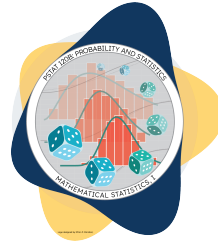


Topic 00: Introduction to the Course

Ethan P. Marzban University of California, Santa Barbara PSTAT 120B

Outline

1. Why Statistics?

2. Alright, enough waffling - Let’s Get Started!

Welcome!

- Welcome to PSTAT 120B!

UC SANTA BARBARA
Department of Statistics
and Applied Probability

Welcome!

- Welcome to PSTAT 120B!
- My name is Ethan Marzban (please feel free to call me "Ethan"), and I use he/him/his pronouns.

Welcome!

- Welcome to PSTAT 120B!
- My name is Ethan Marzban (please feel free to call me "Ethan"), and I use he/him/his pronouns.
  - I just finished by fourth year in the PhD program here in the Department of Statistics and Applied Probability (PSTAT).

Welcome!

- Welcome to PSTAT 120B!
- My name is Ethan Marzban (please feel free to call me "Ethan"), and I use he/him/his pronouns.
  - I just finished by fourth year in the PhD program here in the Department of Statistics and Applied Probability (PSTAT).
  - My research broadly encompasses nonparametric statistics, Gaussian Processes, and some Fourier analysis. I'm also super passionate about Statistics Education!

Welcome!

- Welcome to PSTAT 120B!
- My name is Ethan Marzban (please feel free to call me "Ethan"), and I use he/him/his pronouns.
  - I just finished by fourth year in the PhD program here in the Department of Statistics and Applied Probability (PSTAT).
  - My research broadly encompasses nonparametric statistics, Gaussian Processes, and some Fourier analysis. I'm also super passionate about Statistics Education!
  - In my free time, I like playing the piano, drinking boba, and talking about cats (feel free to ask me about Kitty!)

Welcome!

- Welcome to PSTAT 120B!
- My name is Ethan Marzban (please feel free to call me "Ethan"), and I use he/him/his pronouns.
  - I just finished by fourth year in the PhD program here in the Department of Statistics and Applied Probability (PSTAT).
  - My research broadly encompasses nonparametric statistics, Gaussian Processes, and some Fourier analysis. I'm also super passionate about Statistics Education!
  - In my free time, I like playing the piano, drinking boba, and talking about cats (feel free to ask me about Kitty!)
- **OH:** Wednesdays 3 - 4:30pm in GIRV 2123
  TBD via Zoom

It Takes a Village...

- **Teaching Assistants (TAs):** Hyuk-Jean Choe and Minwoo Park

It Takes a Village...

- **Teaching Assistants (TAs):** Hyuk-Jean Choe and Minwoo Park
  - Hyuk-Jean's Sections: 2pm and 3pm

It Takes a Village...

- **Teaching Assistants (TAs):** Hyuk-Jean Choe and Minwoo Park
  - Hyuk-Jean's Sections: 2pm and 3pm
  - Hyuk-Jean's OH: T 9am - 12pm in Building 434, Room 113

It Takes a Village...

- **Teaching Assistants (TAs):** Hyuk-Jean Choe and Minwoo Park
  - Hyuk-Jean's Sections: 2pm and 3pm
  - Hyuk-Jean's OH: T 9am - 12pm in Building 434, Room 113
  - Minwoo's Sections 4pm and 5pm

It Takes a Village...

- **Teaching Assistants (TAs):** Hyuk-Jean Choe and Minwoo Park
  - Hyuk-Jean's Sections: 2pm and 3pm
  - Hyuk-Jean's OH: T 9am - 12pm in Building 434, Room 113

  - Minwoo's Sections 4pm and 5pm
  - Minwoo's OH: M 8am - 9am (SH 5432S); T, R 6pm - 7pm (SH 5432S)

Some General Information

- **Main Course Website:** https://pstat120b.github.io

UC SANTA BARBARA
Department of Statistics
and Applied Probability

Some General Information

- **Main Course Website:** https://pstat120b.github.io
  - Pretty much everything will be posted here, and you can find information about pretty much everything course-related here.

Some General Information

- **Main Course Website:** https://pstat120b.github.io
  - Pretty much everything will be posted here, and you can find information about pretty much everything course-related here.
  - The syllabus is up as well; please read through it fully!

Some General Information

- **Main Course Website:** https://pstat120b.github.io
  - Pretty much everything will be posted here, and you can find information about pretty much everything course-related here.
  - The syllabus is up as well; please read through it fully!

- Sections take place twice a week, on Tuesdays and Thursdays.

