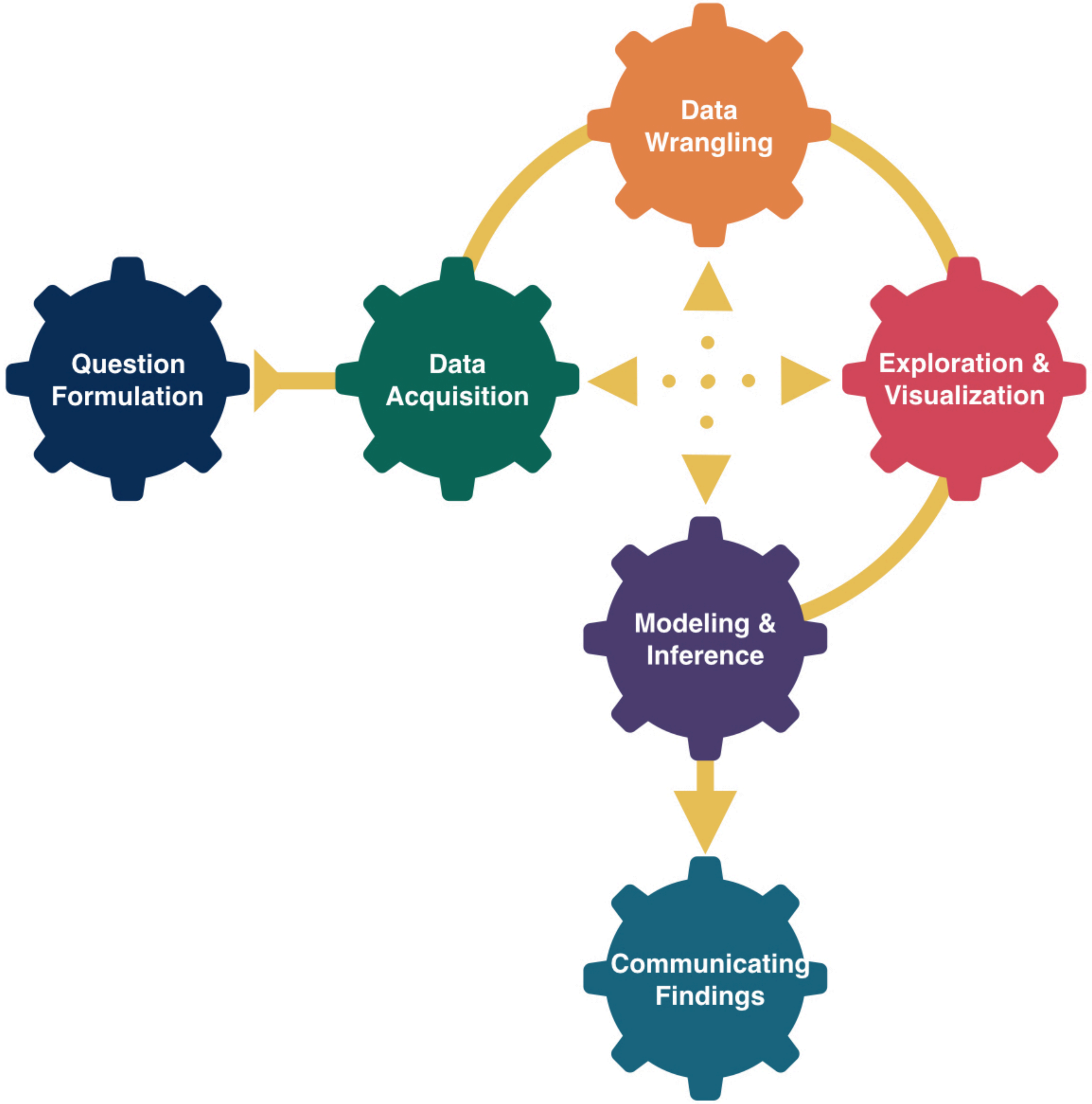
MATH 141: Combined Lecture Slides
Weeks 8-14



Hypothesis Testing II

Megan Ayers

Math 141 | Spring 2026
Monday, Week 8

Goals for today

- Midterm revisions
- Practice framing research questions in terms of hypotheses
- Practice defining null distributions
- 1 vs 2 sided tests

Midterm revisions

If you scored less than 80% on the midterm you have the opportunity to get some points back. Parameters:

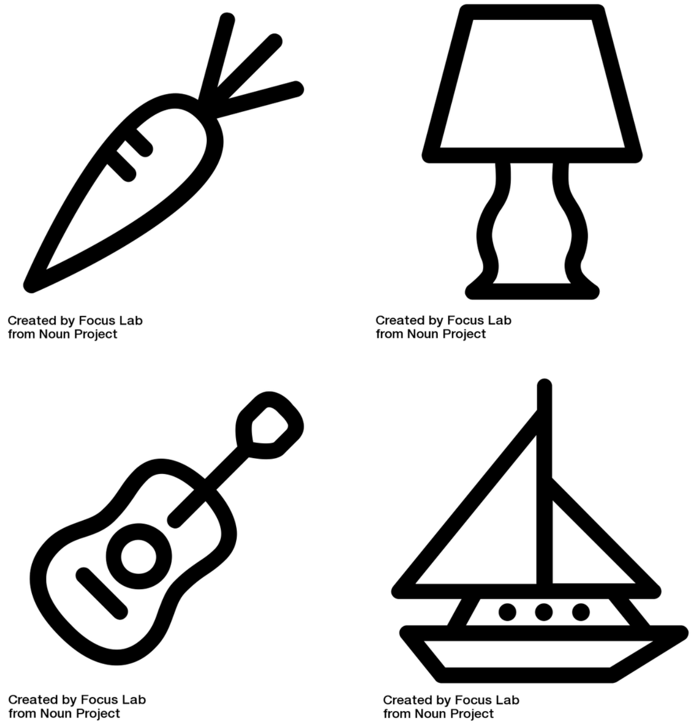
- You can get up to 50% of your missed points back (up to an 80% total score on the midterm).
  - **Example:** If you got a 70% on the midterm, you can get up to an 80% with revisions, but if you got a 50% on the midterm you can get up to a 75% with revisions.
- For each question part you can get half the points you missed.
  - **Example:** If you got a 1 / 3 on Question 1 part (a), an entirely correct solution submitted to me on the midterm makeups will increase your score to 2 / 3 on Question 1 part (a).
- Changes to rules now apply:
  1. You must give a brief explanation of where you went wrong for each revised problem and why your revision fixes this.
  2. You may consult your notes for the course and any of our internal course material. **But outside resources, including friends, peers, the internet, or AI tools, are still prohibited.**
  3. You have no time pressure.
  4. You can come to my (Megans's) office hours to discuss questions you are stuck on, but no other office hours or individual tutoring.
- **Revision instructions are on the course website.**
- **Submit paper copies of your midterm revisions to me by Wednesday 4/1 at 5:00pm. No late submissions!**

Recap: Framework for Hypothesis Testing

1. Present research question and identify hypotheses
   - Null hypothesis (H_0): status quo, random chance, no effect…
   - Alternative hypothesis (H_a): the researchers' conjecture
2. Describe **Null** distribution
3. Obtain data, calculate relevant **Test Statistic**
4. Calculate the **P-value**
   - P-value = likelihood of observing the Test Statistic or something more extreme assuming the Null Hypothesis
5. Use the P-value to make a conclusion on the research question

Let's Practice

Example: Does Extrasensory Perception (ESP) exist?



Psychologists Bem and Honorton conducted extrasensory perception studies:

- A "sender" randomly chooses an object out of 4 possible objects and sends that information to a "receiver".
- The "receiver" is then given a set of 4 possible objects and they must decide which one most resembles the object sent to them.

Out of **329 reported trials**, the "receivers" correctly identified the object **106 times**.

Hypothesis Testing Steps: ESP

Psychologists Bem and Honorton conducted extrasensory perception studies:

- A "sender" randomly chooses an object out of 4 possible objects and sends that information to a "receiver".
- The "receiver" is then given a set of 4 possible objects and they must decide which one most resembles the object sent to them.

Out of **329 reported trials**, the "receivers" correctly identified the object **106 times**.

Discuss with neighbor(s):

1. What is the relevant null hypothesis (H_0) and alternative hypothesis (H_a)?

2. Describe **Null distribution**. How might we simulate it? (*Hint: Recall the card guessing example*)

3. What is our relevant **Test Statistic**?

4. What does the **P-value** represent here? How might we calculate it, given step 2?

5. How would we use the P-value to make a conclusion on the research question?

Hypothesis Testing Steps: ESP

1. What is the relevant null hypothesis (H_0) and alternative hypothesis (H_a)?
    - H_0: Participants were guessing randomly ($p = 0.25$)
    - H_A: Participants were guessing better than random ($p > 0.25$)

Hypothesis Testing Steps: ESP

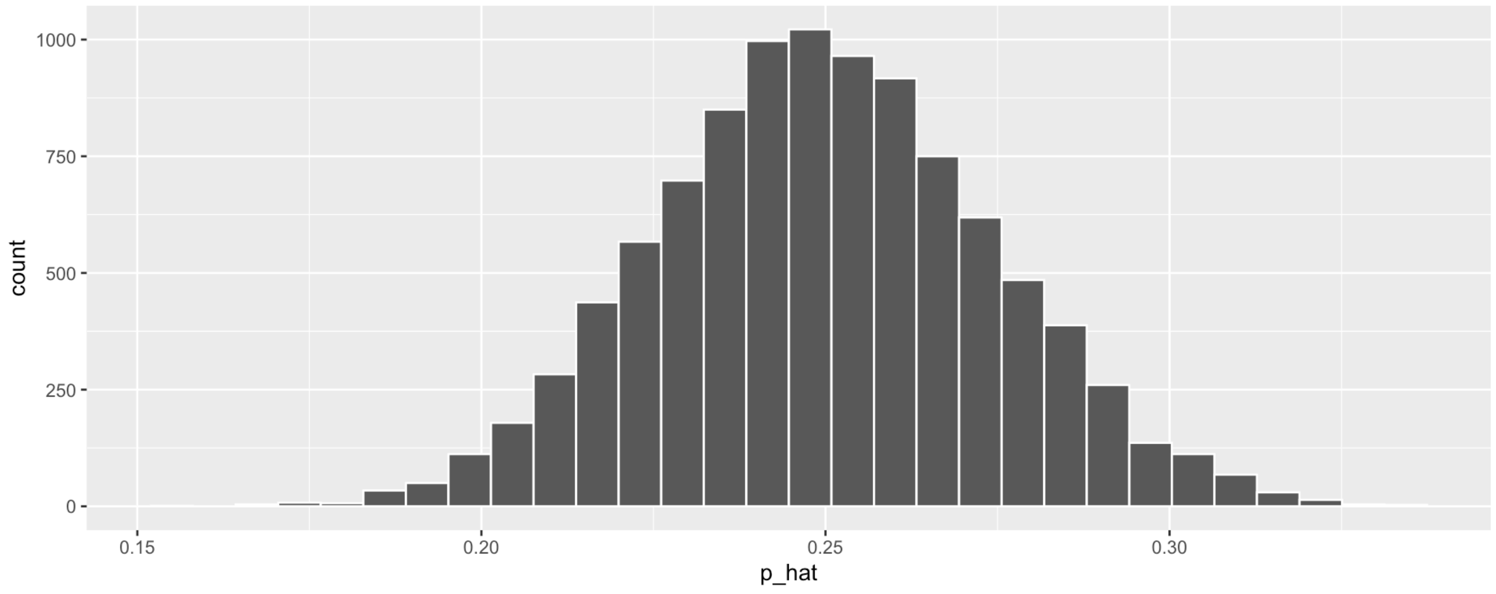
2. Describe "Null" distribution. How might we simulate it?

- The distribution of $\widehat{p}$ (the proportion of correct guesses) we would see if participants guess at random each time.

Steps to simulate this:

1. Sample with replacement from a vector of **0**'s and **1**'s, where we have a 25% chance of sampling **1** each time
2. Compute proportion of **1**'s sampled
3. Repeat 1 and 2 many times.

```
set.seed(123)
guesses <- c(0, 0, 0, 1) # Like a 4-sided dice with sid
null_stats <- data.frame(correct = guesses) %>%
  rep_sample_n(size = 329, replace = TRUE,
               reps = 10000) %>%
  group_by(replicate) %>%
  summarize(p_hat = mean(correct))
ggplot(null_stats, aes(x = p_hat)) + geom_histogram(col
```



Hypothesis Testing Steps: ESP

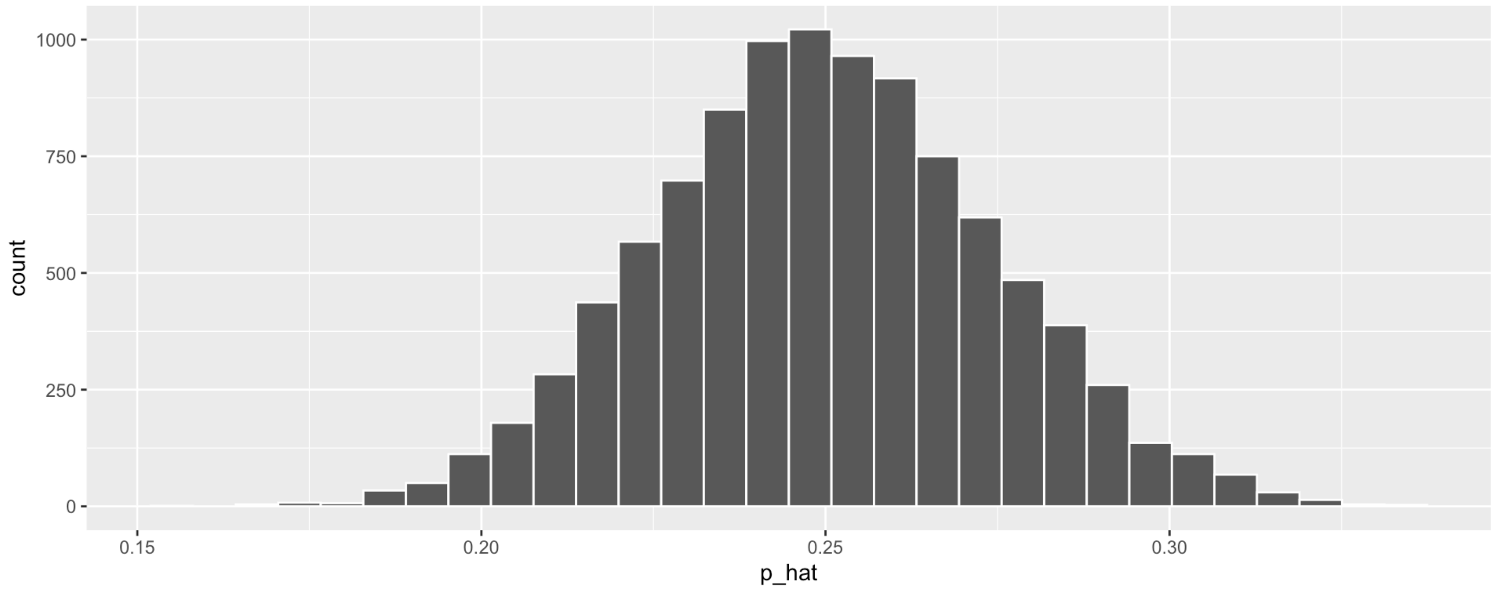
2. Describe “Null” distribution. How might we simulate it?

- The distribution of $\widehat{p}$ (the proportion of correct guesses) we would see if participants guess at random each time.

Steps to simulate this:

1. Sample with replacement from a vector of **0**’s and **1**’s, where we have a 25% chance of sampling **1** each time
2. Compute proportion of **1**’s sampled
3. Repeat 1 and 2 many times.

```r
set.seed(123)
guesses <- c(0, 0, 0, 1) # Like a 4-sided dice with sid
null_stats <- data.frame(correct = guesses) %>%
  rep_sample_n(size = 329, replace = TRUE,
               reps = 10000) %>%
  group_by(replicate) %>%
  summarize(p_hat = mean(correct))
ggplot(null_stats, aes(x = p_hat)) + geom_histogram(col
```



Hypothesis Testing Steps: ESP

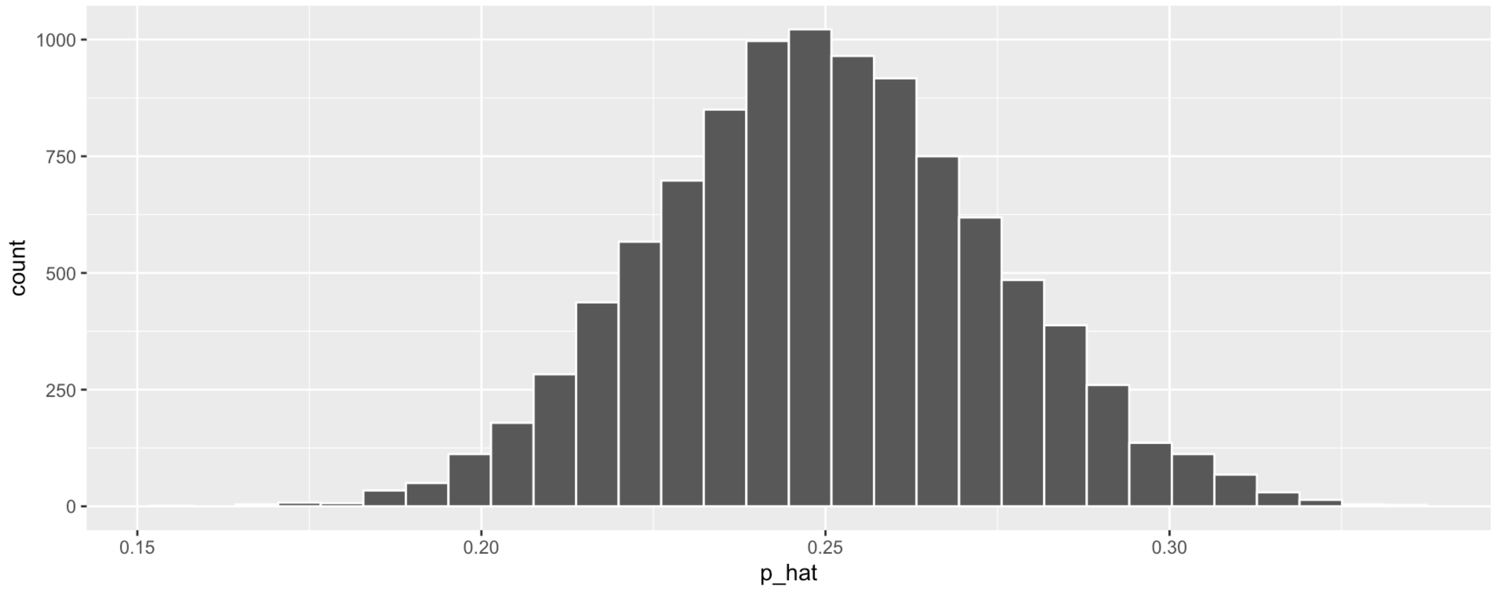
2. Describe “Null” distribution. How might we simulate it?

- The distribution of $\widehat{p}$ (the proportion of correct guesses) we would see if participants guess at random each time.

Steps to simulate this:

1. Sample with replacement from a vector of **0**’s and **1**’s, where we have a 25% chance of sampling **1** each time
2. Compute proportion of **1**’s sampled
3. Repeat 1 and 2 many times.

```r
set.seed(123)
guesses <- c(0, 0, 0, 1) # Like a 4-sided dice with sid
null_stats <- data.frame(correct = guesses) %>%
  rep_sample_n(size = 329, replace = TRUE,
               reps = 10000) %>%
  group_by(replicate) %>%
  summarize(p_hat = mean(correct))
ggplot(null_stats, aes(x = p_hat)) + geom_histogram(col
```



Hypothesis Testing Steps: ESP

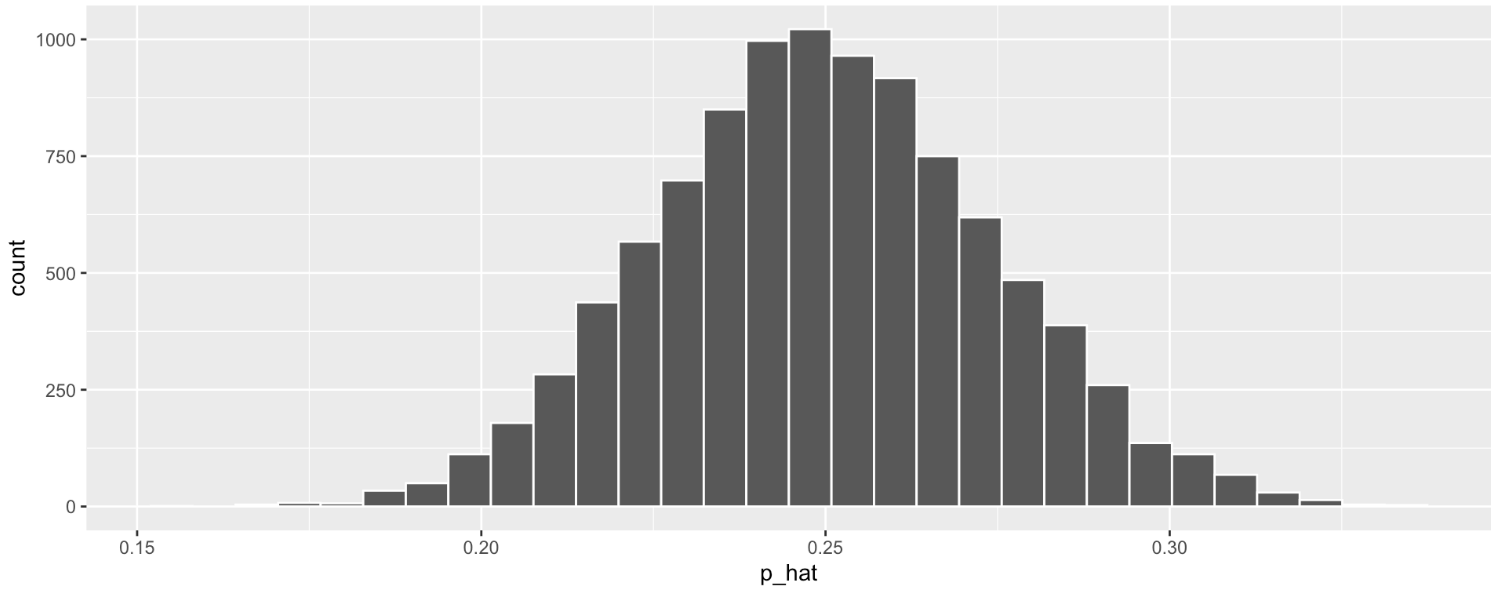
2. Describe “Null” distribution. How might we simulate it?

