Chapter 2: Testing for a Proportion

Investigation: Someone claims to have Extra-sensory perception (ESP). They claim they can perceive things before they happen. We'd like to put their ESP to the test by having them guess the **color** of **16** randomly drawn cards in a row.

In a regular playing deck, half of the cards are red and half of the cards are black. Let's collect some data to see how many cards this individual can guess (perceive?) correctly before we flip them!

Trial	Correct
1	
2	
3	
4	
5	
6	
7	
8	
9	
10	
11	
12	
13	
14	
15	
16	

They guess ____ out of **16** correct.

Based on the data, is there evidence to support this person's claim that they have at least some advantage in guessing the card?

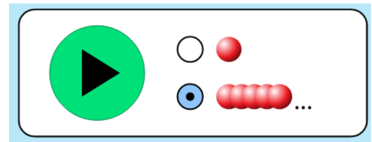
What percent of the time do ***you think*** someone would guess ***at least*** that many cards correctly by chance if they were guessing randomly for each card? Just estimate!

What is the **minimum** number of cards out of 16 that someone would have to guess correctly in order for **you** to believe they really had an advantage in guessing cards today? 11? 12? 13? 14? 15? All 16?...

Let's use "Plinko Probability" to help us navigate this question. *Use this link, or web search "Phet, plinko probability."*

- https://phet.colorado.edu/sims/html/plinko-probability/latest/plinko-probability_en.html
- Choose the "Lab screen" and play around with it for a minute

Mini investigation: Set Rows to 2 (creating bins labeled 0, 1, and 2) and keep binary probability at 0.50. **What percentage of plinko balls land in the 0 bin? The 1 bin? The 2 bin?**



- Note that you can choose to run balls one at a time, or let them run continuously.
- Why do you think the plinko balls land where they do as often as they do?

(outline the 3 bins and write LL, RR, LR, and RL to show which combinations result in which bins)

Mini investigation: Now adjust to 4 rows. What do you notice about where the plinko balls land? **Are they equally likely to land in any bin, or are some more likely than others?**

(outline the 5 bins and write LLLL, RRRR, and some others to show how effect exacerbates with more bins)

Our Investigation: Just as we ran an investigation with 16 cards, let's set a plinko board with 16 rows and a binary probability of 0.50.

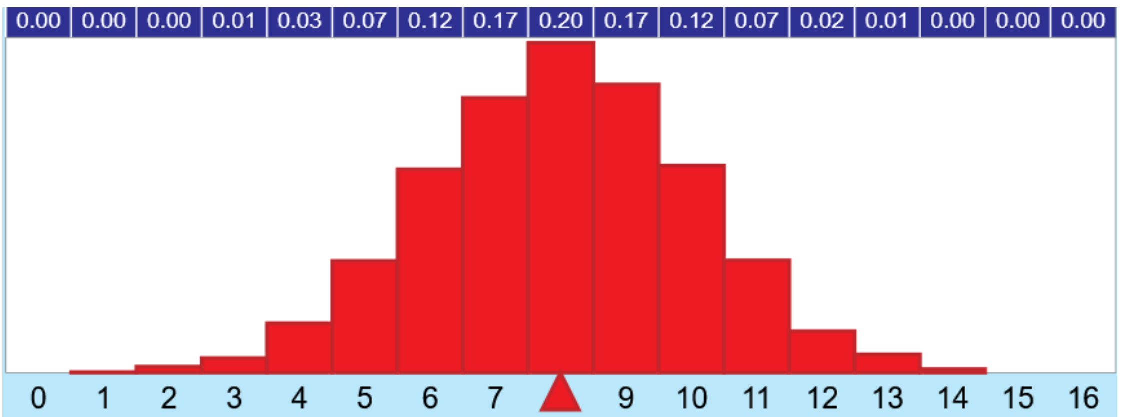
- 16 rows because…we had the student guess 16 cards
- Binary probability of 0.50 because…this shows us what would happen if student had 50% guess rate

Can we use this simulation to help complete our investigation?

