Prob & Stats Review

STSCI/INFO/ILRST 3900: Causal Inference

August 30, 2023

Reminders and Announcements

- HW 1 due tomorrow (August 31) by 5pm
  - Submit a PDF from RMarkdown via Canvas
- Office Hours throughout the week (see Syllabus or website)
  - Mayleen: Fridays 9-10am in Rhodes 657 (Room 2) or Zoom
  - Daniel: Thursdays 1-2pm in Uris 302
  - See Ed Discussion for Zoom links/info

Agenda for Today

- Reminders and Announcements
- Quick Icebreaker
- Probability and Statistics Review
- Homework Check-in and Questions

Icebreaker

Rock-Paper-Scissors Stats Review

- Introduce yourself to the person next to you and play rock-paper-scissors with them, best 2 out of 3
- The person who wins explains to the other person picks one of the topics listed below and explains what they understand about it.
  - Expectation, Variance, Conditional Expectation, Independence, Bernoulli
- The person who lost needs to come up with one follow-up question, and both of you can work together to determine the answer.

Probability and Statistics Review

- Expectation
- Variance
- Conditional Expectation
- Independence
- Bernoulli random variables

Expectation

(Expected Value, Population Mean, Average)

- Notation: $E(X), \mu$
- The **expected value** of a *finite* random variable

$$\mu = E(X) := \sum_{i=1}^{N} x_i \cdot P(x_i) \text{ where } P(x_i) := \text{Prob}(X = x_i)$$

- Can also think of it as a population average; $X = \{x_1, \ldots, x_n\}$

$$E(X) = \frac{1}{N} \sum_{i=1}^{N} x_i$$

Expectation

(Expected Value, Population Mean, Average)

- The **expected value** of a *countable* random variable, i.e. the (long run) average

$$E(X) = \sum_{i=1}^{\infty} x_i \cdot P(x_i)$$

- For n independent and identically distributed (**i.i.d.**) random variables $X_1, \cdots, X_N$

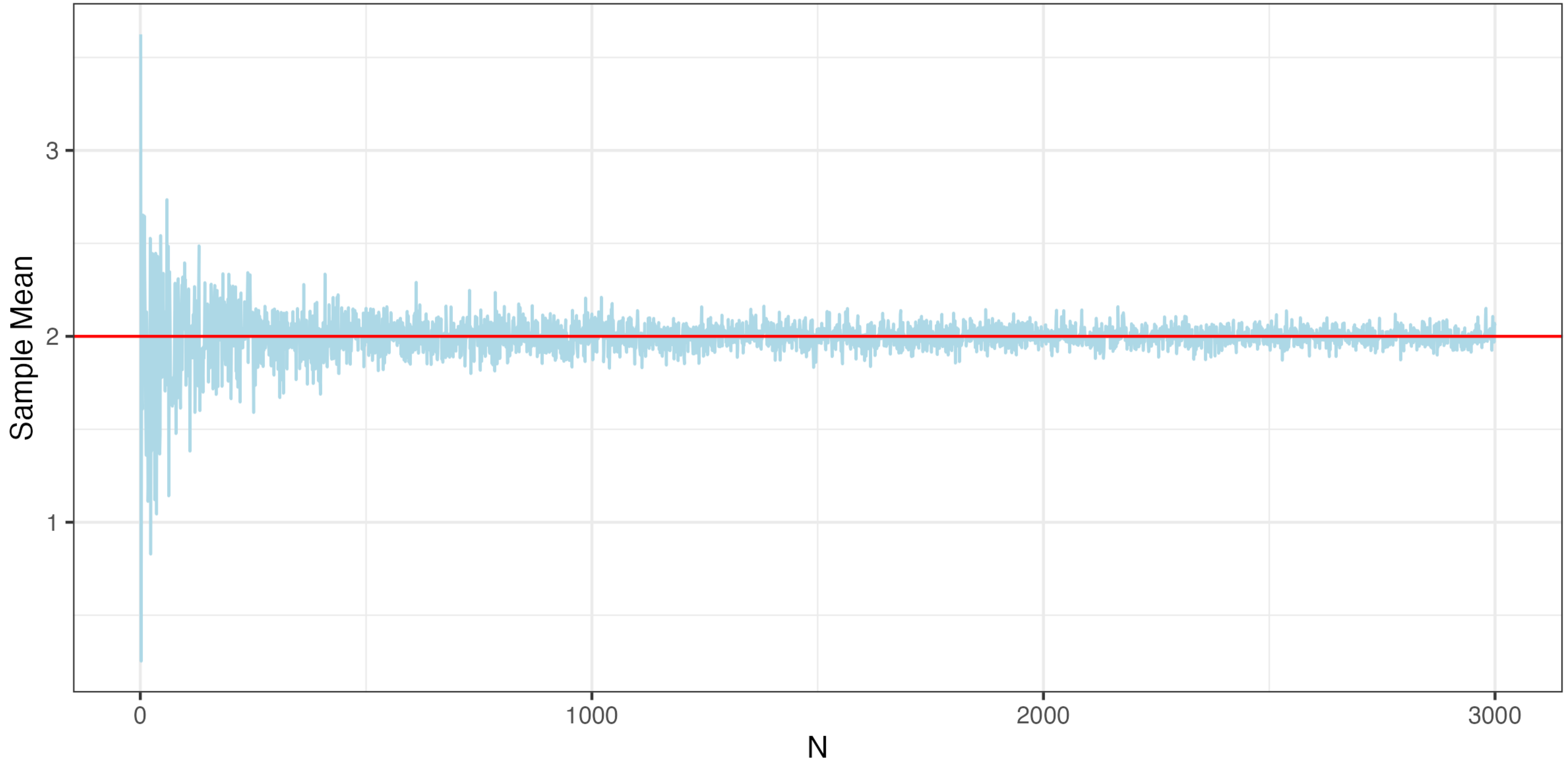
the **sample mean** is $\bar{X} = \frac{1}{N} \sum_{i=1}^{N} X_i$

