

# Intro to Social Science Data Analysis

## Week 14: Statistical Analysis and Visualization of Results

**Christopher Gandrud**

December 3, 2012

- 1 Schedule
- 2 This Week's Goals
- 3 Robust Standard Errors for Dependent Data
- 4 Results tables with xtable

# Schedule

We are dedicating **all** of the class time for the rest of the course to the research project.

## Schedule:

- ▶ Week 13: Research question, design, & data download,
- ▶ Week 14: Statistical Analysis & Results Visualization,
- ▶ Week 15: Write up.
- ▶ Week 16: Presentations.

# Schedule

We are dedicating **all** of the class time for the rest of the course to the research project.

Schedule:

- ▶ Week 13: Research question, design, & data download,
- ▶ Week 14: Statistical Analysis & Results Visualization,
- ▶ Week 15: Write up.
- ▶ Week 16: Presentations.

# Schedule

We are dedicating **all** of the class time for the rest of the course to the research project.

Schedule:

- ▶ Week 13: Research question, design, & data download,
- ▶ Week 14: Statistical Analysis & Results Visualization,
- ▶ Week 15: Write up.
- ▶ Week 16: Presentations.

# Schedule

We are dedicating **all** of the class time for the rest of the course to the research project.

Schedule:

- ▶ Week 13: Research question, design, & data download,
- ▶ Week 14: Statistical Analysis & Results Visualization,
- ▶ Week 15: Write up.
- ▶ Week 16: Presentations.

# Schedule

We are dedicating **all** of the class time for the rest of the course to the research project.

Schedule:

- ▶ Week 13: Research question, design, & data download,
- ▶ Week 14: Statistical Analysis & Results Visualization,
- ▶ Week 15: Write up.
- ▶ Week 16: Presentations.

## Today's Goals:

1. Make sure you data is clean & ready for analysis.
2. Descriptive statistics.



### Today's Goals:

1. Make sure you data is clean & ready for analysis.
2. Descriptive statistics.

### Thursday's Goals:

1. Begin to put together your inferential statistics..
2. Descriptive statistics.

### Thursday's Goals:

1. Begin to put together your inferential statistics..
2. Descriptive statistics.

### Remember:

One of the assumptions of linear regression is that the observations are independent of one another.

Many of you are using data from many countries across many years.

This is called **time-series cross-sectional data**.

## For Example

# Example Time-series Cross Sectional Data

```
## Error: undefined columns selected
```

```
##      iso2c                country year GDPperCapita
## 25      AE United Arab Emirates 2002      33740.9
## 26      AE United Arab Emirates 2003      36562.1
## 27      AE United Arab Emirates 2004      40410.8
## 28      AE United Arab Emirates 2005      44384.7
## 29      AF              Afghanistan 2002        158.0
## 30      AF              Afghanistan 2003        168.7
##      InfantMort FYouthUnemp IMFCode
## 25           9.4           NA      466
## 26           8.9           NA      466
## 27           8.4           NA      466
## 28           8.0          12.9      466
## 29          90.5           NA      512
## 30          88.4           NA      512
```

# One Solution

A common way of handling data like this is to use **robust standard errors**.

They are easy to implement with *Zelig*

```
M1 <- zelig(InfantMortality ~ GDPperCapita,  
            model = "normal",  
            data = MortalityGDP,  
            robust = TRUE,  
            order.by = ~c("country", "year"),  
            cite = FALSE)
```

```
## Error: object 'InfantMortality' not found
```

### Note:

It's (usually) a good thing if the robust and regular standard errors are basically the same.

This indicates you are not violating the model assumption.

For a much more advanced discussion see:

<http://gking.harvard.edu/files/robust.pdf>

There is much more to learn on this topic, which we won't cover in this class.





# Results Tables

Hand typing results tables is really irritating.