Some General Information

- **Main Course Website:** https://pstat120b.github.io
  - Pretty much everything will be posted here, and you can find information about pretty much everything course-related here.
  - The syllabus is up as well; please read through it fully!

- Sections take place twice a week, on Tuesdays and Thursdays.
  - Section attendance is not tracked, but we will have a quiz on Thursdays (in non-exam weeks), and it's also generally a good idea to attend Sections.

Some General Information

- **Main Course Website:** https://pstat120b.github.io
  - Pretty much everything will be posted here, and you can find information about pretty much everything course-related here.
  - The syllabus is up as well; please read through it fully!

- Sections take place twice a week, on Tuesdays and Thursdays.
  - Section attendance is not tracked, but we will have a quiz on Thursdays (in non-exam weeks), and it's also generally a good idea to attend Sections.
  - Due to capacity constraints on the rooms, we cannot allow overenrollment of sections. To unofficially switch Sections, please see the instructions on the Course Website.

Some General Information

- Three exams in total: two midterms and a final.

UC SANTA BARBARA
Department of Statistics
and Applied Probability

Some General Information

- Three exams in total: two midterms and a final.
- Quizzes take place in non-exam weeks, during Sections, on Thursdays.

Some General Information

- Three exams in total: two midterms and a final.
- Quizzes take place in non-exam weeks, during Sections, on Thursdays.
  - Will take place during (roughly) the first 20 minutes of Section, so please try not to be late!

Some General Information

- Three exams in total: two midterms and a final.
- Quizzes take place in non-exam weeks, during Sections, on Thursdays.
  - Will take place during (roughly) the first 20 minutes of Section, so please try not to be late!
  - Unfortunately, make-up quizzes and exams will not be administered for any reasons. (If you end up needing to miss a quiz for an emergency reason, please reach out ASAP to discuss)

Some General Information

- Three exams in total: two midterms and a final.
- Quizzes take place in non-exam weeks, during Sections, on Thursdays.
  - Will take place during (roughly) the first 20 minutes of Section, so please try not to be late!
  - Unfortunately, make-up quizzes and exams will not be administered for any reasons. (If you end up needing to miss a quiz for an emergency reason, please reach out ASAP to discuss)
  - If you need to miss an exam for a university-sanctioned sporting event or a sickness, please reach out to Ethan ASAP to discuss.

Some General Information

Some General Information

- Homework is due every week (Wednesdays in non-exam weeks, and Tuesdays in exam weeks).

Some General Information

- Homework is due every week (Wednesdays in non-exam weeks, and Tuesdays in exam weeks).
  - Because we will be releasing solutions to the homework shortly after the due dates, we unfortunately cannot accommodate late homework submissions. Instead, I'll drop your lowest two (2) homework assignments at the end of the quarter.

Some General Information

- Homework is due every week (Wednesdays in non-exam weeks, and Tuesdays in exam weeks).
  - Because we will be releasing solutions to the homework shortly after the due dates, we unfortunately cannot accommodate late homework submissions. Instead, I'll drop your lowest two (2) homework assignments at the end of the quarter.

- Please see the syllabus for more information about general course policies, as well as information on how your final grades for this course will be calculated.

A Quick Note on Emails

- I ask that you please refrain from emailing me except in true emergency situations (I'll leave it to you to decide what is an emergency).

A Quick Note on Emails

- I ask that you please refrain from emailing me except in true emergency situations (I'll leave it to you to decide what is an emergency).
  - I do this for both of our sakes! I understand how stressful it can be to send an email and not be sure if the instructor received it, etc.

A Quick Note on Emails

- I ask that you please refrain from emailing me except in true emergency situations (I'll leave it to you to decide what is an emergency).
  - I do this for both of our sakes! I understand how stressful it can be to send an email and not be sure if the instructor received it, etc.
  - Instead of email, I ask that you please make use of our Discord Server (link on the next slide) for general course inquiries! (Private matters can be discussed during Office Hours.)

A Quick Note on Emails

- I ask that you please refrain from emailing me except in true emergency situations (I'll leave it to you to decide what is an emergency).
  - I do this for both of our sakes! I understand how stressful it can be to send an email and not be sure if the instructor received it, etc.
  - Instead of email, I ask that you please make use of our Discord Server (link on the next slide) for general course inquiries! (Private matters can be discussed during Office Hours.)
  - Thank you so much for your understanding!

Why Statistics?

