# Experiments and Observational Studies

Javier E. Flores

February 1, 2019

## Name that movie!

So....what exactly does this movie have to do with statistics?

# Well nothing, but...

The real tragedy which inspired the movie, *Titanic*, will be the motivating example for this lecture.

# Some History

The *RMS Titanic* was the product of an intense competition among two rival shipping lines, White Star Line and Cunard, during the early $20^{th}$ century.

Cunard seemed to have an edge over White Star after setting a speed record in 1907 for the fastest average speed during a transatlantic crossing.

However, after White Star's newest product, the *RMS Titanic*, was deemed 'practically unsinkable' by *Shipbuilder* magazine, it seemed there was potential for White Star to establish itself as the superior shipping line.

## *Shipbuilder* was wrong ☹

The *RMS Titanic* departed for its maiden voyage on April 10, 1912 only to tragically sink five days later when the ship struck an iceberg.

Out of the 2,240 passengers and crew on board, fewer than a third survived the disaster. These casualties were recorded by Lord Mersey in his report from the British inquiry into the sinking.

The table below is derived from this report and lists the casualties for third class passengers and crew members.

*Titanic* Survivor Records: Third Class & Crew

| Class | Saved | Lost | Total | Survival Rate (%) |
|---|---|---|---|---|
| **Third** | 151 | 476 | 627 | 24.08 |
| **Crew** | 212 | 673 | 885 | 23.95 |

# Question:

Were third class passengers more likely to survive?

*Titanic* Survivor Records: Third Class & Crew

| **Class** | **Saved** | **Lost** | **Total** | **Survival Rate (%)** |
|---|---|---|---|---|
| **Third** | 151 | 476 | 627 | 24.08 |
| **Crew** | 212 | 673 | 885 | 23.95 |

At first glace, it would certainly seem that third class passengers slightly better off.

However, let's see how this table changes when we consider the sex of each individual.

# Same Data, Different Perspective

*Titanic* Survivor Records: Third Class & Crew, Stratified by Sex

| | Male | | | | Female | | | |
|---|---|---|---|---|---|---|---|---|
| **Class** | **S** | **L** | **T** | **(%)** | **S** | **L** | **T** | **(%)** |
| **Third** | 75 | 387 | 462 | 16.23 | 76 | 89 | 165 | 46.06 |
| **Crew** | 192 | 670 | 862 | 22.27 | 20 | 3 | 23 | 86.96 |

Do any of your conclusions change?

How is it that the crew has better survival when **conditioned** on, or **stratified** by sex, but does worse overall?

This is a classic example of **Simpson's Paradox**.

# Simpson's Paradox

Simply put, Simpson's paradox occurs when the **association** between two variables (e.g. passenger type and survival) changes when stratified by a **confounding** variable (e.g. sex).

An **association** between two variables exists when the values of one variable tend to be related to the values of the other.

Note: Two variables are **causally** associated when changing the value of one variable influences the value of the other.

A **confounding**, or **lurking**, variable (e.g. sex) is a variable associated with both an explanatory (e.g. passenger type) and response (e.g. survival) variable.

## Back to the *RMS Titanic*

How can we explain Simpson's Paradox in the case of the *RMS Titanic*?

"Women and children first!" - **Captain Smith**, *Titanic (1997)*

Note that nearly a third of all third-class passengers were female, whereas less than 3% of crew members were female.

As a result, and with women having survival priority over men (hence the quote), the (higher) survival proportions in women had more of an impact in third-class than among the crew.

This observation, the disparity in male/female composition between third-class and crew members, is what explains the paradox.

## Practice

A 1981 study analyzed murders that took place during felonies committed in Florida between 1972 and 1977. The study recorded the race of the defendant, the race of the victim, and whether the defendant was sentenced to death.

Consider the following two tables...

# Practice

Table 1

| | Death | Total |
|---|---|---|
| **White Defendant** | 46 | 198 |
| **Black Defendant** | 38 | 180 |

Table 2

| | **White Victim** | | **Black Victim** | |
|---|---|---|---|---|
| | **Death** | **Total** | **Death** | **Total** |
| **White Defendant** | 46 | 190 | 0 | 8 |
| **Black Defendant** | 37 | 78 | 1 | 102 |

# Practice

With your group, discuss the following questions:

1) Looking only at Table 1, is there evidence of racial bias in sentencings? In which direction (i.e. biased against whites, blacks, or neither)?
2) Does the information in Table 2 change your response to the first question?
3) Is this an example of Simpson's Paradox? If so, describe which variable is explanatory, which is the response, and which is confounding. Explain why the paradox occurred.

## Confounding Variables

From the previous two examples, it is clear that confounding variables affect our interpretation of the relationship between a given set of explanatory and response variables.

In the presence of confounding it is difficult to properly characterize the true relationship between two variables and even more difficult to make any causal claims.

One solution to these issues can be found in conditioning, or stratifying, on the identified confounder. But how can we trust that there aren't additional confounders?

Is there a way for us to address confounding from multiple sources?

# Randomized Experiments

One way to mitigate confounding (of any degree) is to *randomly assign* values of an explanatory variable to the cases within our data.

*If* we are able to do this, we are performing a **randomized experiment**.

There is a lot of emphasis on the word "if" because randomization is not always possible.

As examples, how exactly would we randomly assign sex (for the titanic example) or victim race (for the death penalty example)? (Answer: we can't).

