - What if the variable changed by 0.5?
  - 2.058028*0.5 = 1.029014
  - As the strength of upper chamber increases by 0.5, party control over candidate selection increases by 1.029014.

IF Significant Coefficient and Continuous (cont.)

- Example (cont.)
  - $\widehat{\text{Candidate Selection}}$ = 2.058028*Strength of Upper Chamber+ 3.180699
  - What if the variable changed by 2?
  - 2.058028*2 = 4.116056
  - As the strength of upper chamber increases by 2, party control over candidate selection increases by 4.116056.

IF Significant Coefficient and a Dummy

- Dummy variable
  - Only 2 values, usually 0 and 1

IF Significant Coefficient and a Dummy (cont.)

- Example
  - Hypothesis: Parties in countries with bicameral legislatures have more control over candidate selection than parties in countries with unicameral legislatures.
  - Dependent variable: Party Control Over Candidate Selection
  - Independent variable: Type of legislature (dummy)
    - Unicameral (0)
    - Bicameral (1)

IF Significant Coefficient and a Dummy (cont.)

- $\widehat{\text{Candidate Selection}}$ = 1.450011*Type of Legislature+ 3.164835
- Divide the data into two groups based on the dummy variable coding
  - Parties in unicameral legislatures (0)
  - Parties in bicameral legislatures (1)
- $\hat{b}$ is a comparison of the dependent variable between the 1 and 0 categories
- $\hat{b}$ applies to the 1 category

IF Significant Coefficient and a Dummy (cont.)

- Interpretation: Parties in countries with bicameral legislatures have 1.450011 units more control over candidate selection than parties in unicameral legislatures.

Constant

- $\widehat{\text{Candidate Selection}}$ = 2.058028*Strength of Upper Chamber+ 3.180699
- $\hat{a}$ = 3.180699
- Negative or positve?

How to Interpret Significance Tests for the Constant in STATA

- The P>t column tells you the results of the significance tests for ($\hat{a}$)
  - If:
  - $P < 0.001$, the constant is significant and different from 0 at the 0.001 level
  - $P < 0.01$, the constant is significant and different from 0 at the 0.01 level
  - $P < 0.05$, the constant is significant and different from 0 at the 0.05 level
  - $P > 0.05$, the constant is not significant
- Note: Anytime you see a P value of 0.000, it is less than 0.001 and significant at the 0.001 level

How to Interpret Significance Tests for the Constant in STATA (cont.)

- Indicate if $\hat{a}$ is significant or not significant
- If significant, whether $\hat{a}$ is positive or negative, and at what level (0.001, 0.01, or 0.05)

How to Interpret Significance Tests for the Constant in STATA (cont.)

- Example: The constant is significant at the 0.001 level and positive.

How to Interpret a Constant

- When the independent variable is 0, the dependent variable is the value of $\hat{a}$.
- $\widehat{\text{Candidate Selection}}$ = 2.058028*Strength of Upper Chamber+ 3.180699
- $\widehat{\text{Candidate Selection}}$ = 2.058028*0 + 3.180699 = 3.180699
  - When Strength of Upper Chamber is 0, Party Control Over Candidate Selection is 3.180699.
  - Realistic? Can the independent variable take a value of 0?

Calculating Estimates of the Dependent Variable

- $\widehat{\text{Dependent Variable}} = \hat{b}$*Independent Variable + $\hat{a}$
- If the $\hat{b}$ is significant, calculate for each observation

Calculating Estimates of the Dependent Variable (cont.)

- Insert the value of “independent variable” and calculate
- $\widehat{\text{Candidate Selection}}$ = 2.058028*Strength of Upper Chamber+ 3.180699
  - Observation 1, Strength of Upper Chamber = 0.67
  - $\widehat{\text{Candidate Selection}}$ = 2.058028*0.67+ 3.180699 = 4.55957776
- If you calculated estimates for each observation, the mean of the estimates would be equal to the mean of the dependent variable.

Calculating Estimates of the Dependent Variable in STATA

- Run your regression
- Type “predict xb” into the command box and press enter
  - New variable called “xb” in your dataset

Graphing Estimates of the Dependent Variable in STATA

- Click on the “Graphics” button
- Choose “twoway graph”
- Click on the “Create” button
- Choose the plot type
- Select xb for the Y option and the independent variable for the X option
- Click “Accept”
- Click “OK”
- Copy and paste into Word

Graphing Estimates of the Dependent Variable in STATA (cont.)

- Editing in the “Graphics” interface
    - See Lecture 1

Intuition Behind the R-Square Value

- Variation in Independent Variable $\rightarrow$ Variation in Dependent Variable
- Percentage of variation in dependent variable explained

R-Square Values

- Range: 0 to 1
  - 1: Perfectly estimates
  - 0: No estimate ability

Equation

$$R^2 = 1 - \frac{\text{Sum of Squared Errors}}{\text{Sum of Squared Deviations from Mean}} = 1 - \frac{\sum(y - \hat{y})^2}{(\sum(y - \bar{y})^2)}$$

Step 1: Estimate Errors

- Difference between estimates and actual values
- Estimate error = Actual Dependent Variable - Estimated Dependent Variable
- $\hat{e} = y - \hat{y}$

Step 1: Calculate Estimate Errors (cont.)

