

Introduction to Statistical Thinking

Felipe José Bravo Márquez

March 22, 2021

Introduction to Statistical Thinking

- Statistical thinking is a systematic way of thinking about how we describe the **world** and use **data** make decisions and predictions.
- Taking into account the inherent **uncertainty** that exists in the real world. [Poldrack, 2019]
- The foundations of statistical thinking come primarily from mathematics and statistics, but also from computer science, psychology, and other fields of study. [Poldrack, 2019]
- **Statistics**, in particular, is the discipline that concerns the collection, organization, analysis, interpretation, and presentation of data [Wikipedia, 2021].



Statistical Thinking and Intuition

- Human **intuition** often tries to answer the same questions that we can answer using statistical thinking, but often gets the answer **wrong**.
- For example, in recent years most Americans have reported that they think that violent crime was worse compared to the previous year (Pew Research Center).
- However, a statistical analysis of the actual crime data shows that in fact violent crime has steadily decreased since the 1990's.
- Intuition fails us because we rely upon **best guesses** (which psychologists refer to as heuristics) that can often get it wrong. [Poldrack, 2019]

What can statistics do for us?

There are three major things that we can do with statistics:

- **Describe:** The world is complex and we often need to describe it in a simplified way that we can understand.
- **Decide:** We often need to make decisions based on data, usually in the face of uncertainty.
- **Predict:** We often wish to make predictions about new situations based on our knowledge of previous situations.

There two radical approaches to teach an introductory course on statistics.

Mathematical Statistics

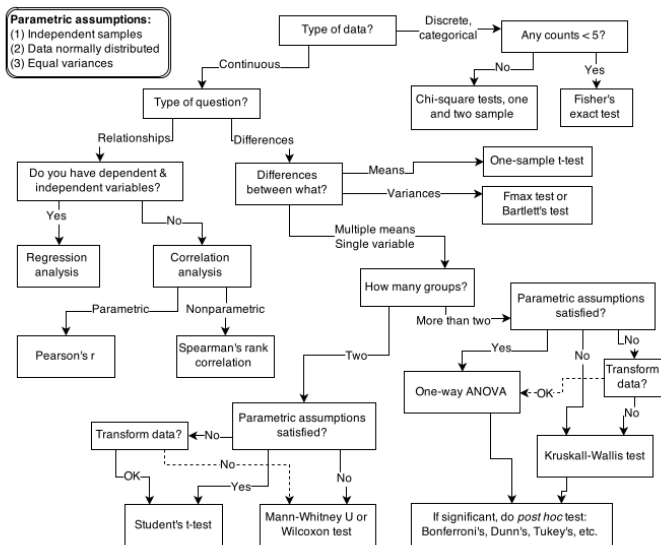
- This is usually taught in mathematics departments.
- The focus is on the mathematical aspects of statistics (e.g., asymptotic properties).

Applied Statistics

- This is usually taught in health and social science departments (e.g., psychology).
- The focus is on the methods (e.g., t-test, ANOVA, linear regression) and how to apply them to the specific field.

- In my personal opinion mathematical courses are in many cases hard to grasp for non-mathematical students and don't pay enough attention to the use of statistics in real world problems.
- On the other hand, applied courses sometimes omit the fundamental assumptions on which the methods are based and end up being a kind of recipe book of methods for ad hoc problems.
- This course distinguishes itself from the two previous paradigms by focusing on the understanding of the basic ideas of statistical thinking.
- We will reflect on the assumptions on which the different models are based, and we will be aware that no method or tool fits perfectly to every problem.

Example of flowchart (recipe book) taken from [McElreath, 2020]



The main topics that will be covered in this course are:

- Statistical Programing in R
- Descriptive Statistics
- Probability
- Frequentist Inference
- Bayesian Inferece

References I



McElreath, R. (2020).

Statistical rethinking: A Bayesian course with examples in R and Stan.

CRC press.



Poldrack, R. A. (2019).

Statistical Thinking for the 21st Century.



Wikipedia (2021).

Statistics — Wikipedia, the free encyclopedia.

<http://en.wikipedia.org/w/index.php?title=Statistics&oldid=1011947657>.

[Online; accessed 19-March-2021].