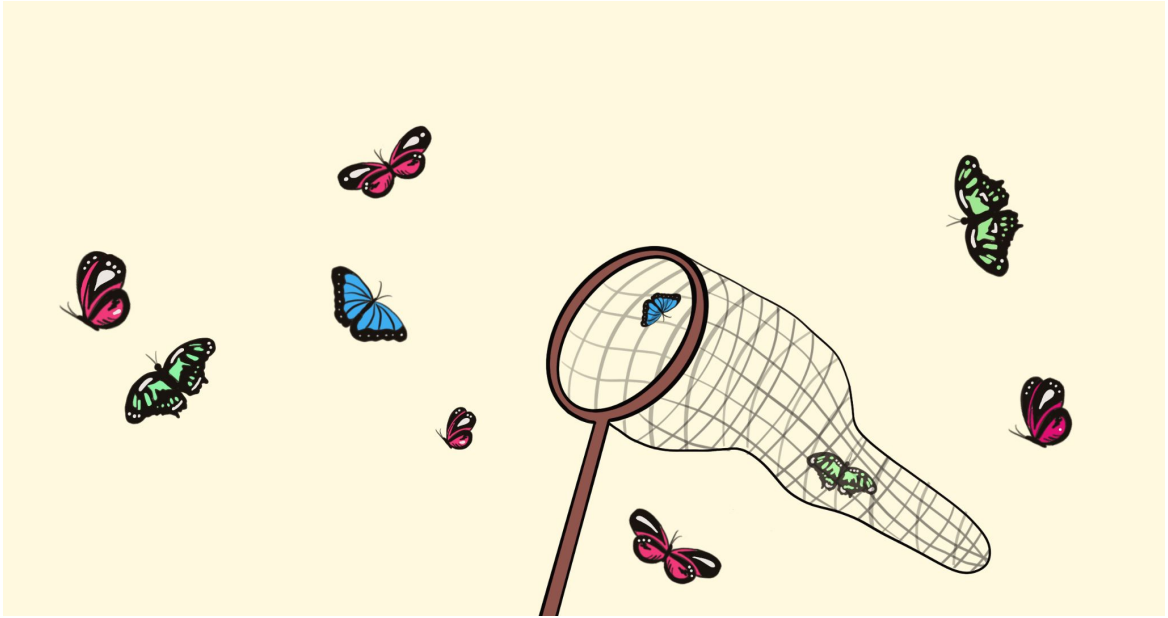
Week 1 Lecture 1:
Populations, samples, and parameters

EDS 222: *Statistics for Environmental Data Science*



Asbestos and mesothelioma

© Wally McNamee

“Science is no inexorable march to truth, mediated by the collection of objective information and the destruction of ancient superstition. Scientists, as ordinary human beings, unconsciously reflect in their theories the social and political constraints of their times.”

-Stephen Jay Gould

Asbestos and mesothelioma

Harvard’s Museum of Comparative Zoology

Building constructed in 1859

Asbestos installed in 1930’s as pipe insulation

Gould hired in 1967

First cancer diagnosis 1982

Asbestos and mesothelioma

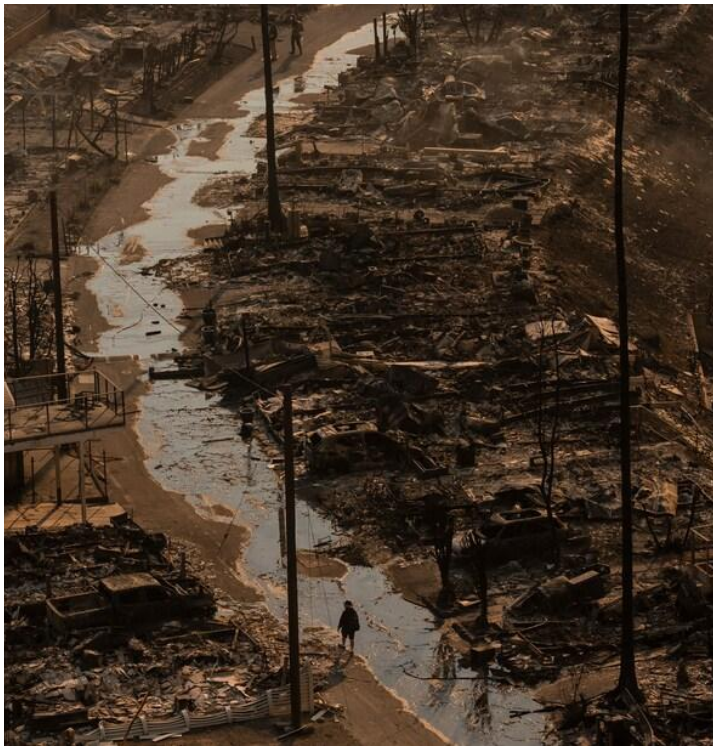


Photo Jae C. Hong/AP

Prevalence of asbestos in Eaton homes (~50%) slowed cleanup times following wildfire

Cleanup contractors improperly dumped asbestos-laden debris in landfills unequipped for toxic waste

Asbestos and mesothelioma

Is prolonged asbestos exposure associated with cases of mesothelioma?