Table :

Obs	Candidate Selection	Estimate	Error
1	6	4.559578	6-4.559578 = 1.440422
2	2	4.518417	2-4.518417 = -2.518417
3	4	5.238727	4-5.238727 = -1.238727
4	6	4.744801	6-4.744801 = 1.255199

Check your math!

- Mean of errors = 0
- $\overline{\hat{e}} = 0$

Step 2: Square Estimate Errors

Table :

Obs	Candidate Selection	Estimate	Error	$Error^2$
1	6	4.559578	1.440422	$1.440422^2 = 2.074816$
2	2	4.518417	-2.518417	$-2.518417^2 = 6.342426$
3	4	5.238727	-1.238727	$-1.238727^2 = 1.534445$
4	6	4.744801	1.255199	$1.255199^2 = 1.575526$

Step 3: Sum Squared Estimate Errors

Table :

Obs	Candidate Selection	Estimate	Error	*Error*^{2}
1	6	4.559578	1.440422	$1.440422^2 = 2.074816$
2	2	4.518417	-2.518417	$-2.518417^2 = 6.342426$
3	4	5.238727	-1.238727	$-1.238727^2 = 1.534445$
4	6	4.744801	1.255199	$1.255199^2 = 1.575526$
Numerator (note) = 2895.367				

Step 4: Subtract the Mean of the Dependent Variable from Each Dependent Variable

Table :

Obs	Candidate Selection	Estimate	Error	$Error^2$	DV-$\bar{DV}$
1	6	4.559578	1.440422	2.074816	6-4.213779 = 1.786221
2	2	4.518417	-2.518417	6.342426	2-4.213779 = -2.213779
3	4	5.238727	-1.238727	1.534445	4-4.213779 = -0.213779
4	6	4.744801	1.255199	1.575526	6-4.213779 = 1.786221
Numerator (note) = 2895.367					
$\overline{Y}$	4.213779				

Step 5: Square the Values from Step 4

Table :

Obs	Candidate Selection	Estimate	Error	$Error^2$	DV-$\bar{DV}$	DV-$\bar{DV}^2$
1	6	4.559578	1.440422	2.074816	1.786221	$1.786221^2 = 3.190586$
2	2	4.518417	-2.518417	6.342426	-2.213779	$-2.213779^2 = 4.900817$
3	4	5.238727	-1.238727	1.534445	-0.213779	$-0.213779^2 = 0.0457015$
4	6	4.744801	1.255199	1.575526	1.786221	$1.786221^2 = 3.190586$
Numerator (note) = 2895.367						
$\overline{Y}$	4.213779					

Step 6: Sum the Values from Step 5

Table :

Obs	Candidate Selection	Estimate	Error	$Error^2$	DV-$\bar{DV}$	DV-$\bar{DV}^2$
1	6	4.559578	1.440422	2.074816	1.786221	$1.786221^2 = 3.190586$
2	2	4.518417	-2.518417	6.342426	-2.213779	$-2.213779^2 = 4.900817$
3	4	5.238727	-1.238727	1.534445	-0.213779	$-0.213779^2 = 0.0457015$
4	6	4.744801	1.255199	1.575526	1.786221	$1.786221^2 = 3.190586$
Numerator (note) = 2895.367						
$\overline{Y}$	4.213779					
Denominator (note) = 3391.892						

Step 7: Divide the Numerator by the Denominator

$$\frac{2895.367}{3391.892} = 0.8536$$

Step 8: Subtract the Value from Step 7 from 1

- 1-0.8536 = 0.1464

Step 9: Interpret

- $R^2 = 0.1464$
- Move the decimal place over twice to the right and use as a percentage- 14.64%
- Variation in our independent variable explains (R^2%) variation in our dependent variable.
- Interpretation: Variation in Strength in Upper Chamber explains 14.64% of the variation in Party Control Over Candidate Selection.

Adjusted R-Square

- Takes sample size and number of independent variables into account
- Adjusted $R^2 = 1 - ((1 - R^2)(\frac{n-1}{n-k-1}))$
  - n = number of observations
  - k = number of independent variables

Adjusted R-Square (cont.)

- R^2 = 0.1464
  - n = number of observations = 987
  - k = number of independent variables = 1
- 1-(1-0.1464)$(\frac{987-1}{987-1-1}) = 1 - (0.8536)\frac{986}{985}$ = 1 - (0.8536)(1) = 1 - 0.8536 = 0.1464
- Adjusted R^2 = 0.1464

Adjusted R-Square (cont.)

- Adjusted R^2 = 0.1464
- Move the decimal place over twice to the right and use as a percentage- 14.64%
- After taking into account the number of observations and the number of independent variables, variation in our independent variable explains (Adjusted R^2%) variation in our dependent variable.
- Interpretation: After taking into account sample size and the number of independent variables, variation in Strength in Upper Chamber explains 14.64% of the variation in Party Control Over Candidate Selection.

R-Square and Adjusted R-Square in STATA

- When you run an OLS regression in STATA, it calculates the R-square and adjusted R-square.

Average Variation

- Range: 0 to 1
  - 1: Perfectly estimates
  - 0: No estimate ability
- Consistent?