

# Introduction to Empirical Analysis and Quantitative Methods<sup>1</sup>

Political Science 3

Jasjeet S. Sekhon

UC Berkeley

January 18, 2011 (02:34)

---

<sup>1</sup>Please do not redistribute these slides without permission. 

## Social Statistics

- Many of the great minds of the 18th and 19th centuries contributed to the development of social statistics: De Moivre, several Bernoullis, Gauss, Laplace, Quetelet, Galton, Pearson, and Yule.
- They searched for a method of statistical calculus that would do for social studies what Leibniz's and Newton's calculus did for physics.
- It quickly came apparent this is would be most difficult.
- For example: deficits crowding out money versus chemotherapy for a rare cancer or a get-out-the-vote mailer

## Social Statistics

- Many of the great minds of the 18th and 19th centuries contributed to the development of social statistics: De Moivre, several Bernoullis, Gauss, Laplace, Quetelet, Galton, Pearson, and Yule.
- They searched for a method of statistical calculus that would do for social studies what Leibniz's and Newton's calculus did for physics.
- It quickly came apparent this is would be most difficult.
- For example: **deficits crowding out money** versus **chemotherapy for a rare cancer** or a **get-out-the-vote mailer**

## Social Statistics

- Many of the great minds of the 18th and 19th centuries contributed to the development of social statistics: De Moivre, several Bernoullis, Gauss, Laplace, Quetelet, Galton, Pearson, and Yule.
- They searched for a method of statistical calculus that would do for social studies what Leibniz's and Newton's calculus did for physics.
- It quickly came apparent this is would be most difficult.
- For example: **deficits crowding out money** versus **chemotherapy for a rare cancer** or a **get-out-the-vote mailer**

## The Experimental Model

- In the early 20th Century, Sir Ronald Fisher (1890-1962) created modern randomized experiments (Fisher 1935).
- And establish randomization as the “reasoned basis for inference”.
- Inference based on observational data remains a difficult challenge.
- Randomization is a powerful tool by which systematic sources of bias are made random.

## The Experimental Model

- In the early 20th Century, Sir Ronald Fisher (1890-1962) created modern randomized experiments (Fisher 1935).
- And establish randomization as the “reasoned basis for inference”.
- Inference based on observational data remains a difficult challenge.
- Randomization is a powerful tool by which systematic sources of bias are made random.

# Randomized Field Experiments

- Randomized Control Trials (RCTs) or field experiments are becoming more common in the social science
- RCTs have a surprisingly long history in political science going back to [Gosnell's 1924](#) Get-Out-the-Vote experiments.
- See, for example, Gosnell (1948).



## Neighbors mailing

3 0 4 2 3 - 3



For more information: (517) 351-1975

email: etov@grebner.com

Practical Political Consulting

P. O. Box 6249

East Lansing, MI 48826

PRSRT STD  
U.S. Postage

**PAID**

Lansing, MI  
Permit # 444

ECRL0T \*\*C050  
THE JACKSON FAMILY  
9999 MAPLE DR  
FLINT MI 48507

Dear Registered Voter:

### WHAT IF YOUR NEIGHBORS KNEW WHETHER YOU VOTED?

Why do so many people fail to vote? We've been talking about the problem for years, but it only seems to get worse. This year, we're taking a new approach. We're sending this mailing to you and your neighbors to publicize who does and does not vote.

The chart shows the names of some of your neighbors, showing which have voted in the past. After the August 8 election, we intend to mail an updated chart. You and your neighbors will all know who voted and who did not.

DO YOUR CIVIC DUTY — VOTE!

MAPLE DR

Aug 04    Nov 04    Aug 06



Dear Registered Voter:

## WHAT IF YOUR NEIGHBORS KNEW WHETHER YOU VOTED?

Why do so many people fail to vote? We've been talking about the problem for years, but it only seems to get worse. This year, we're taking a new approach. We're sending this mailing to you and your neighbors to publicize who does and does not vote.

The chart shows the names of some of your neighbors, showing which have voted in the past. After the August 8 election, we intend to mail an updated chart. You and your neighbors will all know who voted and who did not.

## DO YOUR CIVIC DUTY — VOTE!

---

MAPLE DR	Aug 04	Nov 04	Aug 06
9995 JOSEPH JAMES SMITH	Voted	Voted	_____
9995 JENNIFER KAY SMITH		Voted	_____
9997 RICHARD B JACKSON		Voted	_____
9999 KATHY MARIE JACKSON		Voted	_____
9999 BRIAN JOSEPH JACKSON		Voted	_____
9991 JENNIFER KAY THOMPSON		Voted	_____
9991 BOB R THOMPSON		Voted	_____
9993 BILL S SMITH			_____

- We discuss the **potential outcomes** framework of causal inference.
- This framework originates with **Jerzy Neyman** (1894-1981), the founder of the Berkeley Statistics Department.
- A key insight of the framework is that causal inference is a missing data problem.
- The framework applies regardless of the method used to estimate causal effects, whether it be quantitative or qualitative.