UC SANTA BARBARA
Department of Statistics
and Applied Probability

Rent Burden

More Renters Than Ever Before Are Burdened by the Rent They Pay

A new Harvard report says 22.4 million households in the United States now spend more than 30 percent of their income in rent, with 12.1 million spending more than 50 percent.

https://www.nytimes.com/2024/01/25/realestate/rent-prices-housing.html

UC SANTA BARBARA
Department of Statistics and Applied Probability

Rent Burden

More Renters Than Ever Before Are Burdened by the Rent They Pay

A new Harvard report says 22.4 million households in the United States now spend more than 30 percent of their income in rent, with 12.1 million spending more than 50 percent.

https://www.nytimes.com/2024/01/25/realestate/rent-prices-housing.html

- Where did this 22.4 million come from?

UC SANTA BARBARA
Department of Statistics and Applied Probability

Rent Burden

More Renters Than Ever Before Are Burdened by the Rent They Pay

A new Harvard report says 22.4 million households in the United States now spend more than 30 percent of their income in rent, with 12.1 million spending more than 50 percent.

https://www.nytimes.com/2024/01/25/realestate/rent-prices-housing.html

- Where did this 22.4 million come from?
- Do we trust this statistic? Would small changes to the overall experiment result in drastically different results?

Rent Burden

More Renters Than Ever Before Are Burdened by the Rent They Pay

A new Harvard report says 22.4 million households in the United States now spend more than 30 percent of their income in rent, with 12.1 million spending more than 50 percent.

https://www.nytimes.com/2024/01/25/realestate/rent-prices-housing.html

- Where did this 22.4 million come from?
- Do we trust this statistic? Would small changes to the overall experiment result in drastically different results?
- If we collected data, how likely would it be for our own findings to match the New York Times's?

World Record: Holding Breath

https://www.guinnessworldrecords.com/news/2021/5/freediver-holds-breath-for-almost-25-minutes-breaking-record-660285

UC SANTA BARBARA
Department of Statistics and Applied Probability

World Record: Holding Breath

56-year-old freediver holds breath for almost 25 minutes breaking record

By Connie Suggitt
Published 12 May 2021

https://www.guinnessworldrecords.com/news/2021/5/freediver-holds-breath-for-almost-25-minutes-breaking-record-660285

- What is the true longest time a person can hold their breath?

UC SANTA BARBARA
Department of Statistics and Applied Probability

World Record: Holding Breath

56-year-old freediver holds breath for almost 25 minutes breaking record

By Connie Suggitt
Published 12 May 2021

https://www.guinnessworldrecords.com/news/2021/5/freediver-holds-breath-for-almost-25-minutes-breaking-record-660285

- What is the true longest time a person can hold their breath?
- If we collected data on the times various people held their breaths, how could we construct a "good" estimate of the true longest time a person can hold their breath?

Data Science

- The world is full of claims.

UC SANTA BARBARA
Department of Statistics
and Applied Probability

Data Science

- The world is full of claims.
- The world is also full of data!

UC SANTA BARBARA
Department of Statistics
and Applied Probability

Data Science

- The world is full of claims.
- The world is also full of data!
  - Sometimes this data is intended to assess the validity of claims, other times it exists solely for investigative and descriptive purposes.

Data Science

- The world is full of claims.
- The world is also full of data!
  - Sometimes this data is intended to assess the validity of claims, other times it exists solely for investigative and descriptive purposes.
- Indeed, Data Science has emerged as a field to help process, analyze, and understand data.

Data Science

- The world is full of claims.
- The world is also full of data!
  - Sometimes this data is intended to assess the validity of claims, other times it exists solely for investigative and descriptive purposes.
- Indeed, Data Science has emerged as a field to help process, analyze, and understand data.
- But, Data Science didn't come out of nothing - it is predicated on a strong, formal way of handling uncertainty.

Statistics

- That, in my opinion, is one of the major goals of Statistics as a whole: to provide a rigorous and mathematical framework for discussing uncertainty and variability.

Statistics

- That, in my opinion, is one of the major goals of Statistics as a whole: to provide a rigorous and mathematical framework for discussing uncertainty and variability.
  - I'd like to stress: PSTAT 120B is a bit more of a "theoretical" course, rather than a "data-driven" one.

Statistics

- That, in my opinion, is one of the major goals of Statistics as a whole: to provide a rigorous and mathematical framework for discussing uncertainty and variability.
  - I'd like to stress: PSTAT 120B is a bit more of a "theoretical" course, rather than a "data-driven" one.
  - We'll be doing a fair amount of math, and we'll be developing some pretty formal and careful mathematically-styled arguments.

