

# Ch. 1: Significance: How Strong Is the Evidence?

1 / 86

## Navigation

### By Date

- January 16: [start](#) - [end](#)
- January 21: [start](#) - [end](#)
- January 23: [start](#) - [end](#)
- January 28: [start](#) - [end](#)
- January 30: [Review](#)

### By Section

- [Section 1.1](#)
- [Section 1.2](#)
- [Section 1.3](#)
- [Section 1.4](#)
- [Section 1.5](#)

2 / 86

# 1.1: Introduction to Chance Models

3 / 86

## Statistical Significance

**Statistical significance** - unlikely to have occurred by random chance

| **Helper vs. Hinderer example**, what is  $P(14/16 \text{ infants})$ ? Is it rare?  
What is  $P(8/16 \text{ infants})$ ?

**Probability** - Long run proportion of times an outcome from a random process occurs

| Let's Make a Deal example - result after 100s of simulations

4 / 86

# Statistical Significance

General strategy:

1. Create a model for the process under random chance
2. Use the model to calculate the **probability** of the observation under random chance
3. If that probability is very small, conclude that the effect is likely not due to chance  
e.g. is **statistically significant**

5 / 86

## Vocabulary

- **Statistic** - number summarizing the results from a *sample*
- **Parameter** - long-run numerical property of the *population*

---

**Parameter -> Population**

unknown value

**Statistic -> Sample**

known value

---

We use **statistics** to estimate **parameters**

6 / 86

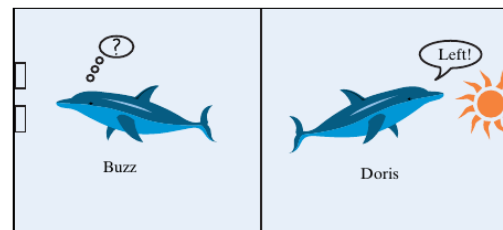
## Example 1.1: Can Dolphins Communicate?

### Training:

1. Shine headlight into tank
2. If light is steady, hit the right-side lever to get a fish
3. If light is blinking, hit the left-side lever to get a fish

### Experiment:

1. Show Doris the light
2. Give Buzz the levers
3. See if Doris could tell Buzz which lever to push



7 / 86

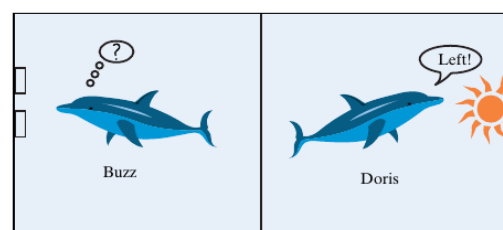
## Example 1.1: Can Dolphins Communicate?

### Training:

1. Shine headlight into tank
2. If light is steady, hit the right-side lever to get a fish
3. If light is blinking, hit the left-side lever to get a fish

### Experiment:

1. Show Doris the light
2. Give Buzz the levers
3. See if Doris could tell Buzz which lever to push



Buzz hit the correct lever 15/16 times

8 / 86

## Example 1.1: Can Dolphins Communicate?

1. Ask a research question:

Can the dolphins communicate ideas?

2. Design a study and collect data:

- Observational Units: Each of Buzz's attempts
- Variable: Correct or incorrect button
- Categorical or quantitative?: Categorical

3. Explore the data: We observed 15 correct responses out of 16 total trials

4. Draw inferences beyond the data:

What would we expect the **statistic** to be if Buzz was guessing?  
Was the observed **statistic** significantly greater than that?

We need to build a chance model to find out!

9 / 86

## Example 1.1: Can Dolphins Communicate?

| Real Study   | Physical Simulation                                   |
|--|---|
| Lever pull by Buzz   | Coin Flip   |
| Correct lever  | Heads   |
| Incorrect lever  | Tails   |
| Probability of correct button<br>when Buzz is guessing (1/2) | Chance of heads (1/2)                                 |
| 16 attempts by Buzz  | 16 coin flips =<br>One <b>Repetition</b> of the study |

If Buzz is just guessing, what is a **typical** value we would see in one repetition of the study?

<http://www.rossmanchance.com/ISIapplets.html>

10 / 86

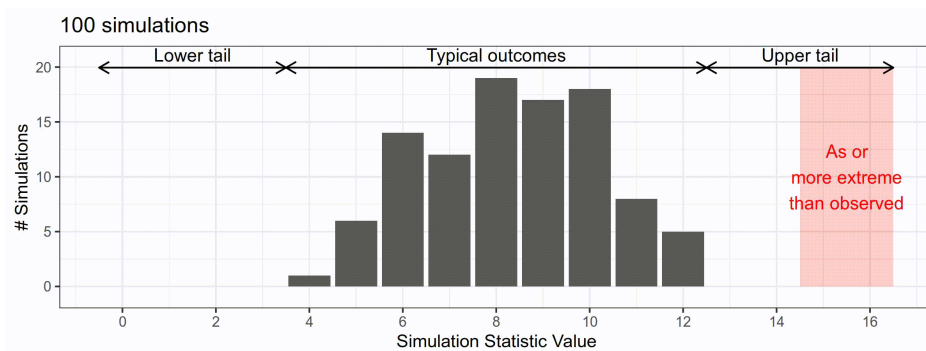
# Determining Statistical Significance: 3 S Strategy

## 1. **Statistic** - numerical summary from sample

- Helper/Hinderer: 14/16
- Doris/Buzz: 15/16

## 2. **Simulate** - identify a "by chance alone" model for the scenario and repeatedly simulate values for the statistic under that model

## 3. **Strength of Evidence** - Is the observed statistic unusual?



11 / 86

## Exploration 1.1: Can Dogs Understand Human Cues?

1. State the research question
2. Observational units
3. Variable
4. Sample size
5. Observed statistic
6. Is the statistic in the direction suggested by the researcher?
7. Is the observed result *possible*?
8. Is the observed result *probable*?
9. If he picked the cup at random, what is the probability (parameter) that he chooses correctly?
10. What would the probability of choosing the correct cup be if Harley understands the experimenter?  
(can be a range of values)

12 / 86

## For next time

- Exploration P.3 due January 17 (Friday) at 6pm
- Ch. P HW due January 20 at 6pm
- Complete Exploration 1.1 and submit it on Canvas by January 20 (Monday) at 6pm
- Read Section 1.2 of the textbook
- Complete the first two questions of Exploration 1.2: Tasting water

13 / 86

## 1.2: Measuring the Strength of Evidence

14 / 86

# Definitions

The **Null Hypothesis**  $H_0$  is the "by random chance alone" explanation

The **Alternative Hypothesis**  $H_A$  is the explanation that consists of a not-random effect.

■ This is usually the thing we want to show - our research hypothesis.

We want to disprove  $H_0$ , forcing us to conclude that  $H_A$  is a better (more probable) explanation.

15 / 86

## Binary Variables

**Binary** variables have only two possible outcomes

$\pi$  is used to represent the *parameter* for a long-run probability or proportion

$n$  is the sample size or number of observational units

$\hat{p}$  represents our *sample statistic* (the observed value)

16 / 86



# Rock Paper Scissors

- Two players
- Three options

Are all options equally likely to be played?

