PSC 400
SYRACUSE UNIVERSITY

DATA ANALYTICS FOR POLITICAL SCIENCE

BIVARIATE RELATIONSHIPS

ASSIGNMENTS

- Data Analysis Memo 2 due on Friday
- Problem Set 3 will be posted after class
  - due Friday next week
  - solution to PS2 online
- Review Exercise 5 due Monday
  - Probably, depending on how far we get today

UNCERTAINTY

```
> reg1 <- lm(d.share ~ d.comp, data=facedata)
> summary(reg1)

Call:
lm(formula = d.share ~ d.comp, data = facedata)

Residuals:
     Min       1Q   Median       3Q      Max
-0.33743 -0.08300  0.00700  0.08871  0.37149

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  0.34389    0.03298  10.427  < 2e-16 ***
d.comp       0.33019    0.06359   5.193 8.85e-07 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 0.1332 on 117 degrees of freedom
Multiple R-squared:  0.1873,    Adjusted R-squared:  0.1803
F-statistic: 26.96 on 1 and 117 DF,  p-value: 8.854e-07
```

PROBLEM

- Want to know: is perceived facial competence correlated with election performance in the *population*?
- We only have data from a *random sample*
- Idea: Use relation between two variables in *sample* to make inference about relation between two variables in *population*
  - Of course, means we can make mistakes

NULL HYPOTHESIS

- In the population, there is *no relationship* between dependent and independent variable
  - H_0

ALTERNATIVE HYPOTHESIS

- There *is* a relationship between the independent and dependent variable in the population
  - H_a or H_1

ERRORS

	There Is A Relation In The Population	There Is No Relation In The Population
We Conclude There Is A Relation	✔	✘ Type I
We Conclude There Is No Relation	✘ Type II	✔

TYPE I ERROR

- We conclude there is a relationship between X and Y when in reality there is not
  - "Type I error"
  - We falsely reject H_0

TYPE II ERROR

- We conclude there is no relationship between X and Y when in reality there is
  - "Type II error"
  - We falsely do not reject H_0

DECISION

- It's really bad if we conclude there is a relationship when in reality there is not
- Type I error: falsely rejecting H_0
- We only want to reject H_0 based on our sample if chance of committing Type I error is relatively small

UNCERTAINTY

```
> reg1 <- lm(d.share ~ d.comp, data=facedata)
> summary(reg1)

Call:
lm(formula = d.share ~ d.comp, data = facedata)

Residuals:
     Min       1Q   Median       3Q      Max
-0.33743 -0.08300  0.00700  0.08871  0.37149

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  0.34389    0.03298  10.427  < 2e-16 ***
d.comp       0.33019    0.06359   5.193 8.85e-07 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 0.1332 on 117 degrees of freedom
Multiple R-squared:  0.1873,    Adjusted R-squared:  0.1803
F-statistic: 26.96 on 1 and 117 DF,  p-value: 8.854e-07
```

- Estimate: Intercept/slope coefficient
- Std. Error: Estimate of random sampling error
- t value: Estimate/Std. Error
- Pr(>|t|): Probability of falsely rejecting H_0

CONFIDENCE INTERVAL

- Related to standard error and p-value: 95 confidence interval
- 95% CI: Estimate ± 1.96 * Std. Err.

CONFIDENCE INTERVAL

- If we run many regressions using random samples and construct 95% CI for each, 95% of those intervals contain the true effect

EXAMPLE

Table 4.5. 2012 US Presidential Election Data.

Variable	*Description*
`state`	abbreviated name of the state
`Obama`	Obama's vote share (percentage)
`Romney`	Romney's vote share (percentage)
`EV`	number of Electoral College votes for the state

- pres12.csv
- How does Obama's vote share in 2012 depend on his 2008 vote share?

EXPLANATORY POWER MEASURE

- Need: measure of how well independent variable explains dependent variable in a linear regression
- Measure is called R^2
- R^2 tells us how much variation of the dependent variable is explained by the independent variable
  - Between 0 and 1
  - 0: The independent variable explains *none* of the variation in the dependent variable
  - 1: The independent variable explains *all* of the variation in the dependent variable