## How Does Randomization Work?

Randomization works by (on average) "balancing" potential confounders across the different values of your explanatory variable.

As a result, we remove the association between these potential confounders and our explanatory variable.

Remember: A confounder is only a confounder if it is associated with both the explanatory and response variables.

If we remove the association with the explanatory variable, the variable is no longer a confounder.

Randomization is not guaranteed to completely remove all confounding. It is only guaranteed to minimize confounding, assuming you have a large enough sample.

## Randomization (Activity)

When class began, you were given a blank notecard. On this notecard, I want you to write values for two variables.

The first variable should correspond to whether you have taken a stats class before. The second should be whether you see yourself as right-brained or left-brained.

Next, I will assign each of you a "treatment group", either "A" or "B". We will do this using a non-randomized approach and a randomized approach.

We will assess and compare the "balance" of the covariates (StatsClass and RLbrain) in each approach.

## Acupuncture

In 2007, the *Archives of Internal Medicine* (now known as *JAMA Internal Medicine*) published a study on the efficacy of acupuncture in treating chronic low back pain.

In this study, 1162 patients were randomized to one of three treatment groups: acupuncture, sham acupuncture, or conventional therapy (a combination of drugs, physical therapy, and exercise).

Patients who received either form of acupuncture were *completely unaware* of whether they were receiving the real or fake version.

# What Did the Study Find?

At the end of the study, it was found that the efficacy of acupuncture was almost twice that of conventional therapy!

The catch? There was virtually no difference in efficacy between the true and sham versions of acupuncture.

So if you know someone who has chronic low back pain, just convince them you're an acupuncturist and act the part - they just might start feeling better! ☺ (JK, Don't actually do that)

Fun Fact: Going into college, my plan was initially to study acupuncture. Oh, how things have changed...

# Design of Experiments

Beyond the takeaways on the efficacy of acupuncture, this study illustrates important aspects of a well-designed experiment that are important to be aware of:

- **Control Group** - some patients were randomly assigned to receive conventional therapy. This comparison group was generally balanced across baseline characteristics with both acupuncture groups.
- **Placebo** - some patients received fake acupuncture.
- **Blinding** - Using a placebo is ineffective if patients are aware of which group they're in. Similarly, staff might act differently if they know the assigned treatment.
    - **Single-blind** - the participants don't know the treatment assignment.
    - **Double-blind** - neither participants or interacting staff know the treatment assignment.

## Observational Studies

The "gold standard" of experimental design is widely considered to be randomized, placebo-controlled, double-blinded experiments.

Unfortunately, there are several practical and ethical considerations that may limit one's ability to meet this standard.

The examples at the beginning of this lecture are certainly illustrations of this, but consider the following additional scenarios:

- Is it ethically justifiable to use a placebo when assessing the efficacy of a treatment for late-stage cancer patients?
- To study the natural progression of a disease, could you withhold treatment from some of your study subjects? (Sadly, this was actually done by the US in 1972)

# Observational Studies

When randomized experiments are not possible, **observational studies** are a useful alternative.

One important limitation to observational studies is the inability to claim that any observed associations are causal.

While causality definitely adds strength to any conclusions drawn from associations, conclusions based on non-causal associations may still carry a lot of weight.

As an example, how is it that smoking was "proven" to be harmful? Conducting a randomized experiment was definitely not possible.

Fun Fact: One of the most famous statisticians ever, R.A. Fischer, was actually an opponent of research linking smoking to lung cancer.

## Minneapolis Police Study

As I mentioned earlier, randomization is not a cure-all solution.

Aside from the practical and ethical considerations which might prevent its use, there are instances when it is used and things still go awry.

Take for example a study conducted by the Minneapolis Police Department.

Police officers wanted to determine which of three response strategies - arrest, advice, or separate - are best in responding to cases of domestic abuse.

## Minneapolis Police Study

Officers were randomly assigned one of these three strategies for each case but were allowed to change strategy should the situation demand it.

Shown below is a table describing officer adherence to each strategy assigned.

Officer Strategy Adherence

| | | Assigned Strategy | | |
|---|---|---|---|---|
| | | **Arrest** | **Advice** | **Separate** |
| Actual | **Arrest** | 91 | 18 | 26 |
| | **Advice** | 0 | 84 | 5 |
| | **Separate** | 1 | 6 | 82 |

## Minneapolis Police Study

Despite the high adherence (about 82%) it seemed that when officers changed strategy, they chose arrest more often than not.

As a result, the advice and separate groups likely lost some of their highest risk members - consequently skewing the perceived efficacy of either strategy.

This demonstrates that even with a well-designed study that incorporates randomization, bias can creep in.

## Intent-to-Treat Principle

Another concept relevant when discussing the policy study is the **intent-to-treat principle** (ITT).

ITT mandates that, regardless of adherence, all randomized subjects must be included in any final analyses.

A consequence of this principle is that the effects of *treatment assignment* are being estimated, as opposed to the effects of treatment.

To learn more about ITT analyses and how they might contrast with non-ITT analyses, read the following article.

## Wrap-Up

Right now, you should...

- Recognize and explain examples of Simpson's paradox.
- Identify and describe potential confounding variables.
- Know the strengths and limitations of randomized experiments and observational studies.

These notes cover section 1.3 of the textbook. Please read through the section and its examples along with any links provided in this lecture.