PSYC 640
Grad Stats

Probability & Inference

FALL 2023

Reminders to Professor

Focus of today will be probability lecture

Have remaining time to review Lab 1

Reminder about final project – next class will have time for group work

Descriptives vs. Inferences

Moving from simply describing our data

With *means* and *standard deviations*

To drawing conclusions about the world

Using inferential statistics

“When you make assumptions you make an…”

It is common to view assumptions in a negative way

- Previously spoke about biases that can exist

"When you make assumptions you make an…"

It is common to view assumptions in a negative way

- Previously spoke about biases that can exist

Statistical inference *relies* on assumptions

- Nearly impossible to use logical reasoning on it's own

Limits of Logical Reasoning

Let’s test out the limits to logical reasoning using a coin flip game!

We will try to predict the coin flip prior and have us determine what the logical side would say vs. including assumptions

Assumptions

Reasonable to think that the unbeatable record is good evidence to predict the next win

- No easy logical justification
- Cannot truly justify betting on the win without an assumption

Relied on an implicit assumption that there was a skill/knowledge difference

- Able to then learn from the assumption that was made
- If a less sensible assumption was used to drive your understanding, it may have been less accurate

Logic will reject that assumption entirely

- All outcomes remain to be equally as plausible
- Trying to rule out what is impossible first did not allow for a prediction to be made

Assumptions & Inferences

You cannot avoid making assumptions with data

Since assumptions are necessary, it becomes important to make sure you make the right ones!

Throughout semester we will try to identify the underlying assumptions within a technique and how to check

Probability & Inferential Statistics

Random Process

- A *random process* is a situation in which we know what outcomes could happen, but we don't know which particular outcome will happen.
- Examples: coin tosses, die rolls, music shuffle, whether the stock market goes up or down tomorrow, etc.
- It can be helpful to model a process as random even if it is not truly random.

Probability – Understanding Randomness

- There are several possible interpretations of probability but they (almost) completely agree on the mathematical rules probability must follow.
  - P(A) = Probability of event A
  - $0 \leq P(A) \leq 1$

Frequentist interpretation:

The probability of an outcome is the proportion of times the outcome would occur if we observed the random process an infinite number of times.

Bayesian interpretation:

A Bayesian interprets probability as a subjective degree of belief: For the same event, two separate people could have different viewpoints and so assign different probabilities.

Largely popularized by revolutionary advance in computational technology and methods during the last twenty years.

Practice

Which of the following events would you be most surprised by?

(a) exactly 3 heads in 10 coin flips

(b) exactly 3 heads in 100 coin flips

(c) exactly 3 heads in 1000 coin flips

Practice

Which of the following events would you be most surprised by?

(a) exactly 3 heads in 10 coin flips

(b) exactly 3 heads in 100 coin flips

(c) exactly 3 heads in 1000 coin flips

Law of Large Numbers

- *Law of large numbers* states that as more observations are collected, the proportion of occurrences with a particular outcome, $\hat{p}_n$, converges to the probability of that outcome, p.
    - For Example:
        - As the sample size increases, the sample mean tends to get closer to the population mean. And as the sample size approaches infinity, the sample mean approaches the population mean

Law of Large Numbers

When tossing a fair coin, if heads comes up on each of the first 10 tosses, what do you think the chance is that another head will come up on the next toss? 0.5, less than 0.5, or more than 0.5?

$\underline{H}\ \underline{H}\ \underline{H}\ \underline{H}\ \underline{H}\ \underline{H}\ \underline{H}\ \underline{H}\ \underline{H}\ \underline{H}\ \underline{?}$

- The probability is still 0.5, or there is still a 50% chance that another head will come up on the next toss.
  - P(H on 11th toss) = P(T on 11th toss) = 0.5
- The coin is not “due” for a tail
- The common misunderstanding of the LLN is that random processes are supposed to compensate for whatever happened in the past; this is just not true and is also called gambler’s fallacy (or law of averages).

Disjoint and non-disjoint outcomes

Disjoint (mutually exclusive) outcomes: Cannot happen at the same time.

- The outcome of a single coin toss cannot be a head and a tail.
- A student both cannot fail and pass a class.
- A single card drawn from a deck cannot be an ace and a queen.

Disjoint and non-disjoint outcomes

Disjoint (mutually exclusive) outcomes: Cannot happen at the same time.

- The outcome of a single coin toss cannot be a head and a tail.
- A student both cannot fail and pass a class.
- A single card drawn from a deck cannot be an ace and a queen.

Non-disjoint outcomes: Can happen at the same time.

- A student can get an A in Stats and A in Econ in the same semester.

Probability distributions

A *probability distribution* lists all possible events and the probabilities with which they occur.

- The probability distribution for the sex assigned at birth of one kid:

Event	Male	Female
Probability	0.5	0.5

Probability distributions

A *probability distribution* lists all possible events and the probabilities with which they occur.

- The probability distribution for the sex assigned at birth of one kid:

Event	Male	Female
Probability	0.5	0.5

- Rules for probability distributions:

1. The events listed must be disjoint
2. Each probability must be between 0 and 1
3. The probabilities must total 1

Probability distributions

A *probability distribution* lists all possible events and the probabilities with which they occur.

- The probability distribution for the sex assigned at birth of one kid:

Event	Male	Female
Probability	0.5	0.5

- Rules for probability distributions:

  1. The events listed must be disjoint
  2. Each probability must be between 0 and 1
  3. The probabilities must total 1

- The probability distribution for the sex assigned at birth of two kids:

Event	*MM*	*FF*	*MF*	*FM*
Probability	*0.25*	*0.25*	*0.25*	*0.25*

Statistics, Inference and Probability

Think about seeing statistics posted on the news

- Often times polling companies survey a large sample to then make a statemebt about the population

We assume the sample is representative of the larger population, but how representative?

