Welcome to MATH 140

Intro to Statistics

Mike Hitchman

February 10, 2026

Welcome

Goals for Today

1. To meet one another
2. Talk briefly about the course
3. Data literacy
4. Get connected with online course resources
   - course resource page
   - WeBWorK
   - RStudio

Welcome

I have a question for you:

Why are you taking this class?

Welcome

A second question:

> *Write down three areas of special interest you have about the world around us (physical world, society, theatre, art, music, sports, hobbies, whatever interests you that you would like to share)*

Welcome

Third question:

Have you done any coding?

About this course

Course Resources

- Blackboard
  - syllabus, announcements, worksheet due dates, quiz reviews, scores on assignments

About this course

Course Resources

- Blackboard
  - syllabus, announcements, worksheet due dates, quiz reviews, scores on assignments
- Course Resource page: `mphitchman.com/stats`
  - presentation slides, in-class worksheet pdfs, RStudio tutorials, RStudio assignments

About this course

Course Resources

- Blackboard
  - syllabus, announcements, worksheet due dates, quiz reviews, scores on assignments
- Course Resource page: `mphitchman.com/stats`
  - presentation slides, in-class worksheet pdfs, RStudio tutorials, RStudio assignments
- WeBWork - online homework

About this course

Course Resources

- Blackboard
  - syllabus, announcements, worksheet due dates, quiz reviews, scores on assignments
- Course Resource page: `mphitchman.com/stats`
  - presentation slides, in-class worksheet pdfs, RStudio tutorials, RStudio assignments
- WeBWork - online homework
- My office! Taylor 212

About this course

Course Resources

- Blackboard
  - syllabus, announcements, worksheet due dates, quiz reviews, scores on assignments
- Course Resource page: `mphitchman.com/stats`
  - presentation slides, in-class worksheet pdfs, RStudio tutorials, RStudio assignments
- WeBWork - online homework
- My office! Taylor 212
- Drop-in Tutoring - evenings in Taylor Hall, 2nd floor ("math alley")

About this course

Course Resources

- Blackboard
  - syllabus, announcements, worksheet due dates, quiz reviews, scores on assignments
- Course Resource page: `mphitchman.com/stats`
  - presentation slides, in-class worksheet pdfs, RStudio tutorials, RStudio assignments
- WeBWork - online homework
- My office! Taylor 212
- Drop-in Tutoring - evenings in Taylor Hall, 2nd floor ("math alley")
- LSS

About this course

Course Resources

- Blackboard
  - syllabus, announcements, worksheet due dates, quiz reviews, scores on assignments
- Course Resource page: `mphitchman.com/stats`
  - presentation slides, in-class worksheet pdfs, RStudio tutorials, RStudio assignments
- WeBWork - online homework
- My office! Taylor 212
- Drop-in Tutoring - evenings in Taylor Hall, 2nd floor ("math alley")
- LSS
- your classmates!

About this course

Course Description

- This class is based on the principle that you learn better by doing statistics than by listening to someone talk about it all the time.

About this course

Course Description

- This class is based on the principle that you learn better by doing statistics than by listening to someone talk about it all the time.
- We will work together to learn some
  - probability

About this course

Course Description

- This class is based on the principle that you learn better by doing statistics than by listening to someone talk about it all the time.
- We will work together to learn some
  - probability
  - quantitative reasoning, and data literacy

About this course

Course Description

- This class is based on the principle that you learn better by doing statistics than by listening to someone talk about it all the time.
- We will work together to learn some
  - probability
  - quantitative reasoning, and data literacy
  - exploratory data analysis

About this course

Course Description

- This class is based on the principle that you learn better by doing statistics than by listening to someone talk about it all the time.
- We will work together to learn some
  - probability
  - quantitative reasoning, and data literacy
  - exploratory data analysis
  - statistical inference

About this course

Course Description

- This class is based on the principle that you learn better by doing statistics than by listening to someone talk about it all the time.
- We will work together to learn some
  - probability
  - quantitative reasoning, and data literacy
  - exploratory data analysis
  - statistical inference
  - coding in RStudio

Quantitative Reasoning Course Designation

Courses with a Quantitative Reasoning (QR) designation such as MATH 140 are designed to develop the student's ability to do the following:

1. Frame contextual questions using mathematical representation.
2. Apply models to deduce consequences or make predictions.
3. Communicate quantitative arguments using clear prose.
4. Critique quantitative arguments with respect to assumptions, constraints, and logical coherence.