- **Law of Large Numbers (LLN)**: the sample mean converges to the expected value (population mean) as $N \to \infty$

- Example: R (compute the sample mean for larger and larger N)

Expectation

- X_i are random draws from $\sim \mathcal{N}(2,5)$ (a Normal r.v. with mean 2, variance 5)
- How quickly does the sample mean converge to the population mean?



Variance

Describes the spread of the data

- Notation: $V(X), Var(X), \sigma^2$
- Variance is the average of the squared differences from the mean
- For a random variable X with expected value $\mu := E(X)$, the variance is

$$\sigma^2 = Var(X) := E\left[\left(X - \mu\right)^2\right] = E\left[X^2\right] - \mu^2$$

More explicitly, $Var(X) = \sum_{i=1}^{n} P(x_i) \cdot (x_i - \mu)^2$ where $P(x_i) :=$ Prob$(X = x_i)$

Sample (Empirical) Variance

For a finite dataset or finite sample

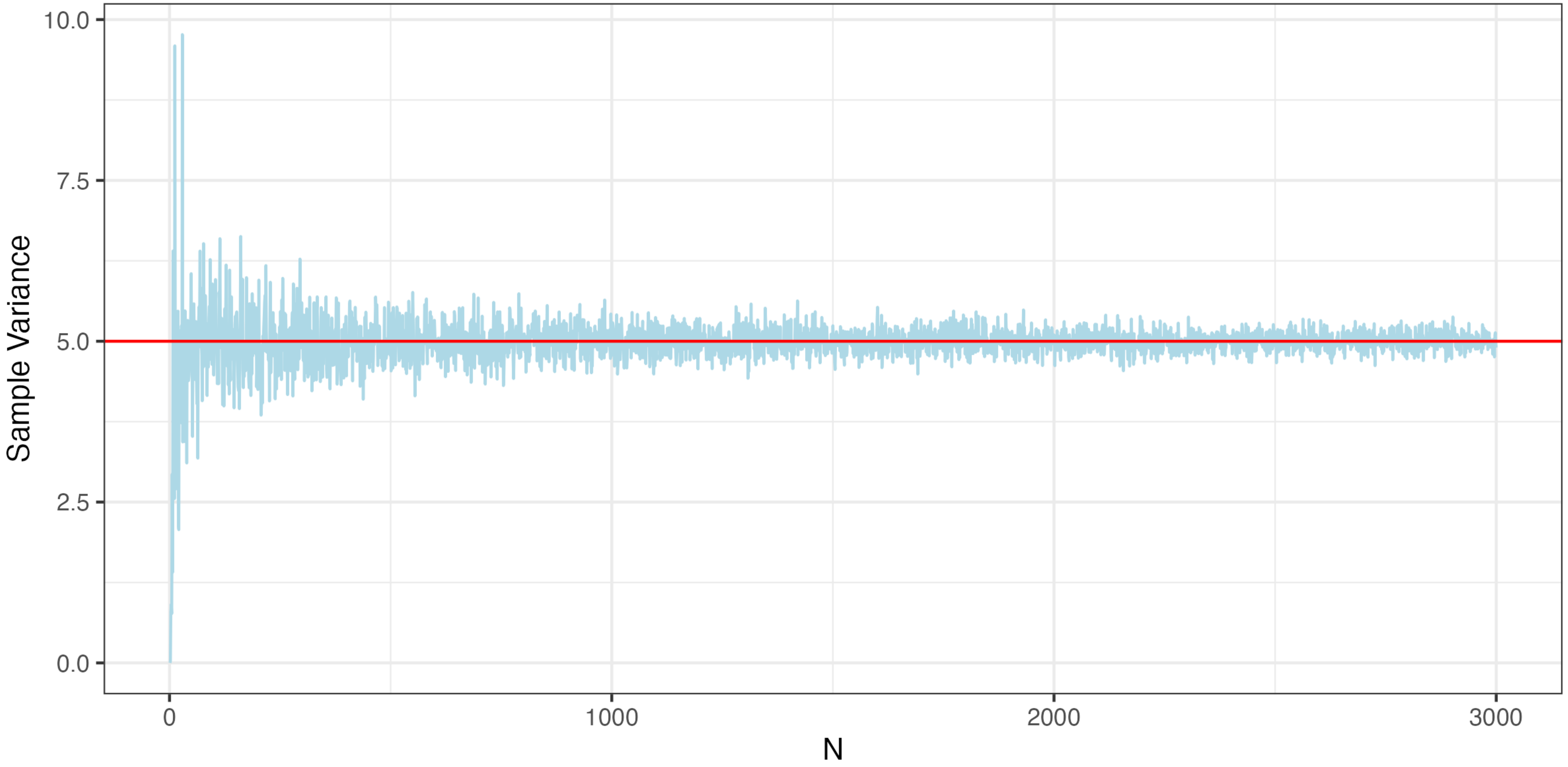
- In practice, you can compute the variance of a finite dataset as