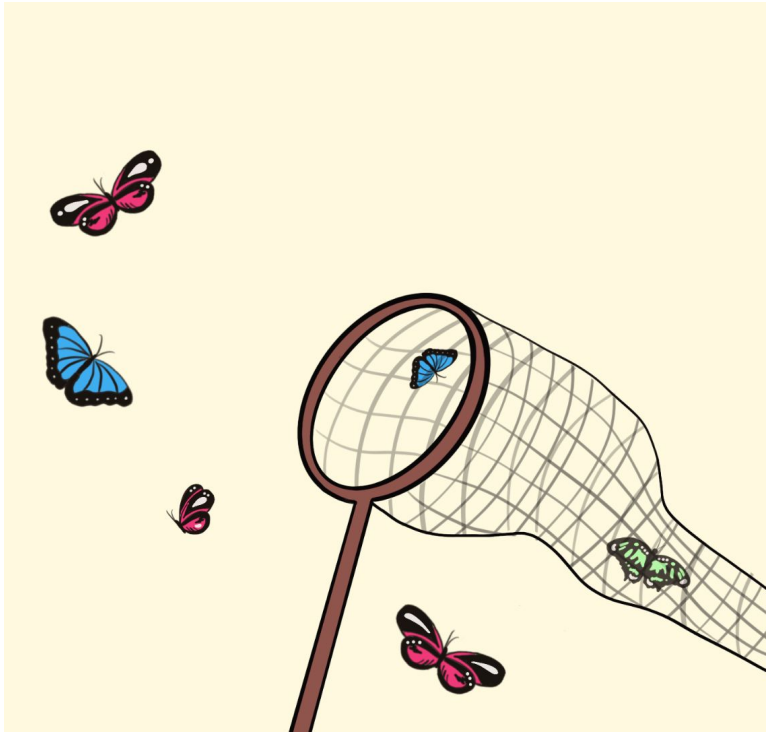
Will increasing rates of extreme weather events release more asbestos into the environment?

Do some communities face relatively more risk of asbestos exposure?

Have policies to protect the public from asbestos reduced exposure risks?

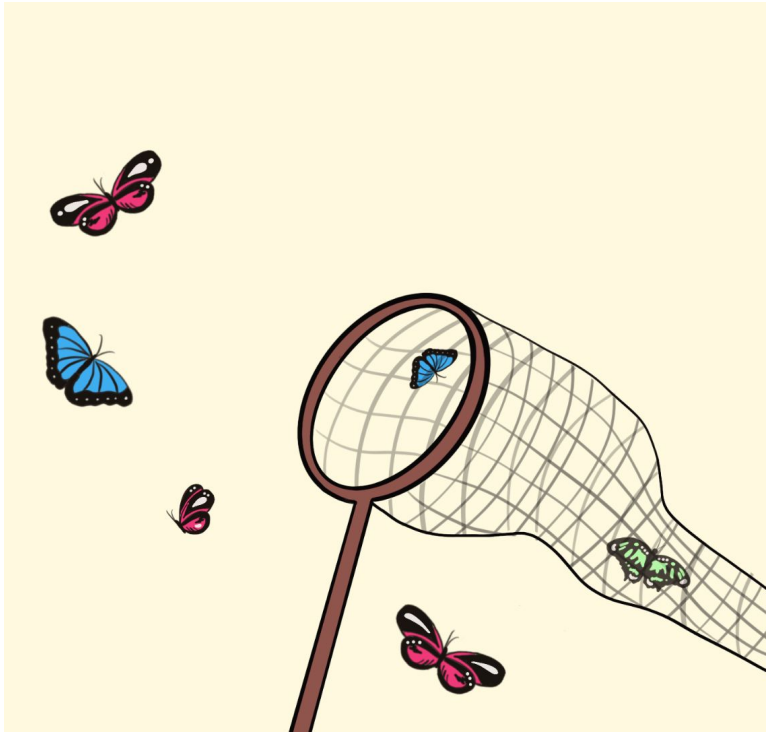
Today’s agenda

- Course structure
  - Learning objectives
  - Schedule
- Statistics anxiety
- Introduce the cast
  - Populations
  - Samples
  - Parameters



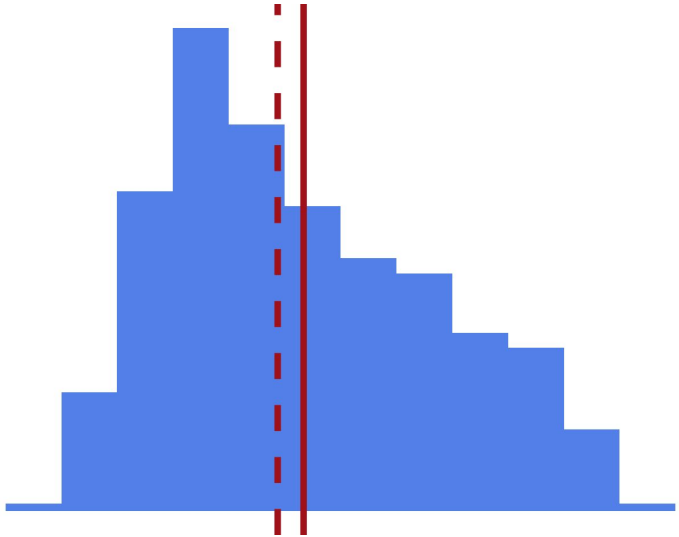
Today's agenda

- **Course structure**
  - **Learning objectives**
  - **Schedule**
- Statistics anxiety
- Introduce the cast
  - Populations
  - Samples
  - Parameters