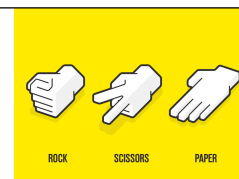
Some evidence\* suggests that novice players are less likely to choose scissors.



\*See Eyler et al., 2009

17 / 86

## Rock Paper Scissors



1. Research question: **Are novice players are less likely to choose scissors?**

2. Design a study and collect data:

Play  $n$  rounds, recording 0 or 1, where 1 = scissors was played

- Observational Unit: **Each round of play**
- Variable: **Outcome of a round (rock, paper, scissors)** Categorical
- Outcomes: **Scissors or not scissors (rock, paper)** Binary
- Statistic: **Let  $x$  = # scissors selected in  $n$  rounds; then  $\hat{p} = \frac{x}{n}$**

| In our study,  $n = 12, x = 2$

- Parameter of interest: **Probability of selecting scissors**

Options:

- Equal preference for all 3 choices, so  $\pi = 1/3$
- $\pi_{\text{scissors}} < 1/3$ , that is, scissors is chosen less frequently

Note: You don't specify a particular value, just a direction from a particular value

3. Explore the data: **2 scissors in 12 trials;  $2/12 = 1/6 \approx 0.1667$**

18 / 86

# Rock Paper Scissors

1. Research question

2. Design a study and collect data

- Outcomes: **Scissors or not scissors (rock, paper)** Binary
- Statistic:  $\hat{p} = \frac{x}{n}$
- Parameter of interest: **Probability of selecting scissors**

Options:

- Equal preference for all 3 choices, so  $\pi = 1/3$
- $\pi_{\text{scissors}} < 1/3$ , that is, scissors is chosen less frequently

Note: You don't specify a particular value, just a direction from a particular value

3. Explore the data

4. Draw inferences beyond the data:

Your parameter options represent the null and alternative hypotheses.  
Which is which?

19 / 86

# Rock Paper Scissors



1. Research question

2. Design a study and collect data

- Outcomes: **Scissors or not scissors (rock, paper)** Binary
- Statistic:  $\hat{p} = \frac{x}{n}$
- Parameter of interest: **Probability of selecting scissors**

Options:

- Equal preference for all 3 choices, so  $\pi = 1/3$
- $\pi_{\text{scissors}} < 1/3$ , that is, scissors is chosen less frequently

Note: You don't specify a particular value, just a direction from a particular value

3. Explore the data

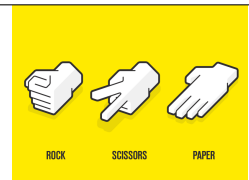
4. Draw inferences beyond the data:

$$H_0 : \pi = 1/3$$

$$H_A : \pi < 1/3$$

20 / 86

# Rock Paper Scissors



- Is  $\hat{p}_{scissors}$  less than the probability specified in the null hypothesis?

Yes, the sample proportion of 1/6 is in the direction of the alternative hypothesis

- Is it possible that  $\hat{p}$  could turn out this small even if  $H_0$  were true?

Yes, it is possible that it could be this small

21 / 86

## P-values

The **p-value** is the probability of obtaining a value of the statistic at least as extreme as the observed statistic *when the null hypothesis is true*

We estimate a p-value by finding the **proportion of simulated statistics** using the random chance model assuming  $H_0$

**that are at least as extreme** (in the direction of  $H_A$ )

**as the value of the observed statistic** from the research study.

Note that the **p-value** is not the same p as the sample proportion,  $\hat{p}$ .  
p-values can be computed using any sample statistic, not just proportions.

22 / 86

# Guidelines for Interpreting p-values

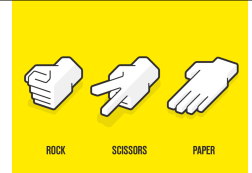
| p-value range        | Interpretation   |
|----------------------|--|
| $0.10 < p$           | Not much evidence against $H_0$<br>( $H_0$ is plausible) |
| $0.05 < p \leq 0.10$ | Moderate evidence against $H_0$                          |
| $0.01 < p \leq 0.05$ | Strong evidence against $H_0$                            |
| $p < 0.01$           | Very strong evidence against $H_0$                       |

If you have done the modeling correctly, you'll still be wrong  $(100 \times p)\%$  of the time when  $H_0$  is actually true

The smaller the p-value, the more evidence against  $H_0$

23 / 86

## Rock Paper Scissors



3 S Strategy:

1. Statistic:  $2/12 \approx 0.1667$
2. Simulation: "coin flip" with  $p(\text{heads}) = 1/3$

Actual coins don't work, but we could use dice - rolls of 1 or 2 represent scissors... or we could let the computer do it for us using the

**One Proportion applet**



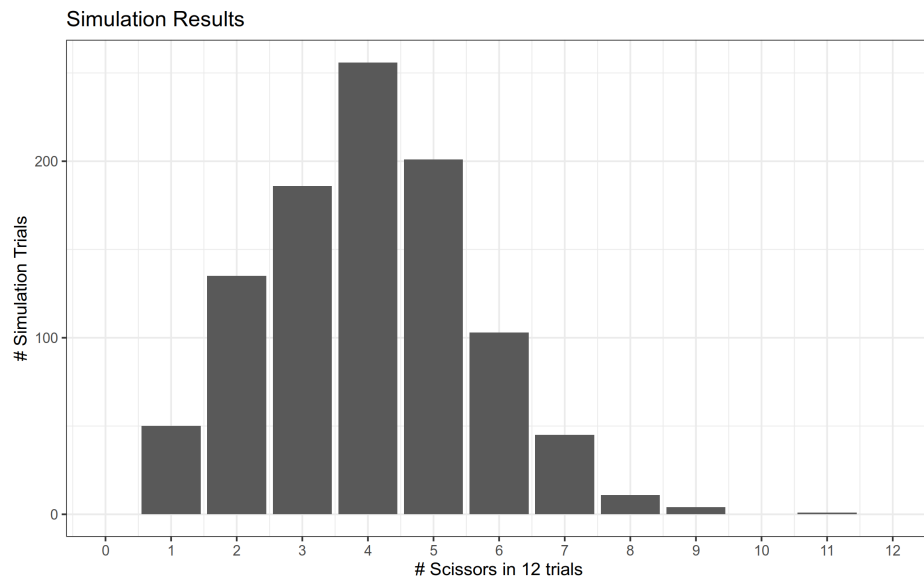
3. Strength of evidence:

193 of 1000 simulation samples showed a value of 0, 1, or 2 scissors (out of 12 games)

~19% of the time, we would see a value as or more extreme than 2 under the random chance model

