# In-person session 6

**September 26, 2022**

PMAP 8521: Program evaluation
Andrew Young School of Policy Studies

# Plan for today

**Exam 1**

**FAQs**

**Confidence intervals, credible intervals, and a crash course on Bayesian statistics**

# Exam 1

# Tell us about Exam 1!

# FAQs

# Are *p*-values really misinterpreted in published research?

# Power calculations and sample size

## Won't we always be able to find a significant effect if the sample size is big enough?

**Yes!**

# Math with computers

andhs.co/live

# Are the results from p-hacking actually a threat to validity?

# Do people actually post their preregistrations?

# Yes!

## OSF

See **this** and **this** for examples

## As Predicted

See **this**

# Do you have any tips for identifying the threats to validity in articles since they're often not super clear?

## Especially things like spillovers, Hawthorne effects, and John Henry effects?

**Using a control group of some kind seems to be the common fix for all of these issues.**

**What happens if you can't do that? Is the study just a lost cause?**

# Confidence intervals, credible intervals, and a crash course on Bayesian statistics

# In the absence of p-values, I'm confused about how we report... significance?

# Imbens and p-values

Nobody really cares about p-values

Decision makers want to know
a number or a range of numbers—
some sort of effect and uncertainty

Nobody cares how likely a number would be
in an imaginary null world!

# Imbens's solution

**Report point estimates and some sort of range**

> "It would be preferable if reporting standards emphasized confidence intervals or standard errors, and, even better, Bayesian posterior intervals."

**Point estimate**

**The single number you calculate (mean, coefficient, etc.)**

**Uncertainty**

**A range of possible values**

# Greek, Latin, and extra markings

**Statistics: use a sample to make inferences about a population**

## Greek

Letters like $\beta_1$ are the ***truth***

Letters with extra markings like $\hat{\beta}_1$ are our ***estimate*** of the truth based on our sample

## Latin

Letters like $X$ are ***actual data*** from our sample

Letters with extra markings like $\bar{X}$ are ***calculations*** from our sample

# Estimating truth

**Data → Calculation → Estimate → Truth**

| | |
|---|---|
| Data | $X$ |
| Calculation | $\bar{X} = \frac{\sum X}{N}$ |
| Estimate | $\hat{\mu}$ |
| Truth | $\mu$ |

$$\bar{X} = \hat{\mu}$$

$$X \rightarrow \bar{X} \rightarrow \hat{\mu} \xrightarrow{\text{🤞 hopefully 🤞}} \mu$$

# Population parameter

**Truth = Greek letter**

**An single unknown number that is true for the entire population**

Proportion of left-handed students at GSU

Median rent of apartments in NYC

Proportion of red M&Ms produced in a factory

ATE of your program

# Samples and estimates

## We take a sample and make a guess

**This single value is a *point estimate***

**(This is the Greek letter with a hat)**

# Variability

**You have an estimate,
but how different might that
estimate be if you take another sample?**

# Left-handedness

You take a random sample of
50 GSU students and 5 are left-handed.

**If you take a different random sample of 50 GSU students, how many would you expect to be left-handed?**

3 are left-handed. Is that surprising?

40 are left-handed. Is that surprising?

# Nets and confidence intervals

**How confident are we that the sample picked up the population parameter?**

**Confidence interval is a net**

**We can be X% confident that our net is picking up that population parameter**

If we took 100 samples, at least 95 of them would have the true population parameter in their 95% confidence intervals

> A city manager wants to know the true average property value of single-value homes in her city. She takes a random sample of 200 houses and builds a 95% confidence interval. The interval is ($180,000, $300,000).