- This is where probability theory comes in

Probability theory provides tools to assess how likely sample results are if they differ from the true population parameter

Role of Probability Theory

- Probability theory allows us to determine if survey results like “67% of respondents prefer Android" plausibly reflect the true population parameter
- While intuitive reasoning can get us part way, probability theory provides powerful mathematical tools to assess sample representativeness
- Though not statistics per se, probability theory forms the foundation for statistical inference

Probability

Frequentist approach

- Probability is a **long-run frequency** of an event

Bayesian approach

- Probability of an event as the **degree of belief**

Probability – Frequentist Approach

The more dominant one in our field

Flipping a coin over and over again

- P(heads) = 0.5
- First 20 flips: T,H,H,H,H,T,T,H,H,H,H,T,H,H,T,T,T,T,T,H
- 11/20 were Heads (55%)
- Calculate the proportion of heads at each flip (Number of Heads / Total Flips)

Probability – Frequentist Approach

Flip Number	Coin Result	Number of Heads	Proportion Heads
1	T	0	0.00

Flip Number	Coin Result	Number of Heads	Proportion Heads
1	T	0	0.00
2	H	1	0.50

Probability – Frequentist Approach

Flip Number	Coin Result	Number of Heads	Proportion Heads
1	T	0	0.00
2	H	1	0.50
3	H	2	0.67

Probability – Frequentist Approach

Probability – Frequentist Approach

$$As\ N \to \infty$$

Flip Number	Coin Result	Number of Heads	Proportion Heads
1	T	0	0.00
2	H	1	0.50
3	H	2	0.67
4	H	3	0.75
5	H	4	0.80
6	T	4	0.67
7	T	4	0.57
8	H	5	0.63
9	H	6	0.67
10	H	7	0.70
11	H	8	0.73
12	T	8	0.67
13	H	9	0.69
14	H	10	0.71
15	T	10	0.67
16	T	10	0.63
17	T	10	0.59
18	T	10	0.56
19	T	10	0.53
20	H	11	0.55

Look to R for simulation

Probability – Frequentist Approach

Pros

- Objective: the probability of an event is grounded in the world & exist in the physical universe
- Unambiguous: two people can watch the same sequence of events and will come up with the same answer

Cons

- Infinity: not possible in the physical world
  - What would happen if we flipped a coin an infinite amount of times?
- Narrow in scope: although we want to make statements about a single event, we typically are forbidden (see above point about infinity)
  - Example: There is a 60% chance that it will rain on Wednesday in Rochester.
  - “There is a category of days for which I predict a 60% chance of rain, and if we look only across those days for which I make this prediction, then on 60% of those days it will actually rain”

Probability – Bayesian Approach

Probability of an event as the degree of belief

Probabilities don't exist in the world, but in the thoughts and assumptions of the people

What do we mean by "degree of belief?"

- "Rational Gambling"

Probability – Bayesian Approach

Would you bet $5 on this? You will lose it all if you are incorrect.

- I have a 75% chance of making the shot.
- I have a 45% chance of making the shot.
- “Subjective probability” in terms of the bets you are willing to take

This allows researchers to bring previous knowledge to assign probabilities to a single event

- They do not only need to be repeatable

However, this means that this approach cannot be purely objective

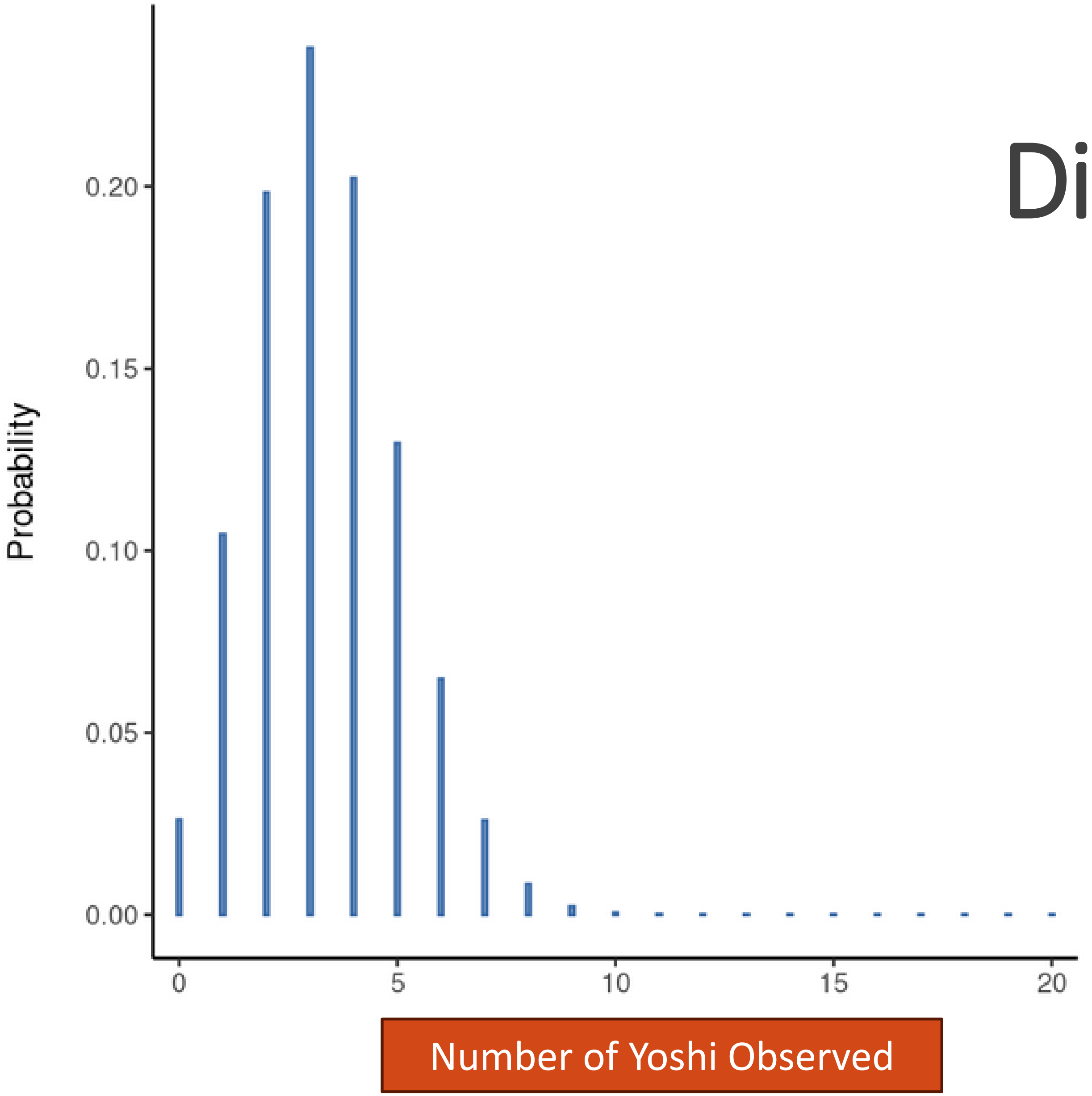
Frequentist vs. Bayesian Approach

Which is *right*?

- Nothing mathematically incorrect for either
- Have actually informed one another
- Some people get upset
  - Relying on frequentist methods could turn you into “a potent but sterile intellectual rake who leaves in his merry path a long train of ravished maidens but no scientific offspring” (Meehl, P. H. (1967). Theory testing in psychology and physics: A methodological paradox. *Philosophy of Science*, *34*, 103–115.)