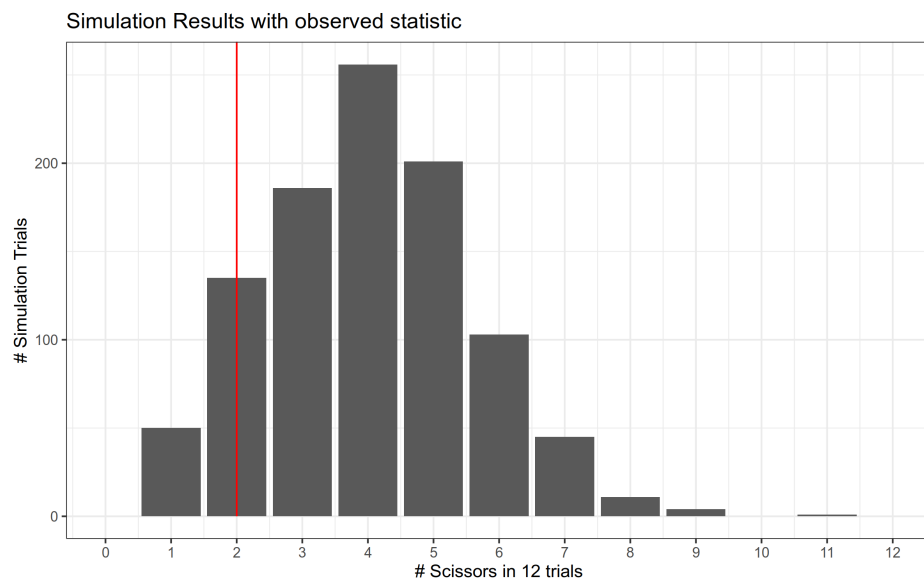
24 / 86

# P-values



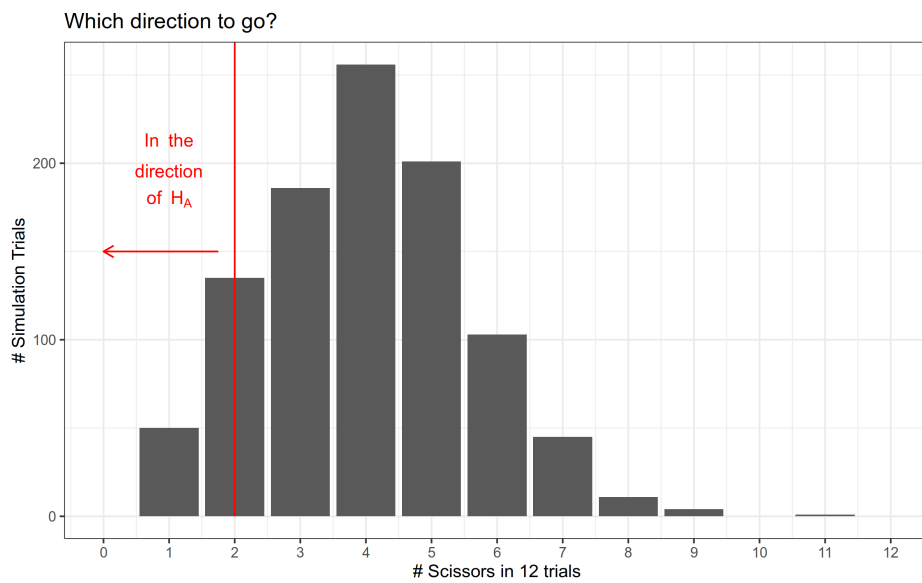
25 / 86

# P-values



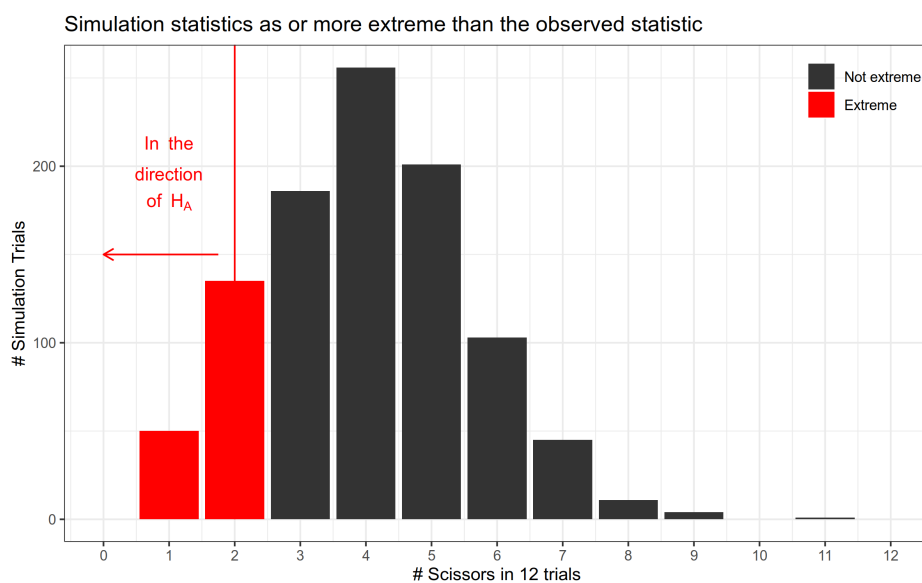
26 / 86

# P-values



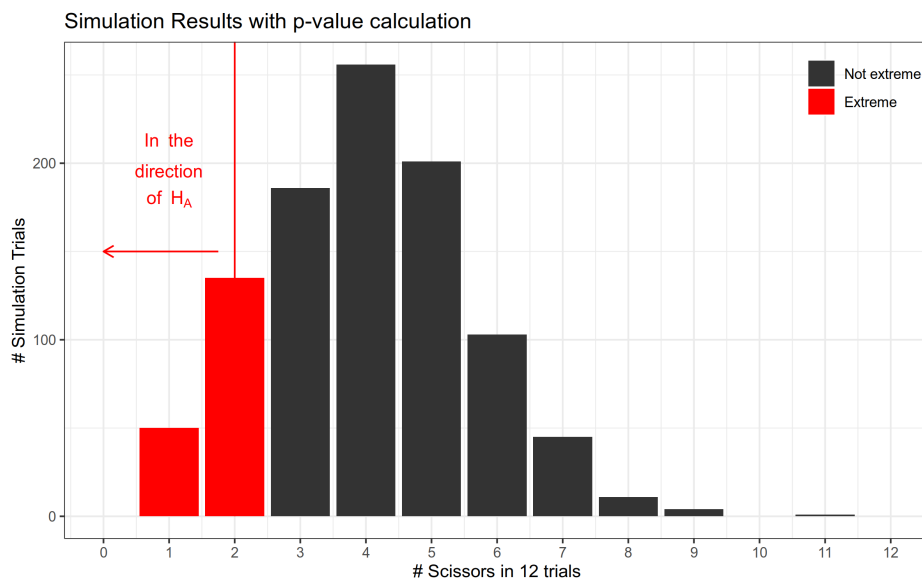
27 / 86

# P-values



28 / 86

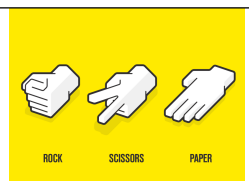
# P-values



$$\underbrace{P(x \leq 2)} \approx P(x^* \leq 2) = \frac{193}{1000} = .193$$

29 / 86

## Rock Paper Scissors



$$\underbrace{P(x \leq 2)} \approx P(x^* \leq 2) = \frac{193}{1000} = .193$$

Is this an unlikely result if  $H_0$  is true?