Data Literacy

- An important skill for all of us to possess.
- Always consider context, beware simplified responses
- A few examples:
  - Highway fatalities in Oregon
  - Testing positive for a disease
  - Simpson's Paradox

Data Literacy

Oregon traffic fatalities in 2018: 414 people died in cars compared to 85 on motorcycles.

2018 Oregon Traffic Fatalities

Is it more dangerous to drive in a car than ride a motorcycle?

Testing Positive for a Disease

Suppose you have taken a test for a deadly disease. The doctor tells you that the test is quite accurate, in that, if you have the disease then the test will correctly tell you that you have the disease 100% of the time. However, if you don't have the disease, the test will very occasionally (say 1 time in 10) mistakenly tell you that you have it.

The test comes back positive (it says you have the disease)! Are you worried!? In particular, can you estimate the probability that you actually have the disease given that the test came back positive?

Testing Positive for a Disease

What is your estimate?

- a 99% probability I have the disease
- b 90% probability I have the disease
- c 50% probability I have the disease
- d 10% probability I have the disease
- e I don't know and I am mad at you for asking me!

Testing Positive for a Disease

This example illustrates the importance of asking the right question.

- I was told the probability of a positive test given that I have the disease

Testing Positive for a Disease

This example illustrates the importance of asking the right question.

- I was told the probability of a positive test given that I have the disease
- I was told the probability of a positive test given that I don't have the disease

Testing Positive for a Disease

This example illustrates the importance of asking the right question.

- I was told the probability of a positive test given that I have the disease
- I was told the probability of a positive test given that I don't have the disease
- I want to know the probability that I have the disease given that I had a positive test.

Testing Positive for a Disease

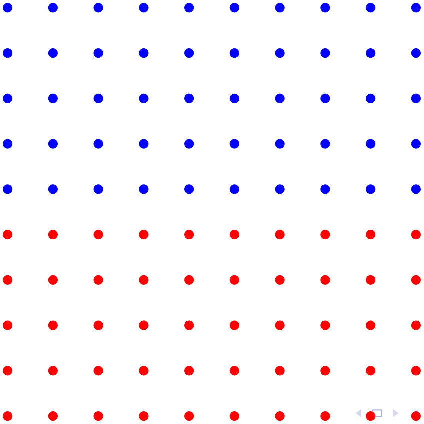
- I want to know the probability that I have the disease given that I had a positive test.

Testing Positive for a Disease

- I want to know the probability that I have the disease given that I had a positive test.
- It turns out I need more information than I've been given to answer this question, as the following examples demonstrate.

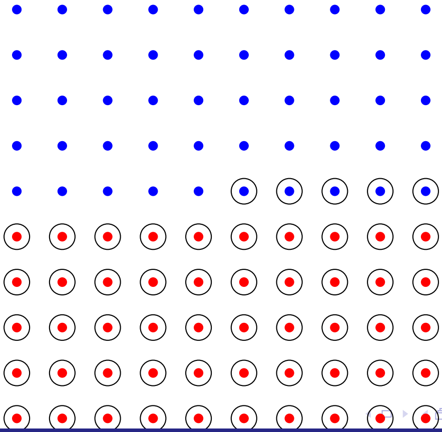
Testing Positive for a Disease

Suppose the population consists of 100 people, and 50 people, in fact, have the disease (blue - healthy, red - sick). About how many people test positive?



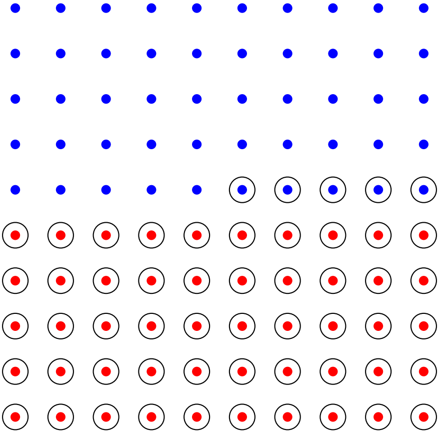
Testing Positive for a Disease

Suppose the population consists of 100 people, and 50 people, in fact, have the disease (blue - healthy, red - sick). About how many people test positive? (circled)



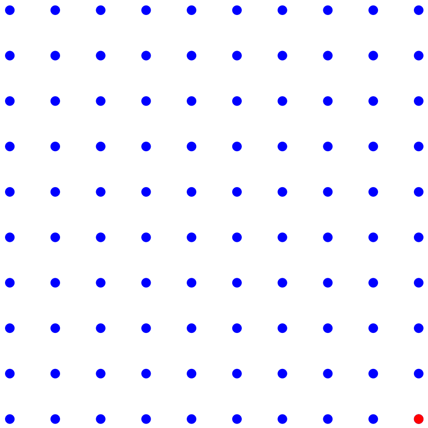
Testing Positive for a Disease

Given you have a positive test, what is the probability you have the disease?



