

Political Science 104

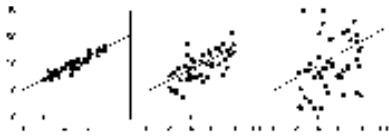
Lecture 11:
Multiple Regression

Review: Interpretation of Regression Results

- $y = a + b \cdot x + e$
- **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.

One More Statistic to Know: Goodness of Fit

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

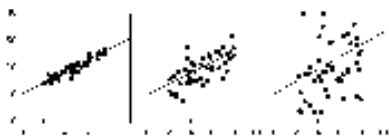


- How can we summarize these differences?

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

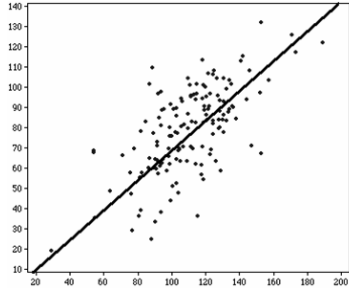


$R^2 = 0.9$ $R^2 = 0.5$ $R^2 = 0.1$

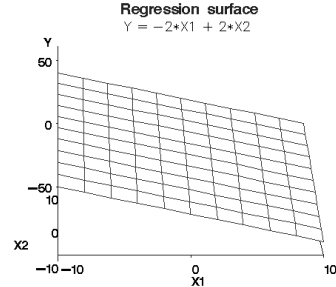
Multiple Regression

- 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.

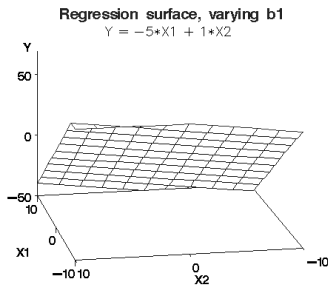
Example: Regression with One Independent Variable



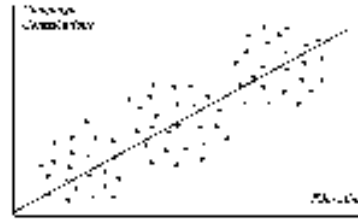
Example: Regression with Two Independent Variables



Changing One Coefficient, Holding Another Constant



Control Variables in Multiple Regression



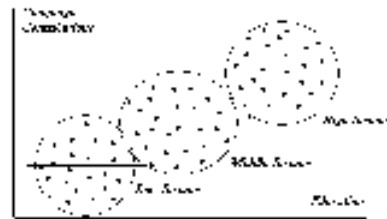
Without controlling for income, a positive relationship between education and campaign contributions.

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).
- $y = a + b*x + c*z + e$
- **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 tells 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 less to political campaigns.” We also want to control for income.
- $y = a + b*x + c*z + e$
- **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.
- $y = a + b*x + c*z + e$
- **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?

Adjusted R²

- Adjusted R² is a measure of the goodness of fit of a regression line when there is more than one independent variable.
- R² always improves as you add more independent variables.
- Adjusted R² is adjusted by the number of independent variables, and only improves if the new variables improve the fit of the model beyond that we would expect by chance.
- Interpret this just as you would R². Note that adjusted R² can be less than 0.