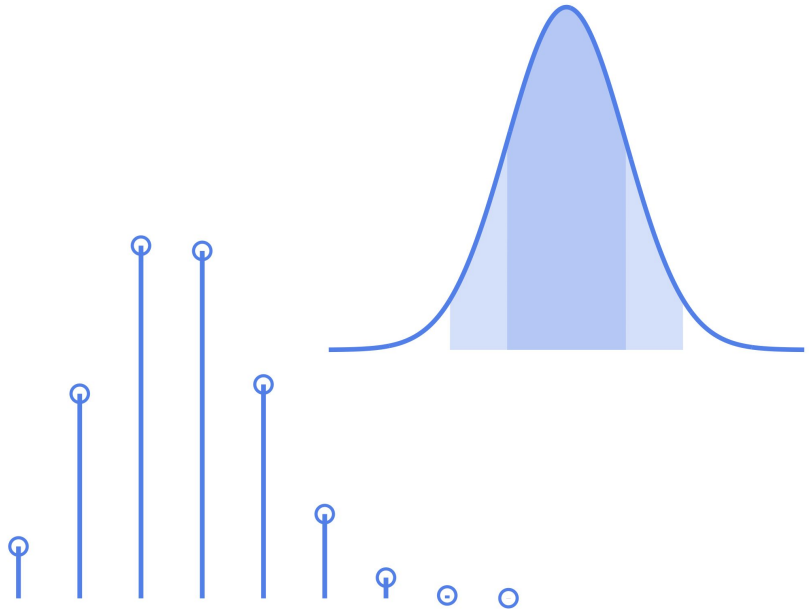
Course structure: Learning objectives

- ➔ **Summarize data** numerically and graphically
- ➔ **Interpret** random variables
- ➔ **Simulate data** to test assumptions
- ➔ **Describe models** using statistical notation
- ➔ **Draw DAGs** to clarify causal relationships
- ➔ **Quantify uncertainty** to perform inference



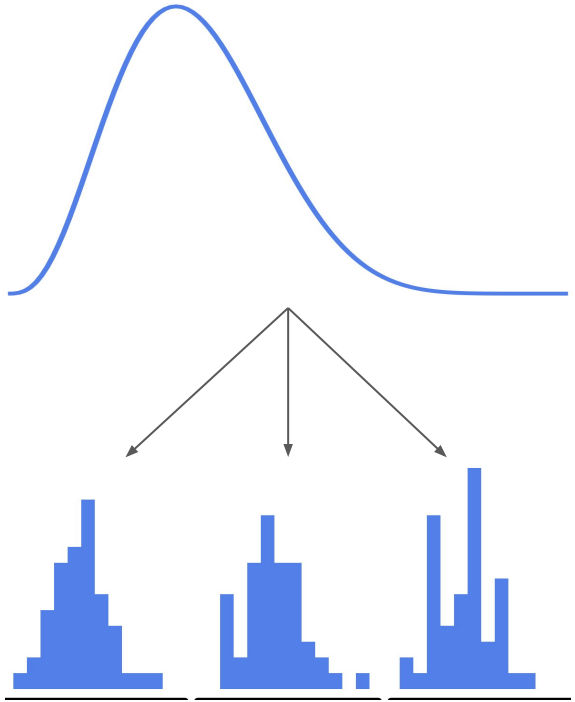
Course structure: Learning objectives

- ➔ **Summarize data numerically and graphically**
- ➔ **Interpret** random variables
- ➔ **Simulate data** to test assumptions
- ➔ **Describe models** using statistical notation
- ➔ **Draw DAGs** to clarify causal relationships
- ➔ **Quantify uncertainty** to perform inference



Course structure: Learning objectives

- ➔ **Summarize data numerically and graphically**
- ➔ **Interpret** random variables
- ➔ **Simulate data** to test assumptions
- ➔ **Describe models** using statistical notation
- ➔ **Draw DAGs** to clarify causal relationships
- ➔ **Quantify uncertainty** to perform inference



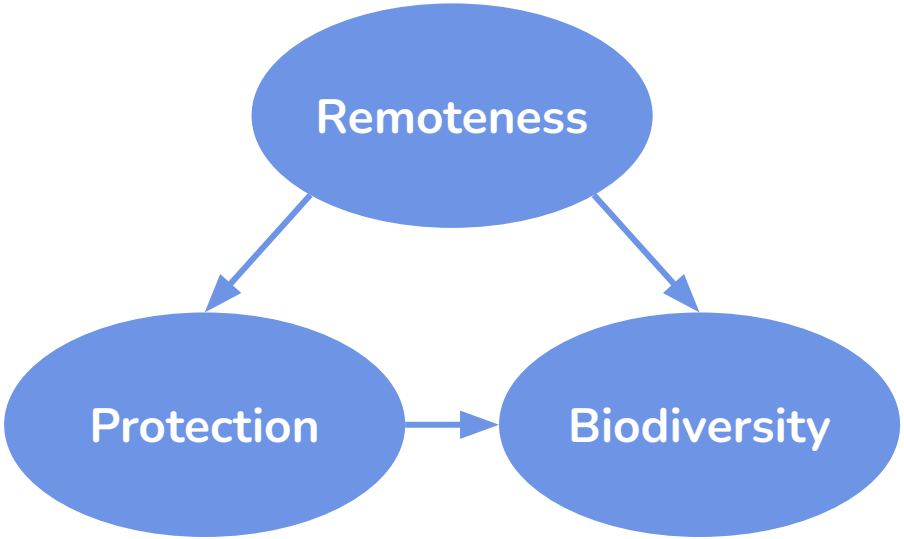
Course structure: Learning objectives

- ➔ **Summarize data numerically and graphically**
- ➔ **Interpret** random variables
- ➔ **Simulate data** to test assumptions
- ➔ **Describe models** using statistical notation
- ➔ **Draw DAGs** to clarify causal relationships
- ➔ **Quantify uncertainty** to perform inference

$$y \sim Binomial(p)$$
$$logit(p) = \beta_0 + \beta_1 x$$

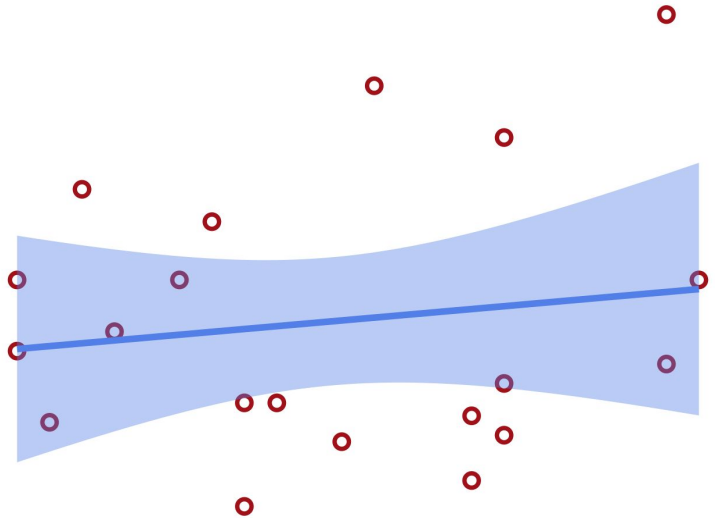
Course structure: Learning objectives

- ➔ **Summarize data numerically and graphically**
- ➔ **Interpret** random variables
- ➔ **Simulate data** to test assumptions
- ➔ **Describe models** using statistical notation
- ➔ **Draw DAGs** to clarify causal relationships
- ➔ **Quantify uncertainty** to perform inference