Most commonly used in Psychology (and what we will focus on) is the frequentist method

Probability Distributions



Distributions - Binomial

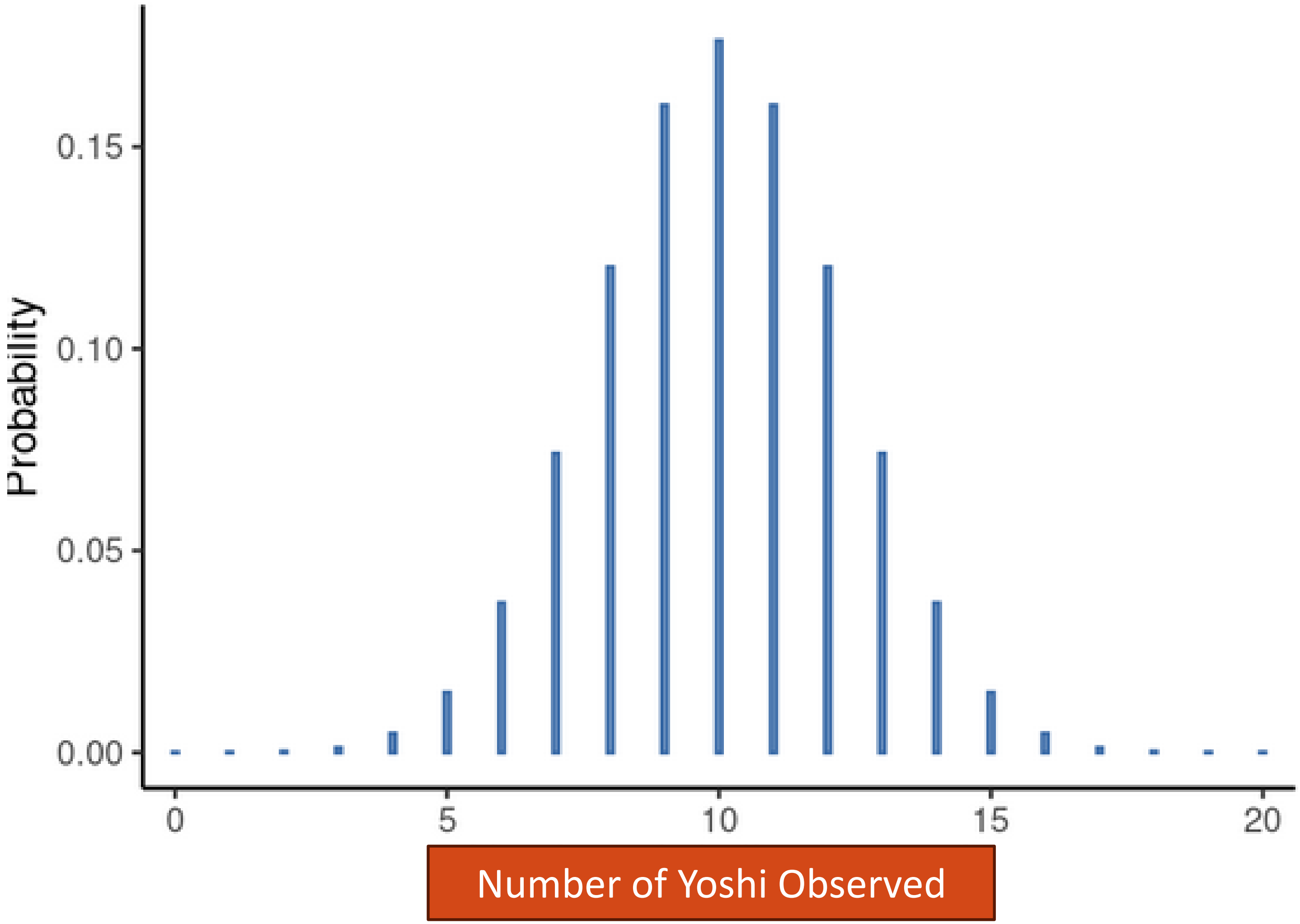
Discrete Variables

Experiment: 20 six-sided dice; all blank with 1 side that has a Yoshi

If we roll all 20, what is the probability that we'll get exactly 4 Yoshi?