**We're 95% confident that the interval ($180,000, $300,000) captured the true mean value**

# WARNING

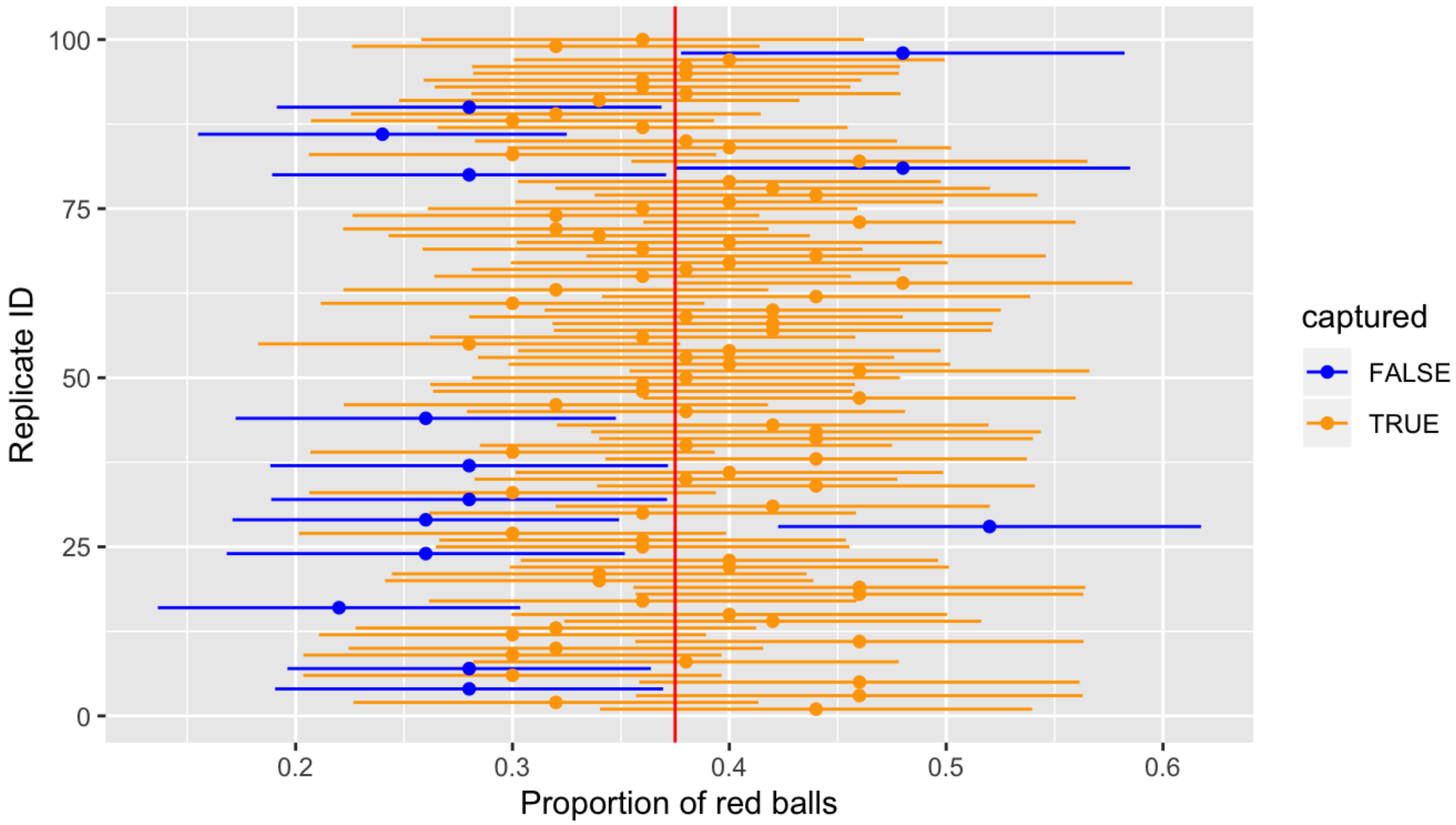
It is way too tempting to say "We're 95% sure that the population parameter is X"

People do this all the time! People with PhDs!

YOU will try to do this too

# Nets

**If you took lots of samples,
95% of their confidence intervals
would have the single true value in them**



100
75
50
25
0
Replicate ID
0.2
0.3
0.4
0.5
0.6
Proportion of red balls
captured
FALSE
TRUE

# Frequentism

## This kind of statistics is called "frequentism"

**The population parameter θ is fixed and singular while the data can vary**

$$P(\text{Data} \mid \theta)$$

**You can do an experiment over and over again; take more and more samples and polls**

# Frequentist confidence intervals

"We are 95% confident that this net captures the true population parameter"