The Language of Hypothesis Testing

- **Null Hypothesis:** Null means <u>nothing.</u> A Null Hypothesis assumes no change, no difference, or no association in the situation we're studying.
    - **In our investigation**…the "Null Hypothesis" would be that…<u>this person guessed cards with no advantage. They are guessing with 50% accuracy.</u>
    -

- **Null Model:** A set of possible results that could happen under a Null Hypothesis.
    - **In our investigation**…Our <u>histogram</u> was a null model, representing how often someone would guess various numbers of cards correctly if each card is a <u>50%</u> guess.



- **P-value:** The probability of getting a **sample** result <u>at least this far</u> from expectation by random chance if the Null Hypothesis is true.
    - **In our investigation**…The probability that someone would guess at least ____ cards correctly by random chance—if we assume a <u>50%</u> accuracy rate—is approximately _______

Challenge Question: If this person had guessed fewer cards correctly, how would the p-value change?

Read on your own

- A **Null Hypothesis** should represent a **general** event rather than a specific sample event
    - This person will guess 50% of these 16 cards correctly (right / <u>wrong</u>)
    - If this person repeatedly guessed cards, they would guess 50% correct (<u>right</u> / wrong)
- A **P-value** is a way for us to judge the **plausibility** of a particular sample result under the Null Hypothesis.
    - A p-value approaching 0 means that these sample results are almost impossible to observe under the Null Hypothesis—like if someone guessed all 16 cards correctly!
    - The closer our sample results are to expectation, the higher the p-value and the more compatible our sample results are with the Null Hypothesis.

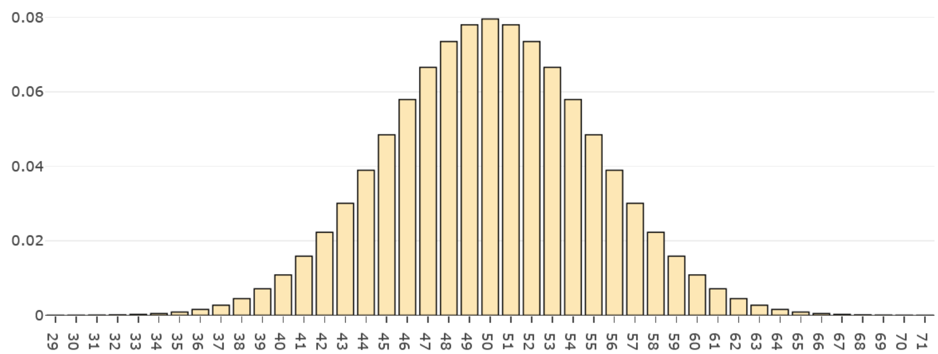
Investigation: We hammer a quarter into a new shape, creating a visible dent. We'd like to test whether this coin's probability of landing heads or tails has changed as a result. We flip this coin 100 times. We observe 58 heads and 42 tails. Would this be evidence that the coin is now biased? Or might this be a reasonable result to observe if the coin were still fair?

What is the Null Hypothesis in this investigation?

Now that we have a sample size of 100, we need to use a larger simulation. Let's use this simulation from "Art of Stat": https://istats.shinyapps.io/BinomialDist/

- Go to the "Find Probabilities" Tab up top
- Adjust "Bernoulli trials" to our sample size
- Adjust "Probability of success" to match our Null Hypothesis
- Set "Type of Probability" to reflect the direction of our investigation
- Set "Value of x" to our particular sample result

According to the Null Model, how often would we see *at least* 58 heads out of 100 by random chance?



Investigation Reconceived: Truthfully, we started this investigation without knowing whether the dent would favor heads or tails.

Should this change how we evaluate the unusual-ness of our sample result?

Binomial Probabilities

If you're curious how we can calculate the probability of observing x "successes" out of "n" trials, check out the "Formulas and Properties" tab on the sim!