Statistics

- That, in my opinion, is one of the major goals of Statistics as a whole: to provide a rigorous and mathematical framework for discussing uncertainty and variability.
  - I'd like to stress: PSTAT 120B is a bit more of a "theoretical" course, rather than a "data-driven" one.
  - We'll be doing a fair amount of math, and we'll be developing some pretty formal and careful mathematically-styled arguments.
  - But, I hope you don't loose sight of the bigger picture!

Probability

- Wait- “discussing uncertainty and variability.” Isn’t that just Probability?

Probability

- Wait- “discussing uncertainty and variability.” Isn’t that just Probability?
- Well, probability is definitely a *part* of things.

Probability

- Wait- “discussing uncertainty and variability.” Isn’t that just Probability?
- Well, probability is definitely a *part* of things.
  - That’s why a strong familiarity with material covered in classes like PSTAT 120A is an absolutely crucial prerequisite for this course.

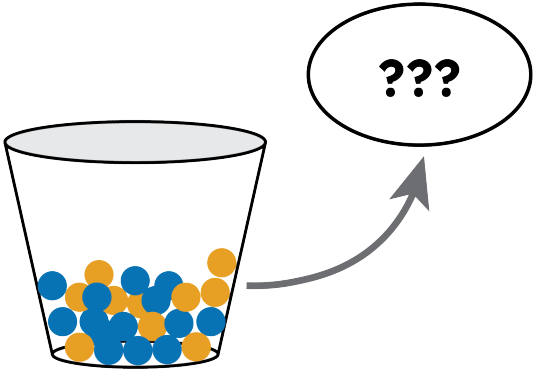
Probability

- Wait- “discussing uncertainty and variability.” Isn’t that just Probability?
- Well, probability is definitely a *part* of things.
  - That’s why a strong familiarity with material covered in classes like PSTAT 120A is an absolutely crucial prerequisite for this course.
- But, probability and statistics exist as two separate (but overlapping) fields.

Probability

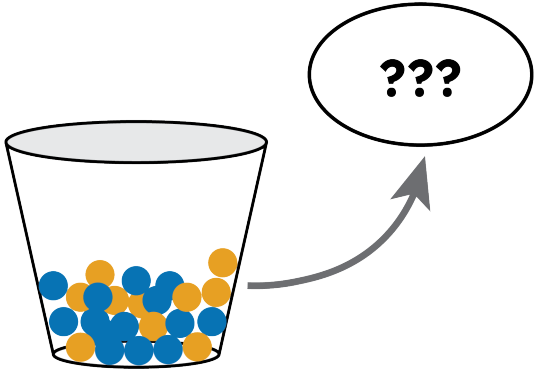
- Wait- “discussing uncertainty and variability.” Isn’t that just Probability?
- Well, probability is definitely a *part* of things.
  - That’s why a strong familiarity with material covered in classes like PSTAT 120A is an absolutely crucial prerequisite for this course.
- But, probability and statistics exist as two separate (but overlapping) fields.
- To highlight the distinction between probability and statistics, allow me to use a metaphor that my own statsitics professor used when I was taking my first mathematical statistics course.

Scenario 1



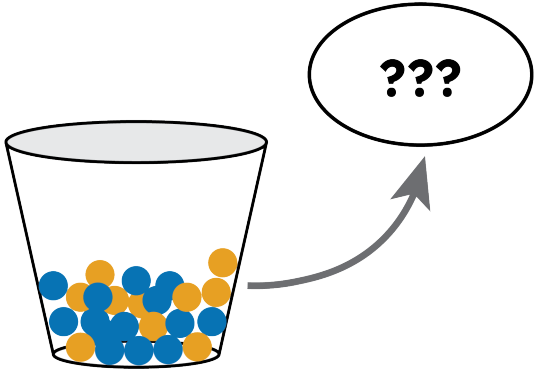
UC SANTA BARBARA
Department of Statistics
and Applied Probability

Scenario 1



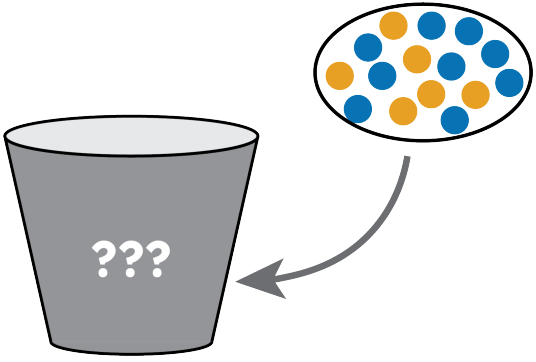
- We know that a bucket contains some (known) number of blue and gold marbles. From this bucket we take a sample.