- The distribution of $\widehat{p}$ (the proportion of correct guesses) we would see if participants guess at random each time.

Steps to simulate this:

1. Sample with replacement from a vector of **0**’s and **1**’s, where we have a 25% chance of sampling **1** each time
2. Compute proportion of **1**’s sampled
3. Repeat 1 and 2 many times.

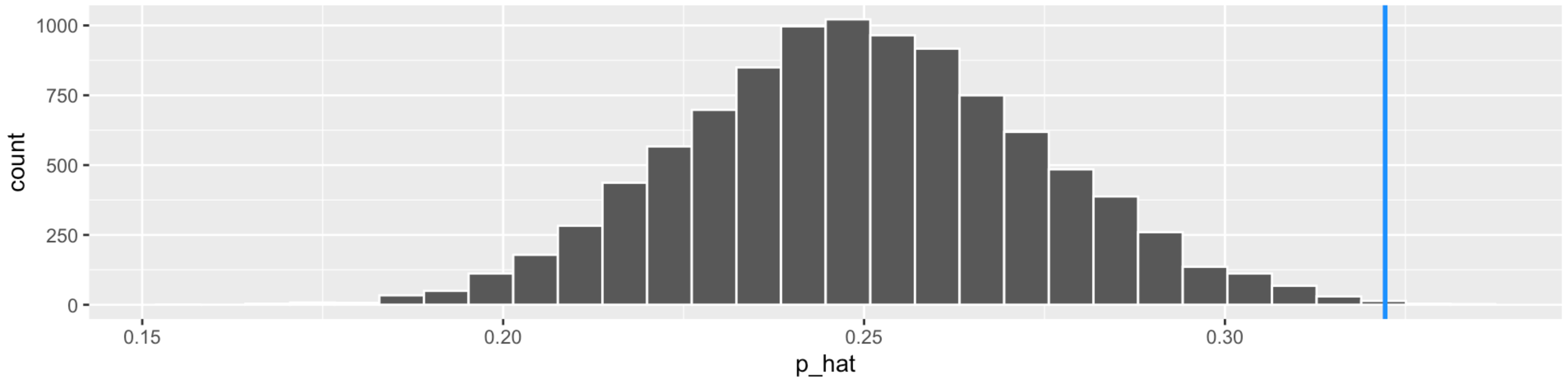
```
set.seed(123)
guesses <- c(0, 0, 0, 1) # Like a 4-sided dice with sid
null_stats <- data.frame(correct = guesses) %>%
  rep_sample_n(size = 329, replace = TRUE,
               reps = 10000) %>%
  group_by(replicate) %>%
  summarize(p_hat = mean(correct))
ggplot(null_stats, aes(x = p_hat)) + geom_histogram(col
```



Hypothesis Testing Steps: ESP

3. What is our relevant “Test Statistic”?

- $\widehat{p}$ = 106 / 329 = 0.32



4. How can we calculate the “P-value”?

- Can use probability theory to find $P(\text{Correct guesses} \geq 106) = \ldots?$
- Or our simulated null distribution

```
mean(null_stats$p_hat >= 106/329)
```

```
[1] 0.0013
```

5. How would we use the P-value to make a conclusion on the research question?

- If the p-value is “small enough”, we reject the null hypothesis of random guessing.

Hypothesis Testing: ESP

- We got a pretty small p-value (0.0013). Hooray! It is reasonable to reject the null hypothesis.
- But really, do we believe that ESP is real?

- Next lecture, we'll talk more about hypothesis testing errors.

Side Note: Reproducibility and Replicatability

- Two important words in data analysis:
  - Reproducibility
  - Replicability

- **Reproducibility**: If I give you the raw data and my write-up, you will get to the exact same final numbers that I did.
  - By using `Quarto` Documents, we are learning a **reproducible** workflow.
- **Replicability**: If you follow my study design but collect new data (i.e. repeat my study on new subjects), you will come to the same conclusions that I did.
  - Sadly, **replication** studies of Bem and Honorton's ESP trials largely failed to find evidence of ESP.

Switching gears: 1 vs 2 sided tests

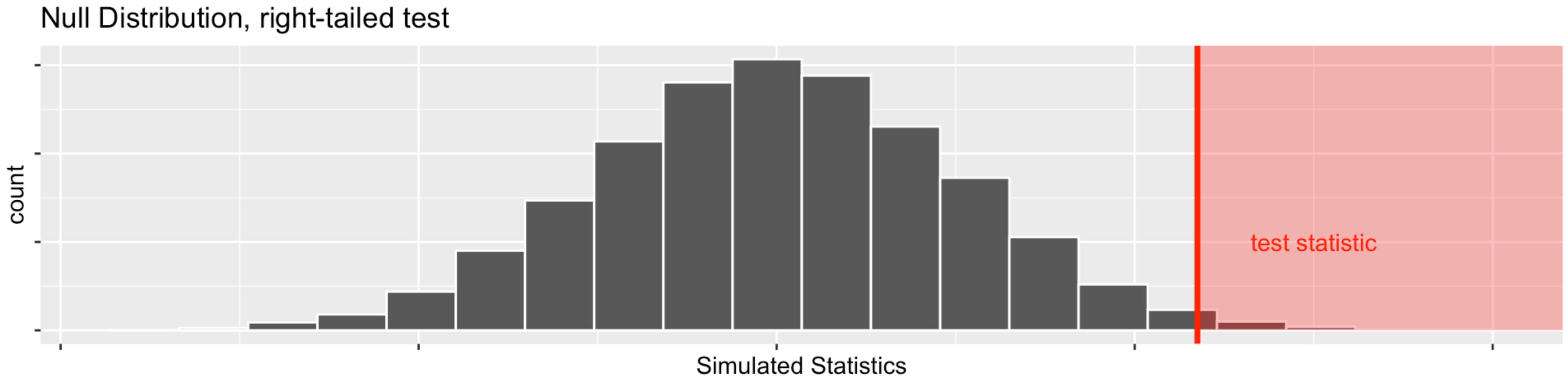
- Say we're **flipping a coin 20 times**, and (for some reason) we are worried that it's not a "fair" coin
  - i.e., The probability of heads is not 0.5
- Let $p =$ **probability of getting a heads with this coin**
- **Q**: What is the null hypothesis?
- Possible alternative hypotheses:
  - $H_a : p \leq 0.5$
  - $H_a : p \geq 0.5$
  - $H_a : p \neq 0.5$
- We've been thinking about the first two cases (1 sided tests). Let's see what changes if $H_a : p \neq 0.5$ (2 sided test).

Recall the hypothesis testing steps again:

1. Present research question and identify hypotheses (H_0 and H_a)
   - $H_a : p \leq 0.5$, or $p \geq 0.5$, or $p \neq 0.5$?
   - $H_0 : p = 0.5$
2. Describe "Null" distribution
   - **This is the same regardless of which H_a we use, depends only on H_0**
3. Obtain data, calculate relevant "Test Statistic"
4. **Calculate the "P-value"**
   - P-value = likelihood of observing the Test Statistic **or something more extreme** assuming the Null Hypothesis
5. Use the P-value to make a conclusion on the research question

1 vs 2 sided tests: what does "more extreme" mean?

- Suppose our test statistic from 20 coin flips is $\widehat{p} = 0.85$.
- If we have $H_a : p > 0.5$ (**right**-tailed test)
  - We'd want $\mathrm{P}(\widehat{p} \geq 0.85)$ under the null hypothesis as our P-value




```
mean(null_stats$p_hat >= 0.85)
```

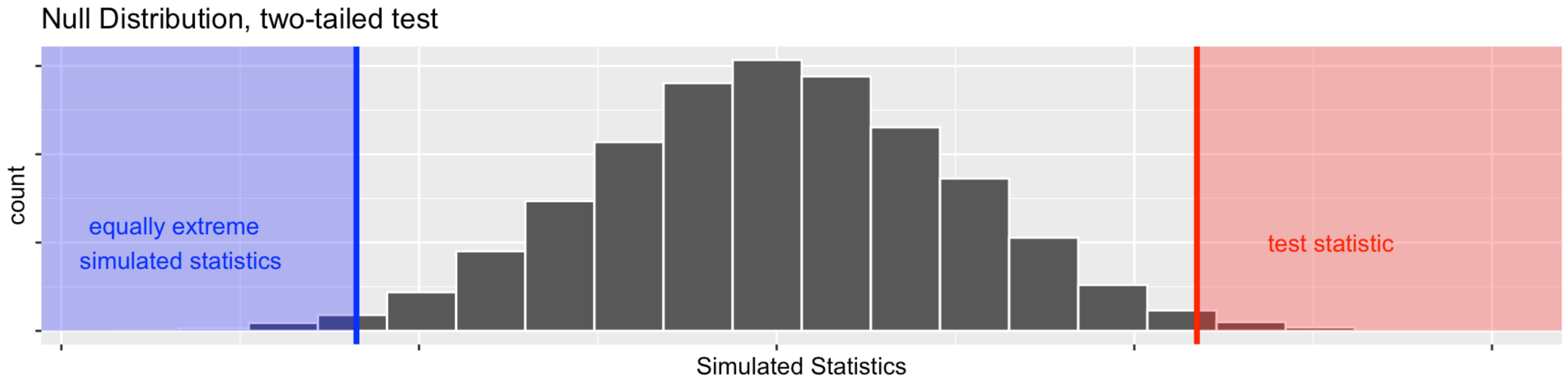
```
[1] 0.0351
```

1 vs 2 sided tests: what does "more extreme" mean?

- Suppose our test statistic from 20 coin flips is $\widehat{p} = 0.85$.
- Suppose we have $H_a : p \neq 0.5$
  - $\widehat{p} = 0.85$ is 0.35 *more* than $p = 0.5$, so we want to include $\mathrm{P}(\widehat{p} \geq 0.85)$
  - $\widehat{p} = 0.15$ is 0.35 *less* than $p = 0.5$, so we'd also want $\mathrm{P}(\widehat{p} \leq 0.15)$ (this would be more extreme in the other direction)
- So our P-value would be the sum of those two things (both tails)!

```
mean(null_stats$p_hat >= 0.85) + mean(null_stats$p_hat <= 0.15)
```

```
[1] 0.0732
```



1 vs 2 sided tests: what does "more extreme" mean?

Note the effect that our choice of H_a had on our p-values!

- Our test statistic was $\widehat{p} = 0.85$
- $H_a : p \leq 0.5$: P-value = 0.9649
- $H_a : p \geq 0.5$: P-value = 0.0351
- $H_a : p \neq 0.5$: P-value = 0.0732

Another Example

Can you tell if a mouse is in pain by looking at its facial expression? A recent study created a "mouse grimace scale" and tested to see if there was a positive correlation between scores on that scale and the degree of pain (based on injections of a weak and mildly painful solution). The study's authors believe that if the scale applies to other mammals as well, it could help veterinarians test how well painkillers and other medications work in animals.

- **Q**: Write out H_0 and H_a qualitatively (in words).
- **Q**: Write out H_0 and H_a in terms of population parameters. (*Hint: recall that we use r for correlation*)
- **Q**: Describe how we'd expect the data to behave under H_0.

Another Example!

Can a simple smile have an effect on punishment assigned following an infraction? In a 1995 study, Hecht and LeFrance examined the effect of a smile on the leniency of disciplinary action for wrongdoers. Participants in the experiment took on the role of members of a college disciplinary panel judging students accused of cheating. For each suspect, along with a description of the offense, a picture was provided with either a smile or neutral facial expression. A leniency score was calculated based on the disciplinary decisions made by the participants.

- **Q**: Write out H_0 and H_a qualitatively (in words).
- **Q**: Write out H_0 and H_a in terms of population parameters (*Hint, we can write the population mean leniency score for smilers as* μ_S).
- **Q**: Describe how we'd expect the data to behave under H_0.

Null distributions: how to generate them for these other examples?

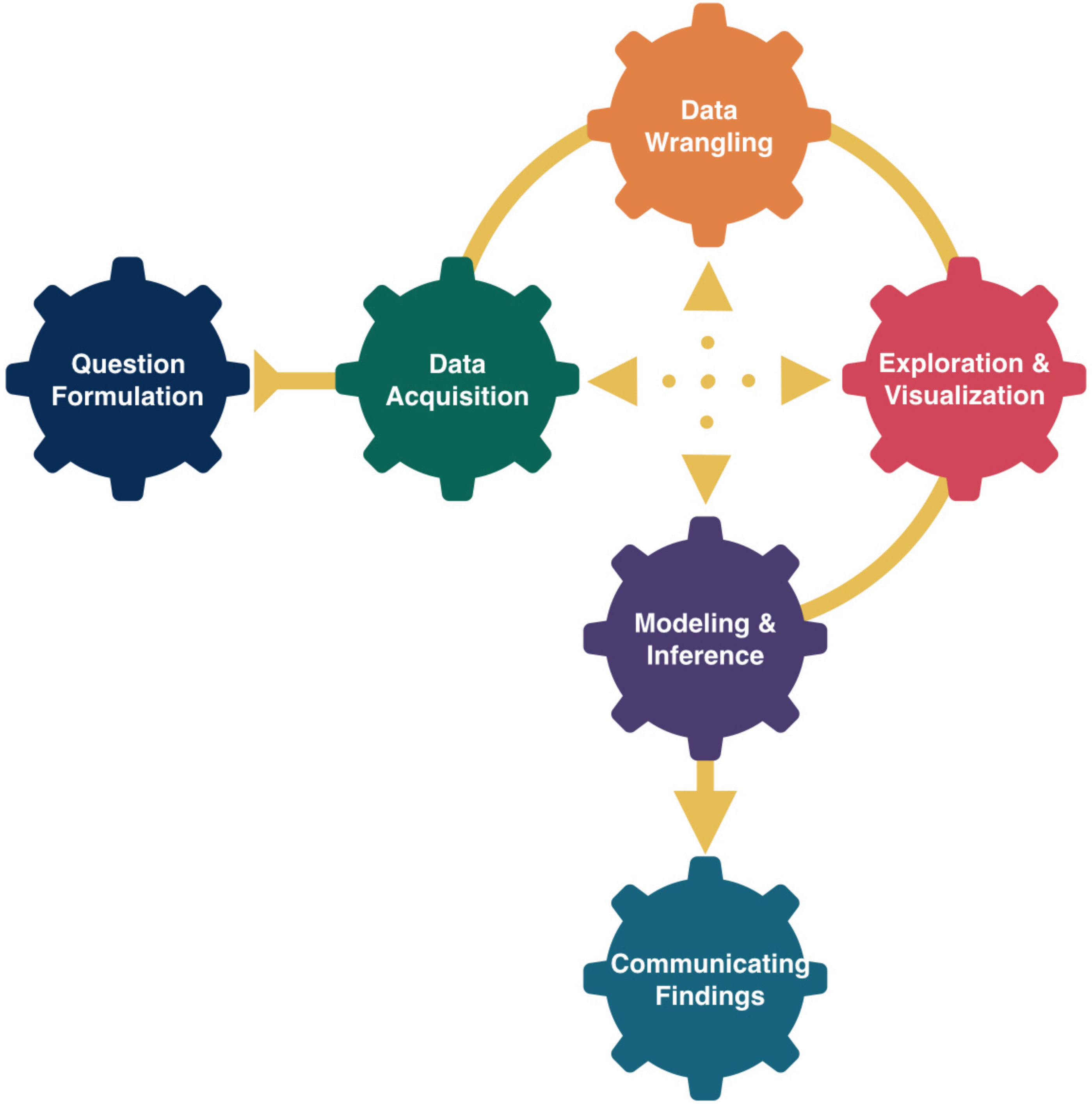
Null distributions: how to generate them for these other examples?

Next time

- Decisions in hypothesis testing

- Types of hypothesis testing errors

- Power



Hypothesis Testing III

Megan Ayers

Math 141 | Spring 2026

Wednesday, Week 8

Goals for today

- Discuss **decisions** in a hypothesis test
  - Types of errors
- Discuss and learn to compute the **power** of a hypothesis test

Hypothesis Testing Framework

Have two competing hypotheses:

- Null Hypothesis (H_0): "Dull" hypothesis, status quo, random chance, no effect…
- Alternative Hypothesis (H_a): The researchers' conjecture.

Must first take those hypotheses and translate them into statements about the **population parameters** so that we can test them with sample data.

Example:

H_0: ESP doesn't exist.

H_a: ESP does exist.

Then translate into a statistical problem!

Q: Using formal statistical/mathematical notation, how should we define these?

$p =$

H_0:

H_a:

New Example: Swimming with Dolphins

In 2005, researchers Antonioli and Reveley posed the question "Does swimming with the dolphins help depression?" They recruited 30 US subjects diagnosed with mild to moderate depression. Participants were randomly assigned to either the treatment group (swimming with dolphins) or the control group (swimming without dolphins). After two weeks, each subject was categorized as "showed substantial improvement" or "did not show substantial improvement".

Here's a contingency table of `improve` and `group`.

```r
dolphins %>%
  count(group, improve)
```

```
      group improve  n
1   Control      no 12
2   Control     yes  3
3 Treatment      no  5
4 Treatment     yes 10
```

H_0:

H_a:

Snapshot of the data:

```
      group improve
1   Control     yes
2 Treatment      no
3   Control      no
4 Treatment     yes
5   Control      no
6   Control      no
7 Treatment     yes
8   Control      no
```

Q: How might we generate the null distribution for this scenario?

Penguins Example

Let's return to the `penguins` data and ask if flipper length varies, on average, by the sex of the penguin.

Research Question: Does flipper length differ by sex?

