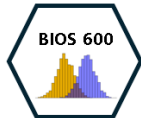


# Lec 1: What is statistical inference?

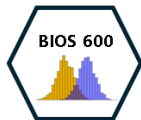
BIOS 600 - Fall 2024

Dr. Kara McCormack

UNC - Chapel Hill



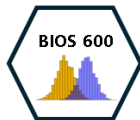
# Biostatistics



# What is biostatistics?

A process that converts data into useful information, whereby practitioners

1. form a question of interest
2. collect and summarize data
3. and interpret the results

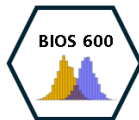


# Identifying the population and question of interest

The population is the group we'd like to learn something about:

- ▶ What is the prevalence of diabetes among US adults, and has it changed over time?
- ▶ Is there a relationship between tumor type and five-year mortality in breast cancer patients?
- ▶ Does the average amount of caffeine vary by vendor in 12 oz cups of coffee at UNC coffee shops?

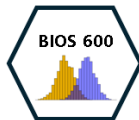
If we had data from every unit in the population, we could just calculate what we wanted and be done!



# Sampling from the population

Unfortunately, we (usually) have to settle with a sample from the population.

- ▶ Ideally, the sample is representative, allowing us to use probability and statistical inference to make conclusions that are generalizable to the broader population of interest.



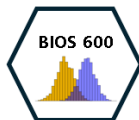
# Sampling methods

Probability sampling (e.g., simple random sampling, stratified, cluster, or multi-stage sampling)

- ▶ All units have a known chance of being selected
- ▶ More likely to be generalizable
- ▶ Can be more expensive and time-consuming

Non-probability sampling (e.g. quota, convenience, or snowball sampling)

- ▶ Some units unable to be selected, with no way of knowing size or effect of sampling errors
- ▶ Less generalizable to population of interest
- ▶ More convenient and less costly



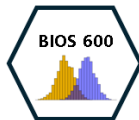
# Study design

## Experimental studies (e.g. RCTs)

- ▶ Researchers directly control exposures or treatments
- ▶ Ability to make causal statements
- ▶ Less real-world applicability and generalizability

## Observational studies (e.g. surveys)

- ▶ Researchers do not assign exposures or treatments
- ▶ Real-world setting with lower burden on participants
- ▶ Inability to prove causality



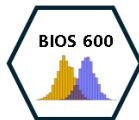
# What can go wrong?

Selection bias, reporting bias, non-response bias, attrition bias, confounding, detection bias, lack of blinding, straight-up falsified data (this happens), ...

## Catalogue of Bias



and so much more





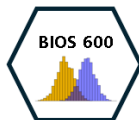
# What is biostatistics good for?

*The NEW ENGLAND JOURNAL of MEDICINE*

## SPECIAL ARTICLE

### Mortality in Puerto Rico after Hurricane Maria

Nishant Kishore, M.P.H., Domingo Marqués, Psy.D., Ayesha Mahmud, Ph.D.,  
Mathew V. Kiang, M.P.H., Irmay Rodriguez, B.A., Arlan Fuller, J.D., M.A.,  
Peggy Ebner, B.A., Cecilia Sorensen, M.D., Fabio Racy, M.D., Jay Lemery, M.D.,  
Leslie Maas, M.H.S., Jennifer Leaning, M.D., S.M.H., Rafael A. Irizarry, Ph.D.,  
Satchit Balsari, M.D., M.P.H., and Caroline O. Buckee, D.Phil.



# What is biostatistics good for?

Research

JAMA Internal Medicine | [Original Investigation](#)

## Comparison of Hospital Mortality and Readmission Rates for Medicare Patients Treated by Male vs Female Physicians

Yusuke Tsugawa, MD, MPH, PhD; Anupam B. Jena, MD, PhD; Jose F. Figueroa, MD, MPH; E. John Orav, PhD; Daniel M. Blumenthal, MD, MBA; Ashish K. Jha, MD, MPH



# What is biostatistics good for?

RESEARCH ARTICLE

Open Access

## Mass incarceration and public health: the association between black jail incarceration and adverse birth outcomes among black women in Louisiana

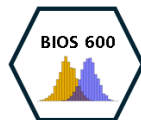


Lauren Dyer<sup>1\*</sup> , Rachel Hardeman<sup>2</sup>, Dovile Vilda<sup>1</sup>, Katherine Theall<sup>1</sup> and Maeve Wallace<sup>1</sup>



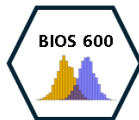
# Course FAQ

- ▶ **What background is assumed for the course?** A basic understanding of algebra and arithmetic. No previous coursework in probability and statistics is required.



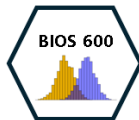
## Course FAQ

- ▶ **What background is assumed for the course?** A basic understanding of algebra and arithmetic. No previous coursework in probability and statistics is required.
- ▶ **Will we be doing computing?** Yes. We will use R. You're welcome to check your work in another language (e.g. SAS), but homework/labs will be turned in using R.



# Course FAQ

- ▶ **What background is assumed for the course?** A basic understanding of algebra and arithmetic. No previous coursework in probability and statistics is required.
- ▶ **Will we be doing computing?** Yes. We will use R. You're welcome to check your work in another language (e.g. SAS), but homework/labs will be turned in using R.
- ▶ **Will we learn the mathematical theory of biostatistics?** Yes and No. The course is primarily focused on application; however, we will discuss some of the mathematics of biostatistics.



# Some of what you'll learn

- ▶ Apply descriptive techniques commonly used to summarize public health data.
- ▶ Learn methods to analyze real-world data to answer research questions in a biomedical setting.
- ▶ Use [Quarto](#) to write reproducible reports.
- ▶ Communicate results from statistical analyses to a general audience.

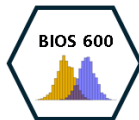


# Reproducibility and replicability

- ▶ **Reproducibility:** being able to take original data and code to reproduce all numerical findings.
- ▶ **Replicability:** being able to independently repeat an entire study without use of the original data (generally with the same methods)

Best practices from the American Statistical Association:

- ▶ End-to-end scripting of research
- ▶ Use of version control and documentation (we will not use this in our class)
- ▶ Publication of code along with data





The current replication crisis

# Drip, drip: Former cell researcher up to 18 retractions



About OIG ▾

Reports ▾

Fraud ▾

Compliance ▾

[Home](#) > [Fraud](#) > [Enforcement Actions](#) > Duke University

Piero Anversa, a former star researcher at Harvard Medical School who left the institution under a cloud, is up to 18 retractions. In barely half of the 31 papers by Anversa that Harvard has requested journals publish, concerns about the integrity of the findings.

## Duke University Agrees to Settle False Claims Act Scientific Research Misconduct



Piero Anversa

The two articles, published in the *Proceedings of the National Academy of Sciences*, appeared in

2008 and 2009. Anversa and a frequent co-author, Annarosa Leri, are