$$\sigma^2 = \left(\frac{1}{N}\sum_{i=1}^{N} x_i^2\right) - \bar{X}^2 \text{ where } \bar{X} := \frac{1}{N}\sum_{i=1}^{N} x_i$$

- You don't need to have the formula memorized, just be aware of it
- Likely you'll never have to explicitly compute it this way, just use an R function

Sample Variance

- X_i are random draws from $\sim \mathcal{N}(2, 5)$ (a Normal r.v. with mean 2, variance 5)
- How quickly does the sample variance converge to the population variance?



Conditional Expectation

- Notation: $E(X \mid Y)$
- The expected value given a set of "conditions"
- Read as "the expectation of X given (or conditioned on) Y"

$$E(X \mid Y) = \sum_{i=1}^{n} x_i \cdot P(X = x_i \mid Y)$$

where $P(X = x_i \mid Y) = \dfrac{P(X = x_i \text{ and } Y)}{P(Y)}$

Conditional Expectation

Example: Roll a fair die

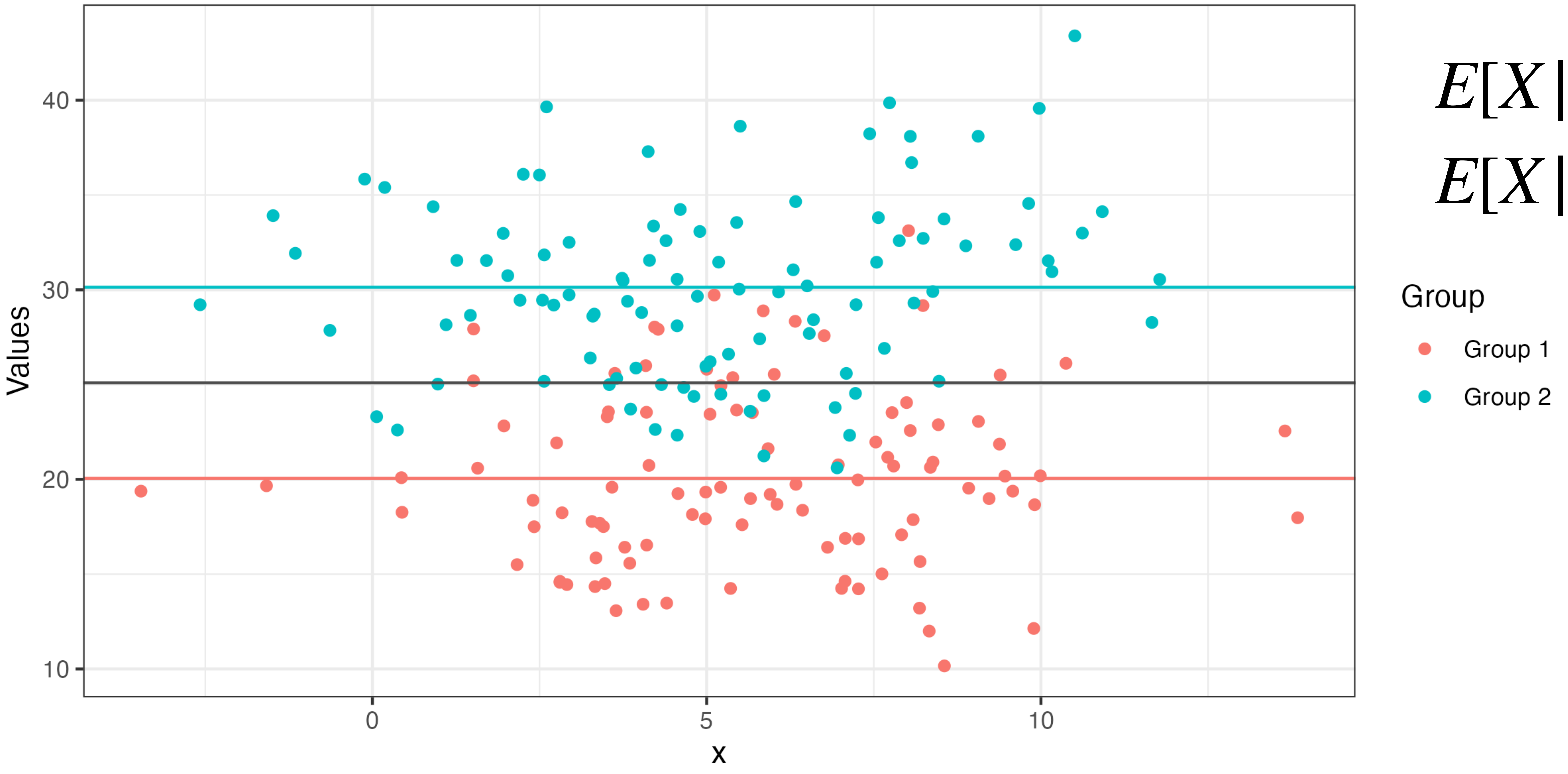
- Let $A = 1$ if you roll an even number, 0 otherwise.
- Let $B = 1$ if you roll a prime number, 0 otherwise. Then,

