Introductory lecture: A brief survey of what to expect in this course

Subhadeep Paul

Dept. Of Statistics, University of Illinois at Urbana-Champaign

August 26, 2013

Lies, Damned lies and Statistics

- British Prime Minister Benjamin Disraeli (18041881) said : "There are three kinds of lies: lies, damned lies, and statistics.". This phrase marks his frustration on how statistics or numbers was used in British politics to manipulate acheivements and failures, public opinion and to bolster weak arguments.
- So why would I start a statistics class saying that Statistics is a lie ? I do that because the goal of this course is not only to teach you what is correct, but also to teach you what is incorrect.
- Statistics is used in industry, academia and businesses to analyze real life datasets by people who are not statisticians by training. So statistics is generally more abused that it is used !

goals/learning outcomes

- you will be able to understand statistical terms used in your respective major fields of study.
- You will be able to judge what result/study/ experiment to trust and what not to trust.
- You will be able to read and understand news in a better way.
- You will be able to make informed data driven decisions and present your thoughts/ideas in a better way

Data and information

So, Why learn statistics?

- Data is raw unorganized collection of facts. When data is processed, structured, organized and given in a context it becomes information. So, information is the story you want to infer from data.
- Statistics a powerful tool to extract information out of data and discard the junk part.
- Scientists, researchers, physicists, chemists, biologists, economists, business professionals, lawyers, doctors, engineers- everyone needs statistics to design their experiments, to make sense of their observational data, to learn from the data, to control quality of analysis by accessing error and to test hypothesis.
- For example. data program. grades data

Graphical inference

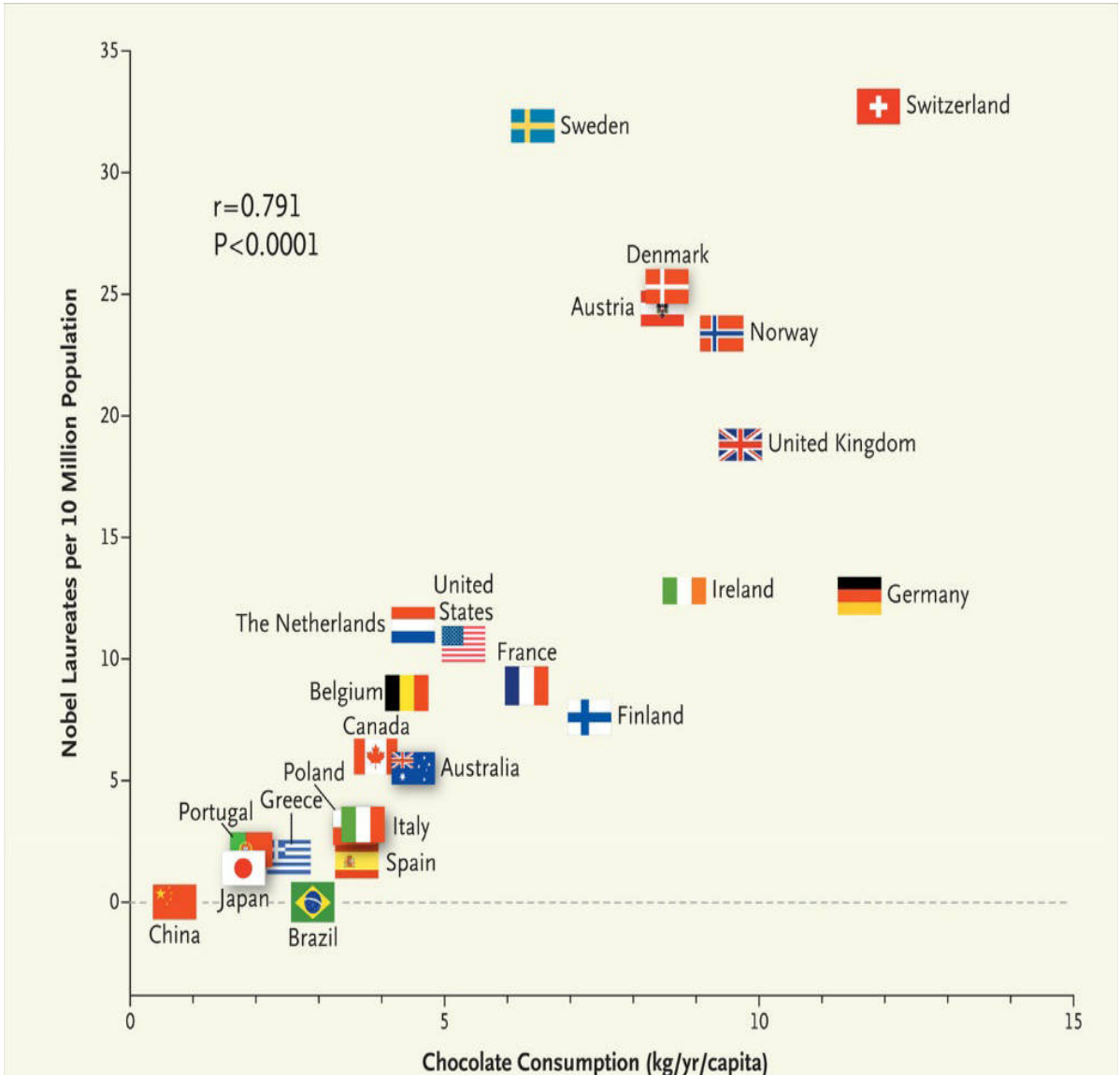
- You will learn about some popular charts and tools of visualizing real life data including scatter plots, bar graphs, pie charts, historgrams.
- We will look into some simple plots that appear regularly in news papers, learn how to read them and how to create our own plots. You will see, why a picture says a 1000 words is actually true in reporting statistics.
- Han's Rosling's video.
- Grades data histogram.