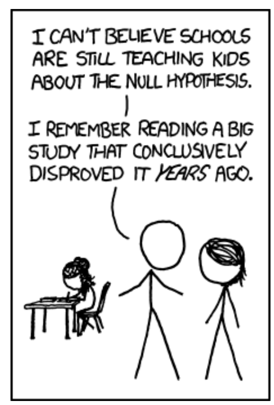
Making sense of Directional and Non-Directional Investigations

- **Directional Investigations:** Our results would only matter (or support our theory) if they are specifically lower or specifically higher than Expectation.  
    - **ESP Example**: If someone claims to have ESP, then we'd only find the sample results compelling if it was <u>higher</u> than expectation.

- **Non-Directional Investigations:** Our results could matter (or support our theory) if they demonstrate a departure in either direction from Expectation. 
    - **Dented Coin Example:** We would have found a result of 42 heads and 58 heads <u>equally</u> compelling in suggesting a possible bias.

Writing Null and Alternative Hypotheses

- When completing a hypothesis test, there are two scenarios we are weighing:
    - The Null Hypothesis is True
    - The Null Hypothesis is False and some Alternative is True



- **The Null Hypothesis:** There is **<u>no</u>** difference or effect.
    - The term "Null" literally means "Nothing." We use that term for a reason! It is the idea that **nothing** of importance **is happening**. Think of it as the status quo.
    - We abbreviate the null hypothesis with $\mathbf{H_0}$ (pronounced "H not")

- **The Alternative Hypothesis:** The parameter is some alternative value. There is at least <u>some difference or effect.</u>
    - In many investigations, the **alternative** hypothesis represents **our theory**. We are testing whether there is evidence for some departure from the status quo.
    - We abbreviate the alternative hypothesis with $\mathbf{H_A}$

- Remember that **Hypotheses** are **statements** about a **parameter.**
    - We are not hypothesizing about the value of our sample proportion, $\hat{p}$ (we know that!)
    - We are hypothesizing about the population proportion, $\boldsymbol{\pi}$

Non-Directional Case: Dented Coin	**Directional Case:** ESP Test

- **P-values for Binary Decisions – Assume the Null unless evidence otherwise!**
  - In some cases, we want to make a simple decision:
    - **Reject the Null:** Our p-value is low. The Null is not compatible with our sample results
    - **Fail to Reject the Null**: Our p-value is not low. Our sample results could reasonably take place under the Null Hypothesis.
  - We often set a significance level (represented as α) to represent our cut-off point.
    - A common choice is $\alpha = 0.05$
    - If the p-value is at or below α…Reject the Null
    - If the p-value is above α…Fail to Reject the Null

Practice: What should we decide in the biased coin example if we use $\alpha = 0.05$ as our benchmark comparison?

Challenge Question: Why wouldn't we "conclude" or "accept" the null hypothesis when our p-value is above α?

- **P-values as insights**
  - There are times to make binary decisions from hypothesis tests, but we can still regard lower p-values as stronger evidence and higher p-values as weaker evidence.
    - https://apnews.com/12cf3d07354c47b3b9bb552776071522
    - The table below provides **suggested interpretations** for different p-value ranges.

P-value	**Suggested Interpretation**
P > 10%	Weak or Little evidence (against the null) / (for the alternative)
5-10%	Modest evidence (against the null) / (for the alternative)
1-5%	Strong evidence (against the null) / (for the alternative)
P < 1%	Very strong evidence (against the null) / (for the alternative)

- Making Errors
  - **Type I Error:** Incorrectly rejecting the null hypothesis (concluding a difference when there really is none).
    - When the Null hypothesis is true, the probability of making a Type I error is what we set our $\underline{\alpha}$ to.
  - **Type II Error:** Incorrectly "failing to reject" the null hypothesis (failing to conclude a difference when there really is a difference).
    - Several things can affect the likelihood of making a Type II error. They are most likely to happen when:
      - Our sample size is small.
      - The true departure from the Null is small.
      - We set a significance level very low.