Course structure: Learning objectives

- ➔ **Summarize data numerically and graphically**
- ➔ **Interpret** random variables
- ➔ **Simulate data** to test assumptions
- ➔ **Describe models** using statistical notation
- ➔ **Draw DAGs** to clarify causal relationships
- ➔ **Quantify uncertainty** to perform inference

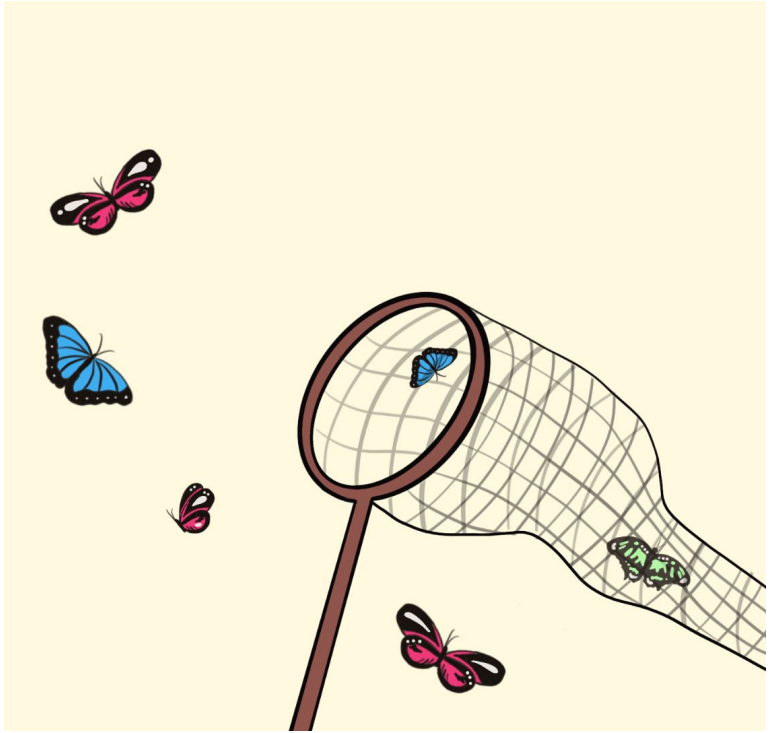


Course structure: Schedule

Weeks 1-3	**Exploring uncertainty**	Homework 1	Midterm 1	Choose data and question
Weeks 4-6	**Conducting inference**	Homework 2	Midterm 2	Exploratory analysis
Weeks 7-9	**Non-normal data**	Homework 3	Midterm 3	Describe hypotheses
				Simulate data
				Fit model to real data

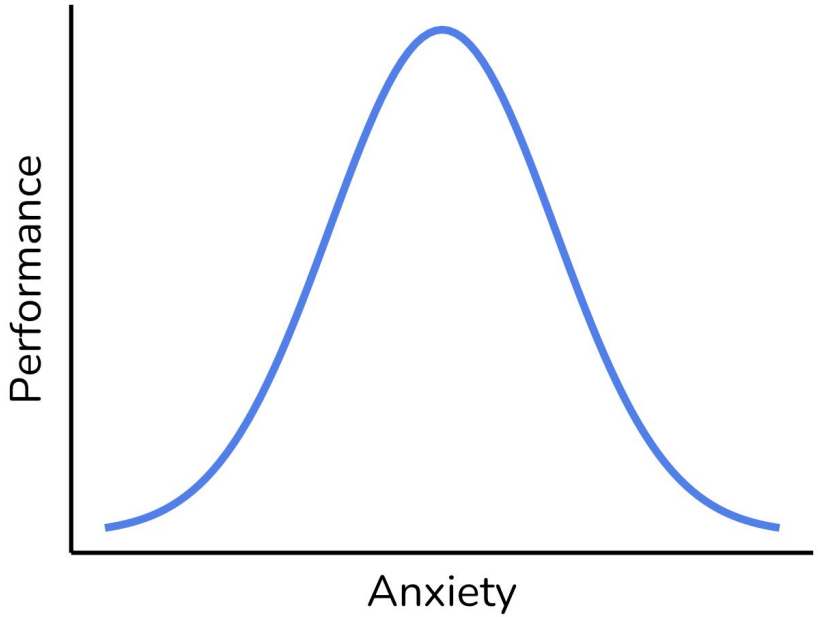
Today’s agenda

- ➔ Course structure
  - ◆ Learning objectives
  - ◆ Schedule
- ➔ **Statistics anxiety**
- ➔ Introduce the cast
  - ◆ Populations
  - ◆ Samples
  - ◆ Parameters