Chance of one dice rolling Yoshi -> 1 in 6

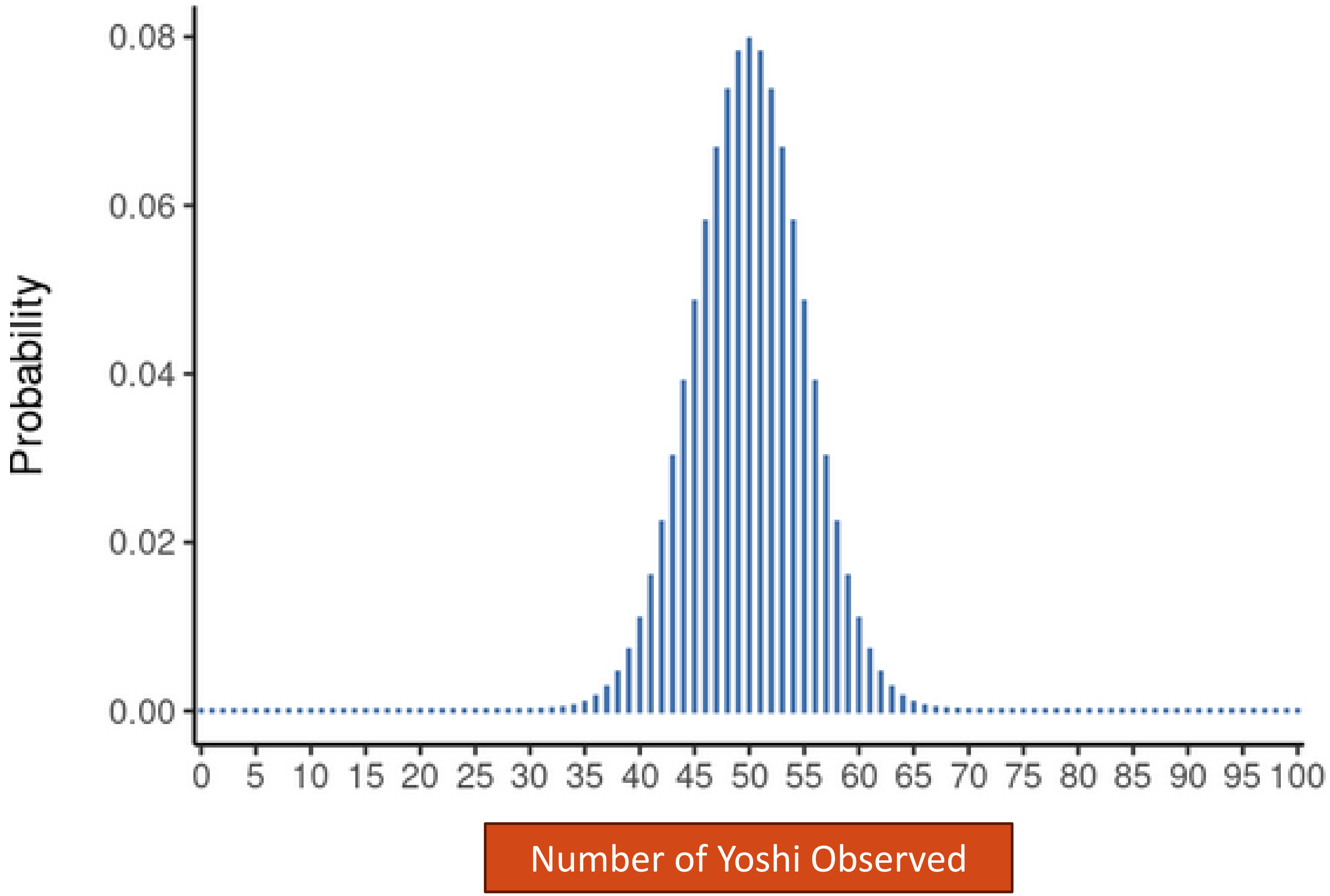
Distributions - Binomial



Experiment: Coin flip; 1 side Yoshi

Flip the coin 20 times

Chance of getting Yoshi -> 1 in 2



Distributions - Binomial

Experiment: Coin flip; 1 side Yoshi

Flip the coin **100** times

Chance of getting Yoshi -> 1 in 2

Approaches a bell curve

https://www.statcrunch.com/applets/type2&binomial

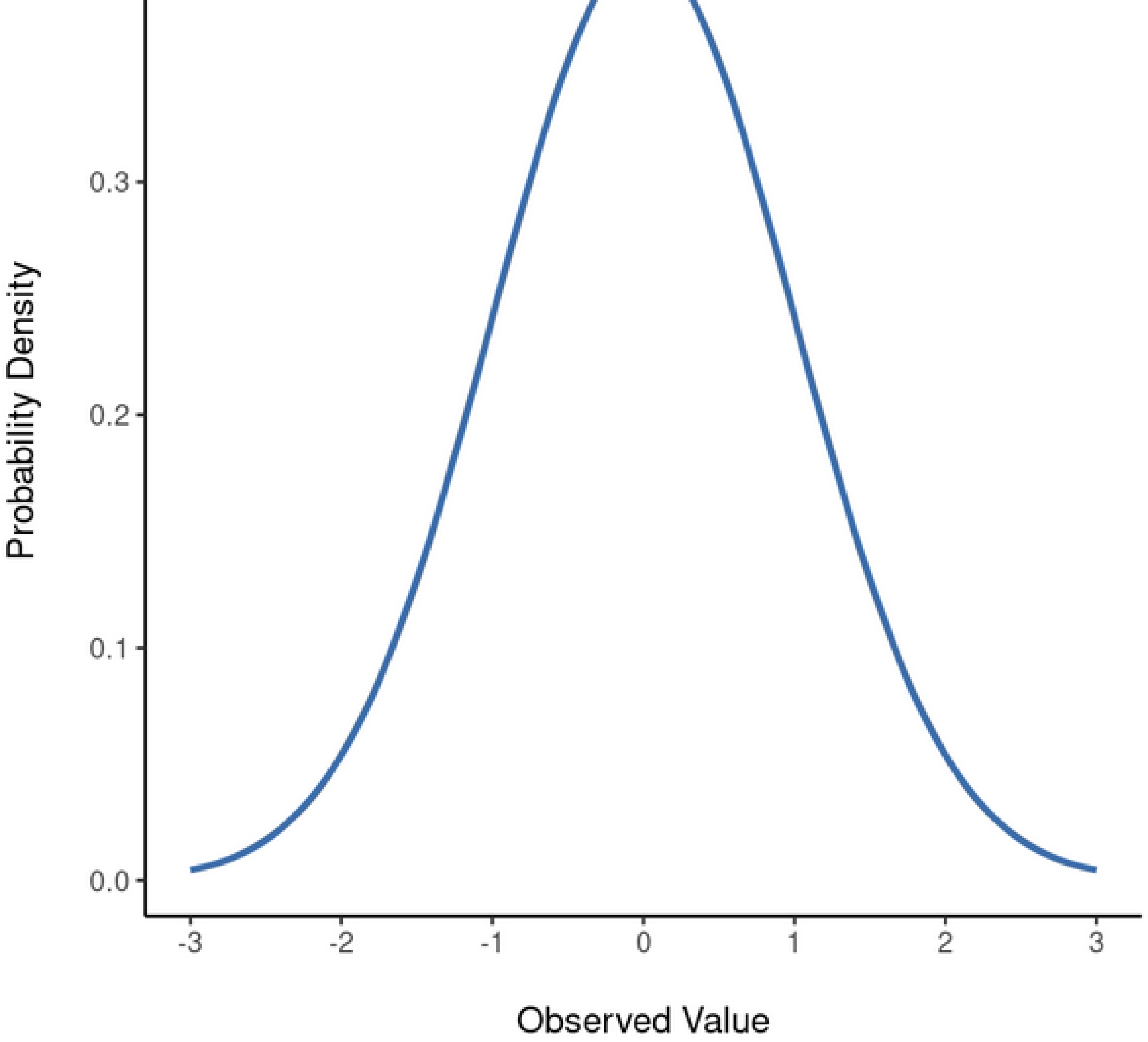
Distributions - Normal

Continuous Variables

Described by 2 parameters, the mean and the standard deviation

Mean = 0, SD = 1

https://www.statcrunch.com/applets/type2&normal

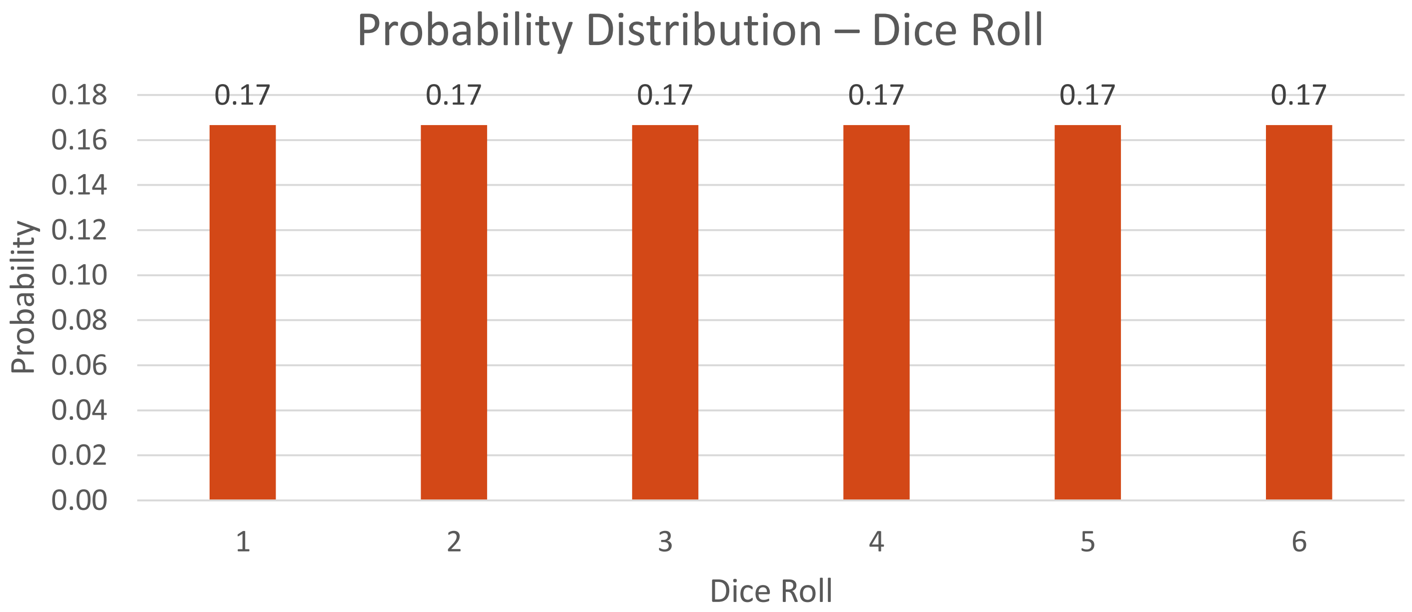