No, our sample statistic is not in the tail of the distribution. There is an almost 20% chance of obtaining a result as or more extreme as this by random chance.

What would you expect to happen to the p-value if we had seen only 1 scissors in the 12 rounds of the original study?

There would be fewer simulation statistics as or more extreme than the observed statistic, so the p-value would be smaller.

How much smaller?

$$P(x \leq 1) \approx P(x^* \leq 1) = \frac{58}{1000} = 0.058$$

If we had originally observed only 1/12 scissors, we would have moderate evidence against the null hypothesis.

30 / 86

## Exploration 1.2: Tasting Water

People spend a lot of money on bottled water. But do they really prefer bottled water to ordinary tap water?

Researchers at Longwood University investigated this question by presenting people who came to a booth at a local festival with four cups of water.

Three cups contained different brands of bottled water, and one cup was filled with tap water.

Each participant (person) was asked which of the four cups of water they most preferred.

Researchers kept track of how many people chose tap water in order to see whether tap water was chosen significantly less often than would be expected by random chance.

31 / 86

## Exploration 1.2: Tasting Water

### Step 1: Ask a research question

What is the question the researcher hoped to answer?

32 / 86



# Exploration 1.2: Tasting Water

## Step 2: Design a study and collect data

Identify the observational units in this study

In groups, continue with Exploration 1.2. Answer questions 1-24.

Hand in one submission per group on Canvas by Jan 24 at 6pm.

33 / 86

Review

34 / 86

## Review

The p-value is the \_\_\_\_\_ of obtaining a value for the statistic which is as or more \_\_\_\_\_ as the observed statistic when \_\_\_\_\_ is true

35 / 86

## Review

If the p-value is  $\leq 0.05$ , we have \_\_\_\_\_ evidence \_\_\_\_\_ the null hypothesis (suggesting the phenomenon is NOT random)

- We are \_\_\_\_\_ likely to see our observed result if the process happens by random chance.
- **Written conclusion:** With a p-value of \_\_\_\_\_, I \_\_\_\_\_ the null hypothesis and conclude the alternative (in the context of the problem)

36 / 86

# Review

If the p-value is  $> 0.05$ , we have \_\_\_\_\_ evidence against the null (suggesting the phenomenon could plausibly occur due to chance).

- **Written conclusion:** With a p-value of \_\_\_\_\_, I \_\_\_\_\_ the null hypothesis. We do not have enough evidence to conclude that the alternative hypothesis is more plausible than the null hypothesis that the effect occurred by chance alone.
- This does not mean we know the \_\_\_\_\_!

37 / 86

# Review

- P-value calculation
  - Want to calculate statistics as or more extreme than our observed statistic, in the direction of the alternative hypothesis
  - Compare our statistic (from the data) to the randomly simulated statistics
- When solving questions
  - Calculate the p-value with the applet
  - Is the p-value less than or greater than 0.05?
  - Should I reject or fail to reject?
  - Make conclusion
  - Interpret in the context of the problem

38 / 86

# 1.3: Alternative Measure of Strength of Evidence

## Standardized Statistics

39 / 86

### Example 1.3 - Heart Transplant Operations

- Sudden spike in heart transplant mortality rates at St. George's Hospital in London
- Of the last 10 transplants, 80% resulted in deaths within 30 days of the transplant
  - Over 5x the national average mortality rate of 15% for the procedure
- **Research question:**
- **Observational Units:**
- **Variable:**
- **Parameter:**
- **Statistic:**

40 / 86

## Example 1.3 - Heart Transplant Operations

- Binary variables are often coded as 0 = failure, 1 = success, where the choice of 0/1 is arbitrary or based on the desired statistic (survival rate vs. mortality rate)
- In medical and epidemiological studies, 1 or success may be used to denote "the patient died" because X-day mortality rate may be more informative



41 / 86

## Example 1.3 - Heart Transplant Operations

- **Null hypothesis**
- **Alternative hypothesis**

42 / 86

## Example 1.3 - Heart Transplant Operations

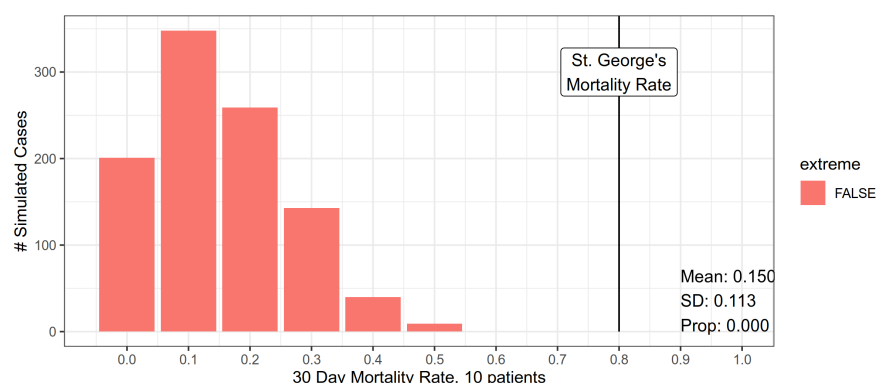
In your groups, use the online applet to conduct a relevant simulation, and apply the 3 S's strategy to the problem:

- Statistic
- Simulate
- Strength

43 / 86

## Example 1.3 - Heart Transplant Operations

- Statistic: 0.80 (8/10)
- Simulate: \_\_\_ trials per simulation, with success probability \_\_\_



Simulated 30 day mortality rate

- Strength: An as or more extreme result than our statistic, assuming the national mortality rate of 15%, occurs with probability  $< 0.001$

We only conducted 1000 simulations, so we can't say the p-value is equal to 0. We can say it is less than 1/1000. Remember that this approach is an *approximation*

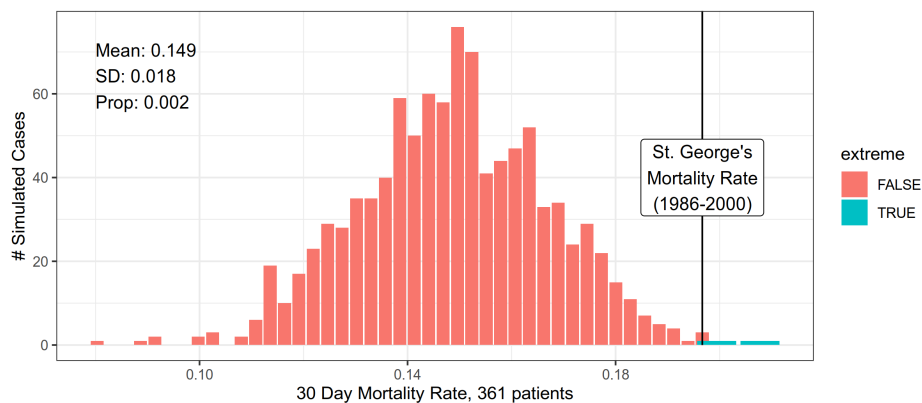
44 / 86

## Example 1.3 - Heart Transplant

- Could it be that people are only paying attention because of a run of bad luck?
- Has something changed in the hospital recently?
- Has St. George's historically had a higher mortality rate than the national average?
  - Previous 361 heart transplants (dating back to 1986) - 71 of the patients died within 30 days of the transplant