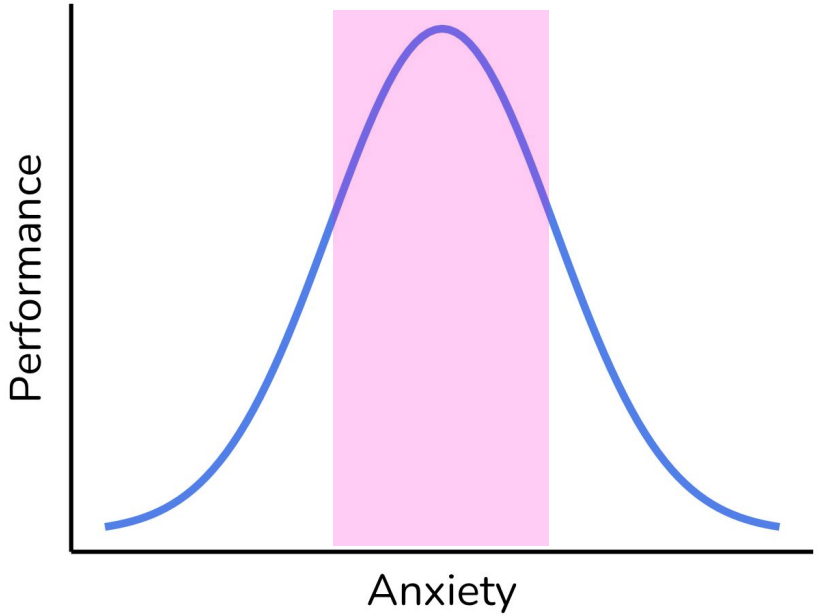
Statistics anxiety

- ➔ Statistics anxiety is widespread among graduate students
- ➔ There is an optimal level of anxiety that helps you succeed!
- ➔ But excessive anxiety is associated with negative outcomes:
  - ◆ Assignment procrastination
  - ◆ Exam freeze



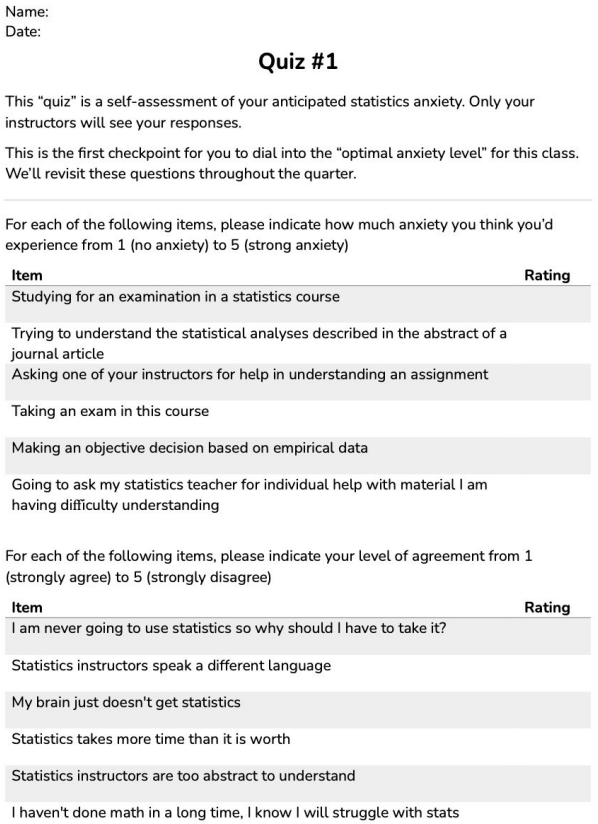
Statistics anxiety

- ➔ Statistics != mathematics
- ➔ Lower stakes for individual assessments
- ➔ Scaffolded final project



Statistics anxiety

- ➔ Statistical Anxiety Rating Scale (STARS)
- ➔ Help your instructors gauge the temperature of the class
- ➔ Encourage you to think about your learning (metacognition)



Name:
Date:

Quiz #1

This “quiz” is a self-assessment of your anticipated statistics anxiety. Only your instructors will see your responses.

This is the first checkpoint for you to dial into the “optimal anxiety level” for this class. We’ll revisit these questions throughout the quarter.

For each of the following items, please indicate how much anxiety you think you’d experience from 1 (no anxiety) to 5 (strong anxiety)

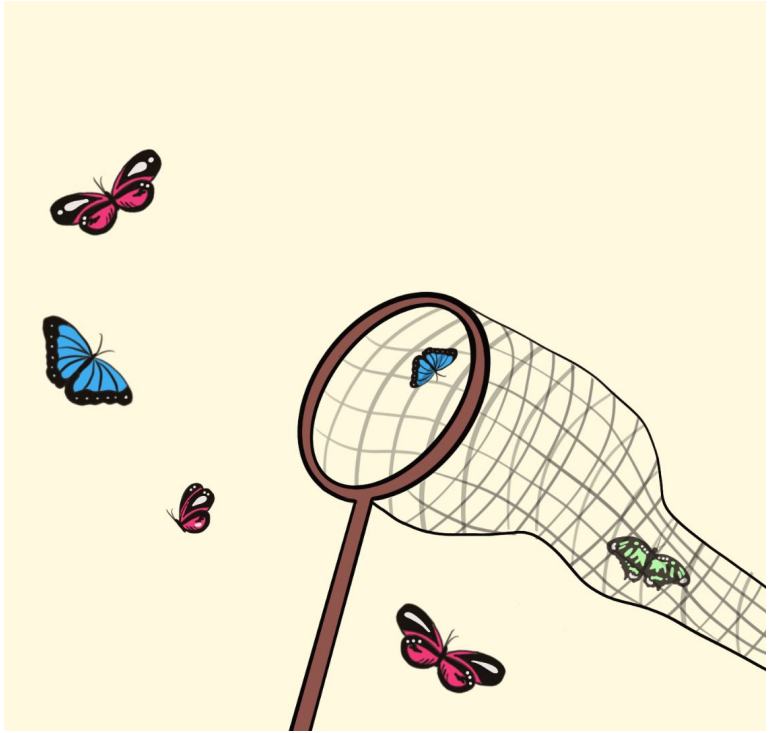
Item	Rating
Studying for an examination in a statistics course	
Trying to understand the statistical analyses described in the abstract of a journal article	
Asking one of your instructors for help in understanding an assignment	
Taking an exam in this course	
Making an objective decision based on empirical data	
Going to ask my statistics teacher for individual help with material I am having difficulty understanding	