Response Variable:

Explanatory Variable:

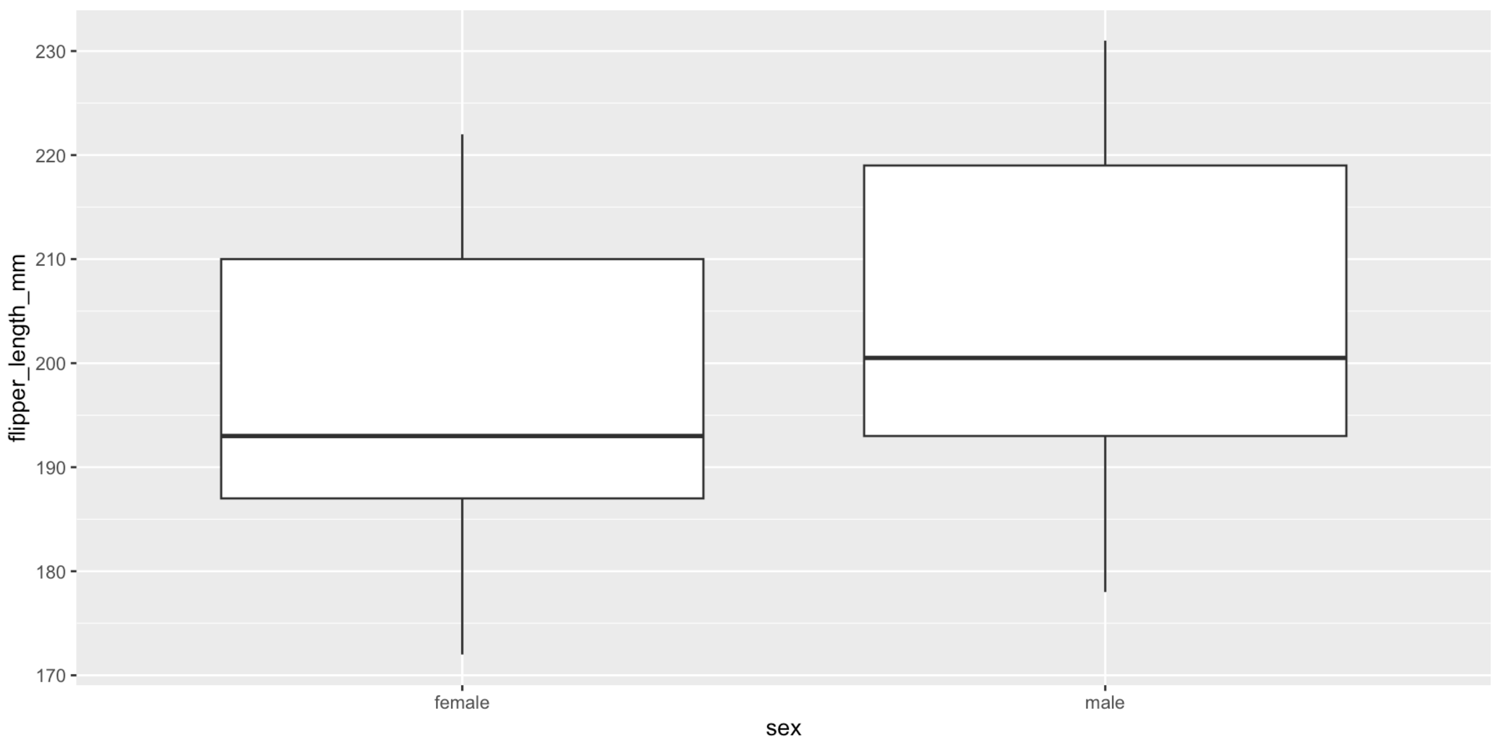
Statistical Hypotheses:

Exploratory Data Analysis

```r
library(palmerpenguins)

penguins <- palmerpenguins::penguins

penguins %>%
  drop_na(sex) %>%
ggplot(mapping = aes(x = sex,
                     y = flipper_length_mm)) +
  geom_boxplot()
```



Two-Sided Hypothesis Test

Compute observed test statistic:

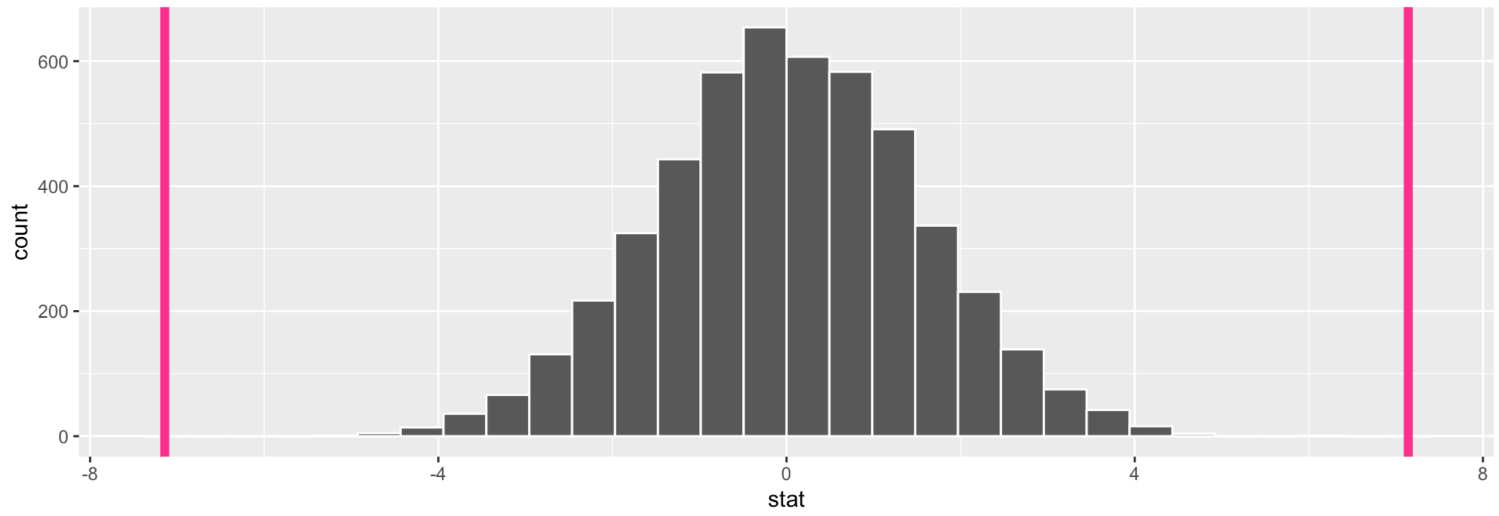
```r
penguins %>% drop_na(sex) %>%
  group_by(sex) %>%
  summarize(avg_length = mean(flipper_length_mm))
```

```
# A tibble: 2 × 2
  sex    avg_length
  <fct>       <dbl>
1 female       197.
2 male         205.
```

→ Our test statistic is: 7.1424

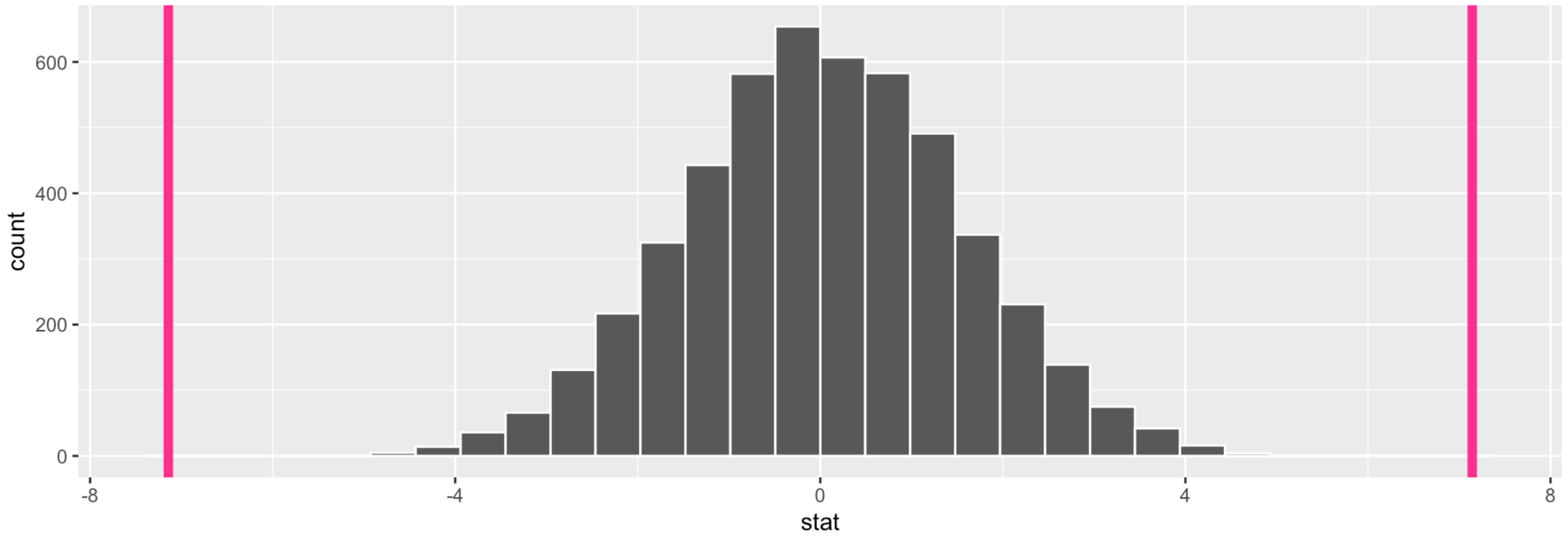
Generate the null distribution by simulating many data sets where gender is shuffled (code not shown for brevity), and visualize null sampling distribution compared to our test statistic:

```r
ggplot(peng_null_stats, aes(x = stat)) +
  geom_histogram(color = "white") +
  geom_vline(xintercept = abs(test_stat),
             color = "deeppink", size = 2) +
  geom_vline(xintercept = -abs(test_stat),
             color = "deeppink", size = 2)
```



Q: Guesses for p-value?

Two-Sided Hypothesis Test



Calculate the p-value for a two-sided test:

```
# Compute p-value
(p_val <- mean(peng_null_stats$stat > abs(test_stat)) +
   mean(peng_null_stats$stat < -abs(test_stat)))
```

```
[1] 0
```

Now: A closer look at the conclusions of a hypothesis test

Hypothesis Testing: Decisions, Decisions

Once you get to the end of a hypothesis test you make one of two decisions:

- P-value is small.
  - I have evidence for H_a. Reject H_0.
- P-value is not small.
  - I don't have evidence for H_a. Fail to reject H_0.

Sometimes we make the correct decision. Sometimes we make a mistake.

Hypothesis Testing: Decisions, Decisions

Let's create a table of potential outcomes on the board.

α = prob of Type I error **under repeated sampling** = prob reject H_0 when it is true

β = prob of Type II error **under repeated sampling** = prob fail to reject H_0 when H_a is true.

Hypothesis Testing: Decisions, Decisions

We should set α level beforehand.

Use α to determine "small" for a p-value.

- P-value ~~is small~~ $< \alpha$.
  - $\rightarrow$ I have evidence for H_a. Reject H_0.
- P-value ~~is not small~~ $\geq \alpha$.
  - $\rightarrow$ I don't have evidence for H_a. Fail to reject H_0.

Hypothesis Testing: Decisions, Decisions

Open Question: How do I select α?

- Will depend on the convention in your field (0.05 is common).
- Want a small α and a small β. But they are related.
  - **The smaller** α is the larger β will be.
- Choose a lower α (e.g., 0.01, 0.001) when the Type I error is worse and a higher α (e.g., 0.1) when the Type II error is worse.

- **Q**: What are some examples of when Type I errors are worse than Type II errors? Examples when the opposite is true?

Hypothesis Testing: Decisions, Decisions

Open Question: How do I select α?

- Will depend on the convention in your field (0.05 is common).
- Want a small α and a small β. But they are related.
  - **The smaller** α is the larger β will be.

- **Q**: Can't easily compute β (probability of failing to reject a false null hypothesis). Why?
- Important related concept:
  - **Power** = probability reject H_0 when the alternative is true.

- **Q**: Why is power important when designing an experiment?

Example

Suppose we want to test whether someone can detect AI-generated text from real text. We show a participant 10 short passages that are each either written by a human or an AI agent, and ask them to identify which are written by AI. Suppose the participant's true (long-run) detection rate is **70%**.

H_0:

H_a:

Example

Suppose we want to test whether someone can detect AI-generated text from real text. We show a participant ~~10~~ 50 short passages that are each either written by a human or an AI agent, and ask them to identify which are written by AI. Suppose the participant's true (long-run) detection rate is **70%**.

Example

Suppose we want to test whether someone can detect AI-generated text from real text. We show a participant ~~10~~ 50 short passages that are each either written by a human or an AI agent, and ask them to identify which are written by AI. Suppose the participant's true (long-run) detection rate is **70%**.

Example

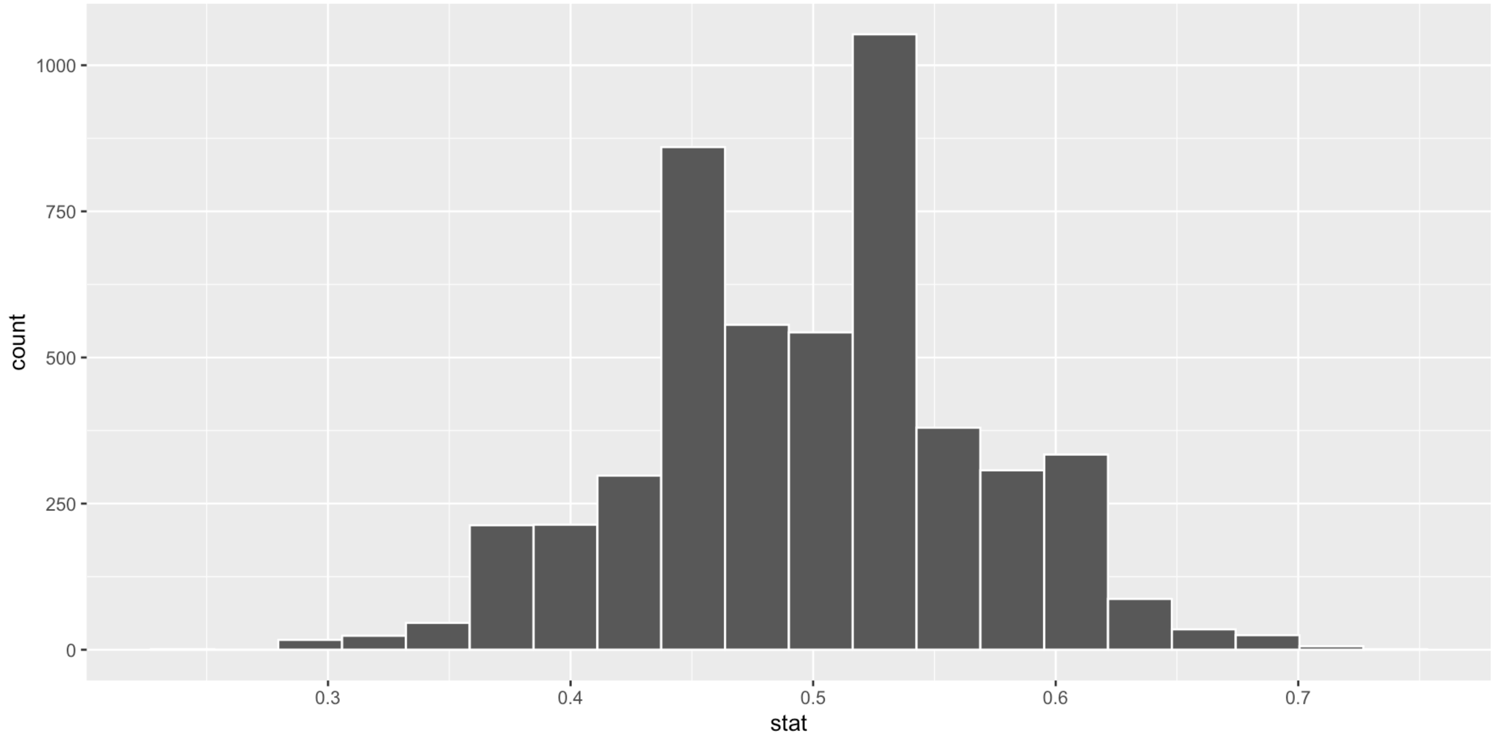
Suppose we want to test whether someone can detect AI-generated text from real text. We show a participant ~~10~~ 50 short passages that are each either written by a human or an AI agent, and ask them to identify which are written by AI. Suppose the participant's true (long-run) detection rate is ~~70%~~ 80%.

Computing Power

1. Generate a null distribution:

```r
# Generate null distribution
null_stats <- data.frame(correct = c(0, 1)) %>%
  rep_sample_n(size = 50, replace = TRUE,
               reps = 5000) %>%
  group_by(replicate) %>%
  summarize(stat = mean(correct))

ggplot(data = null_stats, mapping = aes(x = stat)) +
  geom_histogram(bins = 20, color = "white")
```



Computing Power

2. Determine the "critical value(s)" where $\alpha = 0.05$ (careful with discrete distributions).

```r
quantile(null_stats$stat, 0.95)
```

```
 95% 
0.62 
```

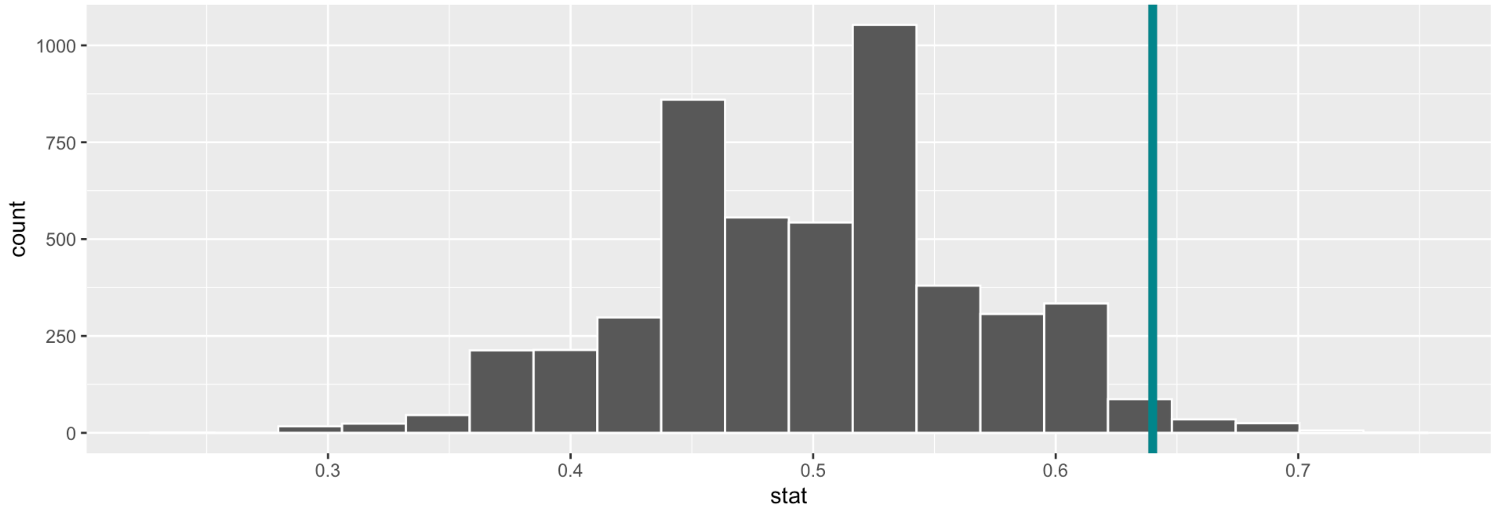
```r
mean(null_stats$stat >= 0.62)  # 31/50
```

```
[1] 0.0518
```

```r
mean(null_stats$stat >= 0.64)  # 32/50
```

```
[1] 0.0246
```

```r
ggplot(data = null_stats, mapping = aes(x = stat)) +
  geom_histogram(bins = 20, color = "white") +
  geom_vline(xintercept = 0.64, size = 2,
             color = "turquoise4")
```

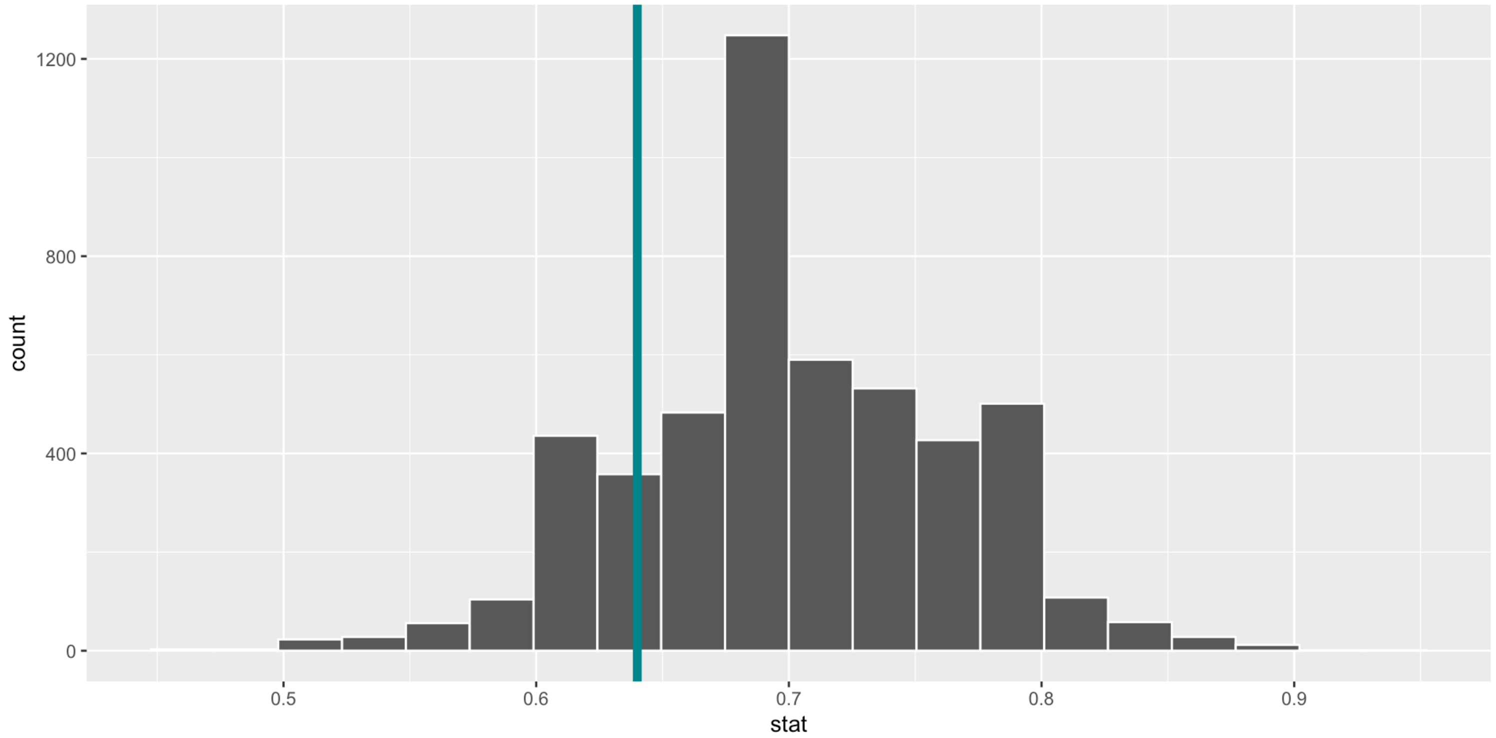


Computing Power

3. Construct the alternative distribution. **This usually requires a leap of faith - we need to assume a parameter under the alternative hypothesis!**
    - Draw on any existing evidence (ex. a pilot study) to justify your assumption

```r
alt_stats <- data.frame(correct = rep(c(0, 1),
                        times = c(3, 7))) %>%
  rep_sample_n(size = 50, replace = TRUE,
               reps = 5000) %>%
  group_by(replicate) %>%
  summarize(stat = mean(correct))

ggplot(data = alt_stats, mapping = aes(x = stat)) +
  geom_histogram(bins = 20, color = "white") +
  geom_vline(xintercept = 0.64,
             size = 2, color = "turquoise4")
```