Scenario 1



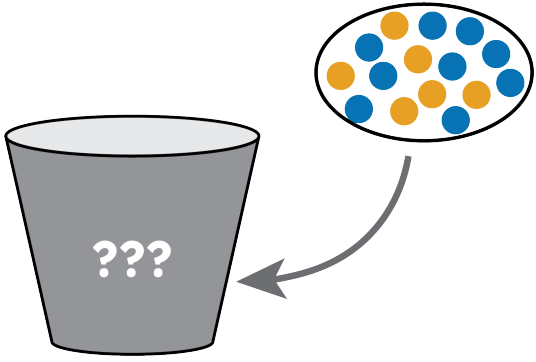
- We know that a bucket contains some (known) number of blue and gold marbles. From this bucket we take a sample.
- Given our knowledge of what's in the bucket, we want to inform what's in our hand (e.g. number of gold marbles, probability of having more than 3 blue marbles, etc.)

Scenario 2



UC SANTA BARBARA
Department of Statistics
and Applied Probability

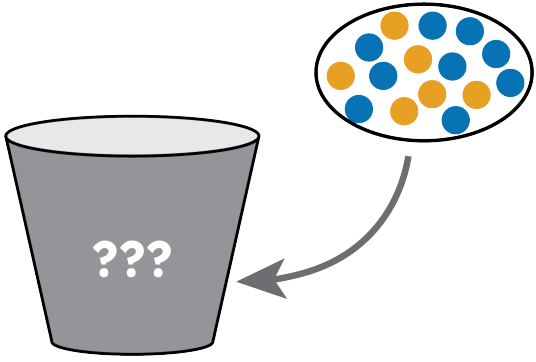
Scenario 2



- We have a sample of blue and gold marbles (and we know how many of each are in our sample), that we know came from a bucket.

UC SANTA BARBARA
Department of Statistics and Applied Probability

Scenario 2



- We have a sample of blue and gold marbles (and we know how many of each are in our sample), that we know came from a bucket.
- Given our knowledge of what's in our hand, we want to inform what's in the bucket (e.g. number of gold marbles, probability of having more than 3 blue marbles, etc.)

Probability vs. Statistics

- Which Scenario, Scenario 1 or Scenario 2, is probability and which is statistics?

Probability vs. Statistics

- Which Scenario, Scenario 1 or Scenario 2, is probability and which is statistics?
- Scenario 1 (in which we use what's in the bucket to inform what's in our hand) is probability.

Probability vs. Statistics

- Which Scenario, Scenario 1 or Scenario 2, is probability and which is statistics?
- Scenario 1 (in which we use what's in the bucket to inform what's in our hand) is probability.
  - We've been there, done that, in PSTAT 120A!

Probability vs. Statistics

- Which Scenario, Scenario 1 or Scenario 2, is probability and which is statistics?
- Scenario 1 (in which we use what's in the bucket to inform what's in our hand) is probability.
  - We've been there, done that, in PSTAT 120A!
- Scenario 2 is statistics (specifically, inferential statistics).

Probability vs. Statistics

- Which Scenario, Scenario 1 or Scenario 2, is probability and which is statistics?
- Scenario 1 (in which we use what's in the bucket to inform what's in our hand) is probability.
  - We've been there, done that, in PSTAT 120A!
- Scenario 2 is statistics (specifically, inferential statistics).
  - Typically, the goal of inferential statistics is to use a sample to make *inferences* (we'll talk more about these later in the course) about a larger population.

PSTAT 120B

- So how does PSTAT 120B (as a course) factor into things?

UC SANTA BARBARA
Department of Statistics
and Applied Probability

PSTAT 120B

- So how does PSTAT 120B (as a course) factor into things?
- We'll start off by finishing up a few probability-related concepts that you might not have seen in PSTAT 120A (conditional distributions and expectations, and transformations).

PSTAT 120B

- So how does PSTAT 120B (as a course) factor into things?
- We'll start off by finishing up a few probability-related concepts that you might not have seen in PSTAT 120A (conditional distributions and expectations, and transformations).
- Then, we'll move into the realm of inferential statistics (specifically, estimation, confidence intervals, and hypothesis testing).

PSTAT 120B

- So how does PSTAT 120B (as a course) factor into things?
- We'll start off by finishing up a few probability-related concepts that you might not have seen in PSTAT 120A (conditional distributions and expectations, and transformations).
- Then, we'll move into the realm of inferential statistics (specifically, estimation, confidence intervals, and hypothesis testing).
- Finally, we'll finish off by comparing across samples and seeing what that tells us about potential differences across populations.

Alright, enough waffling - Let's Get Started!

UC SANTA BARBARA
Department of Statistics
and Applied Probability