~~"There's a 95% chance that the true value falls in this range"~~

# Weekends and restaurant scores

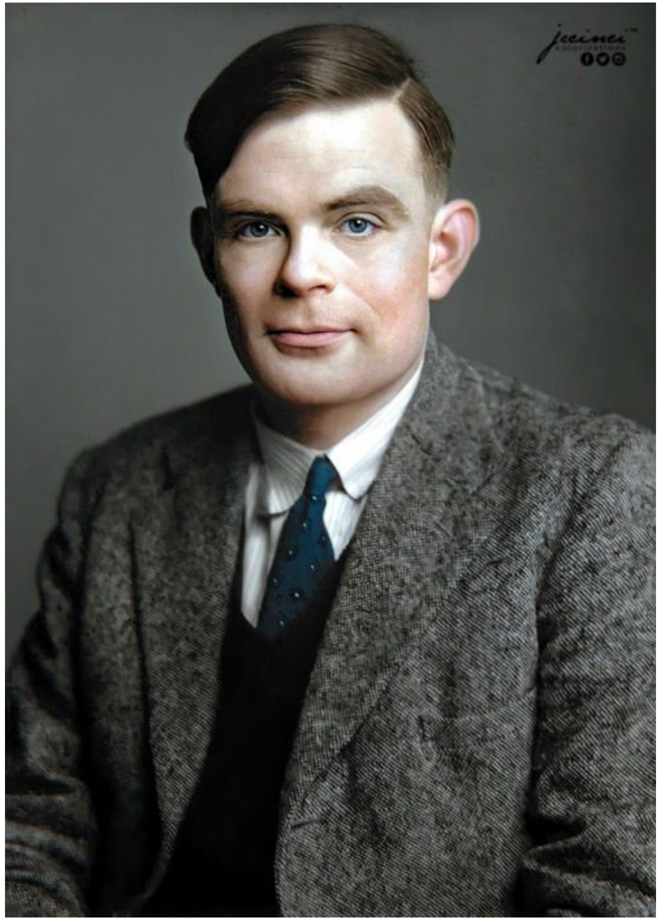
# Bayesian statistics

Rev. Thomas Bayes

$$P(\theta \mid \text{Data})$$

$$P(\text{H} \mid \text{E}) = \frac{P(\text{H}) \times P(\text{E} \mid \text{H})}{P(\text{E})}$$

# Bayesianism in WWII



Alan Turing


An enigma machine

$$P(\mathrm{H} \mid \mathrm{E}) = \frac{P(\mathrm{H}) \times P(\mathrm{E} \mid \mathrm{H})}{P(\mathrm{E})}$$

$$P(\text{Hypothesis} \mid \text{Evidence}) = \frac{P(\text{Hypothesis}) \times P(\text{Evidence} \mid \text{Hypothesis})}{P(\text{Evidence})}$$

$$P(\mathrm{H} \mid \mathrm{E}) = \frac{P(\mathrm{H}) \times P(\mathrm{E} \mid \mathrm{H})}{P(\mathrm{E})}$$

▼

$$\overbrace{P(\text{Unknown} \mid \text{Data})}^{\text{Posterior}} = \frac{\overbrace{P(\text{Unknown})}^{\text{Prior}} \times \overbrace{P(\text{Data} \mid \text{Unknown})}^{\text{Likelihood}}}{\underbrace{P(\text{Data})}_{\text{Average likelihood}}}$$

$$\overbrace{P(\text{Unknown} \mid \text{Data})}^{\text{Posterior}} = \frac{\overbrace{P(\text{Unknown})}^{\text{Prior}} \times \overbrace{P(\text{Data} \mid \text{Unknown})}^{\text{Likelihood}}}{\underbrace{P(\text{Data})}_{\text{Average likelihood}}}$$

▼

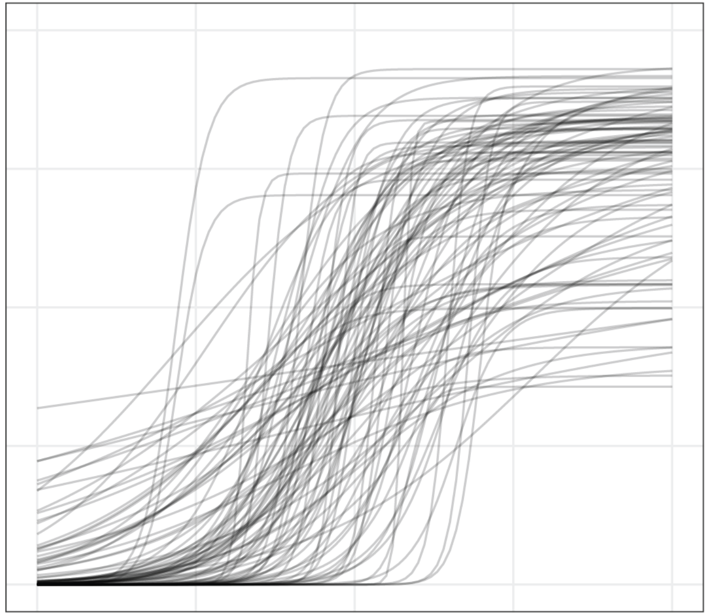
$$\overbrace{P(\text{Unknown})}^{\text{Prior}} \times \overbrace{P(\text{Data} \mid \text{Unknown})}^{\text{Likelihood}} \propto \overbrace{P(\text{Unknown} \mid \text{Data})}^{\text{Posterior}}$$

