Week 1: Scientific Inference & Statistical Goals

ANTH 674: Research Design & Analysis in Anthropology

Professor Andrew Du
Andrew.Du2@colostate.edu

1

Lecture outline

1. Introductions & course description
2. What is science? (the demarcation problem)
3. What is statistics?
   - What role does statistics play in science?

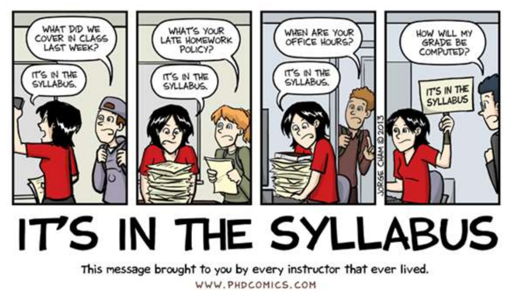
2

Introductions

- What is your name (and pronouns)?
- What department are you in?
- Briefly, what do you study, or what are you interested in?
- What is your background in statistics/data analysis/R coding?
- What are you looking forward to learning about?
- What are you nervous/concerned about?

3

The syllabus



4

Syllabus learning objectives

- Move away from the “cookbook” mentality of statistics to a more **nuanced, philosophical** approach to scientific inference.
- Be able to translate a research question into a **statistical one** that can be addressed with the proper statistical methods.
- Be able to collect and organize data in a manner that is suitable for data analysis.
- Be able to choose the right way to visualize data and results, in a way that is clear and impactful.

5

Syllabus learning objectives

- Develop a **working** knowledge of the breadth of statistical techniques in anthropology.
- Be able to critically evaluate the statistical methods used and reported in the published literature.
- Build a strong enough statistical knowledge base so that you can teach yourself more advanced methods.
- Get hands-on experience analyzing data, and become a proficient and confident coder in R.

6

What this course isn’t

- **NOT** a statistics course per se
- **NOT** a mathematical modeling course
- **NOT** a programming course per se

7

What this course is

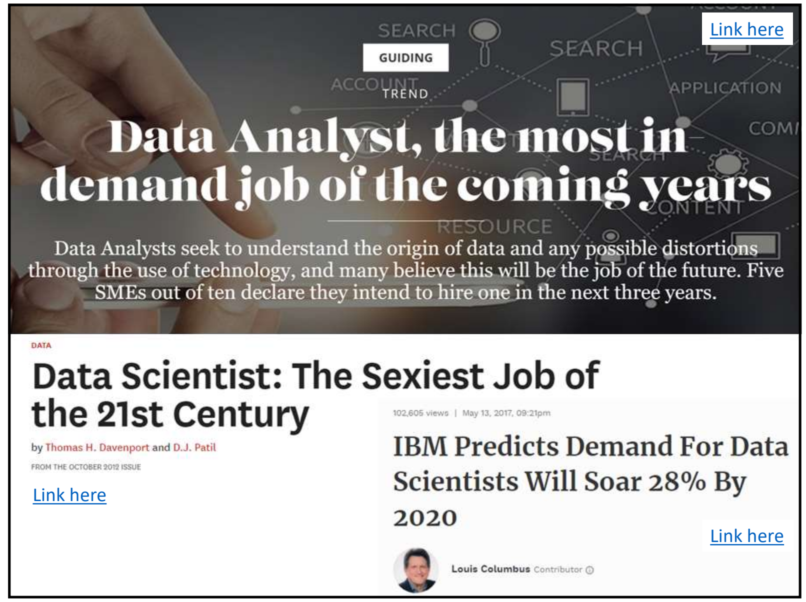
- **Applied** statistics & coding course
- Emphasizes **scientific inference** over mechanics of methods
- Teaches you what you need to know to do research in anthropology **from formulating a research question → interpreting statistical results**
- Emphasizes **breadth** of methods over depth

8

Why this course is important

- Data analysis is an incredibly marketable skill!
  - Cleaning, visualizing, and extracting insights from data (what scientists do too!)

9



10

Why this course is important

- Data analysis is an incredibly marketable skill!
  - Cleaning, visualizing, and extracting insights from data (what scientists do too!)
- Large databases are becoming common in (biological) anthropology
  - Need to know how to organize and analyze large datasets
  - And how to not abuse large datasets!