How does this relate to probability?

- Each statistic we calculate is considered a *random variable*
- Remember the probability distributions?

Event	1	2	3	4	5	6
Probability	1/6	1/6	1/6	1/6	1/6	1/6

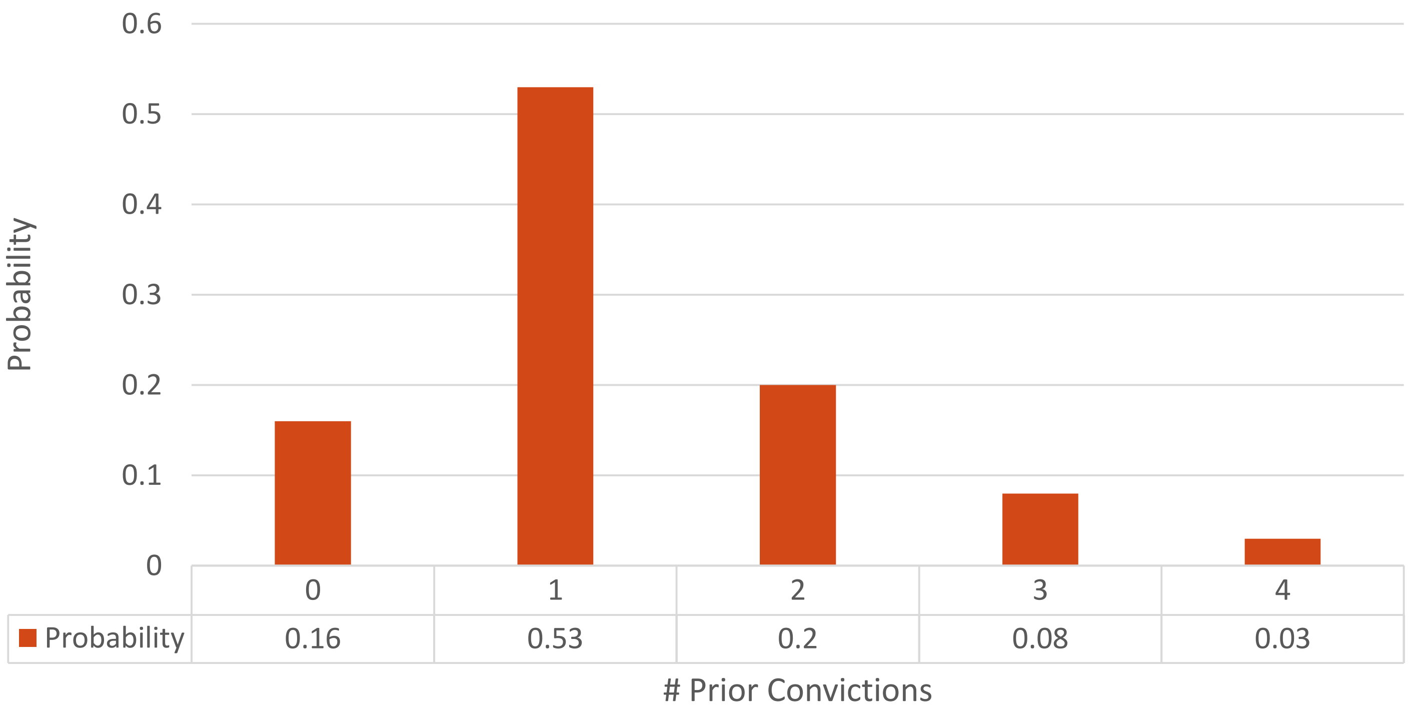
- This is the sampling distribution
- What happens when things aren't equal across groups?



Discrete Variables

Let X = number of prior convictions for prisoners at a state prison at which there are 500 prisoners. ($x = 0, 1, 2, 3, 4$)

$X = x$	0	1	2	3	4
$Number\ of\ Prisoners$	80	265	100	40	15
$f(x) = P(X = x)$	80/500	265/500	100/500	40/500	15/500
$f(x) = P(X = x)$	0.16	0.53	0.2	0.08	0.03

