MA256 Lesson 3 - Significance - Strength of Evidence (1.1-1.2)

Defintions review

Statistical Significance

Statistical Significance indicates the strength of evidence that the observed result was unlikely to have occurred by random chance

What is the 3S Strategy?

Statistic, Simulate, Strength of Evidence

What is the difference between a **parameter** and a **statistic**?

A parameter is the long-run behavior of a random process and it is (generally) not observable. A statistic is the summary of observed results and is used to infer about the parameter.

Define:

H_0 :

Null Hypothesis, the *by-chance-alone* explanation. Use the null model or the chance model to reflect this hypothesis.

H_a : Alternate Hypothesis, the *there is an effect* explanation that contradicts the null hypothesis. Researchers hope this hypothesis is supported by the data they collect.

π : The proportion parameter, the long-run behavior of a random process. It is (typically) unobservable.

$\hat{p}$: The observed proportion statistic from data, used to argue or infer about the parameter.

n : The sample size; The number of observations within each sample to construct the statistic.

p-value :

The probability of observing an observation at least as extreme as $\hat{p}$ assuming the null hypothesis is true.

Classify the strength of evidence for each range of p-values.

0.1	$< p$		Weak Evidence against the null
0.05	$< p \leq$	0.1	Moderate Evidence against the null
0.01	$< p \leq$	0.05	Strong Evidence against the null
	$p \leq$	0.01	Very Strong Evidence against the null

Chips Ahoy vs Keebler Original Chocolate Chip Cookies

Chips Ahoy! conducted a blind taste-test experiment with 50 randomly selected consumers and recorded whether they preferred Chips Ahoy! over Keebler's original chocolate chip cookies. Of those 50 consumers, 45 (90%) of them preferred Chips Ahoy! Chips Ahoy! now claims that 90% of consumers prefer their cookies over Keebler's.

Step 1: Ask a Research Question: Q1) Restate from the given information.

Do people prefer Chips Ahoy! over Keebler's chocolate chip cookies?

Step 2: Design a Study & Collect Data: Q3) Explain what type of study was conducted.

(single) blind study with 50 randomly selected consumers. Each participant was asked if they preferred Chips Ahoy! or Keebler's

Step 3: Explore the Data: Q4) What could we do with the data to help us understand it better?

We could calculate the summary statistics (already done in this case). We could create some figures to get a better idea of the data. Two, side-by-side barcharts or a stacked barchart would probably be best.

Step 4: Draw Inferences: Significance (today's lesson!)

To draw inferences for this or any other research question, we must set the framework by answering the following questions. Do this at boards in teams of 2-3:

Q5) What are the observational units?

The randomly selected consumers who participated in the blind taste test experiment.

Q6) What is the sample size and number of samples?

The sample size (n) = 50. There was only one sample of 50 consumers.

Q7) Identify & classify the variable collected.

The variable is whether a consumer preferred Chips Ahoy! It's a binary categorical variable (Yes or No)

Q8) What is the symbol, definition & value of the sample statistic?

$\hat{p}$ (p-hat) represents the short-run proportion of consumers from the experiment who preferred Chips Ahoy's original chocolate chip cookies over Keebler's original chocolate chip cookies.

NOTE: We can ALWAYS calculate this value from the data using R if not given. $\hat{p} = 45/50 = 90\%$ *or* 0.9.

Q9) What is the symbol, definition, & value of the population parameter?

π represents the long-run proportion of consumers who prefer Chips Ahoy's original chocolate chip cookies over Keebler's original chocolate chip cookies.

NOTE: This is an intuitive random by-chance value based on the context of the problem. For example, if people are just randomly choosing between two cookies, the long-run probability of them choosing the Chips Ahoy! cookie is 1 out of 2, or 50%, so $\pi = 1/2 = 50\%$ or 0.5.

Q10) What are the null & alternate hypotheses in words and symbol notation?

Null Hypothesis: The long run proportion of consumers who prefer Chips Ahoys original chocolate chip cookie over Keeblers original chocolate chip cookie is 50%. $H_0 : \pi = 0.5$

Alternate Hypothesis: The long run proportion of consumers who prefer Chips Ahoy's original chocolate chip cookie over Keebler's original chocolate chip cookie is greater than 50%. $H_a : \pi > 0.5$

Q11) With the framework established, we can use the 3S strategy and *One Proportion* applet (http://www.isi-stats.com/isi2nd/ISIapplets2021.html) to estimate the likelihood that Chips Ahoy got the result they did if consumers could not taste a difference and were randomly choosing one of the two cookies.