11

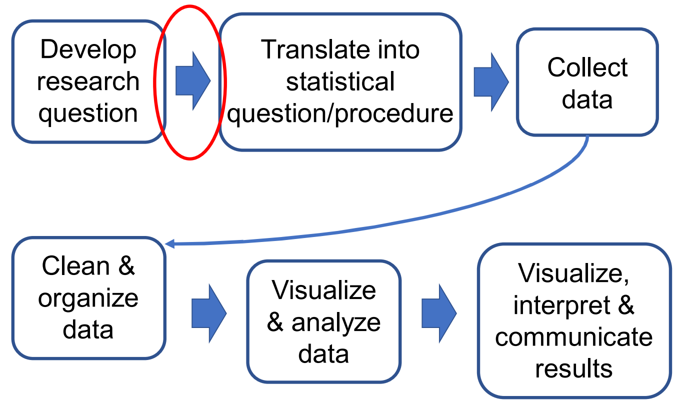
12

Why this course is important

- Data analysis is an incredibly marketable skill!
  - Cleaning, visualizing, and extracting insights from data (what scientists do too!)
- Large databases are becoming common in (bio)anthropology
  - Need to know how to organize and analyze large datasets
  - And how to not abuse large datasets!
- Data analysis is an integral part of the research pipeline

13

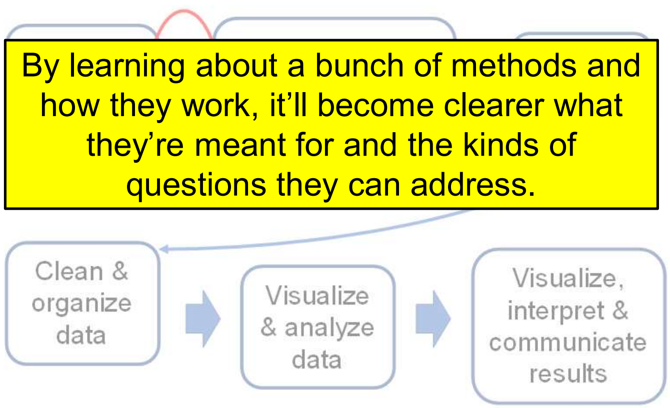
The research pipeline
(broadly speaking)



14

The research pipeline
(broadly speaking)

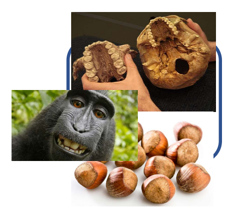
By learning about a bunch of methods and how they work, it'll become clearer what they're meant for and the kinds of questions they can address.



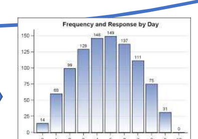
15

A simple example

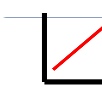
*This is essentially your final paper



Day	Frequency	Response
0	14	0.09983
1	60	0.40242
2	99	0.66562
3	129	0.86366
4	146	0.97716
5	149	0.99500
6	137	0.91545
7	111	0.74629



Slope
P-value



"Relationship is significant…"

16

Questions?

17

Brian McGill

What is science?

18

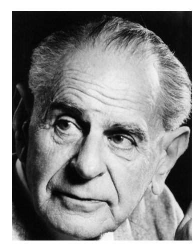
Which of these are scientific?

- The Earth is round
- Gravity attracts bodies in proportion to their mass and distance from each other
- Witchcraft is real
- I know aliens exist

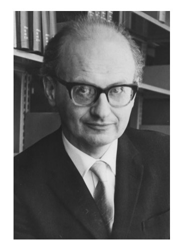
19

The demarcation problem

- Science vs. pseudoscience
- Is there a distinction? What is it?
- Important social and political implications!



Karl Popper



Imre Lakatos

20

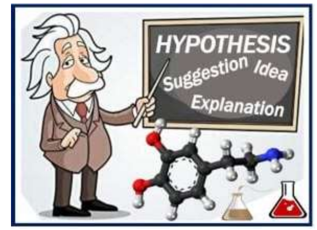
Science usually defined as:

- "Science tests hypotheses against observations or experimental results."

21

What is a hypothesis?

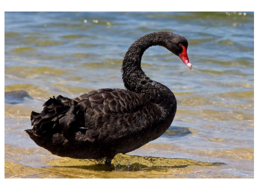
- Google says: "a supposition or proposed explanation made on the basis of limited evidence as a starting point for further investigation."

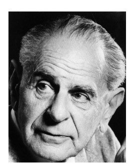


22

What does "testing" mean?

- Testing = hypothesis needs to be **falsifiable** when compared to empirical data
  - Example: "All swans are white."
- Hypothesis **can never be proven**: always the possibility that future data will contradict it
  - Example: Black swans discovered in 1697!