$$E[A] = \sum_{i=1}^{6} a_i \cdot P(a_i) = \frac{0+1+0+1+0+1}{6} = \frac{1}{2}$$

and the conditional expectation of A given $B = 1$ (i.e. we rolled 2, 3, or 5)

$$E[A \,|\, B=1] = \sum_{i=1}^{3} a_i \cdot P(a_i \,|\, B=1) = \frac{1+0+0}{3} = \frac{1}{3}$$

Conditional Expectation - Visualized



$$E[X] = 25$$

$$E[X|\text{group 1}] = 20$$

$$E[X|\text{group 2}] = 30$$

Independence

- Notation: $\perp$, $X \perp Y$
- Two random variables are **independent** if the outcome of one does not give any information about the outcome of the other
- Events A and B are independent if $P(A \cap B) = P(A)P(B)$
- Recall: $P(A \cap B) = P(A \mid B)P(B)$
- If $A \perp B$, then $P(A \mid B) = P(A)$ and $P(B \mid A) = P(B)$

Independence

Example: Dice

- Suppose you roll two fair dice. Let A be the value of the first die and let B be the value of the second die.

- If I say that $A = 3$, does that give you any info about what the value of B is?

- We can show that the **events** $\{A = 3\}$ and $\{B = 3\}$ are independent:

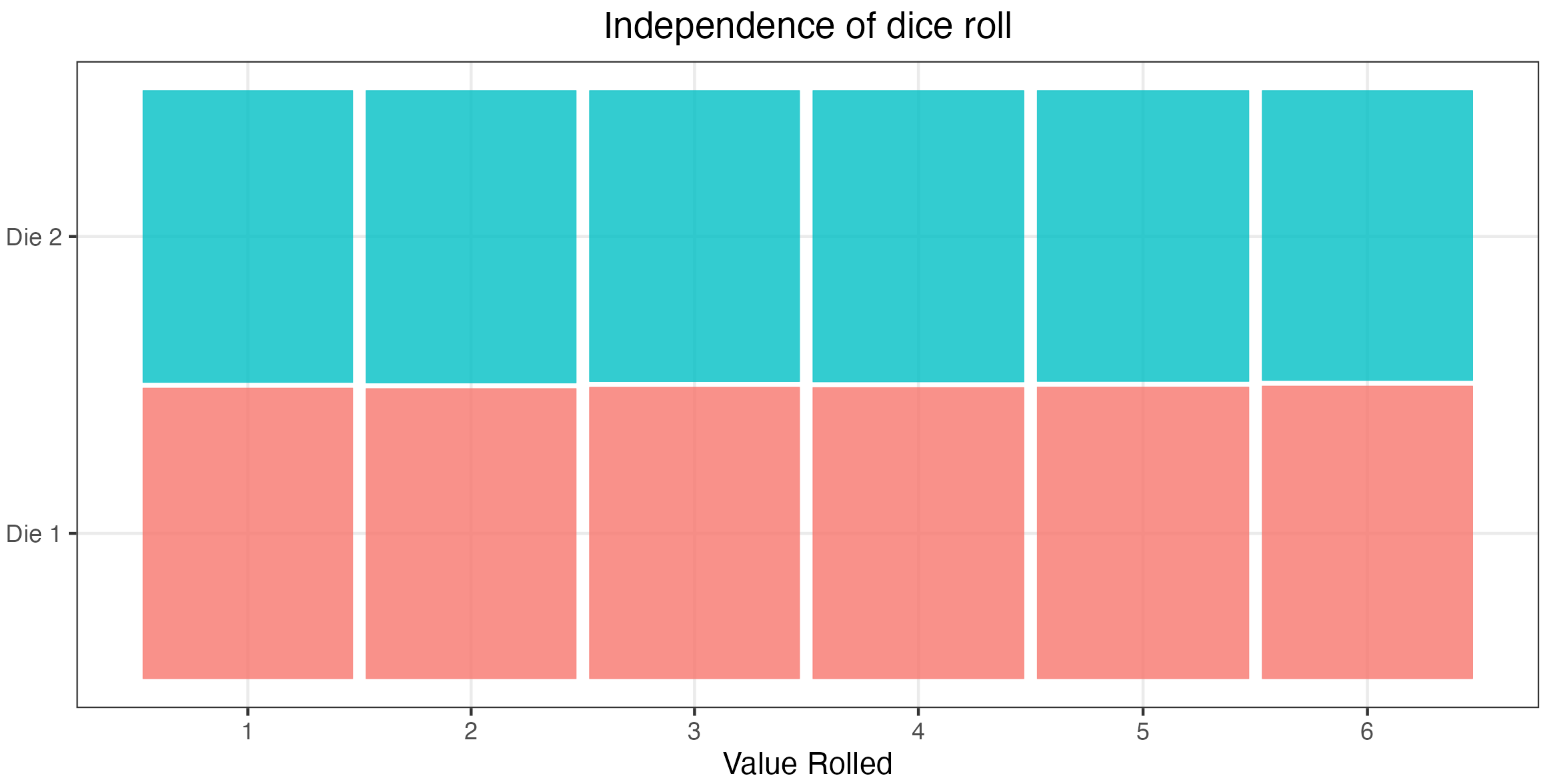
$$\begin{aligned} P(\{A = 3\} \cap \{B = 3\}) &= P(\{A = 3\} \,|\, \{B = 3\}) \cdot P(\{B = 3\}) \\ &= \frac{1}{6} \cdot \frac{1}{6} \\ &= P(\{A = 3\}) \cdot P(\{B = 3\}) \end{aligned}$$

- To show $A \perp B$, you would show this holds for all values of A and B

Independence

Example: Dice

- If we simulate 100k dice rolls, we see that the joint probability of each combination is equal to the individual probabilities multiplied.



Bernoulli Random Variables

A binary/dichotomous random variable

- Notation: $B(p)$, Bernoulli(p), $\mathscr{B}(p)$
- Takes the value 1 with probability (w.p.) p, and the value 0 w.p. $q := 1 - p$
- Let $X \sim B(p)$
  - "Let X be a Bernoulli random variable with mean p"
  - $E(X) = p$ and $Var(X) = p(1 - p) = pq$
- Cool fact: $E(X) = P(X = 1) = p$

Law of Total Expectation
(i.e. law of iterated expectations, tower rule)

- Useful property (or “trick) that will be used in class

$$E(X) = E\Big(E(X \mid Y)\Big)$$

- Don’t worry too much about the technical details, just add to your toolbox :)