1. Statistic: Write out the statistic: $\hat{p} = 45/50 = 90\%$ *or* 0.9

2. Simulate: Run the simulation using the *One Proportion* applet.

Probability of choosing Chips Ahoy! (success/heads) = π (Parameter) = 0.5

Number of tosses = n (Sample size) = 50

Number of repetitions = 1,000

Select "Draw Samples" button.

3. Strength of Evidence:

Q12) How common/uncommon is our observation? Report the p-value. Select the "Count samples" "Count" button to calculate the P-value for this simulation. (What number should you use for the *As extreme as* value?) Did we prove the alternative hypothesis?

We can see subjectively that our statistic (red vertical line) is very uncommon, so there is very strong evidence that our statistic is extremely unlikely to have occurred by random chance. the reported p-value should be around 0. No, we have not proven the alternate hypothesis - it is still possible to get this result from chance, even if unlikely.

2. Simulate (REDUX): Run the simulation using R. For help with a `for` loop, see https://www.rdocumentation.org/packages/openintro/versions/2.4.0/topics/loop or type `?Control` in the console.

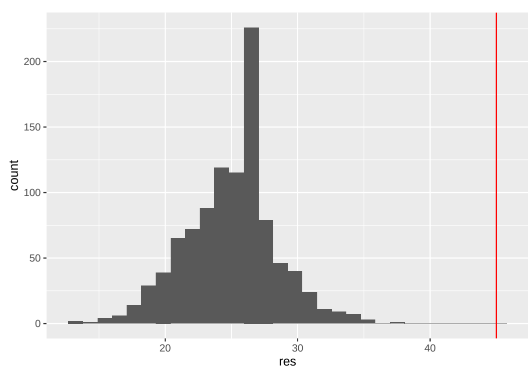
```
set.seed(256)

M <- 1000           # number of replications
mypi <- 0.5         # population parameter (under the null hypothesis)
n <- 50             # sample size
n.succ <- 45        # number of successes
myphat <- n.succ / n    # observed statistic (proportion of "successes")

RES <- data.frame(res = rep(NA, M))  # create a data frame to hold the results of the simulation

for(rep in 1:M){
  myobs <- rbinom(1, n, mypi)
  RES$res[rep] <- myobs
}

# plot a histogram of the results
RES %>% ggplot(aes(x = res)) + geom_histogram() +
  geom_vline(xintercept = n.succ, color = "red")
```




```
# estimate p-value
sum(RES$res >= n.succ)
```

```
## [1] 0
```

Q13) What can you conclude from this study?

We are extremely confident that the long run proportion of consumers who prefer Chips Ahoy! original chocolate chip cookies over Keebler's original chocolate chip cookies is greater than 50% but how much greater do people prefer Chips Ahoy over Keebler? In Lesson 7, we will calculate confidence intervals to estimate the range of values within which the true long run proportion lies based on 90%, 95% or 99% confidence.

Step 5: Formulate Conclusions:

Q14) Can you generalize to other brands? Can you identify any bias in the study?

Consumers seem to prefer Chips Ahoy's original chocolate chip cookies over Keebler's original chocolate chip cookies based on this single experiment, but can we generalize to the entire Chips Ahoy brand versus the entire Keebler brand? We will learn more about generalization in Lessons 5 & 6, but to answer this question – we should not be comfortable generalizing the results to the entire brand. It's possible that Keebler's Chewy or Chocolate Chunk cookies are better than Chips Ahoy's.

Is the cookie the reason people preferred one over the other, or could there have been bias in the research? Examples of intentional or unintentional bias could be:.

1. If researchers lead the experiment by saying, "Folks seem to prefer cookie #1 (Chips Ahoy) over #2 (Keebler), what do you think?", then consumers may be more likely to group-think and prefer cookie #1.

2. If cookie #1 is served on a different/better plate than cookie #2, it may subconsciously influence people to prefer that cookie (because of the plate).

3. Telling people "Cookie #1" vs "Cookie #2" could influence people to select cookie #1, just because we typically think of #1 as first place or the best.

4. Perhaps after taking a bite of the first cookie, people's palate isn't as sensitive to the flavors. Did researchers clean the palate before consumers tasted the 2nd cookie? If not, perhaps everyone thinks the 1st cookie tasted is better due to palate sensitivity to flavors. We will explore what conditions allow us to determine cause and effect in Lesson 8.

Step 6: Look Back and Ahead:

Q15) Identify limitations & suggest ways to overcome in new studies.

Structure the experiment as double-blind, and/or in additional ways to avoid bias examples above. Since Chips Ahoy! has something to gain from the results of the experiment, they are obviously biased. A neutral 3rd party should repeat the experiment to see if they get similar results. In fact, let's do that now.

A third competitor.

Q16) How would the analysis above change if we introduced a third cookie competitor and conducted the experiment in our classroom?