45 / 86

## Example 1.3



Simulated 30 day mortality rate

- Our simulated p-value is  $P(x > \frac{71}{361}) = 0.01$

46 / 86

# An Alternative to the P-value: Standardized Value of a Statistic

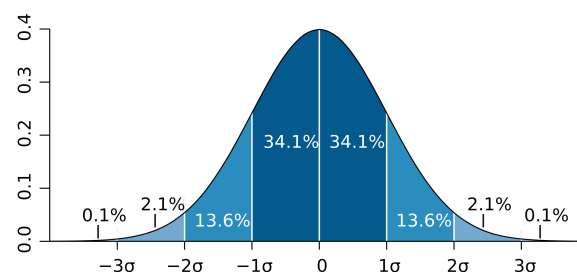
**Standardized statistic** - a measure of how far an observed statistic is from the mean of the distribution

- Commonly denoted by  $z$
- $z = \frac{\text{Statistic} - \text{Mean of null distribution}}{\text{standard deviation of null distribution}}$

47 / 86

# An Alternative to the P-value: Standardized Value of a Statistic

**Standard Deviation** - a measure of the distance a "typical" value is away from the mean



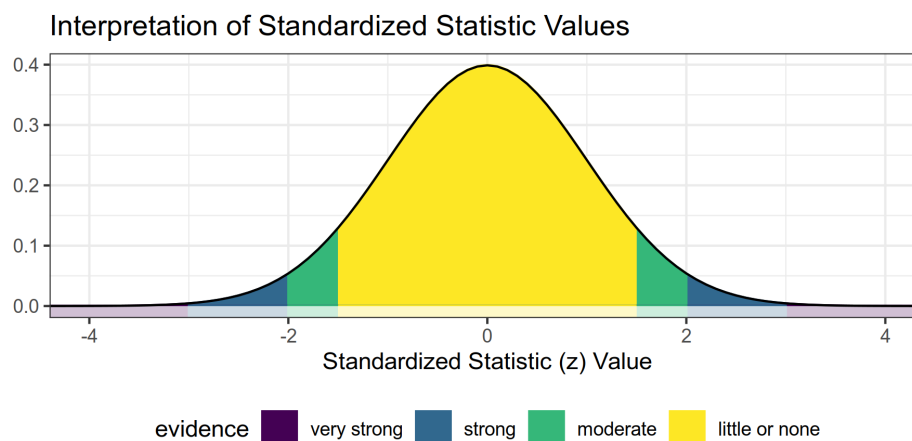
Observations more than 2 or 3 standard deviations from the mean are "in the tail of the distribution" or extreme

48 / 86



# Guidelines for Interpreting Standardized Statistics

| z-value range      | Interpretation                             |
|--------------------|--|
| $ z  < 1.5$        | <b>little or no</b> evidence against $H_0$ |
| $1.5 \leq  z  < 2$ | <b>moderate</b> evidence against $H_0$     |
| $2 \leq  z  < 3$   | <b>strong</b> evidence against $H_0$       |
| $3 \leq  z $       | <b>very strong</b> evidence against $H_0$  |



49 / 86

## An Alternative to the P-value: Standardized Value of a Statistic

**Standardized statistic** - a measure of how far an observed statistic is from the mean of the distribution

- Commonly denoted by  $z$

- $$z = \frac{\text{Statistic} - \text{Mean of null distribution}}{\text{standard deviation of null distribution}} \quad (\text{Equation 1.5})$$

Indicates how many standard deviations from the observed value to the hypothesized process mean

50 / 86

## Ex: St. George

$n=10$

- **Statistic:**
- **Mean of  $H_0$ :**
- **Std. Dev of  $H_0$ :**
- **z**

$n=361$

- **Statistic:**
- **Mean of  $H_0$ :**
- **Std. Dev of  $H_0$ :**
- **z**

51 / 86

## Ex: St. George

$n = 10$

- **Statistic:** 0.8
- **Mean of  $H_0$ :** 0.15
- **Std. Dev of  $H_0$ :** 0.1134587

$$z = \frac{0.8 - 0.15}{0.113} \approx 5.729$$

$n = 361$

- **Statistic:**  $\frac{71}{361} = 0.1966759$
- **Mean of  $H_0$ :** 0.15
- **Std. Dev of  $H_0$ :** 0.0183464

$$z = \frac{0.197 - 0.15}{0.018} \approx 2.544$$

52 / 86

# Standardized Statistic

- Applet will calculate  $z$  as well
- Normally use \_\_\_\_\_ as a single cut-off
  - $z$  can be \_\_\_\_\_
  - The \_\_\_\_\_ the standardized statistic is from zero, the stronger the evidence against the null model

## Summary - Standardized Statistics

Standardized statistic far from 0 → \_\_\_\_\_ p-value → \_\_\_\_\_

Standardized statistic close to 0 → \_\_\_\_\_ p-value → \_\_\_\_\_

53 / 86

# Standardized Statistic

- Can calculate both p-value and standardized statistic
- Quizzes/Exams may have either one, or both
- P-value and standardized statistic should always give the same conclusion
- Standardized Statistic takes into account sample size and spread of the data - helpful to compare multiple samples with different sample sizes

## What you need to find Standardized Statistic

- Observed statistic (from the sample)
- Mean of null distribution
- Standard deviation of null distribution

May be given to you in the question, from the applet, or from equation 1.5

$$z = \frac{\text{Statistic} - \text{Mean of null distribution}}{\text{standard deviation of null distribution}} \quad (\text{Equation 1.5})$$

54 / 86

## 1.4: What Impacts Strength of Evidence?

55 / 86

### Bob vs. Tim



Is the man on the left named Tim or Bob?

# What Impacts the Strength of Evidence?

- Bob/Tim: 85% of people say Tim is on the left
  - Compared to .5 as the null (chance) value
- What happens as the statistic gets farther away from .5?
  - **p-value** gets \_\_\_\_\_
  - **standardized statistic** gets \_\_\_\_\_
- Closer to 0.5?
  - **p-value** gets \_\_\_\_\_
  - **standardized statistic** gets \_\_\_\_\_

57 / 86

# What Impacts the Strength of Evidence?

# 1: Difference between the statistic and the null hypothesis

- Farther apart: Random chance/null is less plausible
  - p-value gets smaller
  - standardized statistic gets larger
- Closer together: Random chance becomes more plausible
  - p-value gets larger
  - standardized statistic gets smaller

