Introduction to Statistics-UA.18

Ercan Karadas

New York University

ercan@nyu.edu

Summer 2016

Overview

What are statistics?

Examples

Practice of Statistics

How to study this course?

What are statistics?

Statistics include numerical facts and figures. For instance:

- The largest earthquake measured 9.2 on the Richter scale.
- By the year 2020, there will be 15 people aged 65 and over for every new baby born.
- Men are at least 10 times more likely than women to commit murder.
- One in every 8 South Africans is HIV positive.

- The study of statistics involves math and relies upon calculations of numbers.
- But it also relies heavily on **how the numbers are chosen** and how the statistics are interpreted.
- For example, consider the following three scenarios and the interpretations based upon the presented statistics.
- You will find that the numbers may be right, but the interpretation may be wrong. Let us try to identify a major flaw with each interpretation.

Example 1: Ice cream?

A new advertisement for Ben and Jerry's ice cream introduced in late May of last year resulted in a 30% increase in ice cream sales for the following three months. Thus, the advertisement was effective.

What do you think? Is there any flaw in the claim?

- Major flaw: ice cream consumption generally increases in the summer regardless of advertisements.
- This effect is called a history or seasonal effect.
- This effect often leads people to interpret outcomes as the result of one variable (here advertisement) when another variable (here weather) is actually responsible.

Example 2: Churches vs Crime?

The more churches in a city, the more crime there is. Thus, churches lead to crime.

What do you think? Is there any flaw in the claim?

- Major flaw: both increased churches and increased crime rates can be explained by larger populations. In bigger cities, there are both more churches and more crime.
- A third variable (larger population) can cause the observed correlation between two variables; however, people erroneously believe that there is a causal relationship between the two.
- Correlation does NOT imply causation!

Example 3: More interracial marriages?

75% more interracial marriages are occurring this year than 25 years ago. Thus, our society accepts interracial marriages.

What do you think? Is there any flaw in the claim?

- Major flaw: we don't have the information that we need.
- For example, what is the rate at which marriages are occurring? Suppose only 1% of marriages 25 years ago were interracial and so now 1.75% of marriages are interracial (1.75 is 75% higher than 1).
- But this latter number is hardly evidence suggesting the acceptability of interracial marriages.
- Not enough information to evaluate the claim

Statistics is more than a pile of numbers

- These examples show that statistics are not only facts and figures; they are something more than that.
- In the broadest sense, "statistics" refers to a range of techniques and procedures for analyzing, interpreting, displaying, and making decisions based on data.
- Ultimate goal: Better Decision Making!

Some more examples ...

Numbers or statistics are often presented in an effort to add credibility to an argument or advice.

- 4 out of 5 dentists recommend Dentine.
- Almost 85% of lung cancers in men and 45% in women are tobacco-related.
- People tend to be more persuasive when they look others directly in the eye and speak loudly and quickly.
- Women make 75 cents to every dollar a man makes when they work the same job.
- A surprising new study shows that eating egg whites can increase one's life span.
- There is an 80% chance that in a room full of 30 people that at least two people will share the same birthday.
- A politician: we have invested in education more than any previous governments.

- All of these claims are statistical in character.
- Notice how diverse the examples are. They come from psychology, health, law, sports, business, etc. Indeed, data and data interpretation show up in discourse from virtually every facet of contemporary life.
- They can be misleading and push you into decisions that you might find cause to regret.
- Learning about statistics is a long step towards taking control of your life.
- Be an intelligent consumer of statistical claims!

There are three kinds of lies: lies, damned lies, and statistics.
(attributed to Benjamin Disraeli by **Mark Twain**)

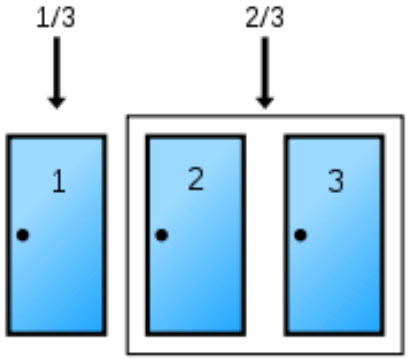
How does statistics work?

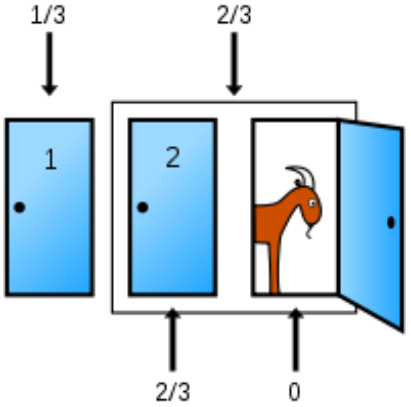
- Main Tool: Probability Theory
- Main Procedure: Inference

Probability? An Example

Month Hall Problem

Suppose you're on a game show, and you're given the choice of three doors: Behind one door is a car; behind the others, goats. You pick a door, say No. 1, and the host, who knows what's behind the doors, opens another door, say No. 3, which has a goat. He then says to you, "Do you want to pick door No. 2?" Is it to your advantage to switch your choice?





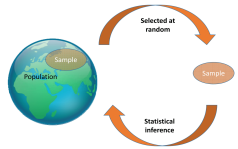
Inference? An Example

Suppose you would like to predict outcome of the next election. What would you do to make the best prediction?

- Of course, just ask voters, right?
- Yes, but how many? Wouldn't it be nice to include all the voters in the poll?
- That would be too costly and sometimes even possible, for example because of legal restrictions. An example to the legal restrictions: only state agencies can collect statistical data fro the whole country.
- Therefore, you have to work with a Sample instead of the whole Population.
- Based on population characteristics, next you determine the sample size. You want your sample to able to *represent* the population with the smallest possible size: trade-off between cost and precision!

Inference? An Example

- After collecting and analyzing your data, you will be able to make a point prediction and/or give an interval estimation about the election outcome.
- These are typical steps that one need to follow in inferential statistics that can be pictured as



- What about how does the probability theory related to the inferential statistics?
- Probability theory pops us basically in everywhere: in determining the sample size we use probability models and in the analysis and prediction stages a wide range of probability methods are indispensable.
- Therefore, roughly the first half of the course will be devoted to the probability theory and the second half to the inferential statistics.

How to Study?

- Main Source: Use first slides and then the textbook
- I will post the slides at least one day before the class, so I strongly recommend to come with print outs in the following format
- Many fully solved problems will be provided
- Very important: do NOT look at the solutions too soon!
- Attend lab sessions
- Solve as many problems as possible