

# Intro to Social Science Data Analysis

## Week 14: Statistical Analysis and Visualization of Results

**Christopher Gandrud**

December 4, 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 your data is clean & ready for analysis.
2. Descriptive statistics/begin to think about linear model assumptions.



### Today's Goals:

1. Make sure your data is clean & ready for analysis.
2. Descriptive statistics/begin to think about linear model assumptions.

### Thursday's Goals:

- ▶ Run your inferential 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**.

This type of data often has **biased standard errors**.

# Example Time-series Cross Sectional Data

##	country	year	GDPperCapita	InfantMort
## 29	Afghanistan	2002	158.0	90.5
## 30	Afghanistan	2003	168.7	88.4
## 31	Afghanistan	2004	196.2	86.4
## 32	Afghanistan	2005	227.9	84.3
## 37	Albania	2002	1440.0	20.9
## 38	Albania	2003	1819.4	19.8

# 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(InfantMort ~ GDPperCapita,  
            model = "normal",  
            data = MortalityGDP,  
            robust = TRUE,  
            cite = FALSE)
```

### 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.

You can use the *xtable* package to automate table creation.

See: <http://bit.ly/TFxDS4>.