- Descriptive Statistics are of general interest
- A vast number of provocative numbers
- Example: U.S. military fatalities since 1980:  
[http://siadapp.dmdc.osd.mil/personnel/CASUALTY/death\\_Rates.pdf](http://siadapp.dmdc.osd.mil/personnel/CASUALTY/death_Rates.pdf)
- There were more military deaths in 1980 [2,392] than 2010 [1,436] or any year during the Iraq/Afghan wars
- More U.S. military deaths during the Reagan administration (17,201) than the H.W. Bush administration (12,568)
- Why? What does this mean? We always want the causal story.

## Descriptive Statistics

- Descriptive Statistics are of general interest
- A vast number of provocative numbers
- Example: U.S. military fatalities since 1980:  
[http://siadapp.dmdc.osd.mil/personnel/CASUALTY/death\\_Rates.pdf](http://siadapp.dmdc.osd.mil/personnel/CASUALTY/death_Rates.pdf)
- There were more military deaths in 1980 [2,392] than 2010 [1,436] or any year during the Iraq/Afghan wars
- More U.S. military deaths during the Reagan administration (17,201) than the H.W. Bush administration (12,568)
- Why? What does this mean? We always want the causal story.

## Descriptive Statistics

- Descriptive Statistics are of general interest
- A vast number of provocative numbers
- Example: U.S. military fatalities since 1980:  
[http://siadapp.dmdc.osd.mil/personnel/CASUALTY/death\\_Rates.pdf](http://siadapp.dmdc.osd.mil/personnel/CASUALTY/death_Rates.pdf)
- There were more military deaths in 1980 [2,392] than 2010 [1,436] or any year during the Iraq/Afghan wars
- More U.S. military deaths during the Reagan administration (17,201) than the H.W. Bush administration (12,568)
- Why? What does this mean? We always want the causal story.

## Descriptive Statistics

- Descriptive Statistics are of general interest
- A vast number of provocative numbers
- Example: U.S. military fatalities since 1980:  
[http://siadapp.dmdc.osd.mil/personnel/CASUALTY/death\\_Rates.pdf](http://siadapp.dmdc.osd.mil/personnel/CASUALTY/death_Rates.pdf)
- There were more military deaths in 1980 [2,392] than 2010 [1,436] or any year during the Iraq/Afghan wars
- More U.S. military deaths during the Reagan administration (17,201) than the H.W. Bush administration (12,568)
- Why? What does this mean? We always want the causal story.

# Formalization of Intuitive Methods

- John Stuart Mill (in his *A System of Logic*) devised a set of five methods (or canons) of inference (Mill 1843/1872).
- They were outlined in Book III, Chapter 8 of his book.



## Two of Mill's Methods of Inductive Inference

- **Of Agreement:** “If two or more instances of the phenomenon under investigation have only one circumstance in common, the circumstance in which alone all the instances agree is the cause (or effect) of the given phenomenon.”
- **Of Difference:** “If an instance in which the phenomenon under investigation occurs, and an instance in which it does not occur, have every circumstance in common save one, that one occurring only in the former; the circumstance in which alone the two instances differ is the effect, or the cause, or an indispensable part of the cause, of the phenomenon.”

## Two of Mill's Methods of Inductive Inference

- **Of Agreement:** “If two or more instances of the phenomenon under investigation have only one circumstance in common, the circumstance in which alone all the instances agree is the cause (or effect) of the given phenomenon.”
- **Of Difference:** “If an instance in which the phenomenon under investigation occurs, and an instance in which it does not occur, have every circumstance in common save one, that one occurring only in the former; the circumstance in which alone the two instances differ is the effect, or the cause, or an indispensable part of the cause, of the phenomenon.”

## Uses of Mills Methods

These methods have been used by a vast number of researchers, including such famous ones as Durkheim and Weber. They are known as the “most similar” and “most different” research designs in comparative politics (Przeworski and Teune 1970).

Here are some examples:

- The Protestant Ethic
- Deficits and interest rates
- Gun control
- Three strikes
- The list goes on, and on....

## Uses of Mills Methods

These methods have been used by a vast number of researchers, including such famous ones as Durkheim and Weber. They are known as the “most similar” and “most different” research designs in comparative politics (Przeworski and Teune 1970).