Computing Power

4. Find the probability of the critical value or more extreme under the **alternative distribution**.

```
alt_stats %>%
  summarize(power = mean(stat >= 0.64))
```

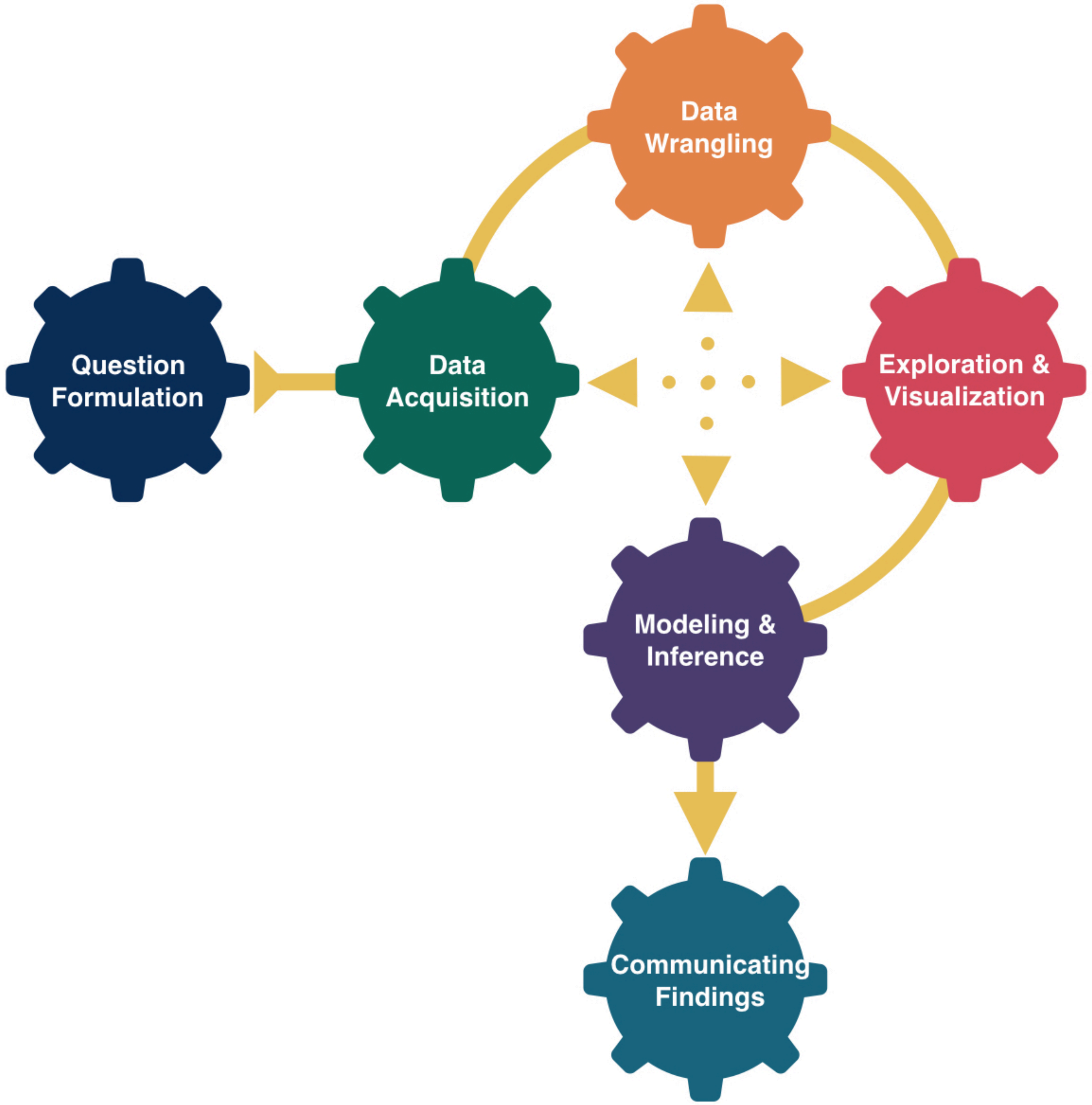
```
# A tibble: 1 × 1
  power
  <dbl>
1 0.849
```

Thoughts on Power

- **Q**: We saw how α, n, and the **effect size** affected power. What aspects of the test do we actually have control over?
- **Q**: Why is it easier to set α than to set β or power?
- Although it can be challenging, considering power before collecting data is very important!
- Under-powered studies carry large risks
  - Grants often require power analyses to justify funding
  - Type II errors can be dangerous depending on the setting

Next time:

- p-value pitfalls



P-Value Pitfalls

Megan Ayers

Math 141 | Spring 2026

Friday, Week 8

Reminders

- Assignment deadlines will resume per usual after the break
  - Lab from yesterday: Due Tuesday March 31st
  - Midterm revisions: Due Wednesday April 1st
  - HW 7: Due Friday April 3rd

Goals for today

- A hearty p-values discussion
- Zoom out on statistical inference so far
- Motivate theory-based inference

Let's Talk About P-values

- The original intention of the p-value was as an **informal** measure to judge whether or not a researcher should take a second look.
- But to create simple statistical manuals for practitioners, this informal measure quickly became a rule: **"p-value < 0.05" = "statistically significant"**.

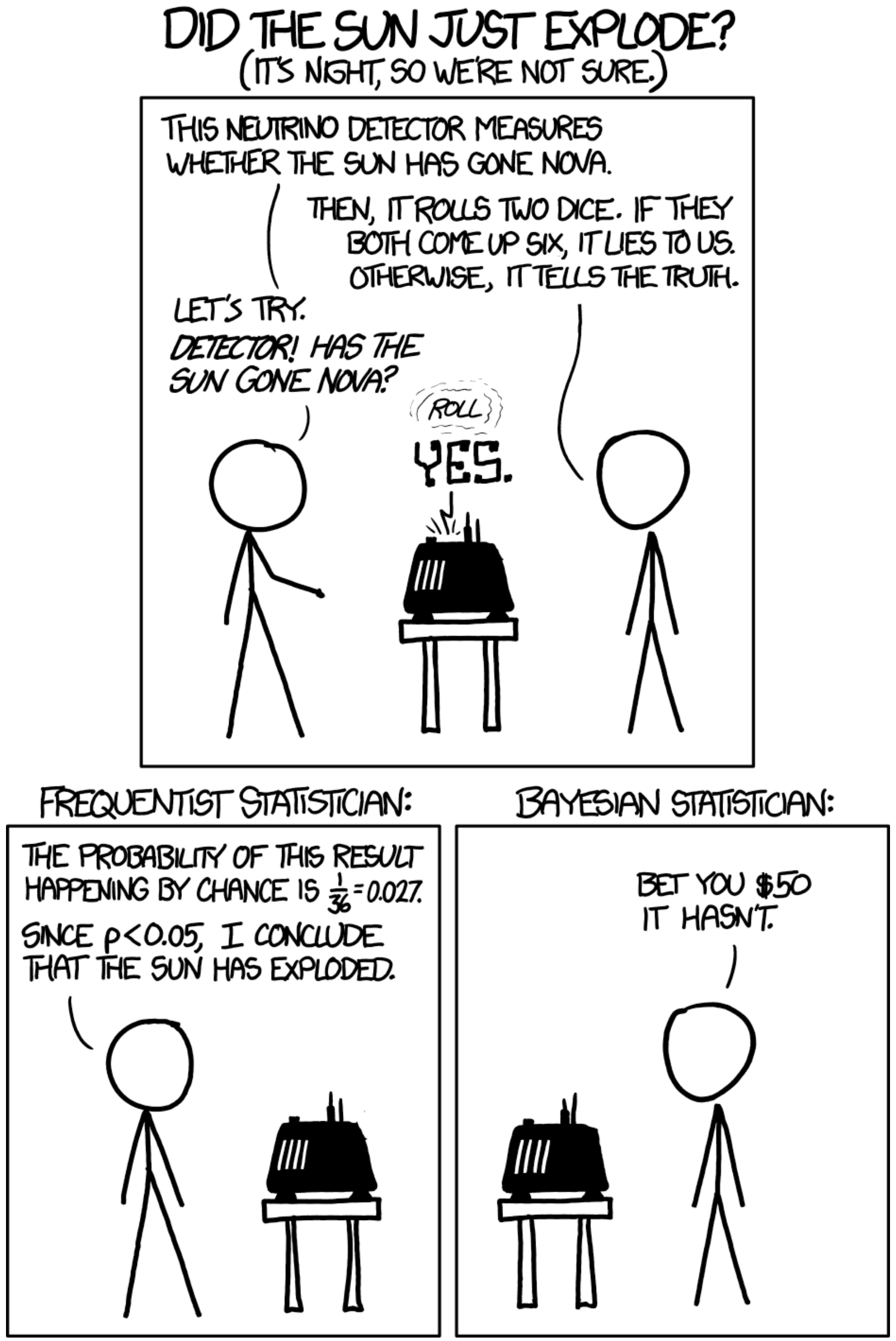
What were/are the consequences of the "p-value < 0.05" = "statistically significant" rule?

Let's Talk About P-values

- **A consequence**: The p-value is often misinterpreted to be the probability the null hypothesis is true.
  - A p-value of 0.003 does not mean there's a 0.3% chance that ESP doesn't exist!
  - By giving people a simple rule, they never learned what the p-value actually measures.

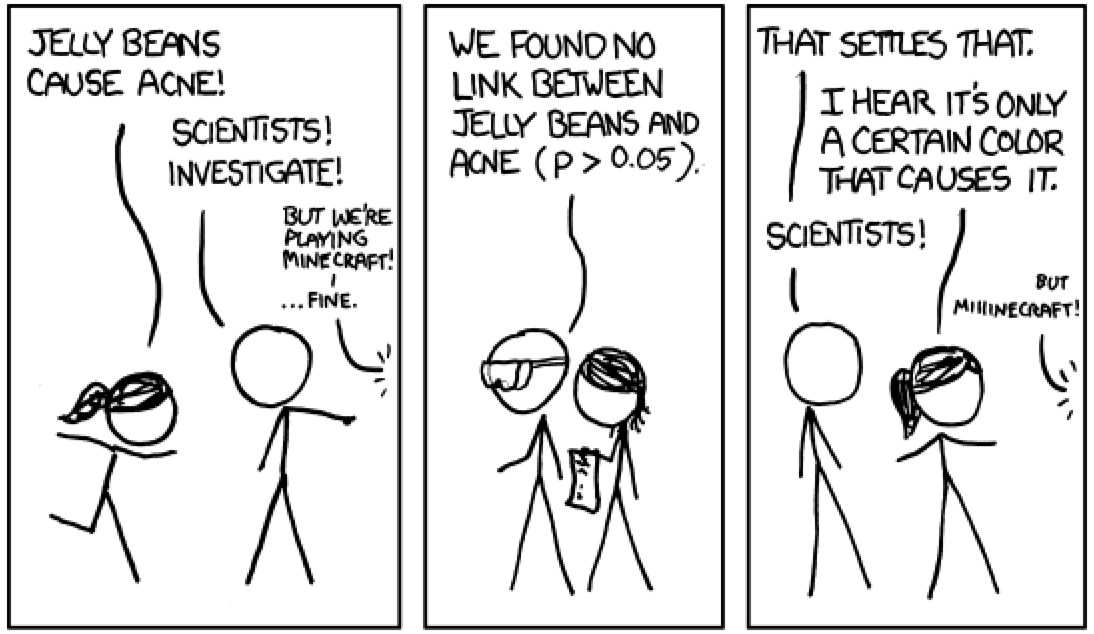
Let's Talk About P-values

- **A consequence**: Researchers often put too much weight on the p-value and not enough weight on their domain knowledge/the plausibility of their conjecture.
- **Read and discuss this xkcd comic with your neighbors**



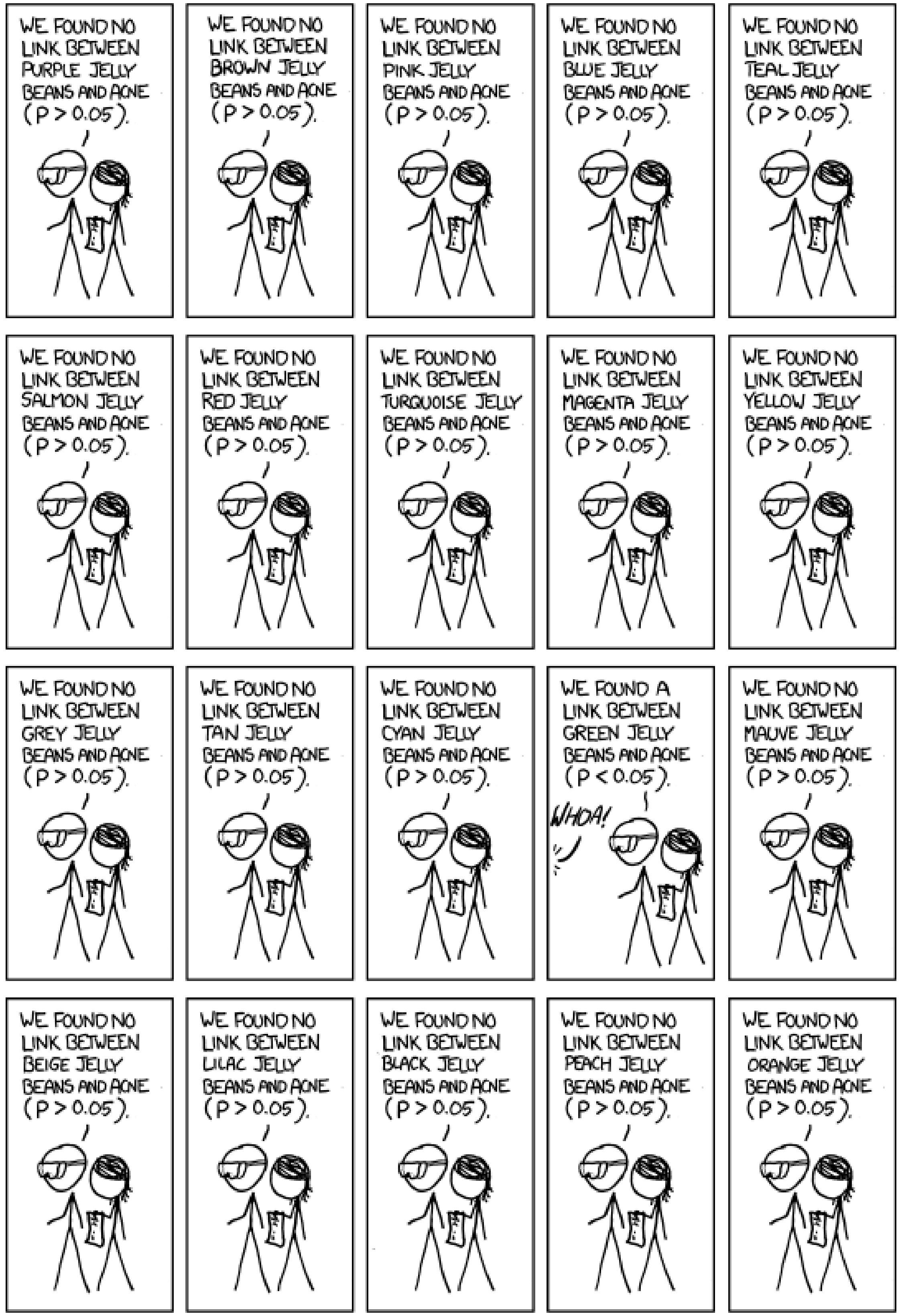
Let's Talk About P-values

- A consequence: P-hacking: Cherry-picking promising findings that are beyond this arbitrary threshold.
- xkcd comic



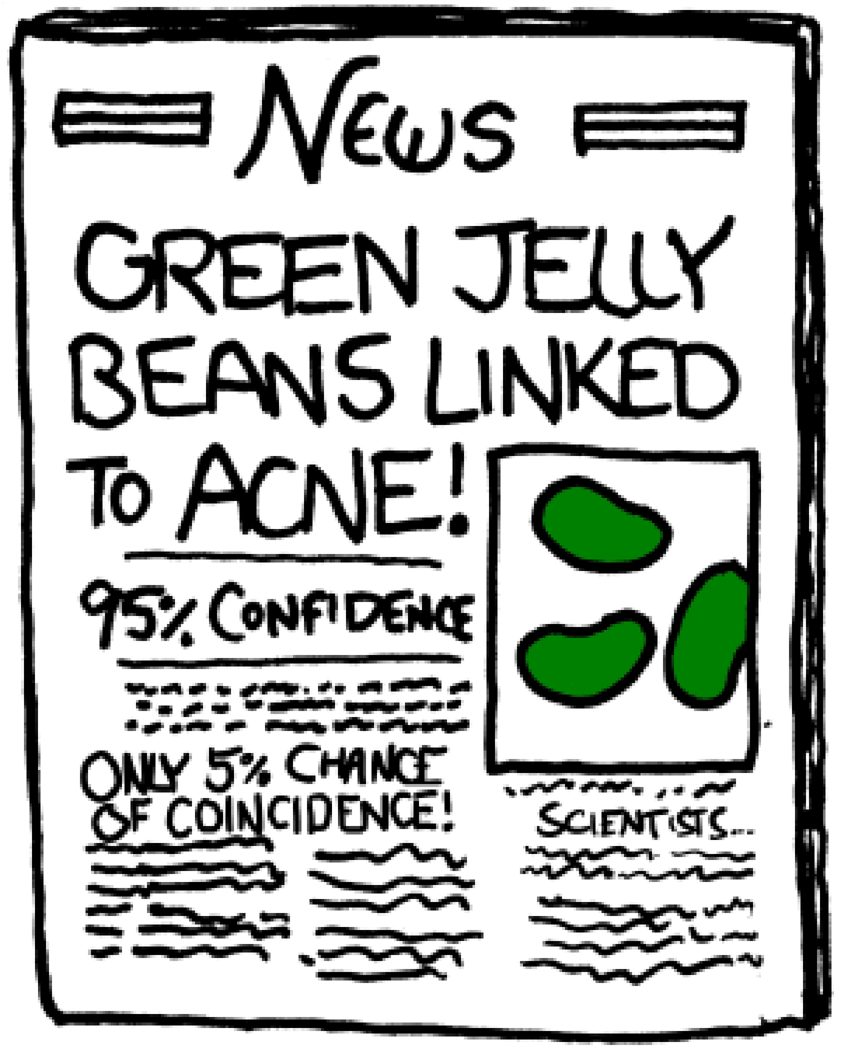
Let's Talk About P-values

- A consequence: P-hacking: Cherry-picking promising findings that are beyond this arbitrary threshold.
- xkcd comic



Let's Talk About P-values

- A consequence: P-hacking: Cherry-picking promising findings that are beyond this arbitrary threshold.
- xkcd comic



- Q: If we run 100 hypothesis tests where all null hypotheses are true, how many do we expect to have a P-value less than 0.05 due to chance?
  - This is also known as the multiple comparisons problem

Where does p-hacking come from?

- Can be an unintentional consequence of data analysis's "garden of forking paths"
- Also partly a byproduct of publication bias/incentives in academia

Moving forward from p-hacking

- Distinguish the purpose of your analysis: **confirmatory vs exploratory**
  - **Confirmatory** analyses: Pre-register your hypotheses and planned analysis before conducting your study
  - **Exploratory** analyses: Less pressure on "statistical significance", more focus on preliminary evidence that requires a confirmatory follow-up
- **Always present statistical context behind your conclusions**: confidence intervals, p-values, analysis decisions, and assumptions
- **Multiple testing corrections**
  - Informally, these inflate your p-values to account for performing many tests
  - Ex. If I perform n hypothesis tests, I change my rejection threshold to α/n (Bonferroni correction)

Let's Talk About P-values

- **Another consequence of p-values**: People conflate *statistical significance* with *practical significance*.

Example: A *Nature* study of 19,000+ recently married people found that those who meet their spouses online…

- Are less likely to divorce (p-value < 0.002)
- Are more likely to have high marital satisfaction (p-value < 0.001)
- BUT the estimated **effect sizes** were tiny.
  - Divorce rate of 5.96% for those who met online versus 7.67% for those who met in-person.
  - On a 7 point scale, happiness value of 5.64 for those who met online versus 5.48 for those who met in-person.

Q: Do these results provide compelling evidence that one should change their dating behavior?

Let's Talk About P-values

The American Statistical Association created a set of principles to address misconceptions and misuse of p-values:

1. P-values can indicate how incompatible the data are with a specified statistical model.
2. P-values do not measure the probability that the studied hypothesis is true.
3. Scientific conclusions and business or policy decisions should not be based only on whether or not a p-value passes a specific threshold (i.e. 0.05).
4. Proper inference requires full reporting and transparency.
5. A p-value, or statistical significance, does not measure the size of an effect or the importance of a result.
6. By itself, a p-value does not provide a good measure of evidence regarding a model or hypothesis.

Let's Talk About P-values

- Despite its issues, p-values are still quite popular and can still be a useful tool **when used properly**.
- In 2014, George Cobb a professor from Mount Holyoke College posed the following questions (and answers):

Q: Why do so many colleges and grad schools teach $p = 0.05$?
A: Because that's still what the scientific community and journal editors use.
Q: Why do so many people still use $p = 0.05$?
A: Because that's what they were taught in college or grad school.

- We can break the cycle!