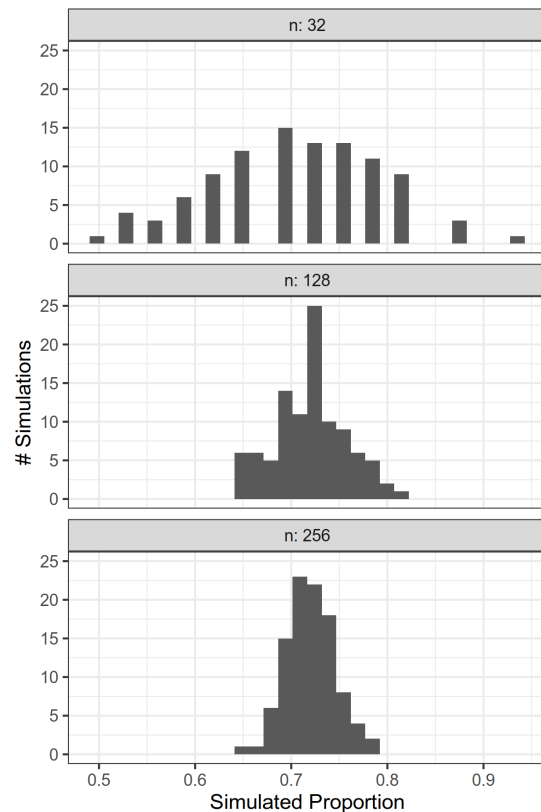
58 / 86

# What Impacts the Strength of Evidence?

## # 2: Sample size

- What happens as sample size increases (holding the statistic constant)?
- Depends on standard deviation, variability, and the spread of the data

Assume the statistic is 0.6. What happens to the strength of evidence against the null as the sample size increases?



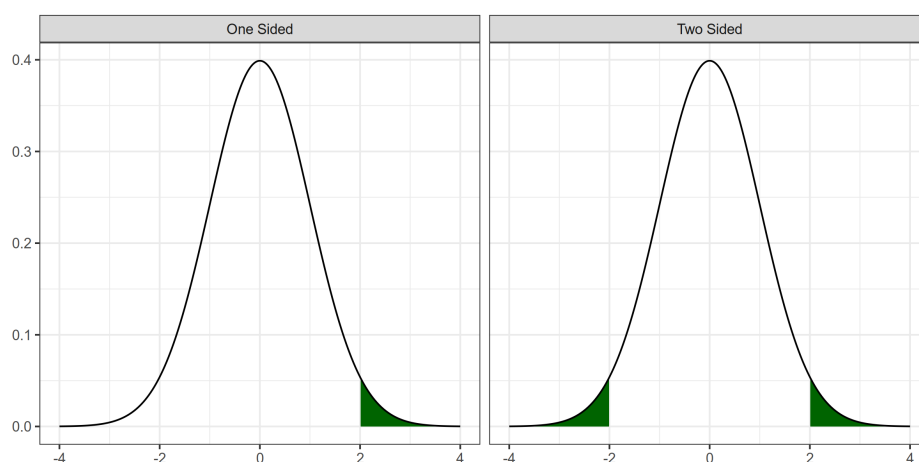
59 / 86

# What Impacts the Strength of Evidence?

## # 3: One-sided vs. Two-sided alternatives

Two-sided alternative:

- don't care about the direction of the effect - only care about the magnitude difference from the null



- Using a two-sided hypothesis causes p-values to \_\_\_\_\_

60 / 86

# Scenarios Worksheet

- Write the hypotheses in symbols for all five scenarios
- Choose 1 to complete a simulation on. Find the p-value, standardized statistic, and draw a conclusion.

This is practice, so ask questions!

## Due dates

- Exploration 1.2 is due at 6pm on Friday, Jan 24
- WileyPlus Chapter 1 Homework is due Feb 3 at 6pm
- Quiz over Ch. P and Ch. 1 is due Feb 5 at 11:59 pm

61 / 86

## Review

62 / 86

# Review

- Standardized Statistic
- What impacts strength of evidence? - Distance between hypothesis and statistic
  - Sample size
  - One sided vs. Two sided tests

63 / 86

## Standardized Statistic

- We can't use raw values to determine significance

Example: Determining if coins are fair

- Flip a coin twice and get heads both times
  - $H_0 : \pi = 0.5$
  - $\hat{p} = 1$
  - $|\hat{p} - \pi_0| = 0.5$
- Flip a coin 100 times and get heads 80 times
  - $H_0 : \pi = 0.5$
  - $\hat{p} = .8$
  - $|\hat{p} - \pi_0| = 0.3$
- Standardization allows us to add some context (the mean and spread of the null distribution) to our estimate

64 / 86



# Standardized Statistic



- Length of I-80 traveling across NE: 455.3 miles
- Distance from LA to Hawaii: 2491 miles

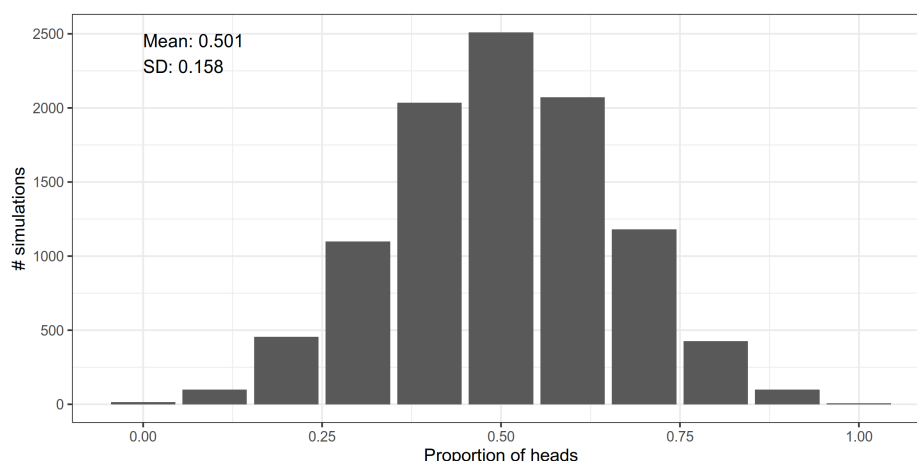
How many times could you fit the state of Nebraska in between LA and Hawaii?

Roughly 5.47 times

65 / 86

## Standardized Statistic Practice

- Flip a coin 10 times and get heads 8 times
  - What value is  $H_0$  centered at?
  - How many standard deviations could fit between the null hypothesis value and the observed value?
  - Is this value significant?



66 / 86

# What Impacts Strength of Evidence?

- Strength of evidence - difference between the observed statistic and the null hypothesis

Which would give us more evidence against  $H_0$ ?

- Flipping a coin 100 times and getting 55 heads
- Flipping a coin 100 times and getting 70 heads

67 / 86

# What Impacts Strength of Evidence?

- Strength of evidence - difference between the observed statistic and the null hypothesis

Which would give us more evidence against  $H_0$ ?

- Flipping a coin 5 times and getting 4 heads
- Flipping a coin 1000 times and getting 800 heads

68 / 86

# What Impacts Strength of Evidence?

## One Sided vs. Two Sided

- One Sided
  - Testing whether we get heads more often than tails
  - $H_0 : \pi = 0.5, H_A : \pi > 0.5$
- Two Sided
  - Testing whether a coin is fair
  - $H_0 : \pi = 0.5, H_A : \pi \neq 0.5$

69 / 86

## One Sided vs. Two Sided Examples

Decide whether the test should be one sided or two sided and determine what  $H_A$  should be

- My favorite outside temperature is 75° F. Has the average temperature this year been significantly different from that?
- In 2009, a dog brought to a shelter had about a 55% chance of being adopted or returned to its owner. In 2017, the chance of adoption or return to owner was almost 70%. Has the probability of a good outcome for shelter pups increased significantly?
- A government official claims that the dropout rate for local schools is 25%. Last year, 190 out of 603 students dropped out. Is there enough evidence to reject the government official's claim?