Popper

23

How falsification is done

The hypothetico-deductive (HD) method

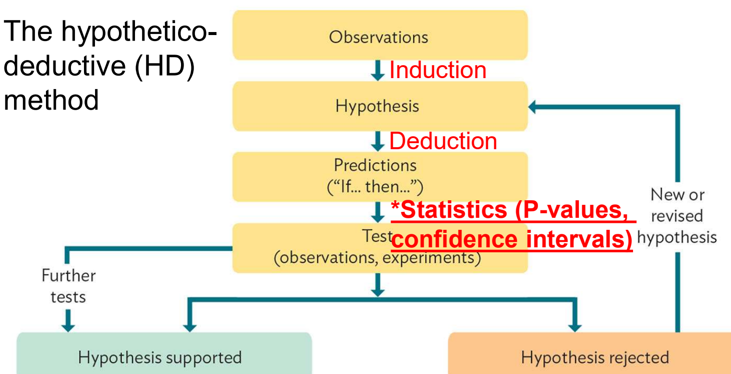


FIGURE 1.8
The Scientific Method: How We Know What We Know

Larsen 2020

24

Competing multiple hypotheses

- Instead of testing hypotheses one at a time, compete multiple against each other
- "Natural selection" of ideas

Observation

H1	H2	H3	H4
Test	Test	Test	Test
✗	✗	✓	✗
0.05	**0.03**	**0.90**	**0.02**


Thomas Chamberlin

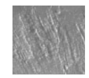

John Platt

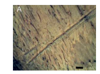
*Statistics formalizes hypothesis selection!

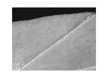
25

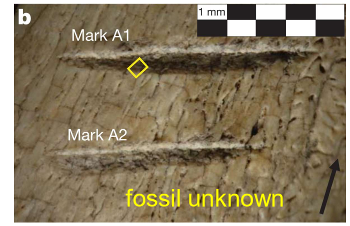
Example: bone surface modifications

- Instead of H_{tooth_mark} →  ✗
- $H_{tooth\text{-}mark}$ →  0.01

  $H_{perc.\text{-}mark}$ →  0.01

  $H_{trample}$ →  0.43

  $H_{cut\text{-}mark}$ →  0.45

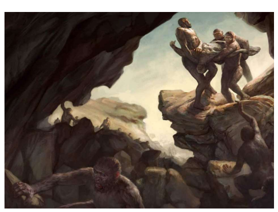


3.4 Myr Dikika "cutmarks"

26

Issues with this paradigm

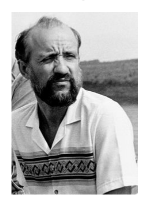
- For any data, there are virtually infinite number of hypotheses explaining it
- Often, patterns are multicausal → hypotheses not mutually exclusive
- What if you exclude the "correct" hypothesis?



27

Example: Winfree's fungi

- Examined spatial pattern of fungi grown in petri dish
- Developed 18 hypotheses to be tested experimentally
- Falsified all 18!
- Left ecology & became world-famous mathematical biologist studying heart defibrillations


Arthur Winfree

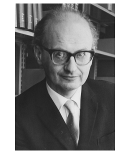
28

Is falsification a reality?

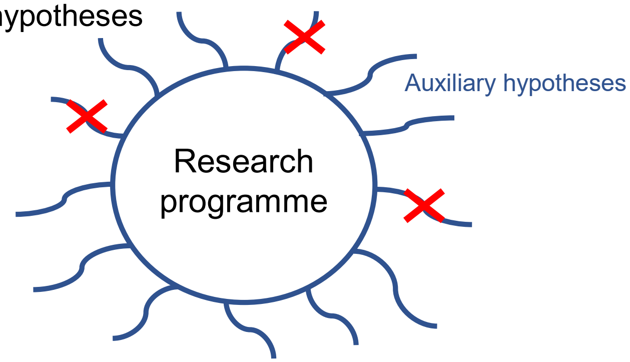
- Are hypotheses (and theories generally) discarded for good when falsified?
- **<u>NO</u>**
  - Researchers are stubborn and hold onto pet hypotheses
  - When you got an unexpected result in high-school chemistry, did you falsify the hypothesis/theory?
  - Copernicus' heliocentrism → stars should move throughout the year, but they do not!
    - Galileo "rescued" heliocentrism by saying stars are *extremely* far away (statement not based on empirical data)

29

Enter Imre Lakatos

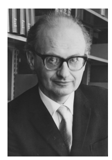


- Fundamental unit of science is the "research programme", protected by a belt of auxiliary hypotheses