Math 141 & P-Values

- Understanding p-values and being able to **interpret a p-value in context** is a learning objective of Math 141.
  - Ex: If ESP doesn't exist, the probability of guessing correctly on at least 106 out of 329 trials is 0.003.
- Understanding that a small p-value means we have **some evidence for H_a** is important.
  - Ex: Because the p-value is small, we have evidence for ESP.
- Understanding that a small p-value alone does not imply practical significance.
  - Create a confidence interval to contextualize the effect size!
- Understanding that what you mean by **small** should depend on your field and whether a Type I Error or Type II Error is worse for **your particular research question**.
- Your ability to tell if a # is less than 0.05 is not a learning objective for Math 141.

Inference: The big picture

Inference: The Big Picture

- **Statistical inference** is the process of drawing conclusions about a population based on sample data.
- **Q**: How do point estimates and confidence intervals help us make inferences?
- **Q**: How does the hypothesis testing framework help us make inferences?

Inference: The Big Picture

- Working with samples provides a window into the population, but we can't see it all
- **Overarching goal**: distinguish between results that arise due to chance vs results that reflect a real, systematic pattern in the population
  - This led us to focus on **sampling distributions**, which we approximate using **bootstrap distributions** and **resampling methods**
  - We used these to accomplish two tasks:
    - **Generate point + interval estimates**: Our best guess + range of plausible values for a parameter
    - **Perform hypothesis tests**: A scientific method for maintaining or rejecting a null hypothesis in favor of an alternative hypothesis

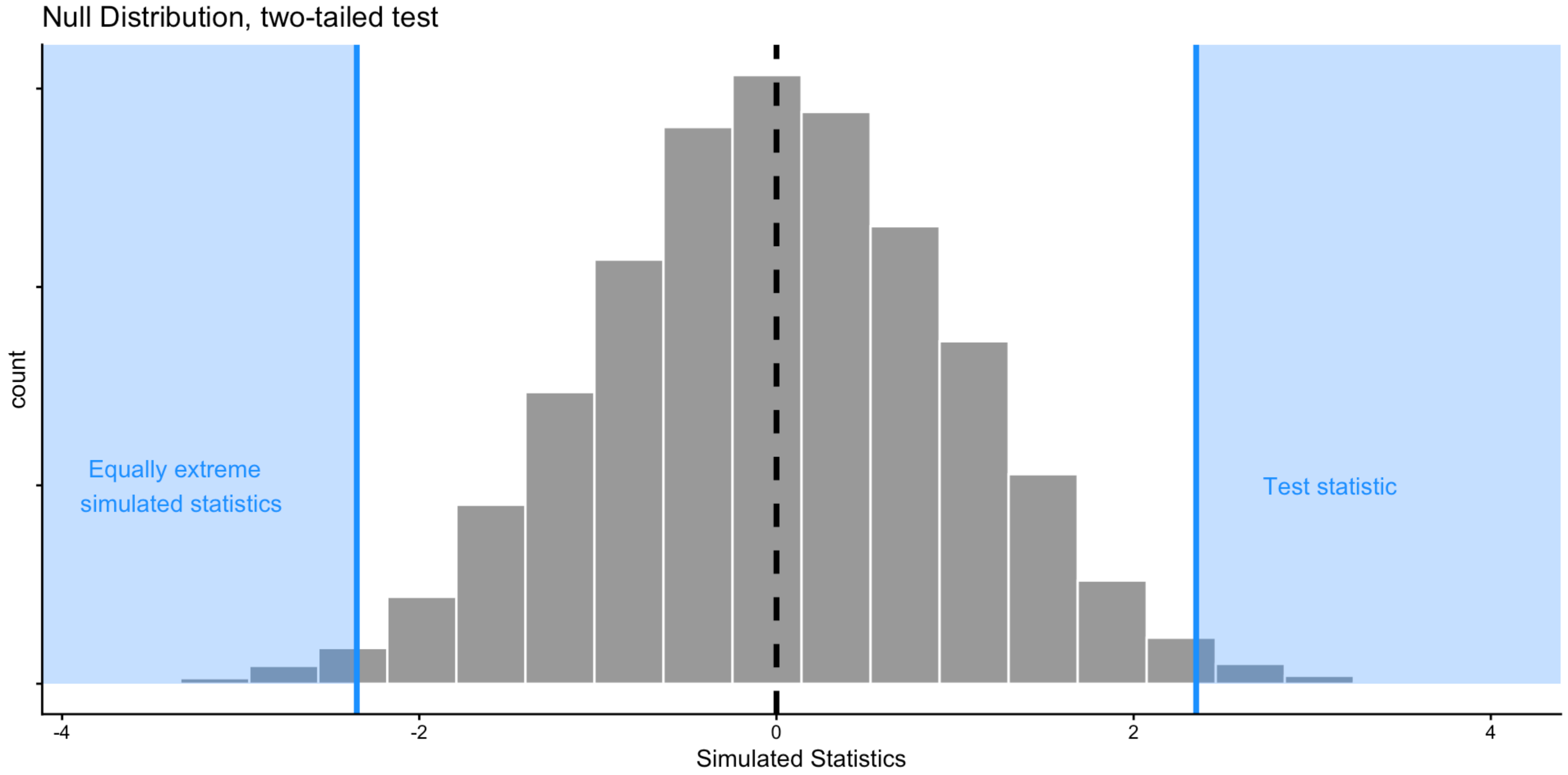
Inference: The Big Picture

- **Point + interval estimates**: Our best guess + range of plausible values for a parameter
- **Hypothesis testing**: A scientific method for maintaining or rejecting a null hypothesis in favor of an alternative hypothesis
- These two approaches are two sides of the same coin!

Inference: The Big Picture

- These two approaches are two sides of the same coin!
- Suppose we have $H_0 : \mu = 0$, and observe a test statistic
- Performing **2-sided** hypothesis test with $\alpha = 0.05$ is equivalent to computing a 95% confidence interval (CI) from our data and seeing if 0 falls inside it.

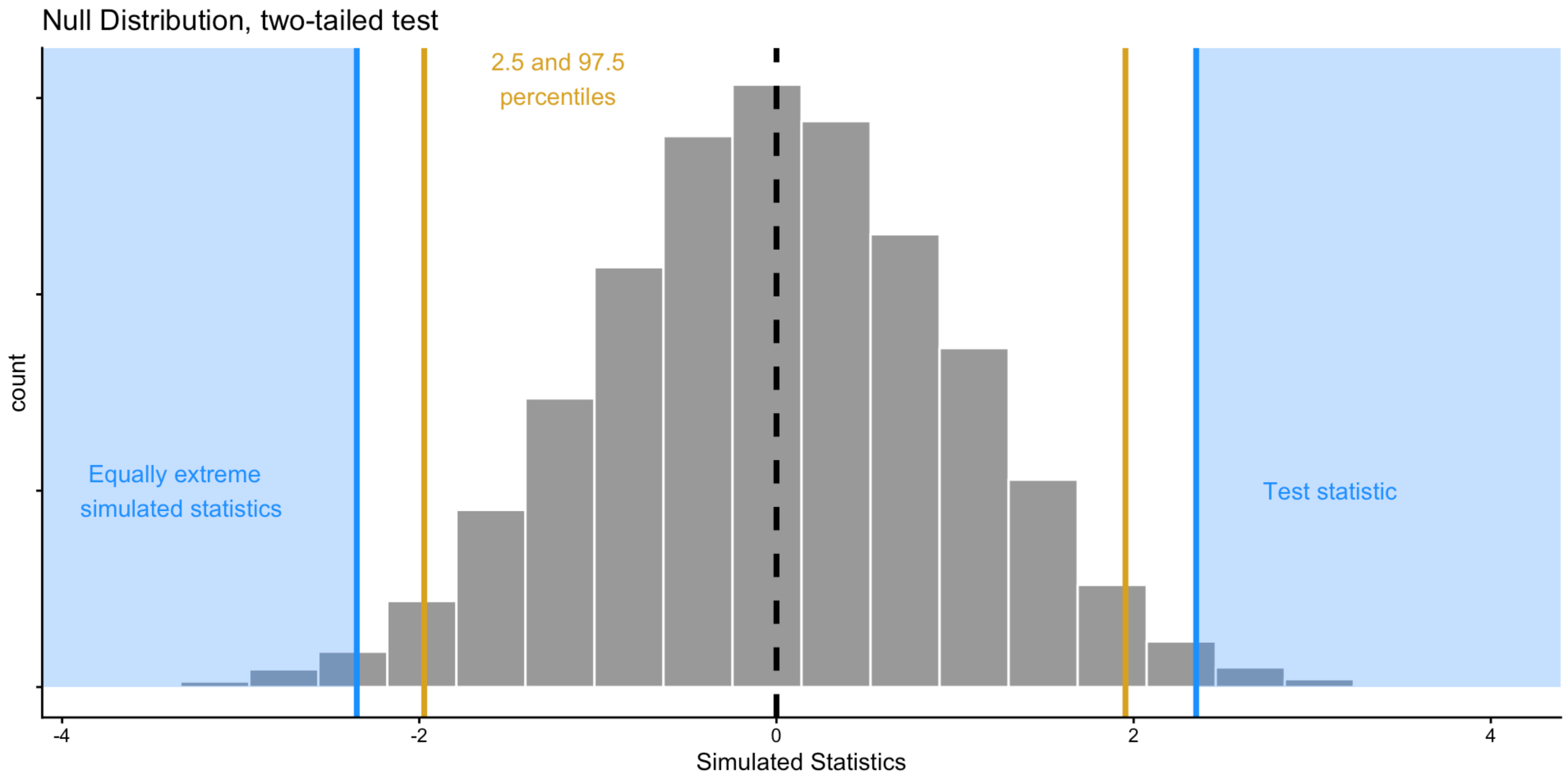
- For any value x_{in} in CI, we **would not** reject a null hypothesis $H_0 : \mu = x_{in}$ given our data.
- For any value x_{out} outside of CI, we **would** reject a null hypothesis $H_0 : \mu = x_{out}$.



Inference: The Big Picture

- These two approaches are two sides of the same coin!
- Suppose we have $H_0 : \mu = 0$, and observe a test statistic
- Performing **2-sided** hypothesis test with $\alpha = 0.05$ is equivalent to computing a 95% confidence interval (CI) from our data and seeing if 0 falls inside it.

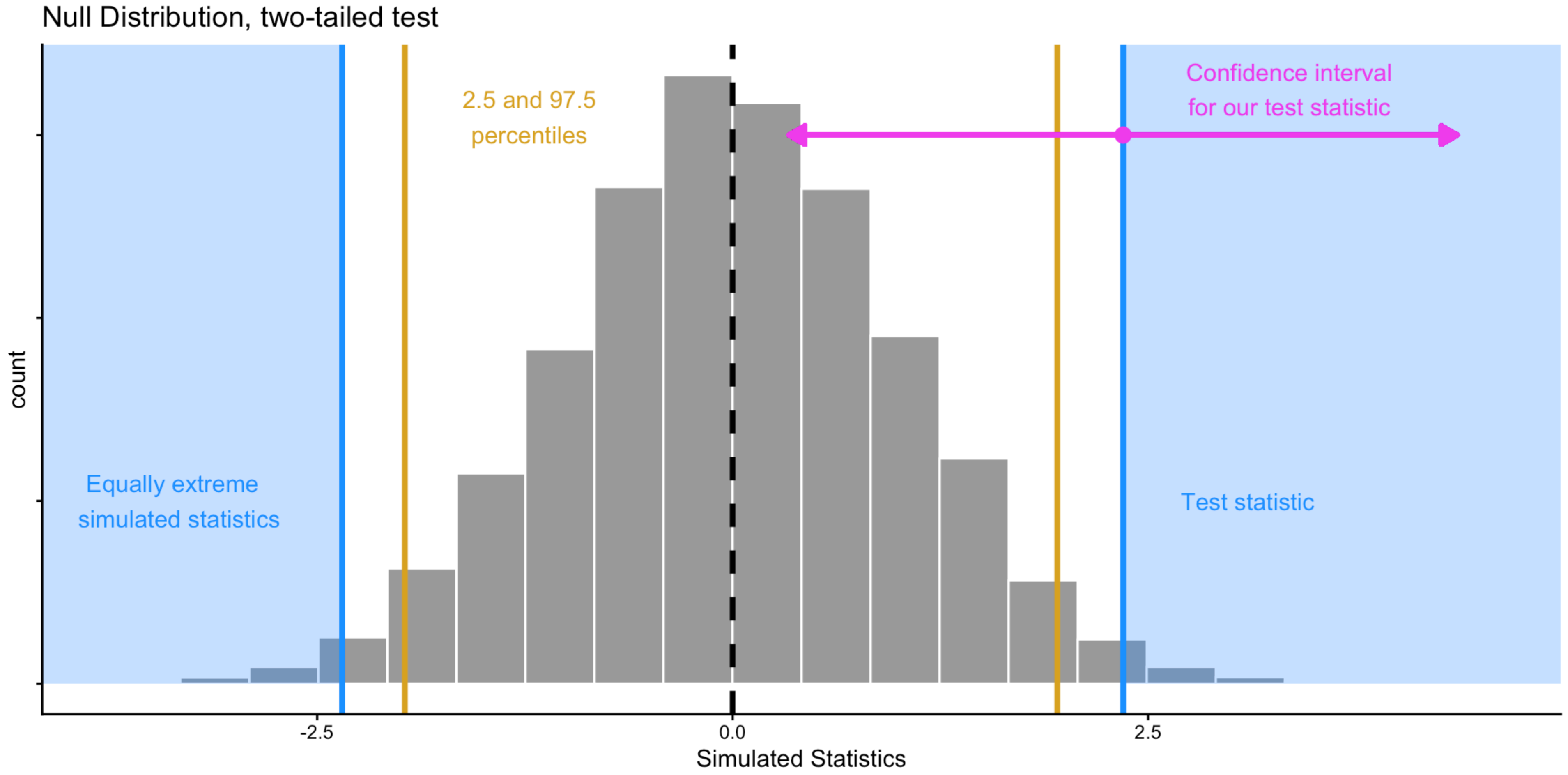
- For any value x_{in} in CI, we **would not** reject a null hypothesis $H_0 : \mu = x_{in}$ given our data.
- For any value x_{out} outside of CI, we **would** reject a null hypothesis $H_0 : \mu = x_{out}$.



Inference: The Big Picture

- These two approaches are two sides of the same coin!
- Suppose we have $H_0 : \mu = 0$, and observe a test statistic
- Performing **2-sided** hypothesis test with $\alpha = 0.05$ is equivalent to computing a 95% confidence interval (CI) from our data and seeing if 0 falls inside it.

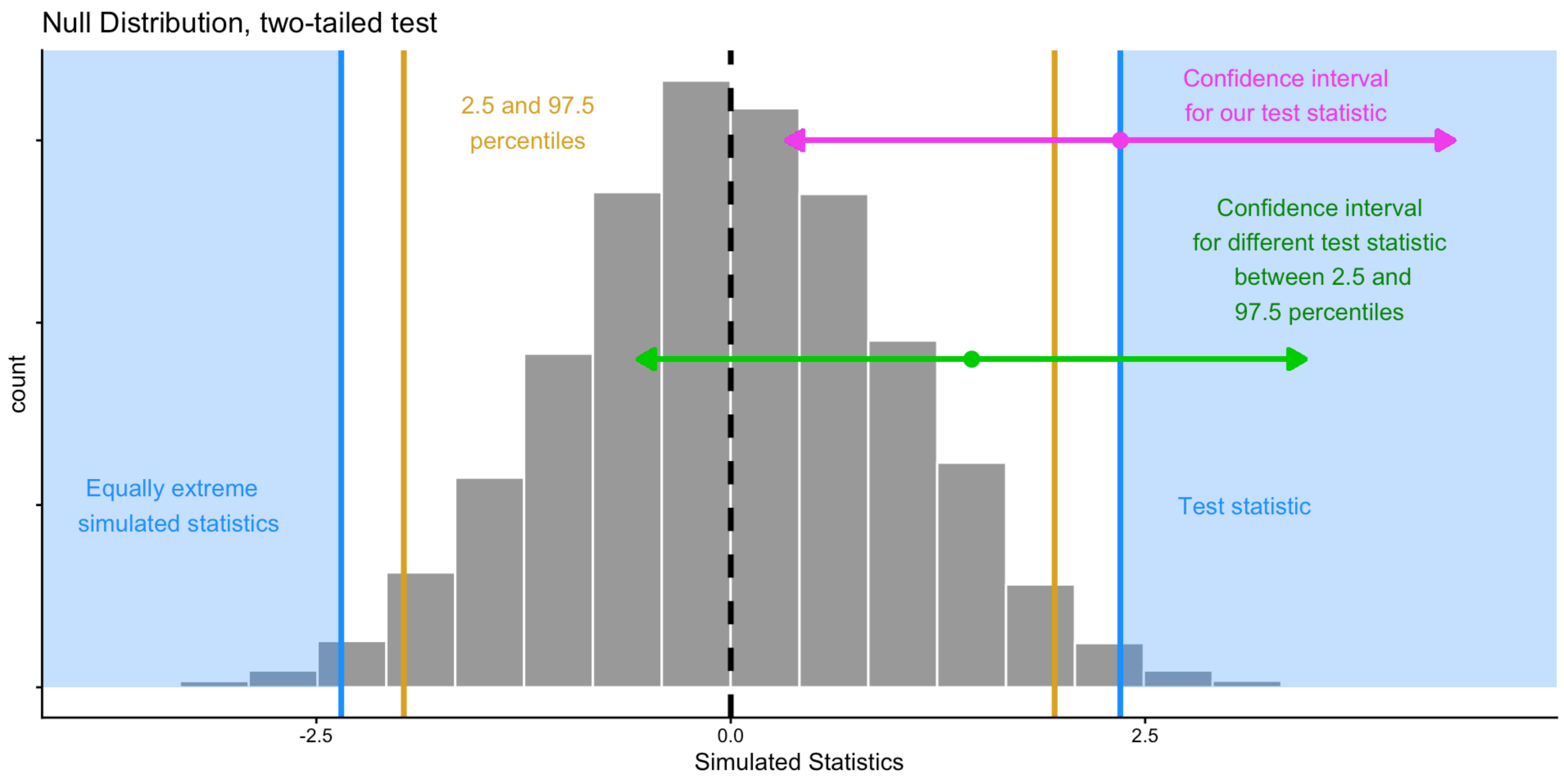
- For any value x_{in} in CI, we **would not** reject a null hypothesis $H_0 : \mu = x_{in}$ given our data.
- For any value x_{out} outside of CI, we **would** reject a null hypothesis $H_0 : \mu = x_{out}$.



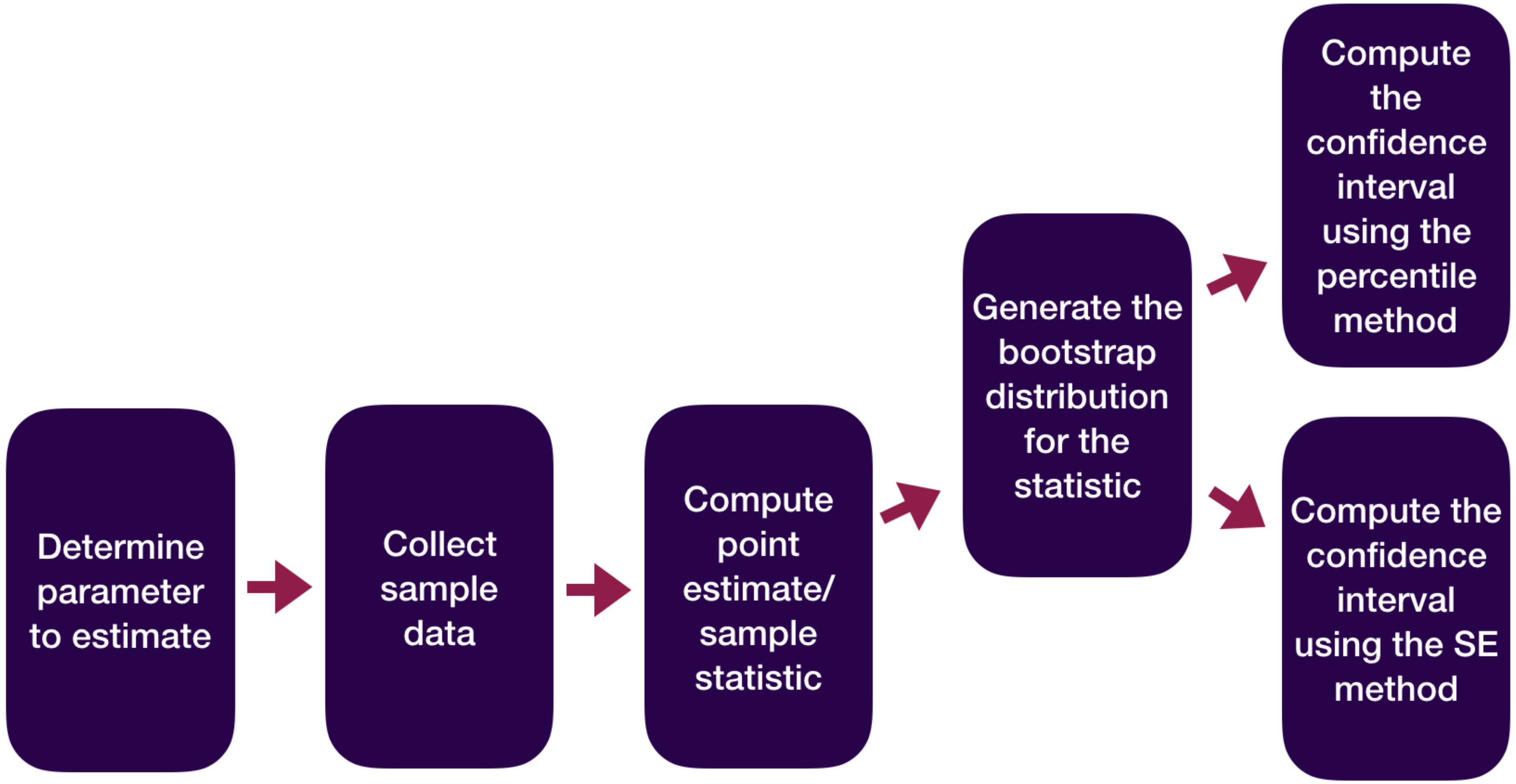
Inference: The Big Picture

- These two approaches are two sides of the same coin!
- Suppose we have $H_0 : \mu = 0$, and observe a test statistic
- Performing **2-sided** hypothesis test with $\alpha = 0.05$ is equivalent to computing a 95% confidence interval (CI) from our data and seeing if 0 falls inside it.