For each of the following items, please indicate your level of agreement from 1 (strongly agree) to 5 (strongly disagree)

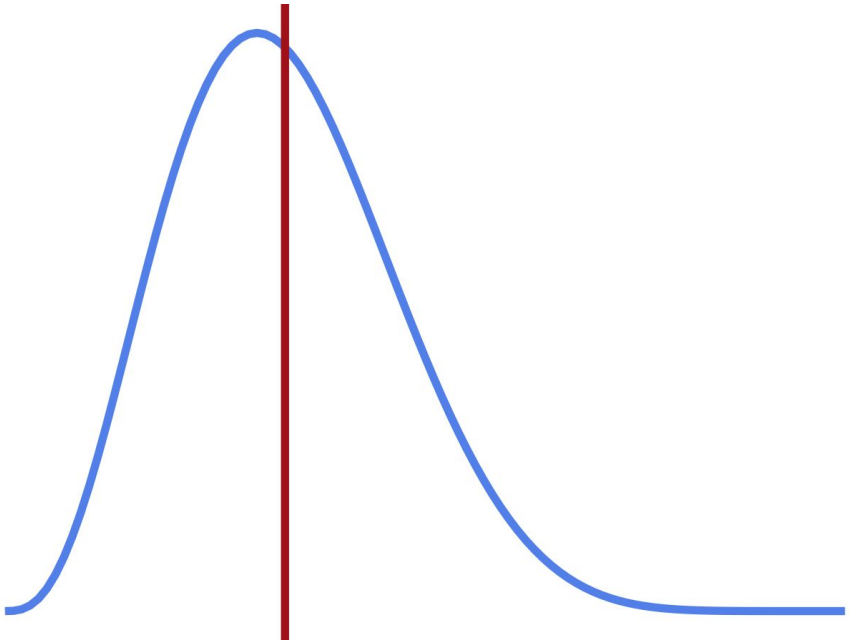
Item	Rating
I am never going to use statistics so why should I have to take it?	
Statistics instructors speak a different language	
My brain just doesn't get statistics	
Statistics takes more time than it is worth	
Statistics instructors are too abstract to understand	
I haven't done math in a long time, I know I will struggle with stats	

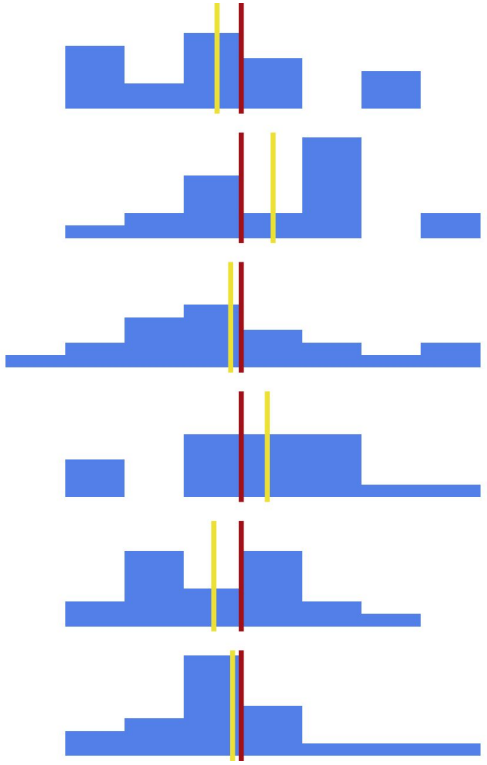
Today's agenda

- ➔ Course structure
  - ◆ Learning objectives
  - ◆ Schedule
- ➔ Statistics anxiety
- ➔ **Introduce the cast**
  - ◆ **Populations**
  - ◆ **Samples**
  - ◆ **Parameters**