70 / 86

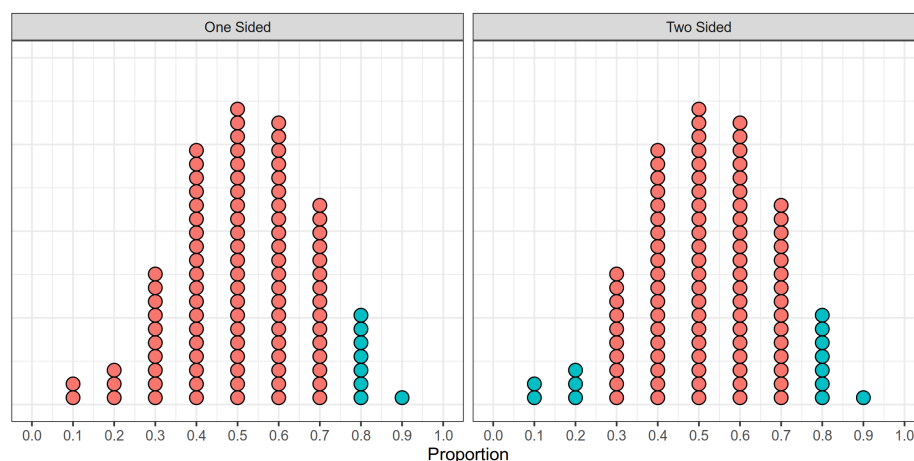
# One Sided vs. Two Sided Examples

Decide whether the test should be one sided or two sided and determine what  $H_A$  should be

- My favorite outside temperature is  $75^\circ$  F. Has the average temperature this year been significantly **different** from that?  
**Different**  $\rightarrow H_A : T \neq 75 \rightarrow$  **two-sided test**
- In 2009, a dog brought to a shelter had about a 55% chance of being adopted or returned to its owner. In 2017, the chance of adoption or return to owner was almost 70%. Has the probability of a good outcome for shelter pups **increased** significantly?  
**increased**  $\rightarrow H_A : \pi > .55 \rightarrow$  **one-sided test**
- A government official claims that the dropout rate for local schools **is** 25%. Last year, 190 out of 603 students dropped out. Is there enough evidence to reject the government official's claim?  
**is (equal to)**  $\rightarrow H_A : \pi \neq .25 \rightarrow$  **two sided test**

71 / 86

## One-Sided vs. Two-Sided tests



- There are 100 dots. To find the p-value, we count the dots which are as or more extreme than what we observed (0.8)
- For a 1-sided test of whether we observed more heads (0.8, 0.9, and 1.0),  $p = 0.03$
- For a 2-sided test of whether the coin is fair (0.8, 0.9, 1.0, and 0.2, 0.1, 0.0),  $p = 0.07$

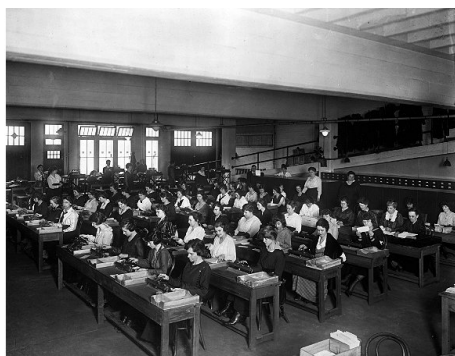
72 / 86

## 1.5: Inference for a Single Proportion: Theory-Based Approach

73 / 86

### Simulations vs. Theory

- Simulation provides an easy method for comparing a statistic to a model representing random chance, but it requires a lot of computer power



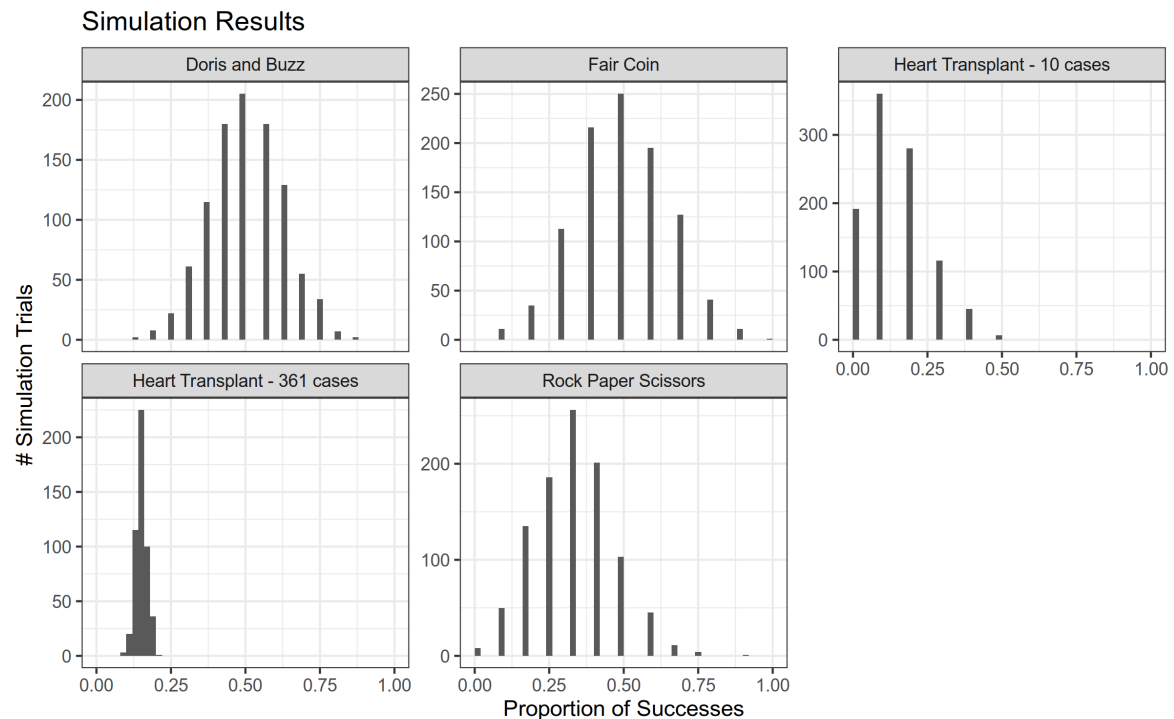
A room full of "computers" during WWII

- Historically, statisticians didn't have enough computing power for simulation, so they developed theory-based approaches to find p-values and standardized statistics