- For any value x_{in} in CI, we **would not** reject a null hypothesis $H_0 : \mu = x_{in}$ given our data.
- For any value x_{out} outside of CI, we **would** reject a null hypothesis $H_0 : \mu = x_{out}$.

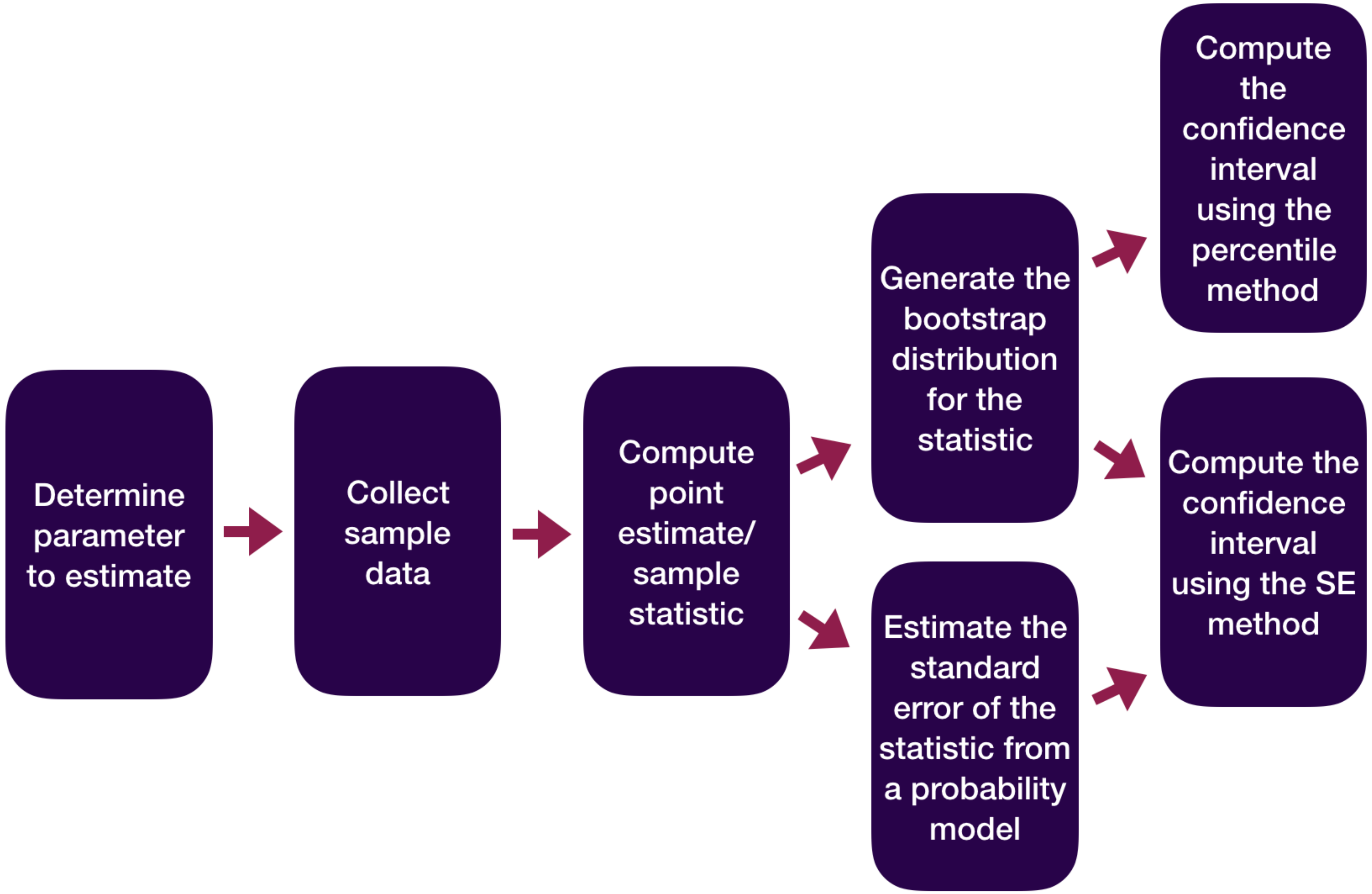


Statistical Inference Zoom Out – Estimation



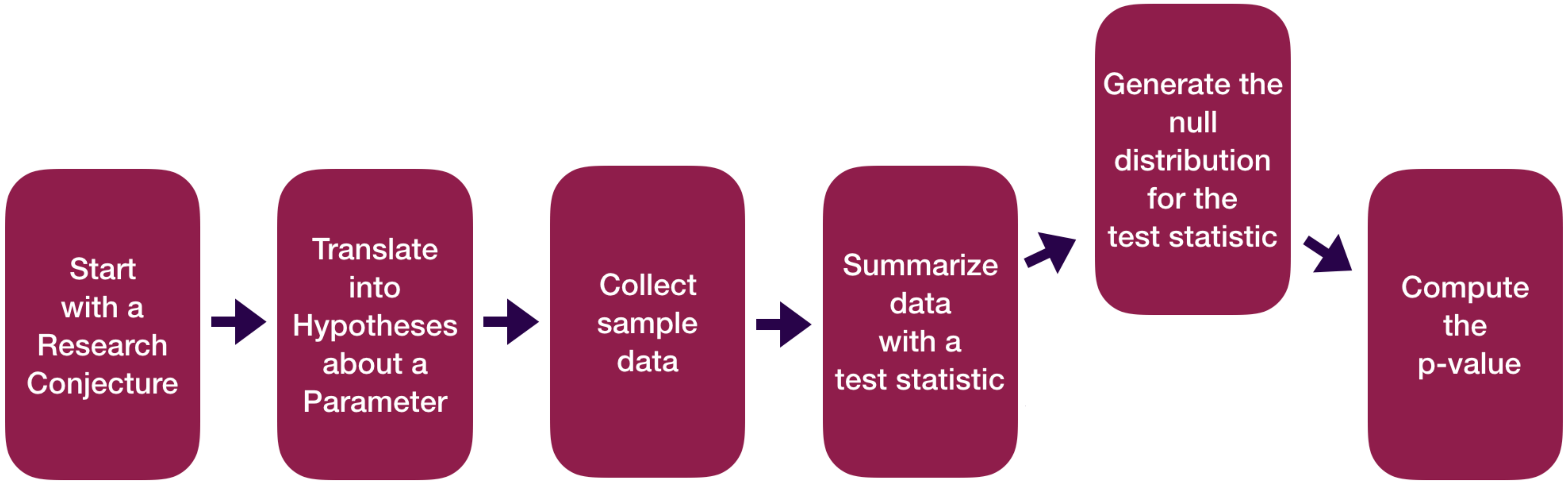
Question: How did folks do inference before computers?

Statistical Inference Zoom Out – Estimation



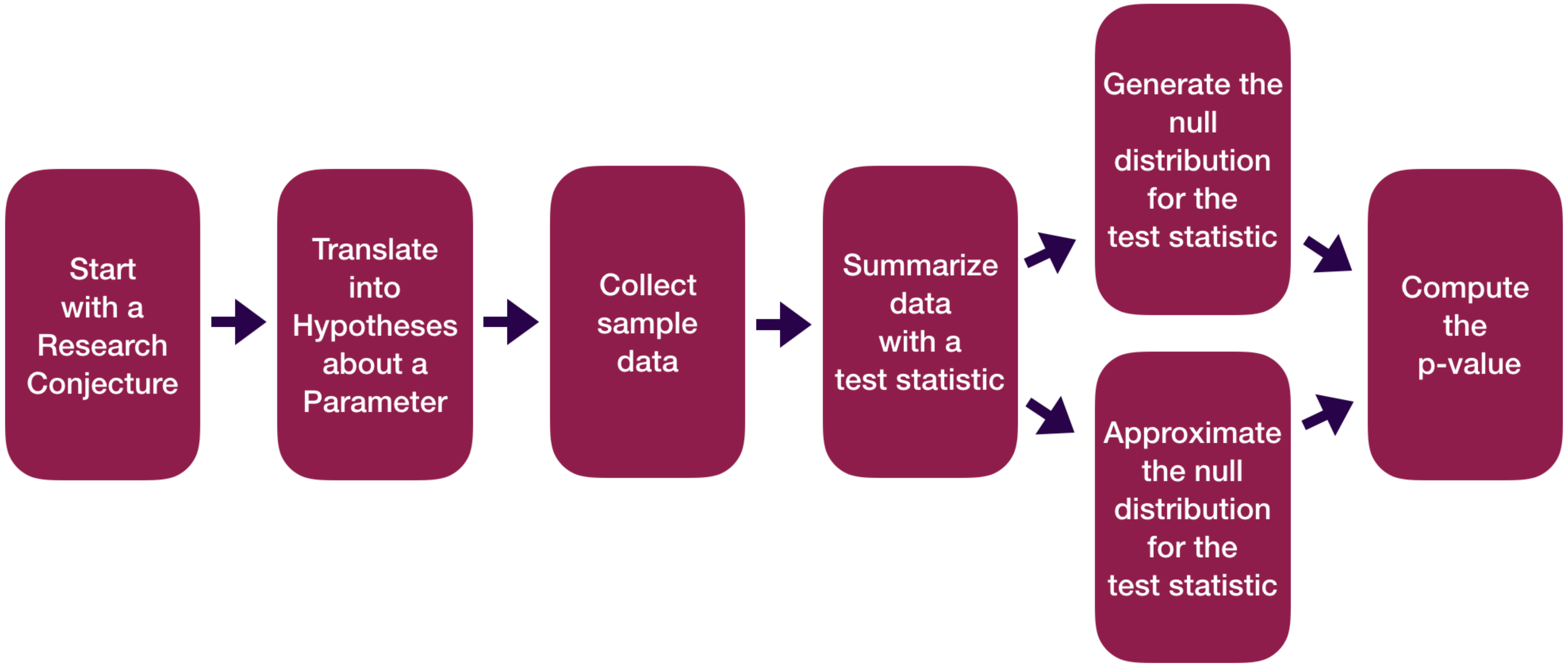
Question: How did folks do inference before computers?

Statistical Inference Zoom Out – Testing



Question: How did folks do inference before computers?

Statistical Inference Zoom Out – Testing



Question: How did folks do inference before computers?

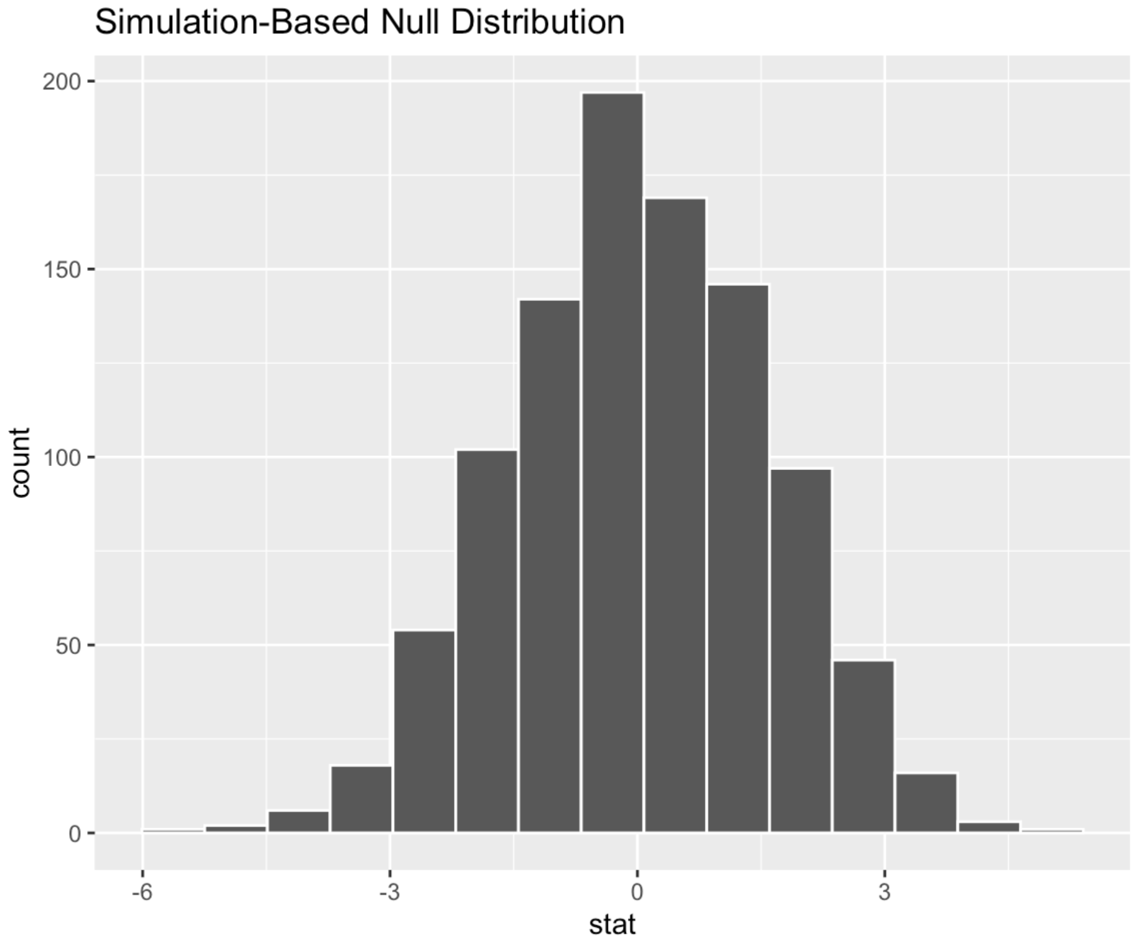
This means we need to learn about probability models!

Probability Models

- We have computers now! Why do we need to know how to do inference without them?
  - It’s important to build literacy for existing and future work that relies on probability models and theory
  - Default settings for many programmed statistical analyses make assumptions based on probability theory
  - Closed-form results build intuition and formalize ideas in ways that simulation methods alone do not

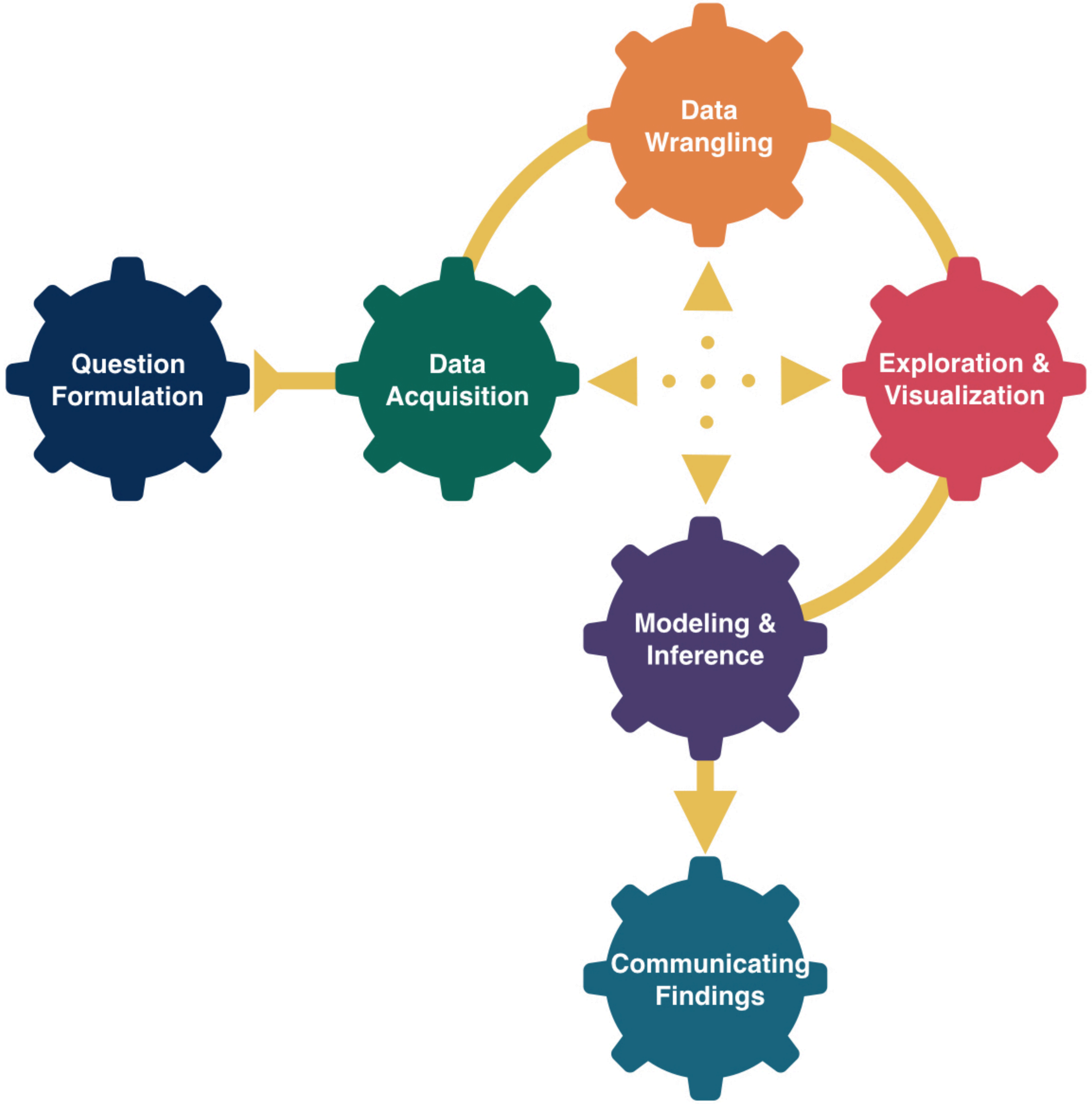
Probability Models

Motivating question: How can we use theoretical probability models to approximate our (sampling) distributions?



Before we can answer that question and apply the models, we need to learn about the theoretical probability models themselves. We'll turn to this **after spring break**!

Enjoy your spring break!!



Random Variables I: Introduction

Megan Ayers

Math 141 | Spring 2026

Monday, Week 10

Announcements

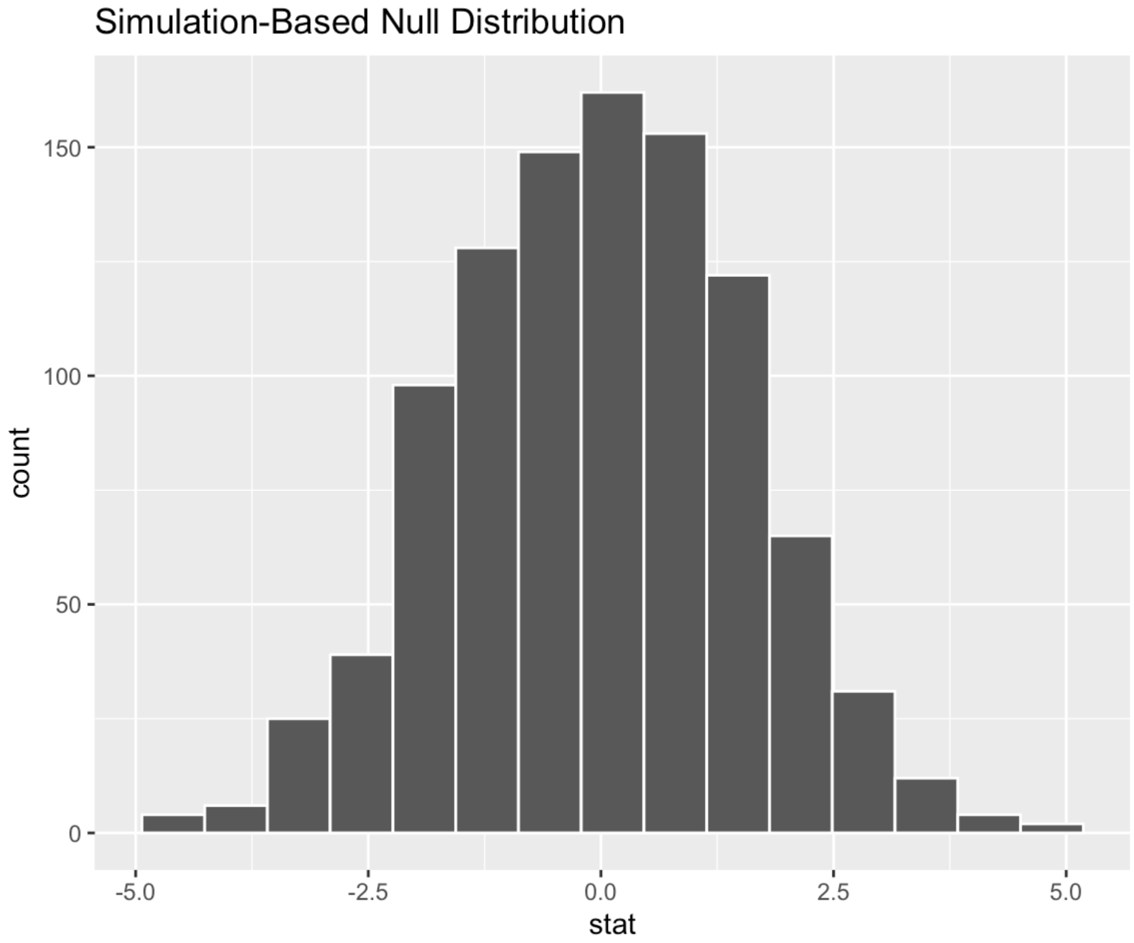
- Midterm revision deadline extended to **Friday April 3rd, at the beginning of lecture**.