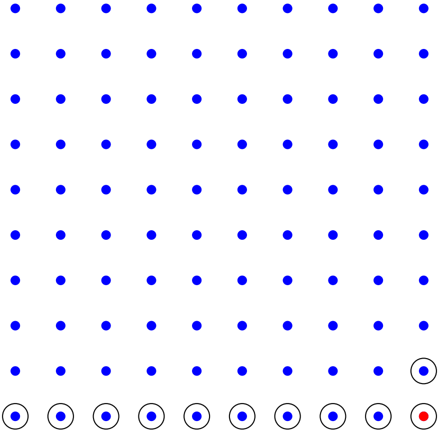
Testing Positive for a Disease

Now suppose only 1 person has the disease. Given you have a positive test, what is the probability you have the disease?



Testing Positive for a Disease

Now suppose only 1 person has the disease. Given you have a positive test, what is the probability you have the disease?



Base Rate Fallacy

- The base rate fallacy is a cognitive bias where people ignore statistically relevant general information (the base rate) in favor of specific, more vivid details, leading to incorrect judgments about probability.

Base Rate Fallacy

- The base rate fallacy is a cognitive bias where people ignore statistically relevant general information (the base rate) in favor of specific, more vivid details, leading to incorrect judgments about probability.
- In our example, the specific detail: tested positive for the disease!
- the relevant underlying info: the proportion of people that have the disease in general

Base Rate Fallacy

- The base rate fallacy is a cognitive bias where people ignore statistically relevant general information (the base rate) in favor of specific, more vivid details, leading to incorrect judgments about probability.
- In our example, the specific detail: tested positive for the disease!
- the relevant underlying info: the proportion of people that have the disease in general
- Our minds jump to the conclusion that we must have the disease, though it is far from likely if it is an extremely rare disease, in this example.

Base Rate Fallacy

Two other examples:

- a person might fear flying more than driving, even though statistically speaking, it is more dangerous to drive.

Base Rate Fallacy

Two other examples:

- a person might fear flying more than driving, even though statistically speaking, it is more dangerous to drive.
- During a disease outbreak, one might find a higher proportion of hospital cases occurring among vaccinated people than unvaccinated people, which might be misinterpreted as a sign that the vaccine is not effective. But it's likely going to be true if the vast majority of people in general are vaccinated against the disease.

Simpson's Paradox

Berkeley Graduate Admissions

In the fall of 1973, the University of California, Berkeley's graduate division admitted about 44% of male applicants and 35% of female applicants. That raised eyebrows among school officials, who feared bias and a possible law suit.

Simpson's Paradox

Here we focus on 4 programs, to illustrate Simpson's Paradox, and the importance of looking at the raw data.

Simpson's Paradox

Here we focus on 4 programs, to illustrate Simpson's Paradox, and the importance of looking at the raw data.

Program	M acc	M den	W acc	W den
A	511	314	89	19
B	352	208	17	8
D	137	270	132	243
F	22	351	24	317
Total	1022	1143	262	587

Simpson's Paradox

Berkeley Graduate Admissions

Program	M apps	M accept rate	W apps	W accept rate
A	825	.62	108	.82
B	560	.63	25	.68
D	407	.34	375	.35
F	373	.06	341	.07
Total	2165	.47	1486	.31

In each of these four programs women accepted at a higher rate, but in the aggregate men were! This is an example of **Simpson's Paradox**
How is this possible?

Dig Deeper!

In this course you are encouraged to dig deeper into things. Go beyond superficial observations. In fact, this effort is expected of you in your term project!

Course Resource Page

mphitchman.com/stats

WeBWorK

- https://athena.cs.linfield.edu/webwork2/MATH140-Hitchman/

RStudio

Two options

- RStudio cloud: `https://posit.cloud/`
- Installing R and RStudio locally: `https://mphitchman.com/stats/rstudio.html`