Our midterm is in one week

- Bring a pen and a bluebook.
- Short answers and interpretations of research.
- No calculations or memorization of formulas required. Just understand basic concepts.
- For example, study X uses this measure for a dependent variable – is this valid? They get this statistical result – does that match up with their hypothesized relationship?
- I’ll put a review sheet up on our class website.
Political Science 15

Lecture 10:
Multiple Regression
Review: Interpretation of Regression Results

- $y = a + b*x$

- $a$ is the intercept (or constant term). This tells us what the value of the dependent variable ($y$) is expected to be when the independent variable ($x$) is equal to zero.

- $b$ is the slope coefficient. This tells us how the value of the dependent variable is expected to change as the independent variable increases by 1 unit.
Interpreting Regressions: Example

- Hypothesis: “Political instability in a country increases as the amount of IMF loans increases.”

- $y = a + b \cdot x$

- $y =$ political instability (measured as # deaths in political protests/riots)

- $x =$ IMF loans (measured in millions of dollars)

- If the hypothesis is correct, what should we expect to see for $b$?

- Suppose $b = 20$. What does this tell us?

- Suppose $a = 10$. What does this tell us?
One More Statistic to Know: Goodness of Fit

- All three scatterplots below have the same regression line \( y = 10 + 3x \), but obviously some lines fit the data better than others.

- How can we summarize these differences?
R² (R-Squared)

- R² is a measure of the goodness of fit of a regression line.
- With one independent variable, R² is simply the correlation (r) squared.
- R² is always between 0 and 1, with higher numbers indicating a better fit of the regression line to the data.
- It is interpreted as the proportion of variance in the dependent variable explained by the independent variable.
Goodness of Fit

$R^2 = 0.9$  $R^2 = 0.5$  $R^2 = 0.1$
Multiple Regression

- Regression with more than one independent variable is known as *multiple regression*.

- Multiple regression is a powerful technique because it allows researchers to simultaneously consider multiple explanations.

- Typically these regressions have the dependent and independent variables from the hypothesis as well as a number of control variables.
Example: Regression with One Independent Variable
Example: Regression with Two Independent Variables
Example: Regression with Two Independent Variables
Example: Rotating 3D Graph
Without controlling for income, a positive relationship between education and campaign contributions.
Control Variables in Multiple Regression

Controlling for income.
Control Variables in Multiple Regression

After controlling for income, no relationship between education and campaign contributions.
Multiple Regression

- Multiple regression is when you have more than one independent variable (independent variable plus control variable).

- \[ y = a + b*x + c*z \]

- **a** is the intercept (or constant term). This tells us what the value of the dependent variable (\(y\)) is expected to be when all independent variables (\(x\) and \(z\)) are equal to zero.

- **b** and **c** are the slope coefficients. These tell us how the value of the dependent variable is expected to change as one independent variable increases by 1 unit, holding the other variable constant.
Interpreting Multiple Regression: Example #1

- Hypothesis: “Individuals with higher education levels donate more to political campaigns.” We also want to control for income.

- \( y = a + b\times + c\times z \)

- \( y = \) donations to campaigns (measured as $)

- \( x = \) education (measured in years)

- \( z = \) income (measured in $10,000s)

- If the hypothesis is correct, what should we expect to see for \( b \)? What do you expect to see for \( c \)?

- Suppose \( b = 55 \). What does this tell us?

- Suppose \( c = 40 \). What does this tell us?

- Suppose \( a = 10 \). What does this tell us?
Interpreting Multiple Regression: Example #2

- Hypothesis: “Political instability in a country increases as the amount of IMF loans increases.” We also want to control for droughts/bad weather.

- \( y = a + b \times x + c \times z \)
- \( y = \) political instability (measured as \# deaths in riots)
- \( x = \) IMF loans (measured in $ millions)
- \( z = \) drought conditions (measured in days per year)

If the hypothesis is correct, what should we expect to see for \( b \)? What do you expect to see for \( c \)?

- Suppose \( b = 0.8 \). What does this tell us?
- Suppose \( c = 0.2 \). What does this tell us?
- Suppose \( a = 1 \). What does this tell us?
# Regression in SPSS

<table>
<thead>
<tr>
<th>Model</th>
<th>Unstandardized Coefficients</th>
<th>Standardized Coefficients</th>
<th>t</th>
<th>Sig.</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>B</td>
<td>Std. Error</td>
<td>Beta</td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>(Constant)</td>
<td>88.881</td>
<td>3.785</td>
<td>18.194</td>
</tr>
<tr>
<td></td>
<td>PayTaxes</td>
<td>3.625</td>
<td>1.494</td>
<td>.079</td>
</tr>
<tr>
<td></td>
<td>ElectUnfair</td>
<td>-30.789</td>
<td>5.648</td>
<td>-.568</td>
</tr>
</tbody>
</table>

a. Dependent Variable: GWBTherm