Goals for today

- Define and introduce **random variables**

Probability Models

Motivating question: How can we use theoretical probability models to approximate our (sampling) distributions?



Before we can answer that question and apply the models, we need to learn about the theoretical probability models themselves.

Random Variables

Definitions

A **random variable** is a numeric quantity whose value depends on the result of a random process.

- Use capital letters at the end of the alphabet (W, X, Y, Z) to denote **random variables**
- Use lowercase letters (w, x, y, z) to denote the **particular values** of a random variable
  - e.g., $X = x$ means "we observed the random variable X take the value x"
- Use equations to express **events** associated to random variables.
  - e.g., "$X = 5$" represents the event "The random variable X takes the value 5".
- Events associated to variables have probabilities of occurring.
  - e.g., $P(X = 5) = 0.5$ means X has probability 0.5 of taking the value 5.

Types of Random Variables

There are two main types of random variables:

1. **Discrete** variables can take only specific values (often a finite number of values).
2. **Continuous** variables can take any value in an interval of real numbers.

Examples of **discrete** variables:

- The number of credits a randomly chosen Reed student is taking.
- The number of vegetarians in a random sample of 10 people.
- The result of a coin flip

Examples of **continuous** variables:

- The temperature of my office at a particular time of the day.
- The amount of time it takes a radioactive particle to decay.

Today, we'll focus on discrete random variables

For a discrete random variable, care about its:

- Distribution: $p(x) = P(X = x)$
- Center (Mean):

$$\mu = \sum_x x p(x)$$

- Spread (Variance & Standard Deviation):

$$\sigma^2 = \sum_x (x - \mu)^2 p(x)$$

$$\sigma = \sqrt{\sum_x (x - \mu)^2 p(x)}$$

The Distribution of a Random Variable

Random variables have distributions, which tell us…

- the values the variable can take, and the *probability* the variable takes those values.

Example: I play a casino game, where that the amount of money I win (in dollars) has the following distribution:

Winnings	**\$5**	**\$10**	**\$20**	**\$50**
Probability	.3	.4	.2	.1

Suppose instead that I have a purse filled with the following 100 bills:

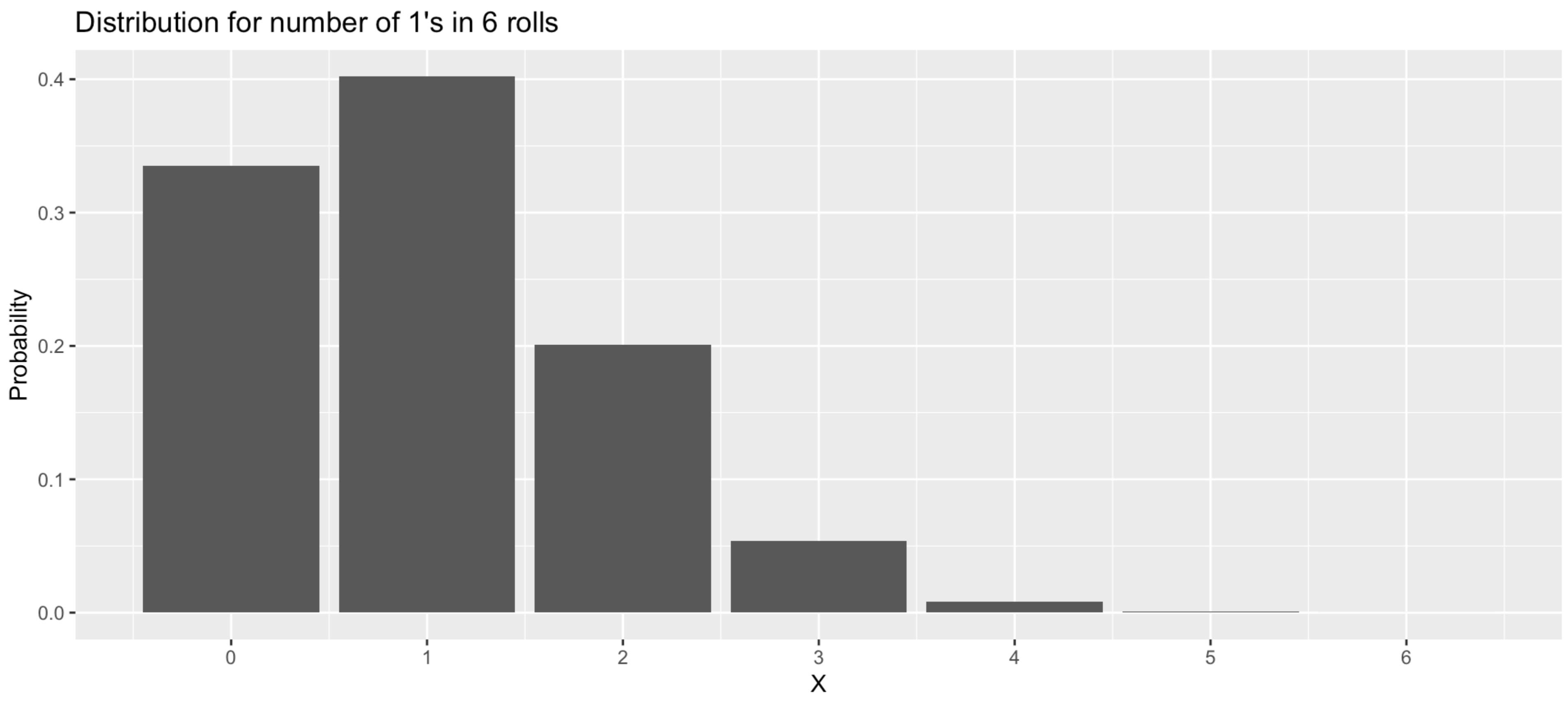
Type	**\$5**	**\$10**	**\$20**	**\$50**
Frequency	30	40	20	10

Playing this casino game is very similar to drawing a random bill from the purse!

Visualizing Discrete Distributions

We often use bar charts to visualize the distribution of discrete random variables.

- Suppose a fair 6-sided die is rolled 6 times. Let X be the number of 1s rolled. The distribution of X is given by:



Expected Value and Variance

Expected Value

Informally: The **expected value** is the average value the random variable takes.

- If X represents the number of Heads when flipping a fair coin, its expected value is 0.5

Formally: The **expected value** (or mean) of a discrete random variable X is

$$E[X] = x_1 P(X = x_1) + x_2 P(X = x_2) + \ldots x_n P(X = x_n) = \sum_{i=1}^{n} x_i P(X = x_i)$$

where $x_1, \ldots, x_n$ are *all the values X could potentially take*.

- The expected value of X is a **weighted average** of the values X can take, where weights are probabilities.
- We often use μ to denote expected values (i.e. μ = E[X])

Expected Value

Recall the example: if X represents the number of Heads when flipping a fair coin, its expected value is 0.5

$$\begin{aligned} E[X] &= 0P(X=0) + 1P(X=1) \\ &= 0(0.5) + 1(0.5) \\ &= 0.5 \end{aligned}$$

- The expected value of X is a **weighted average** of the values X can take, where weights are probabilities.

Practice

$$E[X] = x_1 P(X = x_1) + x_2 P(X = x_2) + \ldots x_n P(X = x_n) = \sum_{i=1}^{n} x_i P(X = x_i)$$

Suppose we have a data set consisting of 10 values: $\{1, 1, 2, 2, 2, 2, 3, 4, 5, 5\}$.

- Let X be a value chosen from this data set randomly.
- What is the expected value of X?

Practice (Answers)

$$E[X] = x_1 P(X = x_1) + x_2 P(X = x_2) + \dots x_n P(X = x_n) = \sum_{i=1}^{n} x_i P(X = x_i)$$

Suppose we have a data set consisting of 10 values: $\{1, 1, 2, 2, 2, 2, 3, 4, 5, 5\}$.

- Let X be a value chosen from this data set randomly.
- What is the expected value of X?

$$\begin{aligned} E[X] =& 1P(X = 1) + 2P(X = 2) + 3P(X = 3) + 4P(X = 4) + 5P(X = 5) \\ =& 1\frac{2}{10} + 2\frac{4}{10} + 3\frac{1}{10} + 4\frac{1}{10} + 5\frac{2}{10} \\ =& \frac{27}{10} \end{aligned}$$

The Law of Large Numbers, again

Previously, we said that by the **Law of Large numbers** (LLN), the proportion of times an outcome occurs in a long sequence of trials is close to the probability for that outcome.

This is a generalization:

Theorem: The Law of Large Numbers

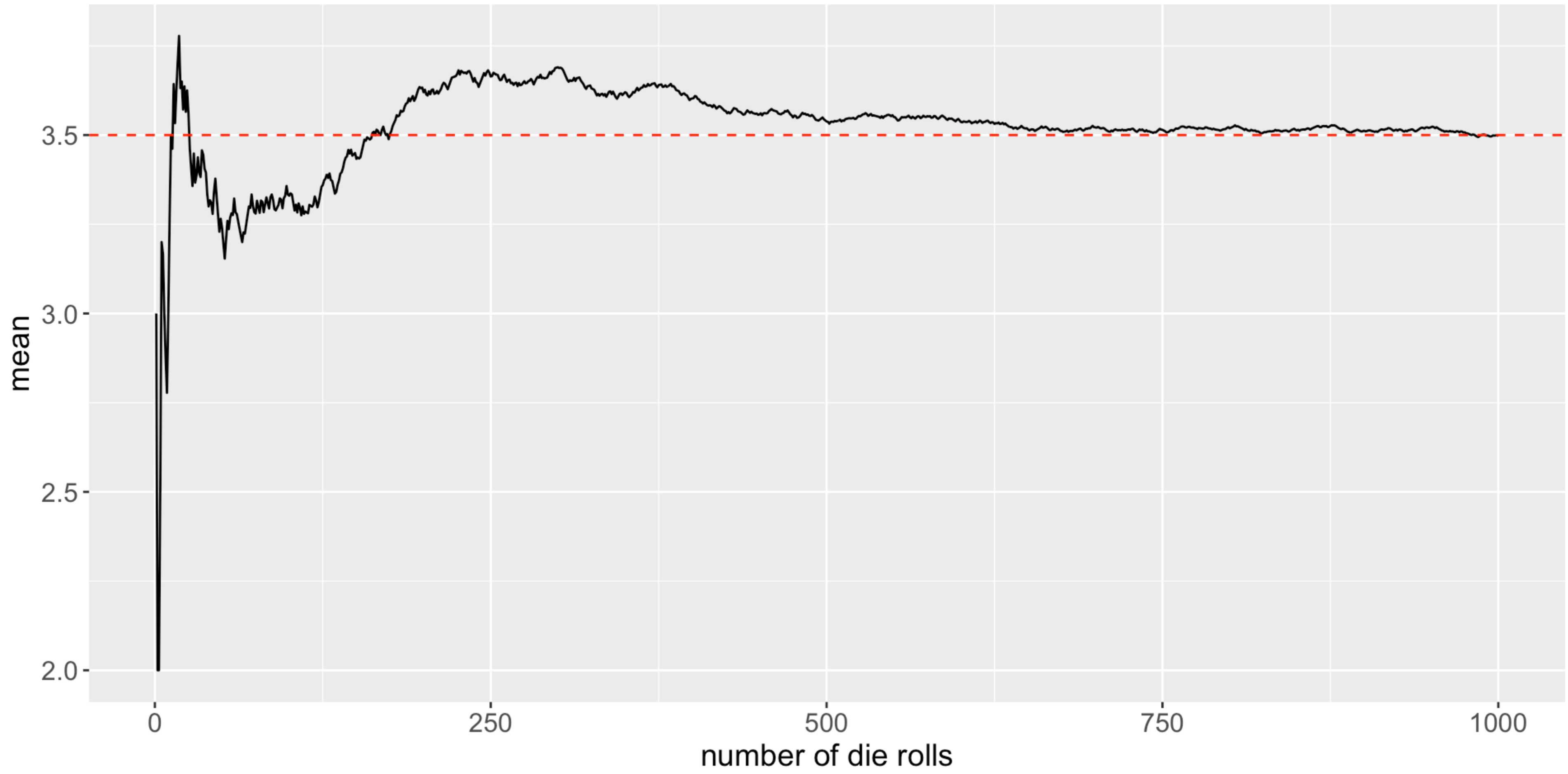
Let X be a random variable.

1. Suppose we observe the random variable n times, $x_1, x_2, \ldots, x_n$.
2. Let $\bar{x}_n$ denote the mean of our n observations of X.
3. Then, as n becomes large, $\bar{x}_n$ will approach the expected value $E[X]$ of the random variable X.

A Roll of the Die

Suppose we roll a fair 6-sided die. What is the expected value of the result?

- Let's roll the same die 1000 times and keep track of the running mean of the results…



We can see that the **expected value** is 3.5

Variance and Standard Deviation

The **variance** of a discrete random variable X with mean $E(X) = \mu$ is

$$\begin{aligned}\mathrm{Var}(X) =& (x_1 - \mu)^2 P(X = x_1) + (x_2 - \mu)^2 P(X = x_2) + \cdots + (x_n - \mu)^2 P(X = x_n) \\ =& \sum_{i=1}^{n} (x_i - \mu)^2 P(X = x_i)\end{aligned}$$

The variance of X is the **sum the squared deviations of X from its mean μ, weighted by the corresponding probabilities.**

We also define the **standard deviation** of a random variable X to be

$$\mathrm{SD}(X) = \sqrt{\mathrm{Var}(X)}$$

We often use:

Variance and Standard Deviation

$$\begin{aligned}\mathrm{Var}(X) =& (x_1-\mu)^2 P(X=x_1) + (x_2-\mu)^2 P(X=x_2) + \cdots + (x_n-\mu)^2 P(X=x_n)\\ =& \sum_{i=1}^{n} (x_i-\mu)^2 P(X=x_i)\end{aligned}$$

Suppose we have a data set consisting of 5 values: $\{1,1,3,5,5\}$. Let X be a value chosen from this data set randomly. What is the variance of X?

- Note that $E(X)=\mu=3$

$$\begin{aligned}\mathrm{Var}(X) =& (1-3)^2 P(X=1) + (3-3)^2 P(X=3) + (5-3)^2 P(X=5)\\ =& (-2)^2\frac{2}{5} + (0)^2\frac{1}{5} + 2^2\frac{2}{5} = \frac{16}{5}\end{aligned}$$

Practice: One Coin Flip

Consider a random variable X which is the number of heads in a single **fair** coin flip.

Q: What are the possible values for X? What are the probabilities each value of X?

- $X = 1$ (H) with $P(X = 1) = \frac{1}{2}$
- $X = 0$ (T) with $P(X = 0) = \frac{1}{2}$

Q: Compute the expected value and variance for X (the number of heads in a single coin flip)

$$E[X] = (0)\frac{1}{2} + (1)\frac{1}{2} = \frac{1}{2}$$

$$Var[X] = (0 - 1/2)^2\frac{1}{2} + (1 - 1/2)^2\frac{1}{2} = \frac{1}{4}$$

Linearity of Expected Value

Theorem: Expectation of Sum

Let X and Y be random variables. Then

$$E(X+Y) = E(X) + E(Y)$$

- e.g., If X is the face of one die roll, and Y is the face of another die roll, then $X+Y$ would be the sum of the faces. And:

$$E(X+Y) = E(X) + E(Y) = 3.5 + 3.5 = 7$$

Linearity of Expected Value

Theorem: Scalar Multiplication with Expectation

Let X be a random variable, and let c be a number. Then

$$E(cX) = cE(X)$$

- e.g., We're rolling a die, and we'll make \$5 times the number we roll – what are our expected earnings?

$$X = \text{face of die roll} \qquad E(5X) = 5E(X) = 5(3.5) = 17.5$$

Variance of Sums and Scalar Multiplication

Theorem: Variance of Sum

Let X and Y be random variables. Additionally, let X and Y be *independent*. Then

$$\mathrm{Var}(X+Y) = \mathrm{Var}(X) + \mathrm{Var}(Y)$$

- Note that X and Y **must be independent** for the above!

Theorem: Scalar Multiplication with Variance

Let X be a random variable, and let c be a number. Then

$$\mathrm{Var}(cX) = c^2\mathrm{Var}(X)$$

- We get the c^2 because variances consider squared differences from the mean

Practice: Two Coin Flips

Now let Y be the number of heads in two coin flips.

Q: What are the possible values for Y? What are the probabilities each value of Y?

- $Y = 0$ (TT) with $P(Y = 0) = \frac{1}{4}$
- $Y = 1$ (HT or TH) with $P(Y = 1) = \frac{1}{2}$
- $Y = 2$ (HH) with $P(Y = 2) = \frac{1}{4}$

Q: Compute the expected value and variance for Y (the number of heads in two coin flips.)

We could use the definitions of expected value and variance directly…

$$E[Y] = (0)\frac{1}{4} + (1)\frac{1}{2} + (2)\frac{1}{4} = 1$$

$$Var[Y] = (0-1)^2\frac{1}{4} + (1-1)^2\frac{1}{2} + (2-1)^2\frac{1}{4} = \frac{1}{2}$$

Practice: Two Coin Flips

…or, we could use our rules for expected value and variance:

$$Y = X_1 + X_2$$

where X_1 and X_2 are two independent versions of our random variable, X.

Thus,

$$E[Y] = E[X_1] + E[X_2] = 0.5 + 0.5 = 1$$

$$Var[Y] = Var[X_1] + Var[X_2] = 0.25 + 0.25 = 0.5$$

That's a lot simpler!

Activity

What about 23 coin tosses?

Let Z be the number of heads in 23 independent coin tosses.

Q: What is $E(Z)$?

Q: What is $\mathrm{Var}(Z)$?

Q: Suppose W represents the number of heads in n independent coin tosses. What is $E(W)$ and $\mathrm{Var}(W)$?

Activity: Answers

Let Z be the number of heads in 23 independent coin tosses.

Q: What is $E(Z)$?

- Let X_1 = Heads on 1st toss, $\ldots$, X_{23} = Heads on 23rd toss:

$$E(Z) = E(X_1) + \cdots + E(X_{23}) = 0.5 + \cdots + 0.5 = 23(0.5)$$

Q: What is $\text{Var}(Z)$?

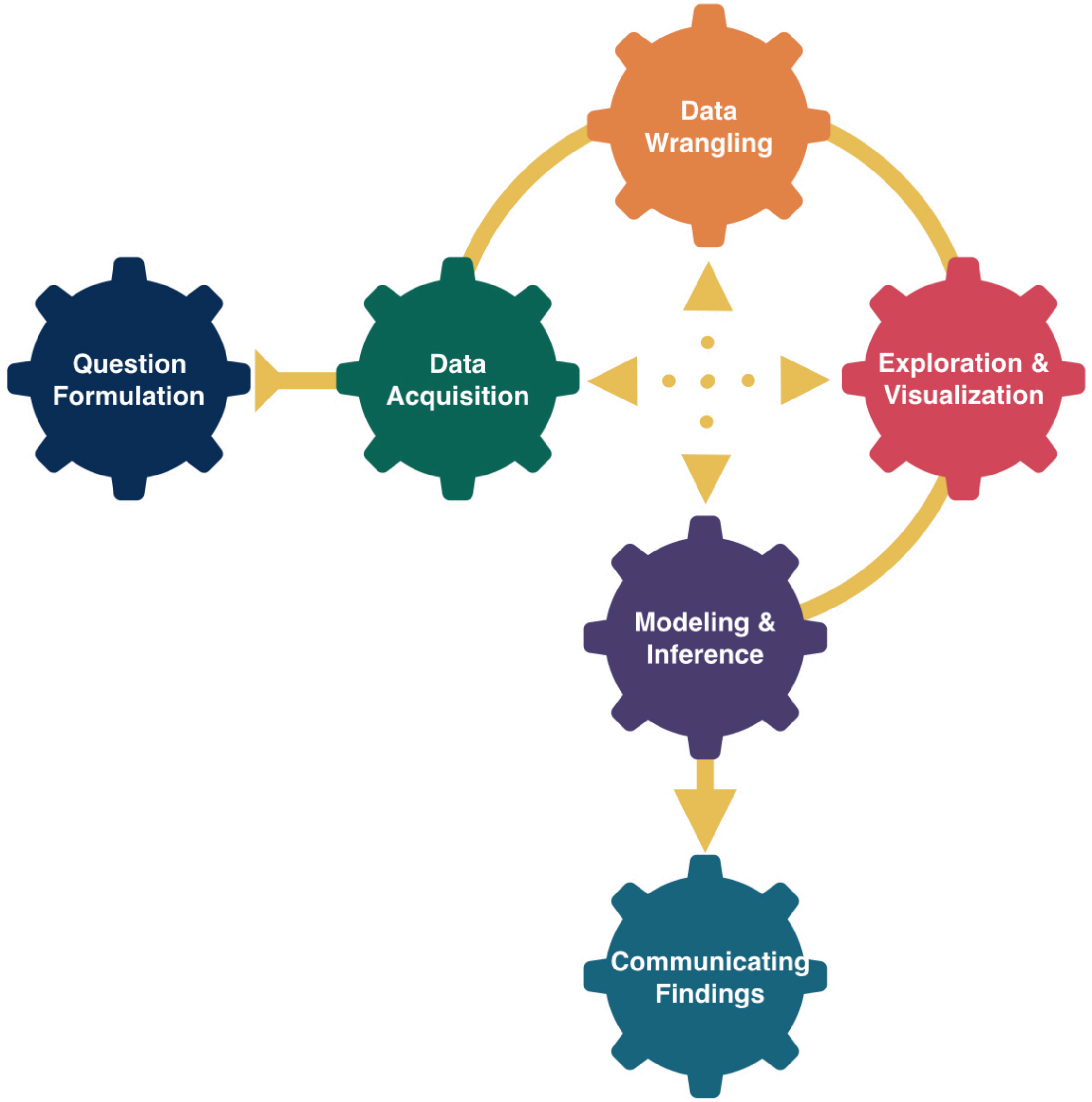
$$\text{Var}(Z) = \text{Var}(X_1) + \cdots + \text{Var}(X_{23}) = 0.25 + \cdots + 0.25 = 23(0.25)$$

Q: Suppose W represents the number of heads in n independent coin tosses. What is $E(W)$ and $\text{Var}(W)$?