74 / 86

# Distributions

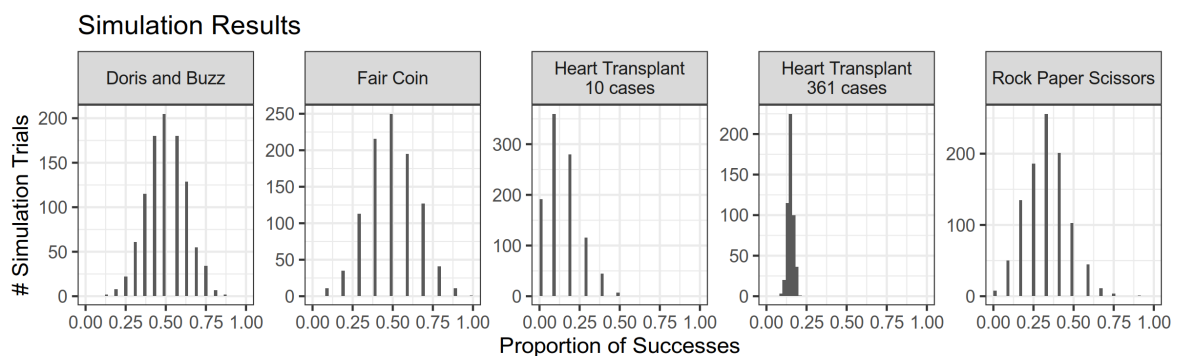
What do the following pictures have in common? What makes the exceptions different?



75 / 86

## Theoretical Null Distributions

Most of the null distributions we've seen so far have been fairly symmetric and "bell shaped"



The not symmetric distribution has  $p = 0.15$  and only 10 cases. In the other examples with small  $n$ ,  $p$  is 0.333 or 0.5.

76 / 86

# Theoretical Null Distribution

When  $n$  is "big enough"\*, we can use theoretical approximations instead of simulations.

Using a theoretical distribution, we can calculate p-values and standardized statistics, if we have

1. the mean of the distribution
2. the theoretical standard deviation of the distribution
3. a good idea of whether the assumptions hold

2 (and sometimes 1) will require formulas and calculations.

\*More on this in a couple of slides

77 / 86

## Theory Based approach

Based on the **Central Limit Theorem** (see [Wikipedia](#) for an illustration)

If the sample size  $n$  is large enough, the distribution of the *sample proportion* will be approximately bell shaped, centered at the long-run probability  $\pi$  with the standard deviation of the null distribution.

The standard deviation of the null distribution,  $\sigma$ , will also be used in the standardized statistic calculation.

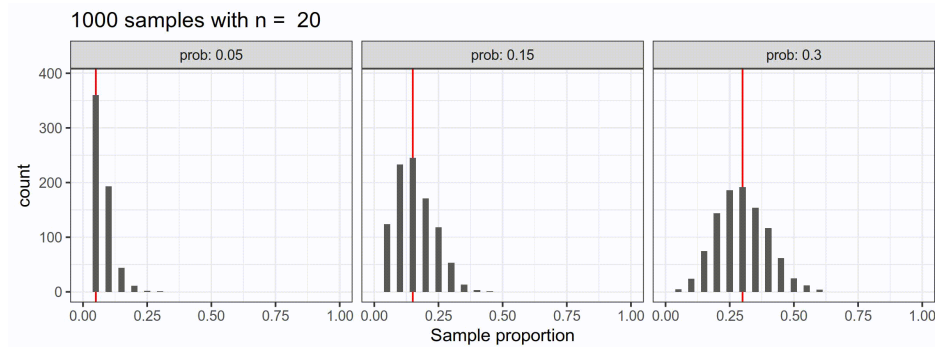
We need a formula for  $\sigma$  because we are no longer doing simulation.

78 / 86

# How large is large enough?

- Our book: if the number of successes and failures are at least 10
- Others:  $n > 30$ ,  $n > 50$ , ...
- Depends on the discipline (and likely values for  $p$  within the discipline)

In the animation below, when does each distribution appear to be centered around the red line?



79 / 86

## Example

St George:  $\pi = 0.15$ ,  $\hat{p} = 0.197$ ,  $n = 361$

New SD? Standardized Statistic?

Simulation approach: SD = 0.018 and standardized statistic = 2.544

Simulation results could be different for everyone.

80 / 86



# Normal Distribution

- Replaces the simulated null distribution
- Fits curve to what we'd expect to see in the data given random chance
- Used to find p-values

Technically, we'd need calculus to find p-values, but the applet will do it for us.

- We can see if it's a good approximation or not by comparing to the simulation results

The prediction is valid if  $n\pi \geq 10$  and  $n(1 - \pi) \geq 10$   
(**Validity condition** for 1-proportion problems)

For proportions, the normal distribution has a mean of  $\pi$  and a standard deviation of  $\sqrt{\pi(1 - \pi)/n}$

81 / 86

## Example 1.5: Halloween Treats

Do children show any preference for candy vs. toys?

Measure the proportion of children choosing candy, compare to  $\pi = 0.5$

- $H_0 :$
- $H_A :$

Data:

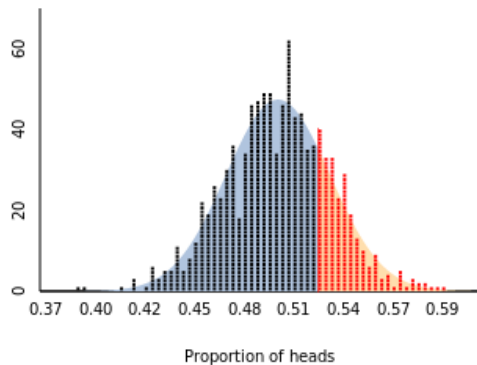
- Sample Size  $n = 283$
- Sample Statistic  $\hat{p} = 148/283 = 0.523$

Use the app to find the standard deviation of the null distribution. How does it compare to the theoretical standard deviation?

82 / 86

## Example 1.5: Halloween Treats

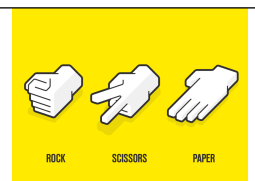
Does the distribution look like one where a bell shaped curve will approximate it?



| Quantity | Theory | Simulation |
|----------|--------|------------|
| SD       |        |            |
| Z-score  |        |            |
| Decision |        |            |

83 / 86

## Example: Rock Paper Scissors



$n = 12$ , with 10 successes and 2 failures

- Are the validity conditions met? Why or why not?
- Normal distribution predictions:
  - bell shaped, approximately normal
  - null centered at .333
  - standard deviation will be:
- What happens with more/less skew?
- What happens with a smaller/larger sample size?

84 / 86

## Summary: Chapter 1.5

- If you want to use a theory based approach:
  - need a sample size large enough to say  $H_0$  follows a normal distribution
  - Applet: Normal approximation button (calculates p-value for you)
  - Can always calculate standardized statistic by hand
- Theory based approach doesn't always work
  - if the sample size is small, null distribution will not be approximately bell shaped
  - if the distribution is not bell shaped, the normal approximation to the p-value and the standardized statistic are not accurate
- Simulation ALWAYS works

85 / 86

## Reminders

- Ch 1 HW due on Canvas Feb 3 at 6pm
- Quiz over Ch P and Ch 1 is due Feb 5. Please work independently on the quizzes.
- Exam 1 is Feb 13 during class and covers Ch P, Ch 1, and Ch 2.
- Read Section 2.1 for Feb. 4

86 / 86