Here are some examples:

- The Protestant Ethic
  - Deficits and interest rates
  - Gun control
  - Three strikes
  - The list goes on, and on....

## Uses of Mills Methods

These methods have been used by a vast number of researchers, including such famous ones as Durkheim and Weber. They are known as the “most similar” and “most different” research designs in comparative politics (Przeworski and Teune 1970).

Here are some examples:

- The Protestant Ethic
- Deficits and interest rates
- Gun control
- Three strikes
- The list goes on, and on....

## Uses of Mills Methods

These methods have been used by a vast number of researchers, including such famous ones as Durkheim and Weber. They are known as the “most similar” and “most different” research designs in comparative politics (Przeworski and Teune 1970).

Here are some examples:

- The Protestant Ethic
- Deficits and interest rates
- Gun control
- Three strikes
- The list goes on, and on....

## Uses of Mills Methods

These methods have been used by a vast number of researchers, including such famous ones as Durkheim and Weber. They are known as the “most similar” and “most different” research designs in comparative politics (Przeworski and Teune 1970).

Here are some examples:

- The Protestant Ethic
- Deficits and interest rates
- Gun control
- Three strikes
- The list goes on, and on....

## Uses of Mills Methods

Mill himself thought they were inappropriate for the study of social questions (Sekhon 2004).

*“Nothing can be more ludicrous than the sort of parodies on experimental reasoning which one is accustomed to meet with, not in popular discussion only, but in grave treatises, when the affairs of nations are the theme. “How,” it is asked, “can an institution be bad, when the country has prospered under it?” “How can such or such causes have contributed to the prosperity of one country, when another has prospered without them?” Whoever makes use of an argument of this kind, not intending to deceive, should be sent back to learn the elements of some one of the more easy physical sciences” (Mill 1873, 346–7).*



## Potential Outcomes

- Fundamental problem: not observing all of the potential outcomes or counterfactuals
- This is the [Neyman-Rubin-Holland Causal Model](#) (Neyman 1923/1990; Rubin 1974; Holland 1986).
- Can be used to describe problems of causal inference for both experimental work and observational studies.

## Potential Outcomes

- Fundamental problem: not observing all of the potential outcomes or counterfactuals
- This is the [Neyman-Rubin-Holland Causal Model](#) (Neyman 1923/1990; Rubin 1974; Holland 1986).
- Can be used to describe problems of causal inference for both experimental work and observational studies.

# Observational Study

An observational study concerns

- cause-and-effect relationships
- treatments, interventions or policies and
- the effects they cause

## A Thought Experiment

An observational study could in principle have been an experiment but for ethical concerns or logistical issues.

You are probably don't understand the causal question being asked if you can't answer: “what experiment would you have run if you were dictator and has infinite resources?”

E.G.: Can we estimate the causal effect of race on SAT scores?

Descriptive and predictive work is something else and can be interesting.

## A Thought Experiment

An observational study could in principle have been an experiment but for ethical concerns or logistical issues.

You are probably don't understand the causal question being asked if you can't answer: “**what experiment would you have run if you were dictator and has infinite resources?**”

E.G.: Can we estimate the causal effect of race on SAT scores?

Descriptive and predictive work is something else and can be interesting.

## A Thought Experiment

An observational study could in principle have been an experiment but for ethical concerns or logistical issues.

You are probably don't understand the causal question being asked if you can't answer: “**what experiment would you have run if you were dictator and has infinite resources?**”

E.G.: **Can we estimate the causal effect of race on SAT scores?**

Descriptive and predictive work is something else and can be interesting.

## A Thought Experiment

An observational study could in principle have been an experiment but for ethical concerns or logistical issues.

You are probably don't understand the causal question being asked if you can't answer: “**what experiment would you have run if you were dictator and has infinite resources?**”

E.G.: **Can we estimate the causal effect of race on SAT scores?**

Descriptive and predictive work is something else and can be interesting.

# Effect of a Cause vs. Cause of an Effect

It is **far** easier to estimate the effect of a cause than the cause of an effect. Why?

Is it easier to find out the effect of smoking on lung cancer or if someone with lung cancer got it because of smoking?



## Effect of a Cause vs. Cause of an Effect

It is **far** easier to estimate the effect of a cause than the cause of an effect. Why?

Is it easier to find out the effect of smoking on lung cancer or if someone with lung cancer got it because of smoking?

# References

Full references are available in the following document:

[http://sekhon.berkeley.edu/ps3/ps3\\_references.pdf](http://sekhon.berkeley.edu/ps3/ps3_references.pdf).