- $E(W) = nE(X_1) = (n)(0.5)$
- $\text{Var}(W) = n\text{Var}(X_1) = (n)(0.25)$

Next time:

- We'll define some specific named random variables!



Goals for Today

- Continue our discussion of expected value and variance
- Introduce the **Bernoulli distribution**
- Introduce the **Binomial distribution**

Review

Review: Expected Value, Variance, and Standard Deviation

The **expected value** (or mean) of a discrete random variable X is

$$\mu = E[X] = \sum_{i=1}^{n} x_i P(X = x_i)$$

where $x_1, \ldots, x_n$ are *all the values X could potentially take.*

The **variance** and **standard deviation** of a discrete random variable X with mean $E(X) = \mu$ is

$$\sigma^2 = \mathrm{Var}(X) = \sum_{i=1}^{n} (x_i - \mu)^2 P(X = x_i)$$

$$\sigma = \mathrm{SD}(X) = \sqrt{\mathrm{Var}(X)}$$

Review: Rules for Expected Value and Variance

Theorem: Expectation of Sum

Let X and Y be random variables. Then

$$E(X+Y) = E(X) + E(Y)$$

Theorem: Scalar Multiplication with Expectation

Let X be a random variable, and let c be a number. Then

$$E(cX) = cE(X)$$

Theorem: Variance of Sum for Independent Random Variables

Let X and Y be random variables. Additionally, let X and Y be *independent*. Then

$$\text{Var}(X+Y) = \text{Var}(X) + \text{Var}(Y)$$

Theorem: Scalar Multiplication with Variance

Let X be a random variable, and let c be a number. Then

$$\text{Var}(cX) = c^2\text{Var}(X)$$

Recall: One Coin Flip

Consider a random variable X which is the number of heads in a single coin flip.

Q: What are the possible values for X? What are the probabilities each value of X?

- $X = 1$ (H) with $P(X = 1) = \frac{1}{2}$
- $X = 0$ (T) with $P(X = 0) = \frac{1}{2}$

Q: Compute the expected value and variance for X (the number of heads in a single coin flip)

$$E[X] = (0)\frac{1}{2} + (1)\frac{1}{2} = \frac{1}{2}$$

$$Var[X] = (0 - 1/2)^2\frac{1}{2} + (1 - 1/2)^2\frac{1}{2} = \frac{1}{4}$$

This distribution has a name!

- X is called a **Bernoulli** random variable!

Specific Named Random Variables

Specific Named Random Variables

- There is a vast array of random variables out there.
- But there are a few particular ones that we will find useful.
  - Because these ones are used often, they have been given names.
- Will identify these named RVs using the following format:

$$X \sim \text{Name}(\text{values of key parameters})$$

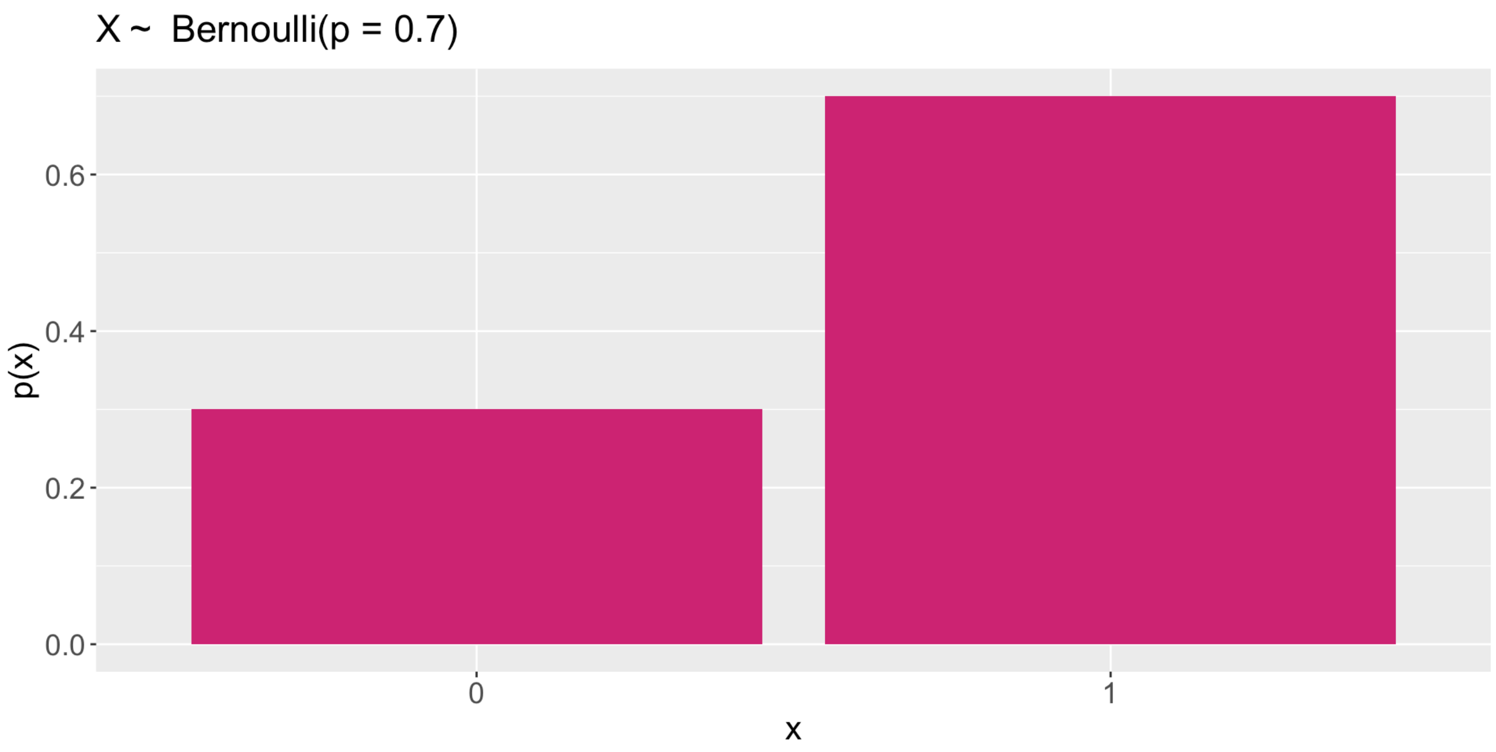
Bernoulli Random Variables

$X \sim \text{Bernoulli}\,(p)$

$$X = \begin{cases} 1 & \text{success} \\ 0 & \text{failure} \end{cases}$$

Important parameter:

$$\begin{aligned} p &= \text{probability of success} \\ &= P(X = 1) \end{aligned}$$



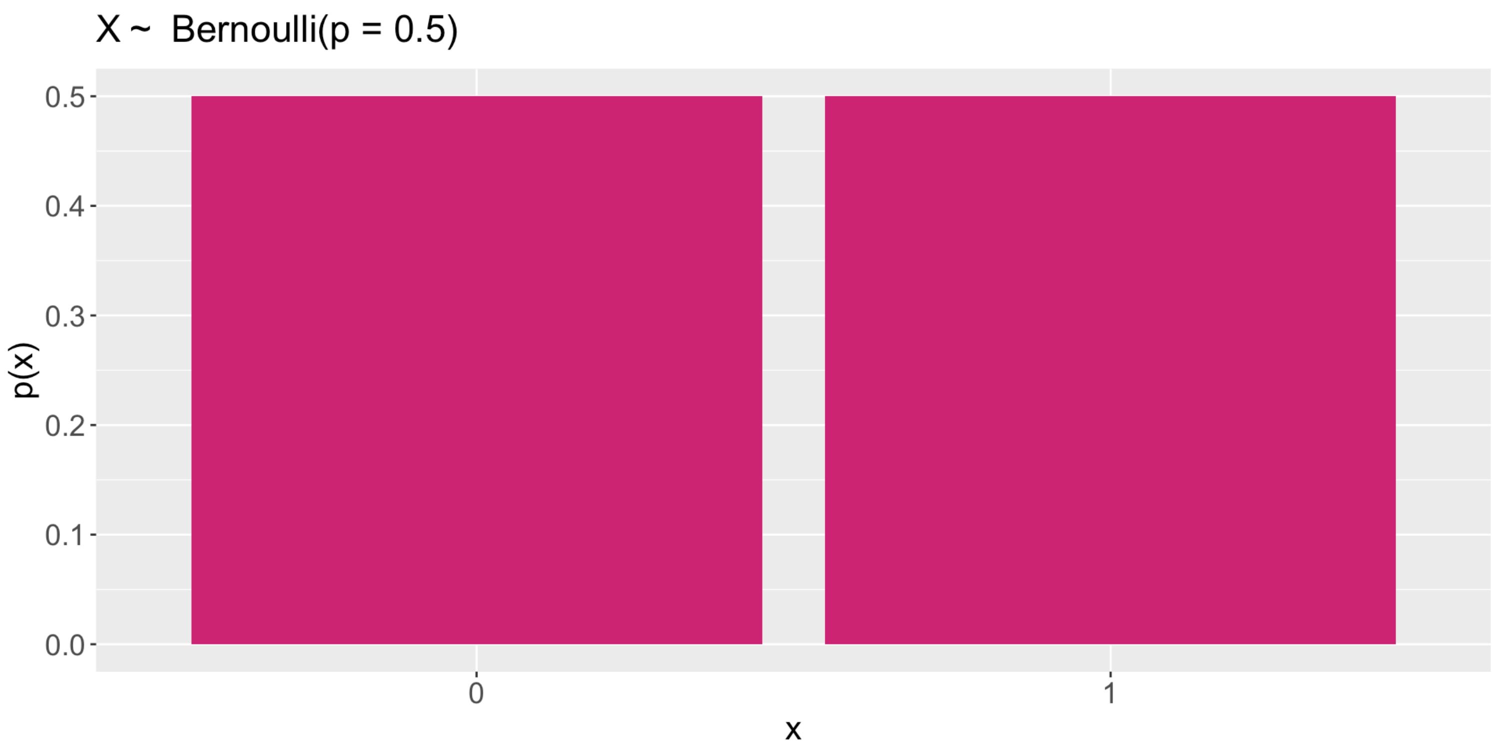
Distribution:

x	**0**	**1**
$p(x)$	$1 - p$	p

Bernoulli Random Variables

$X \sim \text{Bernoulli}\,(p = 0.5)$

$$X = \begin{cases} 1 & \text{success} \\ 0 & \text{failure} \end{cases}$$



Distribution:

x	**0**	**1**
$p(x)$	0.5	0.5

Bernoulli Random Variables

$X \sim$ Bernoulli (p)

$$X = \begin{cases} 1 & \text{success} \\ 0 & \text{failure} \end{cases}$$

Distribution:

x	**0**	**1**
$p(x)$	1 - p	p

Question: What's the expected value of a Bernoulli(p) random variable?

Bernoulli Random Variables

$X \sim$ Bernoulli (p)

$$X = \begin{cases} 1 & \text{success} \\ 0 & \text{failure} \end{cases}$$

Distribution:

x	0	1
$p(x)$	1 - p	p

Question: What's the expected value of a Bernoulli(p) random variable?

$$\begin{aligned} \mu &= \sum xp(x) \\ &= 1 * p + 0 * (1 - p) \\ &= p \end{aligned}$$

Bernoulli Random Variables

$X \sim$ Bernoulli (p)

$$X = \begin{cases} 1 & \text{success} \\ 0 & \text{failure} \end{cases}$$

Distribution:

x	**0**	**1**
$p(x)$	1 - p	p

Question: What's the standard deviation of a Bernoulli(p) random variable?

Bernoulli Random Variables

$X \sim$ Bernoulli (p)

$$X = \begin{cases} 1 & \text{success} \\ 0 & \text{failure} \end{cases}$$

Distribution:

x	0	1
$p(x)$	1 - p	p

Question: What's the standard deviation of a Bernoulli(p) random variable?

$$\begin{aligned} \sigma &= \sqrt{\sum (x-\mu)^2 p(x)} \\ &= \sqrt{(1-p)^2 * p + (0-p)^2 * (1-p)} \\ &= \sqrt{p(1-p)} \end{aligned}$$

What about the sum of many Bernoulli random variables?

n "weighted coin tosses": Expected Value and Variance

Let X be the number of successes in n **independent** Bernoulli(p) random variables.

Q: What are $E(X)$ and $\mathrm{Var}(X)$?

Hint: We know, from the Bernoulli distribution, that if X_1 is the number of heads in a *single* weighted coin flip with success probability p,

$$E(X_1) = p \quad \text{and} \quad \mathrm{Var}(X_1) = p(1-p)$$

n "weighted coin tosses": Expected Value and Variance

Let X be the number of heads in n **independent** Bernoulli(p) random variables.

Answer: Let X_1 = Heads in 1st toss, $\ldots$, X_n = Heads in nth toss:

$$X = X_1 + \cdots + X_n = \sum_{i=1}^{n} X_i$$

$$E(X) = E\left(\sum_{i=1}^{n} X_i\right) = \sum_{i=1}^{n} E(X_i) = \sum_{i=1}^{n} p = np$$

$$\mathrm{Var}(X) = \mathrm{Var}\left(\sum_{i=1}^{n} X_i\right) = \sum_{i=1}^{n} \mathrm{Var}(X_i) = \sum_{i=1}^{n} p(1-p) = np(1-p)$$

What about the distribution of n weighted coin tosses?

We'll start with the unweighted case, i.e., $p = 0.5$

n coin toss probabilities?

We've calculated Expectation and Variance for the number of heads in n coin tosses.

- But what about the **probability** that there are k heads in the n tosses?

For example, **What's the probability of $k = 2$ heads in $n = 3$ coin tosses?**

- Let's break down the problem: **What are the possible outcomes?**

$$\begin{aligned} \text{TTT} &\implies X = 0 & \qquad \text{TTH or THT or HTT} &\implies X = 1 \\ \text{HHH} &\implies X = 3 & \qquad \text{THH or HTH or HHT} &\implies X = 2 \end{aligned}$$

- There are 8 possible sequences from **3 coin flips**, each with equal probability.

The random variable X has 4 possible outcomes: 0, 1, 2, or 3. Those probabilities are:

$$\begin{aligned} P(X = 0) &= \frac{1}{8} & \qquad P(X = 1) &= \frac{3}{8} \\ P(X = 3) &= \frac{1}{8} & \qquad P(X = 2) &= \frac{3}{8} \end{aligned}$$

3 coin tosses probabilities

The random variable X has 4 possible outcomes: 0, 1, 2, or 3. Those probabilities are:

$$P(X=0)=\frac{1}{8} \qquad P(X=1)=\frac{3}{8}$$
$$P(X=3)=\frac{1}{8} \qquad P(X=2)=\frac{3}{8}$$

Notice that the probability of k heads in n coin flips is

$$P[X=k]=\left(\text{Number of ways to get k heads}\right)\times\left(\text{Prob of each flip sequence}\right)$$

For example, we found that in $n=3$ coin tosses,

- 3 ways to get 2 heads
- Each sequence of 3 coin flips has probability $\frac{1}{2}\times\frac{1}{2}\times\frac{1}{2}=\frac{1}{8}$
- Thus, $P(X=2)=\frac{3}{8}$

Generalization 1

First: **What if n and k are bigger?**

- In this case, it's hard to individually count all the different flip sequences
- Instead, we can rely on a **handy formula**: The number of ways to get k "successes" in n "trials" is

$$\binom{n}{k} = \frac{n!}{k!(n-k)!}$$

We read this as **"n choose k"**

The exclamation point is a **factorial**:

$$\begin{aligned} 0! &= 1 \\ 1! &= 1 \\ 2! &= 2 \times 1 = 2 \\ 3! &= 3 \times 2 \times 1 = 6 \\ 4! &= 4 \times 3 \times 2 \times 1 = 24 \end{aligned}$$

Generalization 1

First: **What if n and k are bigger?**

- In this case, it's hard to individually count all the different flip sequences
- Instead, we can rely on a **handy formula**: The number of ways to get k "successes" in n "trials" is

$$\binom{n}{k} = \frac{n!}{k!(n-k)!}$$

We read this as **"n choose k"**

So in our example of 2 heads in 3 flips:

$$\#\text{ of ways to choose 2 things out of 3 things} = \binom{3}{2} = \frac{3!}{2! \cdot 1!} = \frac{3 \cdot 2}{2} = 3$$

Generalization 2

Second: **What if the probability of "success" changes?**

- In other words, how can we generalize to the sum of Bernoulli(p) case from the sum of the Bernoulli($p = 0.5$) case?

- In this case, the probability of getting *a particular sequence* of k "successes" in n **independent** trials is:

$$p^k \times (1-p)^{n-k}$$

- There are k "successes", each with probability $p \ \rightarrow \ p^k$
- There are $n-k$ "failures", each with probability $1-p \ \rightarrow \ (1-p)^{n-k}$

Binomial Distribution: Putting it all together

Thus, the probability of observing k successes in n independent Bernoulli trials with probability of success p is

$$P[X = k] = \binom{n}{k} p^k (1-p)^{n-k}$$

In this case, we say $X \sim \mathrm{Binomial}(n,\ p)$. Or, in words: X follows a **binomial distribution** with n trials each with p probability of success

How to tell if your distribution is a Binomial?

- Number of trials, n, must be **fixed**.
- Each trial must be **Bernoulli** and independent, with success probability p (p must stay the same for each trial!).
- We must be counting the **number of successes**.

Properties

Expectation and Variance of Binomial

If $X \sim \text{Binomial}(n,\ p)$, then:

$$E(X) = np$$
$$Var(X) = np(1-p)$$

Note that we derived this earlier before we knew what a Binomial was, just by using Theorems about sums of random variables!

Calculating Binomial Probabilities: `dbinom()`

If X is a binomial random variable with n trials and probability of success p, we can use R to calculate the probabilities with the `dbinom()` function

- e.g., $P(X = 2)$ where $n = 3$ and $p = 0.5$ (2 heads in 3 fair coin tosses)

```
dbinom(x = 2, size = 3, prob = 0.5)
```

```
[1] 0.375
```

- e.g., $P(X = 10)$ where $n = 15$ and $p = 0.7$

```
dbinom(x = 10, size = 15, prob = 0.7)
```

```
[1] 0.2061304
```

- `dbinom()` takes in an observable value of a Binomial random variable, and returns the probability of observing it.

Calculating Binomial Probabilities: `pbinom()`

We can also calculate $P(X \leq k)$ with `pbinom()`

- e.g., $P(X \leq 2)$ where $n = 3$ and $p = 0.5$ (2 or fewer heads in 3 fair coin tosses)

```r
pbinom(q = 2, size = 3, prob = 0.5)
```

```
[1] 0.875
```

Can also calculate $P(X > k)$ by doing `1 - pbinom()`:

- e.g., $P(X > 2)$ where $n = 3$ and $p = 0.5$ (more than 2 heads in 3 fair coin tosses)

```r
1 - pbinom(q = 2, size = 3, prob = 0.5)
```

```
[1] 0.125
```

- e.g., $P(X \geq 2)$ where $n = 3$ and $p = 0.5$

```r
1 - pbinom(q = 1, size = 3, prob = 0.5)
```

```
[1] 0.5
```

Activity: Survey of Portlanders

Suppose you believe 40% of Portlanders think the city should install more bike lanes. You will take a simple random survey of 50 Portlanders to test your belief.

1. Assume your belief is true. Let X be the number of survey respondents who think the city should install more bike lanes. Does X follow a Binomial distribution? If so, what are n and p?
2. Calculate the expected value and standard deviation of X (still assuming your belief is true)
3. Suppose you conducted the survey and found 30 respondents wanted more bike lanes. What's the probability that $X \geq 30$? (use R!)
4. Draw the connection between the probability you calculated and a hypothesis test.

Next time

- Continuous random variables
- The Central Limit Theorem!