# Bayesian statistics and more complex questions

$$\underbrace{P(\text{Unknown})}_{\text{Prior}} \times \underbrace{P(\text{Data} \mid \text{Unknown})}_{\text{Likelihood}} \propto \underbrace{P(\text{Unknown} \mid \text{Data})}_{\text{Posterior}}$$

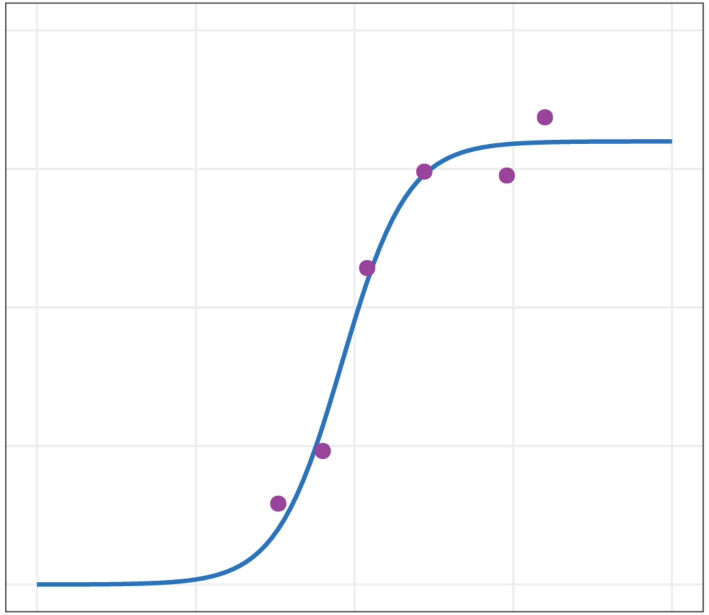
**Plausible curves before seeing the data**

The prior: **P(Unknown)**



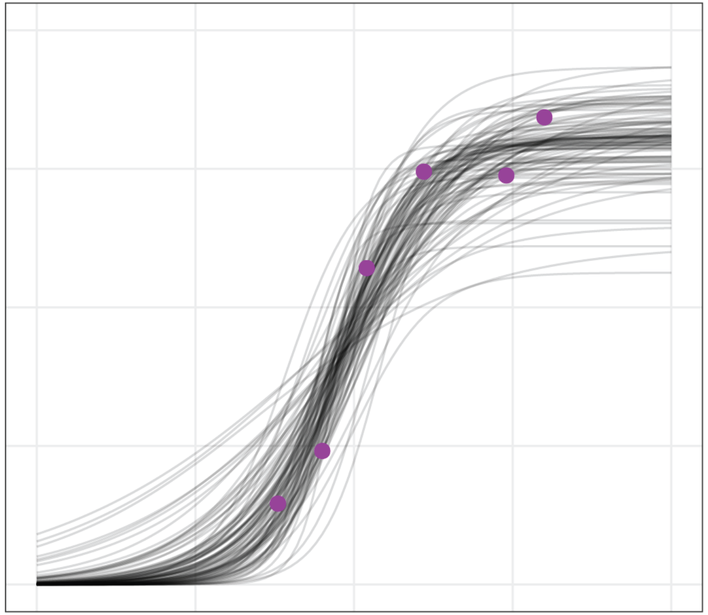
**How well the curves fit the data**

The likelihood: **P(Data | Unknown)**



**Plausible curves after seeing the data**

The posterior: **P(Unknown | Data)**



# But the math is too hard!

So we simulate!

(Monte Carlo Markov Chains, or MCMC)

# Weekends and restaurant scores again

# Bayesianism and parameters

**In the world of frequentism, there's a fixed population parameter and the data can hypothetically vary**

$$P(\text{Data} \mid \theta)$$

**In the world of Bayesianism, the data is fixed (you collected it just once!) and the population parameter can vary**

$$P(\theta \mid \text{Data})$$

# Bayesian credible intervals

**(AKA posterior intervals)**

**"Given the data, there is a 95% probability that the true population parameter falls in the credible interval"**

# Intervals

## Frequentism

There's a 95% probability that the range contains the true value

Probability of the range

Few people naturally think like this

## Bayesianism

There's a 95% probability that the true value falls in this range

Probability of the actual value

People *do* naturally think like this!