	Null is really True	*Null is really False*
Fail to Reject Null	Correctly "failing to reject"	**Type II Error**
Reject Null	**Type I Error**	Correctly reject

Practice: For each example, identify whether a Type I error, Type II error, or neither error was made.

An early study looked at the effectiveness of Remdesivir in lowering the mortality rate among those hospitalized with COVID-19. The small sample study did not have a low enough p-value to conclude it was more effective than standard treatment, but a larger study conclusively found Remdesivir truly lowered the mortality rate.

The null hypothesis for the small sample study was…

Did they make an error?

Many researchers have studied the use of Chicken Noodle Soup as a potential cure for the common cold. One researcher found that those eating chicken noodle soup had faster recovery times than the general population of cold sufferers, finding a p-value below $\alpha = 0.05$. But in reality, chicken noodle soup offers no benefit.

The null hypothesis for the small sample study was…

Did they make an error?

In the dented coin example, let's say the dent has not actually changed the probability of getting heads or tails. In our investigation, we failed to reject the null hypothesis

Did we make an error?

Chapter 2 Additional Practice

Bottlenose Dolphins are considered among the most intelligent animals. A group of researchers would like to test their color recognition and memory.

19 dolphins were shown a panel of four buttons of different colors. One of the four would light up, and then the dolphin would be guided to swim to the other side of the pool to an identical panel of four colored buttons. The dolphin was given a treat if it pressed the same colored button from the other side. This was repeated several times to help the dolphin make the association.

The next day, they repeated this activity with the 19 dolphins—for each dolphin, one of the four colors lit up, and then the dolphin was directed to the other identical panel to see if the first button pressed was of the same color. The biologist would like to know if the dolphins are performing *better* than what we would expect by random guessing.

The sample result showed that 12 out of 19 dolphins guessed on the first time correctly.

Write the null and alternative hypotheses for this investigation? *Is it directional or non-directional?*

Use the "Art of Stat" simulation to simulate the null model: https://istats.shinyapps.io/BinomialDist/

- Go to the "Find Probabilities" Tab up top
- Adjust "Bernoulli trials" to our sample size. Our sample size is…<u>19</u>
- Adjust "Probability of success" to match our Null Hypothesis. The null hypothesized proportion is…<u>0.25</u>
- Set "Type of Probability" to reflect the direction of our investigation. Type of probability should be…<u>upper tail</u>
- Set "Value of x" to our particular sample result. The sample number correct is…<u>12</u>

What is the p-value for our investigation? What would be an appropriate conclusion to make?

Now, imagine that we made a mistake, and it was actually 13 dolphins, not 12, who guessed correctly. Would this change our p-value? *(i.e., is our evidence stronger or weaker now?)*

Chapter 2 Learning Goals

After this chapter, you should be able to…

- Identify a Null Hypothesis in a particular situation as representing the *general* result of no difference, no change, or no association at the population level
- Use a Null Model to interpret the likelihood of different sample results under a Null Hypothesis
- Recognize a p-value as the probability of observing a particular sample result that is at least this far from expectation under the Null Hypothesis
- Recognize π as a symbol representing the proportion of "successes" at the population level
- Recognize $\hat{p}$ as the proportion of "successes" observed in a sample of results
- Write a Null and Alternative hypothesis involving a test for a proportion
- Distinguish between directional and non-directional questions, including how to write hypotheses for each and how to identify/interpret a p-value for each
- Use p-values to judge whether a Null Hypothesis is reasonable or not based on observed data
    - Using a significance level (α) to make a binary decision (Reject Null or Fail to Reject the Null)
    - As a stand-alone probability value, where the lower the value, the stronger the evidence against the Null
- Identify whether a Type I or Type II error might have occurred based on a described result, or whether no error has occurred. *Note that you don't need to memorize the difference. The table will be provided on the exam guide!*