Definitions





Definitions

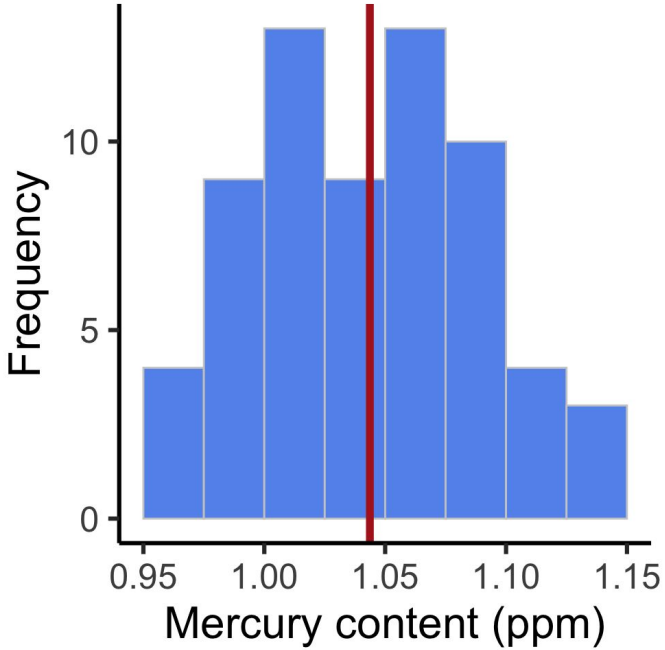
Population

Parameter

Sample

Statistic

Examples of statistics



Question

Statistic

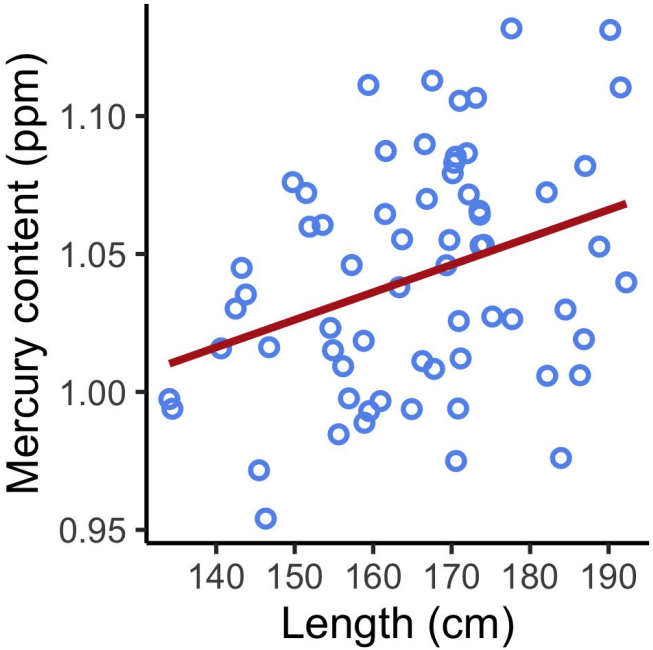
Examples of statistics

Species	Safe	Unsafe	Total
Swordfish	96	19	115
Tuna	72	10	82
Total	168	29	197

Question

Statistic

Examples of statistics



Question

Statistic

Asbestos and mesothelioma

1. Is prolonged asbestos exposure associated with cases of mesothelioma?

2. Do some communities face relatively more risk of asbestos exposure?

3. Will increasing rates of extreme weather events release more asbestos into the environment?

4. Have policies to protect the public from asbestos reduced exposure risks?

Asbestos and mesothelioma

Question #:

Population

Parameter/Statistic

Sample

Asbestos and mesothelioma

Question #:

Sketch your figure/table here

Asbestos and mesothelioma

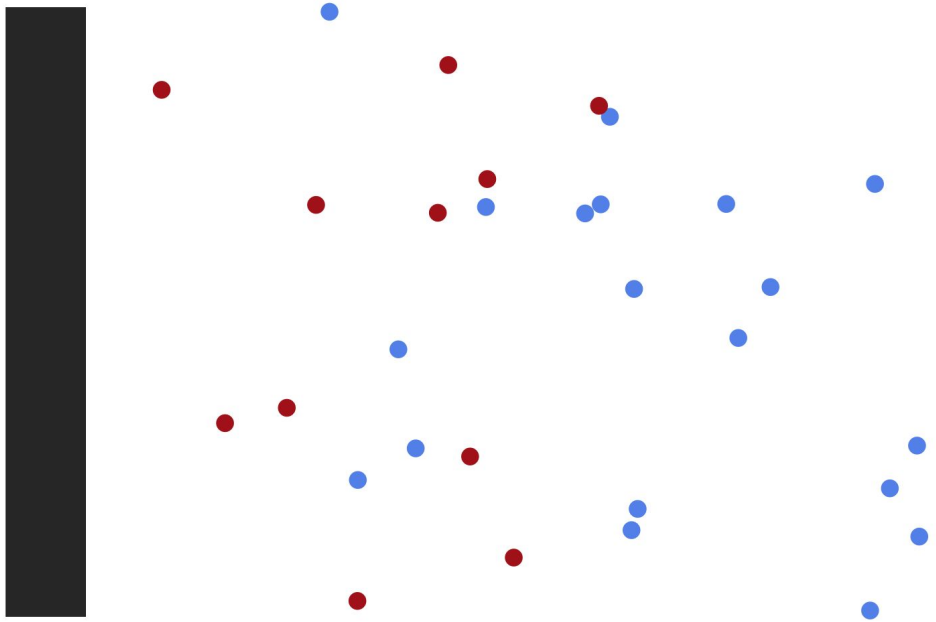
1. Is prolonged asbestos exposure associated with cases of mesothelioma?

2. Do some communities face relatively more risk of asbestos exposure?

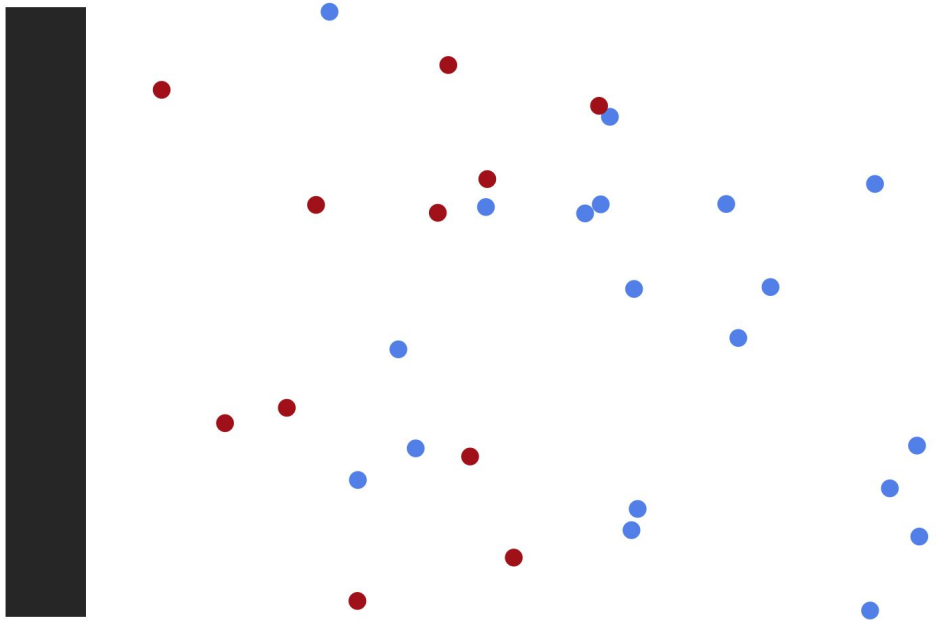
3. Will increasing rates of extreme weather events release more asbestos into the environment?