# Thinking Bayesianly

**We all think Bayesianly, even if you've never heard of Bayesian stats**

**Every time you look at a confidence interval, you inherently think that the parameter is around that value, but that's wrong!**

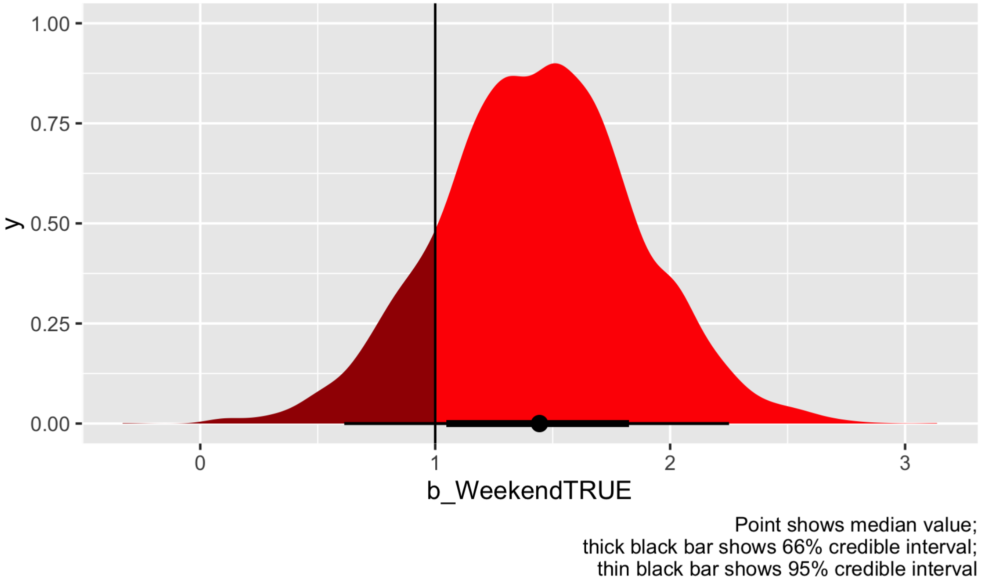
**BUT Imbens cites research that that's actually generally okay**

**Often credible intervals are super similar to confidence intervals**

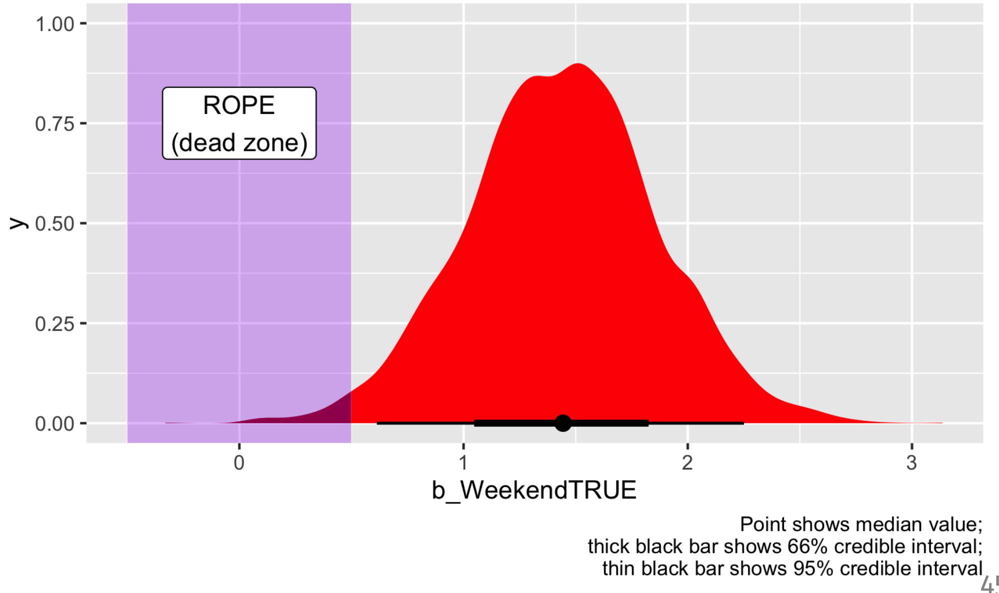
# Bayesian inference

## What do you do without p-values then?

### Probability of direction




### Region of practical equivalence (ROPE)



# Weekends and restaurant scores once more