Experimental design

- You may have heard about the terms statistically significant improvement, double-blinded randomized clinical trial etc in newspapers.
- FDA has stringent requirements on the quality of experiments performed in clinical trials involving a new drug (and rightly so). The reason being that some drug manufacturers may want success from their researchers, so that they can introduce the new drug. The goal might be to show a statistically significant improvement of a disease or a condition over existing therapy or drugs.
- But what about the quality of the study performed? Are they justified in their claims that the new drug is better?
- We will learn about what constitutes a good trustworthy experiment that the FDA can accept.

Example of a bizzare conclusion

Secret formula for More Nobel prizes is simple! Eat more



Probability

- The theory of chance. If I am to pick someone randomly(meaning I am not biased, everyone eligible has equal chance of selection) for a job, What do you think is your chance of selection?

Probability

- The theory of chance. If I am to pick someone randomly(meaning I am not biased, everyone eligible has equal chance of selection) for a job, What do you think is your chance of selection?
- if, the entire class is eligible (44 students)

Probability

- The theory of chance. If I am to pick someone randomly(meaning I am not biased, everyone eligible has equal chance of selection) for a job, What do you think is your chance of selection?
- if, the entire class is eligible (44 students)
- if your column is eligible

Probability

- The theory of chance. If I am to pick someone randomly(meaning I am not biased, everyone eligible has equal chance of selection) for a job, What do you think is your chance of selection?
- if, the entire class is eligible (44 students)
- if your column is eligible
- if only girls/boys in your column are eligible.

Probability

- Now, say you are not interested in just yourself, but want to know, what is the chance that someone from your column will get selected?

Probability

- Now, say you are not interested in just yourself, but want to know, what is the chance that someone from your column will get selected?
- someone from your column will get selected if a girl is to be selected.

Probability

- Now, say you are not interested in just yourself, but want to know, what is the chance that someone from your column will get selected?
- someone from your column will get selected if a girl is to be selected.
- Informal Definition (classical probability)

$$P = \frac{favorable}{total}$$

Probability

- Now, say you are not interested in just yourself, but want to know, what is the chance that someone from your column will get selected?
- someone from your column will get selected if a girl is to be selected.
- Informal Definition (classical probability)

$$P = \frac{favorable}{total}$$

- We also defined conditional probability in the process. We will do the math later, but as of now, if you impose some condition(on possible cases you would consider), then probability given that condition is conditional probability

Chap 1: Design a good experiment

- Definition: Empirical study: Its a way of gaining knowledge by means of direct or indirect observation or experience. So, basically learning from data.
- 2 types of empirical study- experiment and observational study
- Lets look into one of the largest empirical study ever done in history of medicine-Effect of smoking on Lung Cancer by Richard Doll.
- This study presented overwhelming evidence in favor of smoking being a lead "cause" of lung cancer. But for years, people have produced counter arguments.

Where your study can be questioned

- May be those two groups were different, the ones who smoked happened to be more sick or lead different lifestyle compared to the other group.

Where your study can be questioned

- May be those two groups were different, the ones who smoked happened to be more sick or lead different lifestyle compared to the other group.
- May be it was some other factor, environment or so, just like the chocolate and nobel prize study

Where your study can be questioned

- May be those two groups were different, the ones who smoked happened to be more sick or lead different lifestyle compared to the other group.
- May be it was some other factor, environment or so, just like the chocolate and nobel prize study
- May be its just psychological! The people who are smoking know they are smoking and hence they are causing harm to themselves and their lungs. So, they happen to develop cancer.

Where your study can be questioned

- May be those two groups were different, the ones who smoked happened to be more sick or lead different lifestyle compared to the other group.
- May be it was some other factor, environment or so, just like the chocolate and nobel prize study
- May be its just psychological! The people who are smoking know they are smoking and hence they are causing harm to themselves and their lungs. So, they happen to develop cancer.
- we will see next class how to design a study to address those concerns- randomized, controlled, double blinded study !