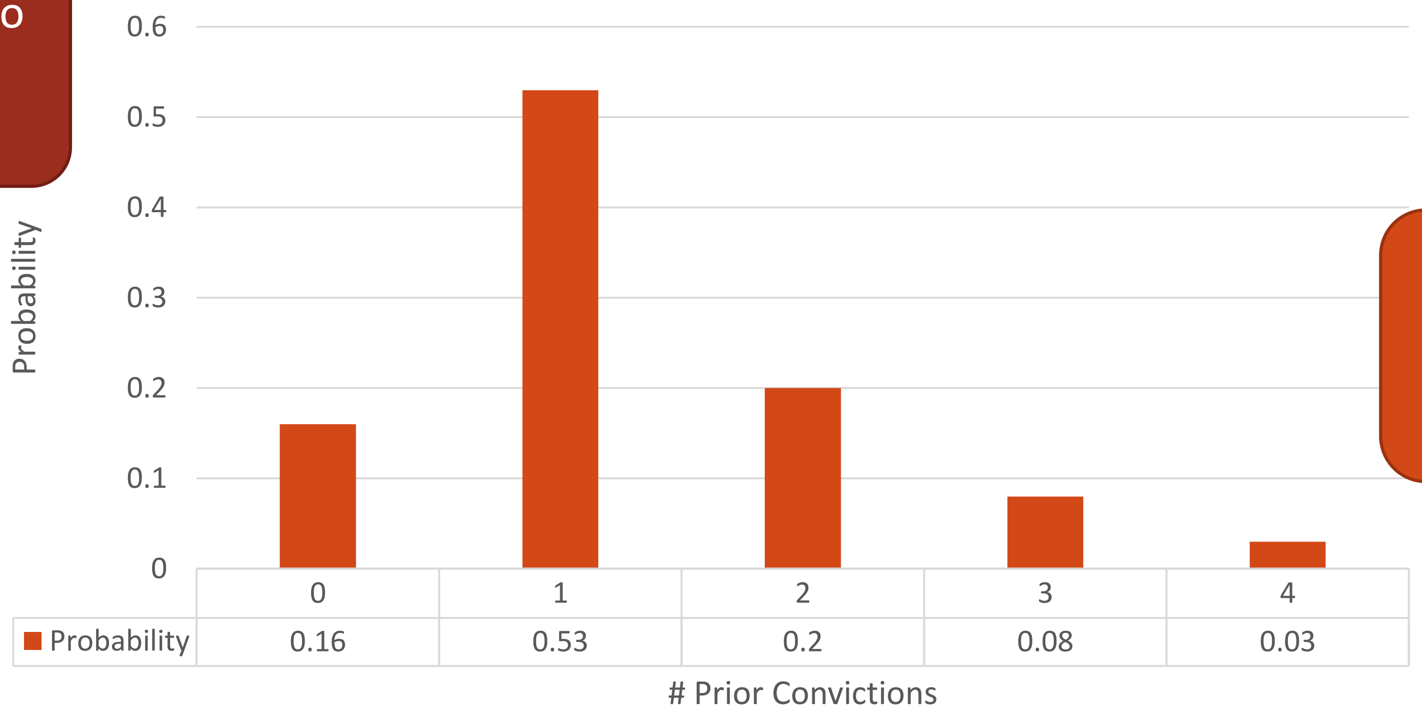


Discrete Variables

Let X = number of prior convictions for prisoners at a state prison at which there are 500 prisoners. ($x = 0, 1, 2, 3, 4$)

$X = x$	0	1	2	3	4
$Number\ of\ Prisoners$	80	265	100	40	15
$f(x) = P(X = x)$	80/500	265/500	100/500	40/500	15/500
$f(x) = P(X = x)$	0.16	0.53	0.2	0.08	0.03

Won't usually be able to sample the whole population



But this type of distribution can help inform us moving forward!

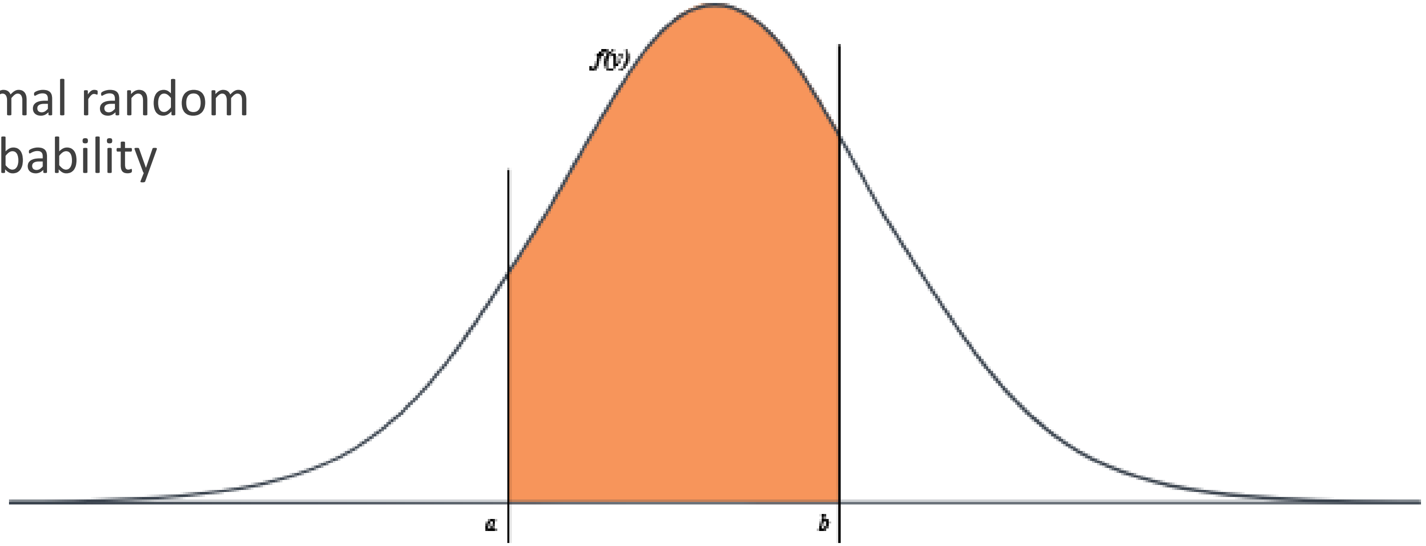
Continuous Variables – Probability Density Function

- What is the probability of getting the value ***exactly*** right?
  - $P(Y=y) = 0$
  - Turtles all the way down…