We would have a sample size of $n = 16 or 17$; p-hat = (changes based on outcomes) ; $H_0 = 1/3; H_a > 1/3$

Q17)Is it possible for the observed proportion to match the null hypothesis if the null hypothesis is false? If the null is true?

Yes, it is possible. x 2

Extra Problems:

1.1.10 Kevin Durant of the Golden State Warriors hit 721 of his 1,383 field goal attempts in the 2018/2019 season for a shooting percentage of 52.1%. Over the lifetime of Kevin's career, can we say that Kevin is more likely than not to make a field goal?

a. Is the long-run proportion of Kevin making a field goal a parameter or a statistic?

parameter

b. Is 52.1% a parameter or a statistic?

statistic

c. When simulating possible outcomes assuming the chance model, how many times would you flip a coin for one repetition of the 2018/2019 season?

1383

d. With each repetition, what would you keep track of?

The number of heads (or proportion of heads) out of 1,383 flips

e. What would be a typical value from a repetition of 1,383 coin flips? Justify your answer.

1/2 of 1,383 or about 691 or 692

1.1.11 If Kevin Durant of the Golden State Warriors hits 52 out of his first 100 field goals in the 2018/2019 season, let's see how we might investigate if he is more likely than not to make a field goal?

a. Based on these first 100 field goals, we want to find out what Kevin's long-run proportion of making a field goal is. Will this value be a statistic or a parameter?

parameter

b. Is 52 out of 100 a statistic or a parameter?

statistic

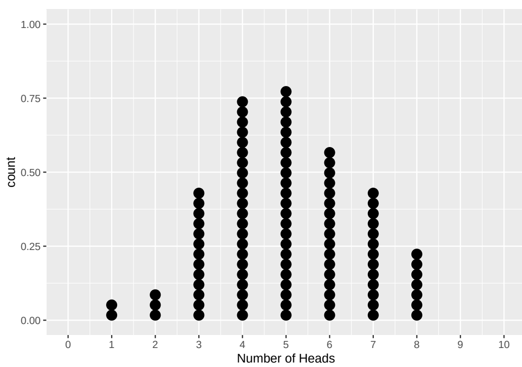
c. We can use the **One Proportion applet** (or use R to simulate) to generate 1,000 possible values of Kevin's last 100 field goals under the chance model that he has a long-run proportion of 0.50 of making a field goal. Match the aspects of the simulation in Column A to their equivalent aspects in the actual study listed in Column B.

Column A	Column B
Coin flip	Kevin misses his field goal
Heads	Long-run proportion of field goals Kevin makes
Tails	One set of 100 field goal shots by Kevin
Chance of heads	Kevin shoots a field goal
One repetition	Kevin makes his field goal

Column A	Column B
Coin flip	Kevin shoots a field goal
Heads	Kevin makes his field goal
Tails	Kevin misses his field goal
Chance of heads	Long-run proportion of field goals Kevin makes
One repetition	One set of 100 field goal shots by Kevin

1.1.14. Buttered Toast If you drop a piece of buttered toast on the floor, is it just as likely to land buttered side up as buttered side down? It sure seems like mine always lands buttered side down! Suppose that 7 of the last 10 times I dropped toast it landed buttered side down. In order to carry out a statistical analysis, the One Proportion applet was used to see whether my toast fell buttered side down a majority (more than 50%) of the time. Use the dotplot generated by the applet to answer (a)–(e).

```
## Bin width defaults to 1/30 of the range of the data. Pick better value with
## `binwidth`.
```



a. How many dots are in the dotplot?

100 dots

b. What does each dot represent in terms of dropped toast and buttered side down?

Each dot represents the number of times out of 10 attempts the toast lands buttered side down when the probability that the toast lands buttered side down is 0.50.

c. At what number is the dotplot centered? Could you have determined that before running the applet simulation? Why or why not?

5, because that is what will happen on average if the toast is dropped 10 times and 50% of the drops it lands buttered side down.

d. Based on 7 of the last 10 slices of toast landing buttered side down and the dotplot, are you convinced that the long-run proportion of times my toast lands buttered side down is greater than 0.50? Explain how you are deciding.

No, we are not convinced that the long-run proportion of times the toast lands buttered side down is above 0.50 because 7 is a fairly typical outcome for the number of times landing buttered side down out of 10 drops of toast when the long-run proportion of times it lands buttered side down is 0.50. Stated another way, 0.50 is a plausible value for the long-run proportion of times that the toast lands buttered side down based on getting 7 times landing buttered side down in 10 drops.

e. Does this prove that my long-run proportion of dropping toast buttered side down is 0.50? No, 0.50 is just a plausible (reasonable) explanation for the data. Other explanations are possible (e.g., the long-run proportion of times the toast lands buttered side down could be 0.60).