30

Enter Imre Lakatos

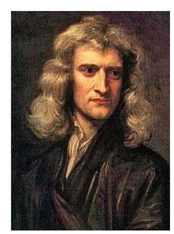


- Fundamental unit of science is the "research programme", protected by a belt of auxiliary hypotheses
- Research programme not discarded at first sign of falsification (of auxiliary hypotheses)
  - Perhaps measurements are faulty
  - Falsification can make programmes stronger, e.g., Galileo rescuing heliocentrism

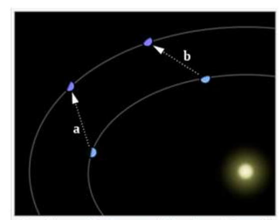
31

Another example: Newton

- Uranus' orbit not predicted by Newton's law of universal gravitation
- Predicted orbit perturbed by undiscovered planet → Neptune



Isaac Newton



At position *a*, Neptune gravitationally perturbs the orbit of Uranus, pulling it ahead of the predicted location. The reverse is true at *b*, where the perturbation retards the orbital motion of Uranus.

32

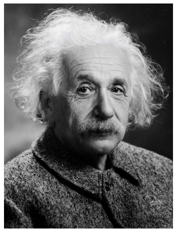
Competing research programmes

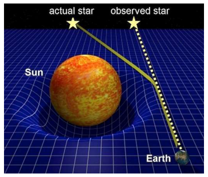
- Degenerate programme – needs too much rescuing (cf. pseudoscience)
- Progressive programme – makes risky, novel, stunning, and precise predictions supported by data
  - "The hallmark of empirical progress is not trivial verifications…It is no success for Newtonian theory that stones, when dropped, fall towards the earth, no matter how often this is repeated…What really count are dramatic, unexpected, stunning predictions."
  - *Statistics provides precise, numerical predictions and tells us whether data matches them or not

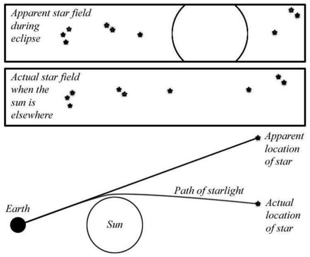
33

An example: Einstein

- Einstein's general relativity predicted Mercury orbit & ***exact*** amount light is bent by gravity
- Example of a new research programme superseding Newton's?


Albert Einstein

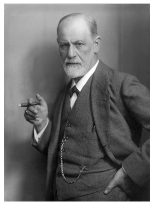


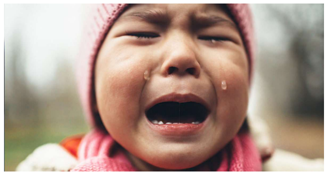


34

What then does pseudoscience look like?

- Hypotheses/predictions are so vague that they can fit any data (not falsifiable)
- Rescues hypotheses by modifying them to fit the data (ad hoc explanation)


Sigmund Freud



35

Class activity

36

My key elements of science

- Competition among hypotheses (though single hypothesis testing has its place)
- Superior hypotheses replace inferior ones
- Ultimate decider is empirical reality

37

Questions?

38

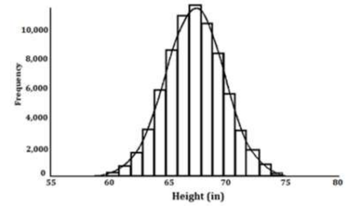
What is statistics?

What role does statistics play in science?

39

What is statistics?

- From Wikipedia: "**Statistics** is the discipline that concerns the collection, organization, analysis, interpretation and presentation of data."
- To me, the key point of statistics is the study of **variation** in data (characterized by distributions)



40

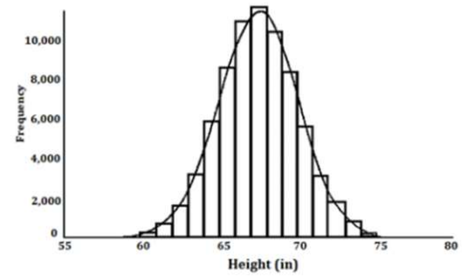
What are sources of variation?