Event	1	2	3	4	5	6
Probability	1/6	1/6	1/[illegible]	1/6	1/6	1/6

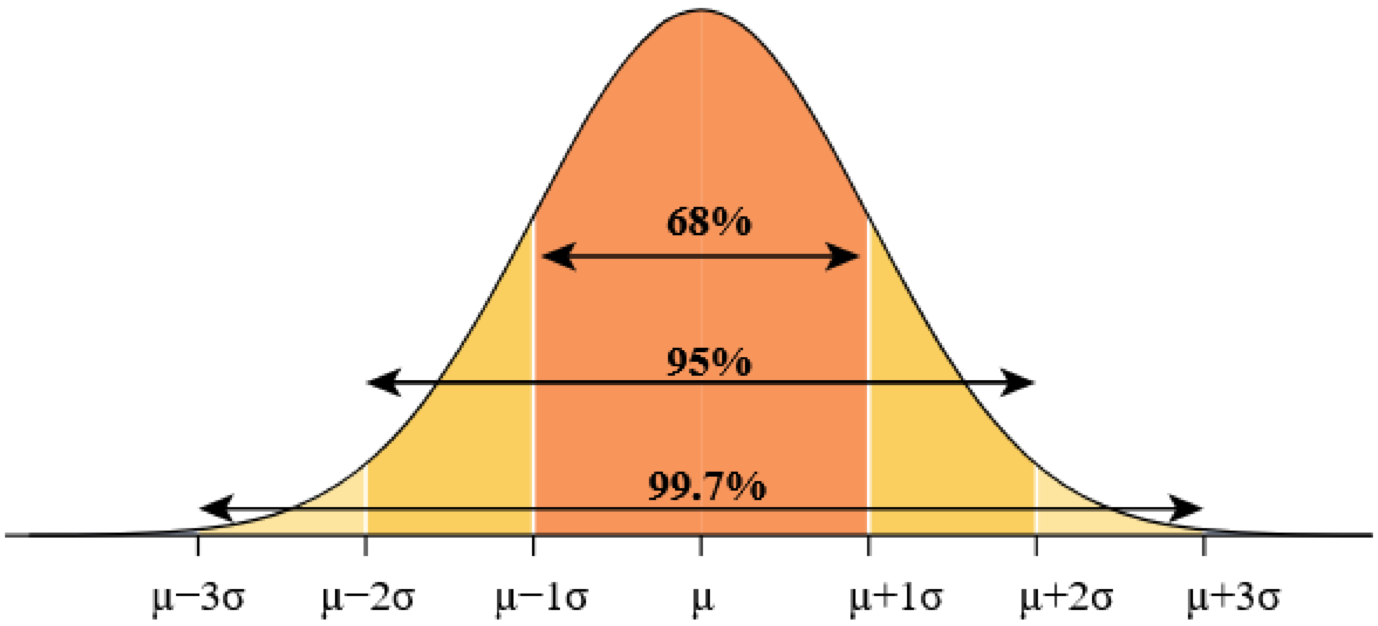
- Look at probability of an interval by examine the area under the curve

- Can calculate z-scores for normal random variables and identify the probability



Empirical Rule

- In any normal or bell-shaped distribution, roughly...
  - **68%** of the observations lie within **one** standard deviation to either side of the mean.
  - **95%** of the observations lie within **two** standard deviations to either side of the mean.
  - **99.7%** of the observations lie within **three** standard deviations to either side of the mean.



Estimating the unknown from a sample – Inferential Statistics

A LITTLE BIT OF METHODS IN THIS SECTION…

Parameter Estimation

- We are often interested in *population parameters*.
- Since complete populations are difficult (or impossible) to collect data on, we use *sample statistics* as *point estimates* for the unknown population parameters of interest.
- Sample statistics vary from sample to sample.
- Quantifying how sample statistics vary provides a way to estimate the *margin of error* associated with our point estimate.

Suppose we randomly sample 1,000 adults from each state in the US. Would you expect the sample means of their heights to be the same, somewhat different, or very different?

What if we took another random sample? Would they be identical?

Samples, populations and sampling

Sampling theory

- Drawing inferences from the sample about the population
- Defining a population
- Creating an appropriate sample
- LLN & CLT
- Population Parameters and Sample Statistics

Outlining a simulation: https://onlinestatbook.com/stat_sim/sampling_dist/

Defining a population

More of an abstract idea about who you want to make statements

Example: Running a study with 100 undergraduate students as participants. Who is the population?

- All of the undergraduate psychology students at RIT?
- Undergraduate psychology students in general, anywhere in the world?
- Americans currently living?
- Americans of similar ages to my sample?
- Anyone currently alive?
- Any human being, past, present or future?
- Any biological organism with a sufficient degree of intelligence operating in a terrestrial environment?
- Any intelligent being?

Critical point is that the sample is a subset of the population with the goal of drawing inferences about the larger population