4. Have policies to protect the public from asbestos reduced exposure risks?

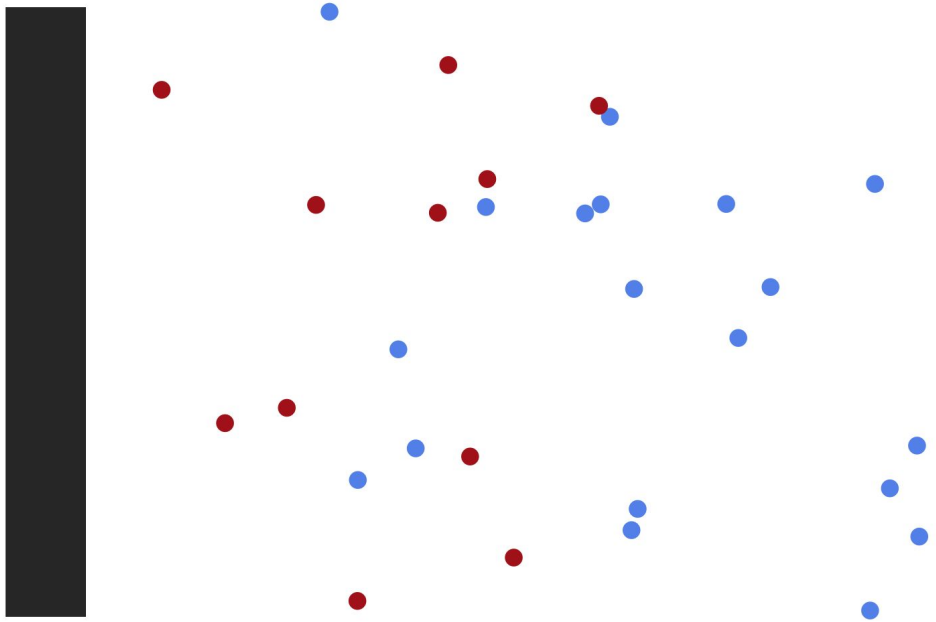
How do we get samples?



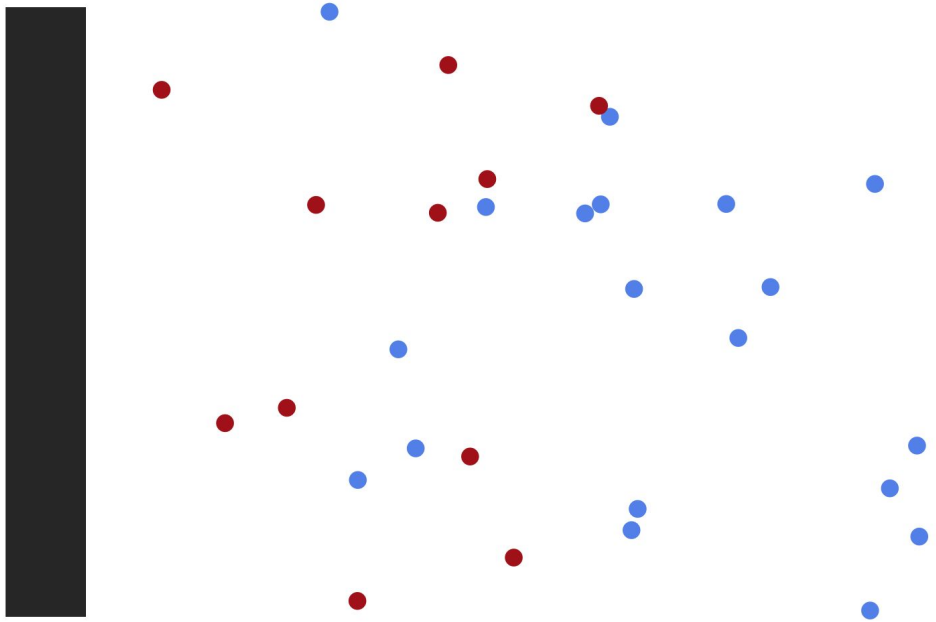
How do we get samples?



How do we get samples?



How do we get samples?



How do we get samples?

What do we do with samples?

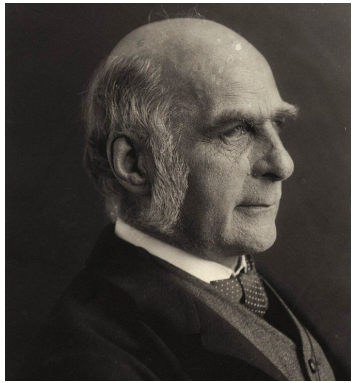
Exploring uncertainty

Conducting inference

Non-normal data

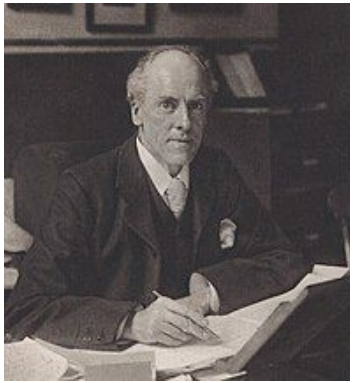
The unethical origins of statistics

Francis Galton



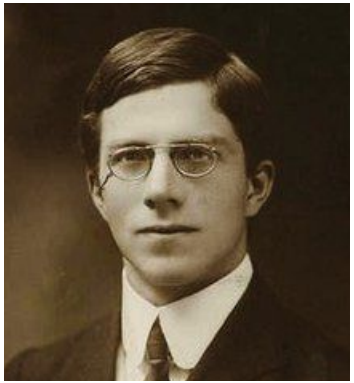
Standard deviation
First weather map

Karl Pearson



Correlation
First statistics department

Ronald Fischer



p-values
Maximum likelihood

What can we learn from statistics' roots in eugenics?

- ➔ The myth of objectivity
- ➔ Correlation is not causation
- ➔ The data may not mean what you think they mean