- Individual variation (genetically based)
- Phenotypic plasticity (environmentally based)
- Ontogeny
- Temporal heterogeneity
- Spatial heterogeneity
- Many, many more sources
- **Variation is a fact of life in anthropology!**

41

Variation = distributions

- Variation can be viewed as distributions and how **likely/probable** certain values are



42

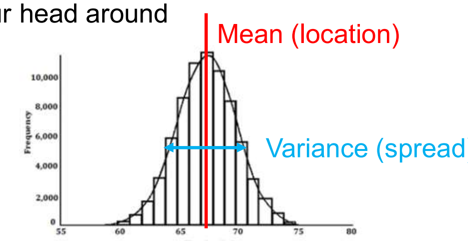
But, humans are terrible at probability

- Fear of flying
  - Annual risk = 1 in 11 million
- Not getting a flu shot
  - 1 death per 50,000 people (2017)
- What is the probability that at least two people in this room have the same birthday?
  - $1 - \frac{365!}{(365-n)! \cdot 365^n}$

43

Statistics helps us with this

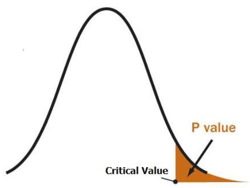
- Offers a formalized, quantitative, & objective way of making inferences from distributions
- For example:
  - Summarize distributions, which can be difficult to get your head around



44

Statistics helps us with this

- Offers a formalized, quantitative, & objective way of making inferences from distributions
- For example:
  - Summarize distributions, which can be difficult to get your head around
  - Estimate probability that data comes from a specified distribution (e.g., P-value)



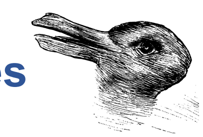
45

Statistics helps us with this

- Offers a formalized, quantitative, & objective way of making inferences from distributions
- For example:
  - Summarize distributions, which can be difficult to get your head around
  - Estimate probability that data comes from a specified distribution (e.g., P-value)
- These are key elements in scientific inference!

46

Humans also have biases



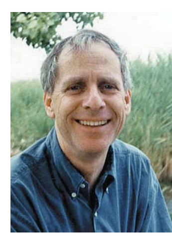
- When data are too few and the results aren't clear, biases dominate
  - E.g., every hominin taxonomist ever

- When data are too numerous and we become overwhelmed, we cherry-pick data according to our biases
  - E.g., social media

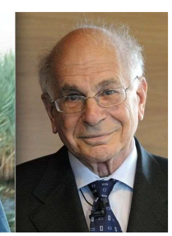
47

AKA behavioral economics

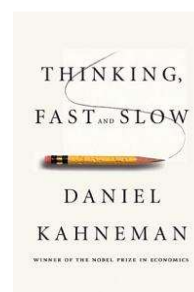
- The study of why humans are bad at making decisions because they're bad at probability and have biases

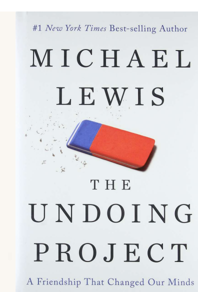


Amos Tversky



Daniel Kahneman





48

Statistics helps us ask the right questions

https://www.youtube.com/watch?v=TpBcwGOvO80

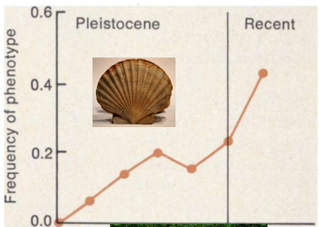
49

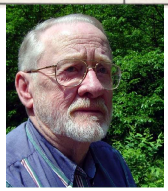
Statistics cuts through biases

www.youtube.com/watch?v=KWPhV6PUr9o

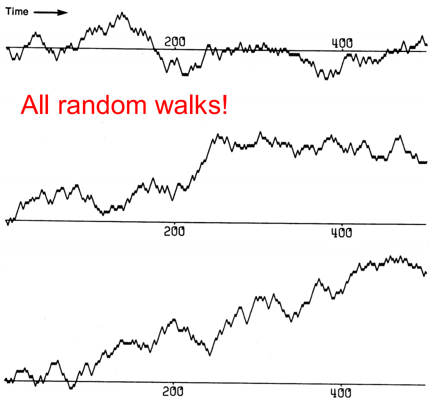
50

Statistics cuts through biases





Raup 1977



51

Statistics is key to scientific inference!

“Statistics is the grammar of science.”

- Karl Pearson

52

Questions?

Summary

- Science is messy! Not 100% demarcated, and problems with all inferential paradigms (choose the right one for your question):
  - Popper's falsifiability
  - Competing multiple hypotheses
  - Lakatos' risky, precise predictions
- Statistics – study of variation in data. Helps us:
  - Interpret probability → distributions → variation
  - Ask the right questions (translating research question into statistical question)
  - Stay objective and cut through biases
- Statistics & scientific inference are intertwined!