The relationship depends upon the procedure of selection for the sample

Sampling – Simple random samples

Let’s say I have a bag containing 10 chips

Note about with vs. without replacement

- Most statistical theory is based on *with* replacement

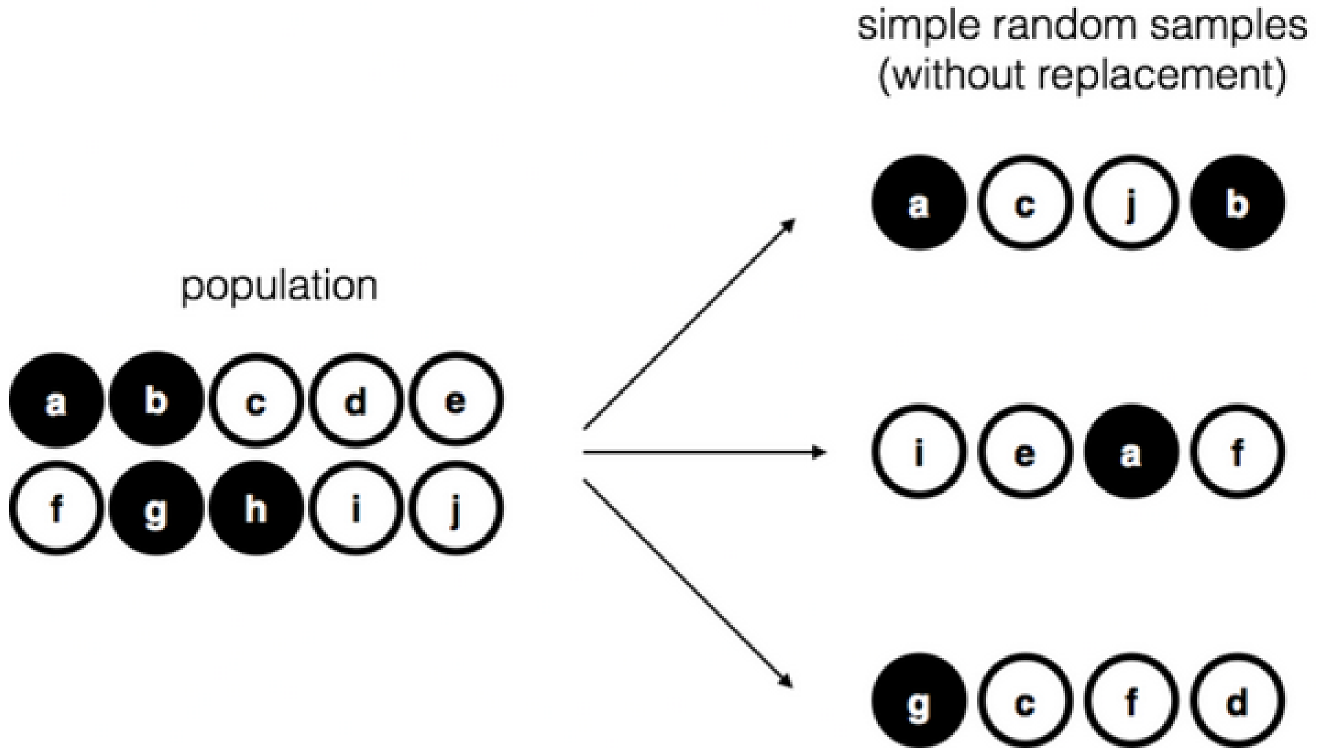
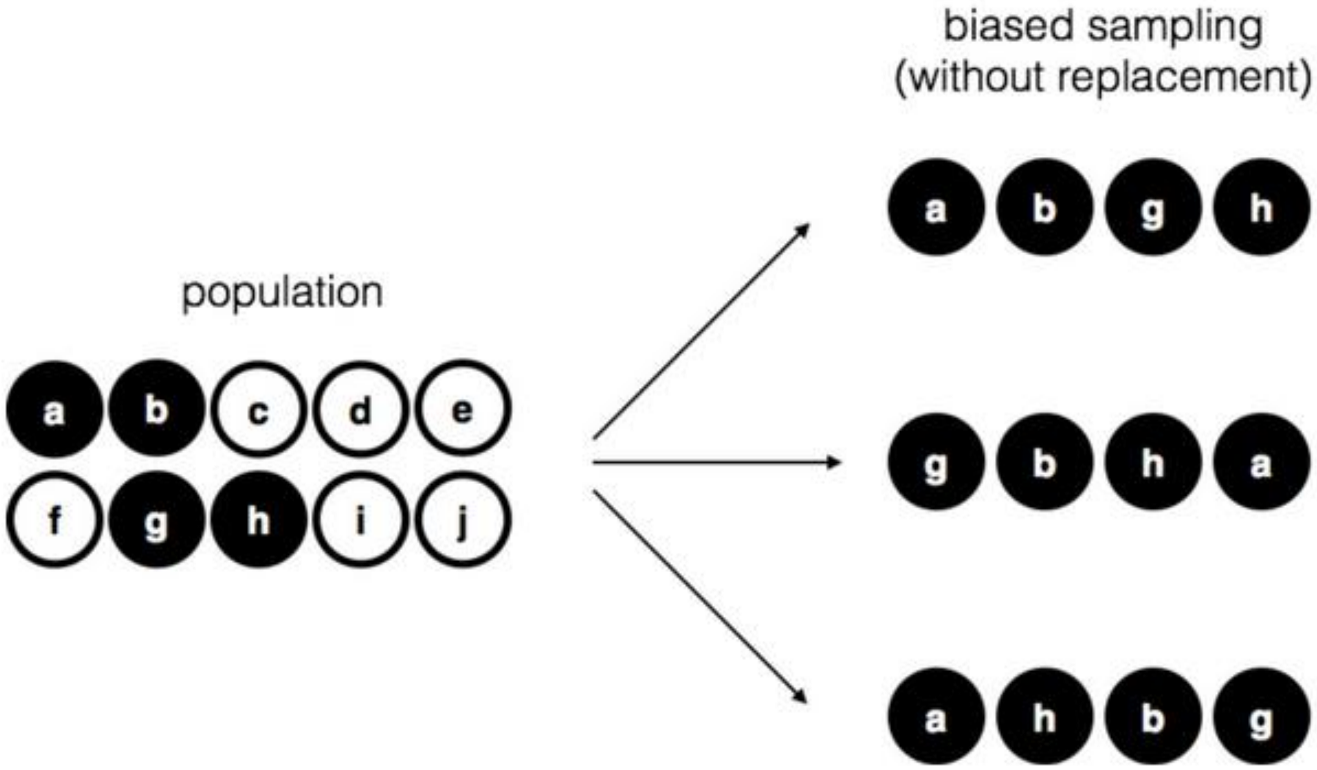


Figure 8.1: Simple random sampling without replacement from a finite population

Sampling – Biased Sampling



Sampling – How much does it matter

Difficult to identify a simple random sample in the real world

- Think about the differences between the two sampling techniques we just saw. Should it matter?

We can adjust for biases

Two Key Points

- Designing studies – think about the population you want to make inferences about
- Critiquing studies – Think about how their sample may impact the conclusions that are being made

Population Parameters & Sample Statistics

Let's return to the simulation
https://onlinestatbook.com/stat_sim/sampling_dist/

Parameters of the population

- Mean = 16; Standard Deviation = 5

Sample is taken – can calculate the mean and standard deviation of the sample

- These are the **sample statistics**

With this procedure, we will be able to estimate the population parameters and estimate confidence intervals

- We need to do a little more with Sampling Theory first

Law of Large Numbers

We may need to increase our precision of our estimates

- Increase our sample size; Why does this work?

Law of Large Numbers (a law about averages)

As the sample size increases, the sample mean tends to get closer to the population mean

- As the sample size approaches infinity, the sample mean approaches the population mean

Sampling Distribution & Central Limit Theorem

LLN – a long run guarantee...not super useful IRL

Create the Sampling Distribution Simulation

Run a study with N = 5

Replicate the study 10 times

- Collect 5 scores, calculate mean of sample
- Repeat
- Look at the simulated distribution

Create a sampling distribution of the means

- Sampling distributions exist for other statistics

Sampling Distribution & Central Limit Theorem

What happens to sampling distribution when the sample size increases?

- Standard Deviation of sampling distribution = **Standard Error**

What happens to sampling distribution when population is non-normal?

Central Limit Theorem

- The mean of the sampling distribution is the same as the mean of the population
- The standard deviation of the sampling distribution (i.e., the standard error) gets smaller as the sample size increases
- The shape of the sampling distribution becomes normal as the sample size increases

Estimating Population Parameter

Calculated from the sample

Sample Statistic ≠ Estimate Population Parameter (conceptually

- Sample statistic – Description of your data
- Estimate – best guess about the population

Most commonly used are Mean and Standard Deviation

- We will get to when talking about descriptives

Confidence Intervals

Every dataset has some level of uncertainty

- Need to quantify amount of uncertainty
- We want to say - “There is a 95% chance that the true mean lies between 12 and 23”
  - Confidence Interval for the mean

Confidence Intervals

CLT – Sampling distribution of the man is approximately normal

Normal Distribution – 95% chance that an observation will fall within 2 standard deviations of the mean

- 1.96 Standard Deviations (Standard Error of Sampling Distribution)

Can repeat the calculation of the confidence interval (frequentist approach)

Look to Simulation: https://rpsychologist.com/d3/ci/

What we can say "There is a 95% chance that the confidence interval captures the true mean"

Summary

Probability Theory

- Frequentist and Bayesian

Inferential Statistics

- Taking data and applying probability theory to determine the likelihood of the data

Distributions

- Normal & Binomial

Estimation

- Making inferences about the population
  - Sampling, Law of Large Numbers, Central Limit Theorem & Sampling Distributions