MA206, Lesson 6 - Strength of Evidence

Review: How is the standardized statistic of a single proportion calculated?

$$\mathbf{z} = \frac{\hat{p} - mean(null)}{SD(null)}$$

When can we use a theory-based approach to calculate a p-value?

When our sample size is large enough. For a one-proportion z-test, we must have at least 10 successes and 10 failures in our sample.

Using the theoretical approach, what is the expected standard deviation of a null distribution?

$$SD = \sqrt{\frac{\pi(1-\pi)}{n}}$$

Using the theoretical approach, how do we calculate our standardized statistic?

$$\mathbf{z} = \frac{\hat{p} - \pi}{\sqrt{\frac{\pi(1-\pi)}{n}}}$$

How do we convert our theoretical z-score into a p-value?

We could integrate, or use R depending on our null and alternate hypothesis.

$$H_a : \pi < \# \qquad \text{pnorm(z)}$$
$$H_a : \pi > \# \qquad \text{1 - pnorm(z)}$$
$$H_a : \pi \neq \# \qquad \text{2*(1 - pnorm(abs(z)))}$$

1) Two firsties miss recall formation because they partied hard over the weekend, but blamed their lateness on a flat tire. The TAC team brings them both into separate offices and asks them both one question which will determine if they get hours or not. Which tire went flat? This question works if we assume that each tire is equally likely to be chosen, but it has been proposed that people tend to answer "right front" more often.

To test this, we asked 28 cadets, if they were in this situation, which tire would they say had gone flat. Our results are shown below.

Left Front	Left Rear	Right Front	Right Rear
6	4	14	4

    a) What is the research question?

Do cadets pick the right front tire more often than other tires?

    b) Identify the observational units in this study.

Each of the 28 cadets asked

c) Describe the parameter of interest (in words).

The parameter π is the long run proportion of cadets who choose the right front tire.

d) State the appropriate null and alternate hypotheses to be tested, both in words and symbols.

$H_0 : \pi = \frac{1}{4}$ - Our null hypothesis is that the long run proportion of cadets who choose the right front tire is one out of four.
$H_a : \pi > \frac{1}{4}$ - Our alternate hypothesis is that the long run proportion of cadets who choose the right front tire is greater than one in four

e) What is our observed statistic? What is our sample size?

$\hat{p} = \frac{14}{28} = 0.5$
$n = 28$

f) Using the one-proportion applet, list the simulated p-value, standardized statistic, and interpret the strength of evidence.

Answers may vary.
p-value = 0.003; This is the probability of observing a result at least as extreme as 14/28 assuming our null hypothesis is true.
z = 3.125. The observed result of 0.5 is 3.125 standard deviations above the hypothesized long-run proportion of 0.25
Both of these indicate very strong strength of evidence against the null hypothesis that the long run proportion of cadets who choose front right is $\frac{1}{4}$.

    g) Does our sample meet the validity conditions to use a theory-based test?

Yes, we have 14 successes and 14 failures, both of which are above 10

$z = \dfrac{0.5 - 0.25}{\sqrt{\frac{0.25 \times (1 - 0.25)}{28}}} = 3.05505$

p-value = 1 - pnorm(3.05505) = 0.001125

i) Summarize the conclusion that you draw from this study and your analysis. Explain your reasoning.

We have very strong evidence that cadets choose the right front tire more than 25% of the time. Through simulation, we would only expect to see our observation of $\frac{14}{28}$ about 1% of the time if the true long-run proportion of this process was 0.25.

Suppose this study were repeated with only 14 cadets and 7 of them answered "front right." Use this reduced sample scenario to answer parts j through l.

We would expect weaker strength of evidence because the proportion stays the same but the sample size is smaller, so our p-value should be higher.

k) Does our reduced sample meet the validity conditions to use a theory-based test?

No, here we have 7 successes and 7 failures, both less than the 10 required

We do not meet validity conditions, so we must simulate.
Answers will vary, should be about
p-value = 0.0370
z = 2.21

2) A taste test was run on 4 sections of MA206 cadets between three types of chewy chocolate chip cookies - MAJ McD's homemade cookies, Chips Ahoy, and Keebler. The results are consolidated in the Cookies.csv file found on Teams. You may use the course guide as a reference. Here, we wish to assess if homemade cookies are preferred among cadets, as has been implied by various internet surveys. As a reference, we will assume that there is no difference between preferences for cookie maker.

```
library(tidyverse)
library(janitor)

Results <- read_csv("Cookies.csv")
```

   a) List the null and alternate hypothesis, as given in this scenario.

$H_0 : \pi = \frac{1}{3}$. The true long-run proportion of cadets who prefer MAJ McD's cookies when given the choice between homemade, Chips Ahoy, and Keebler is $\frac{1}{3}$
$H_a : \pi > \frac{1}{3}$. The true long-run proportion of cadets who prefer MAJ McD's cookies when given the choice between homemade, Chips Ahoy, and Keebler is greater than $\frac{1}{3}$

   <mark>**b)** List the number of successes, number of failures, sample size, and observed statistic as a proportion.</mark>

Number of "successes" $= 50$
Number of "failures" $= 17$
$n = 67$
$\hat{p} = \frac{50}{67} = 0.746$

```
Results %>%
   tabyl(Best) %>%
   adorn_totals()
```

   <mark>**c)** Do we meet the validity conditions to use theoretical methods?</mark>

Yes, we have at least 10 successes and 10 failures.

   <mark>**d)** Calculate the p-value and standardized statistic. Justify why you used simulation or theoretical methods.</mark>

$z = 7.170108$
$p - value = 3.747e^{-13}$
We used theoretical methods because we met our validity conditions.

```
z <- ((50/67) - (1/3))/sqrt((1/3) * (2/3) / 67); z
1 - pnorm(z)
```

   e) Interpret your results.

With a p-value of 0.0000000000003747 and a standardized statistic of 7.17, we have **very strong** evidence against the null hypothesis that cadets have no preference between the homemade, Chips Ahoy, and Keebler makers of chocolate chip cookies.

3) A survey was run on 4 sections of MA206 cadets asking which of two senatorial candidates appeared more competent. The results are consolidated in the Faces.csv file found on Teams. You may use the course guide as a reference. Here, we wish to assess if cadets agree, given one second to look at side-by-side black and white photographs, which candidate appears more competent. We assume that if there is no difference and cadets choose randomly, then either candidate has an equal chance of being selected.

```
library(tidyverse)
library(janitor)

FacesResults <- read_csv("Faces.csv")
```

a) List the null and alternate hypothesis, as given in this scenario.

$H_0 : \pi = \frac{1}{2}$. The true long-run proportion of cadets who choose the candidate on the left (or right) is equal to $\frac{1}{2}$

$H_a : \pi \neq \frac{1}{2}$. The true long-run proportion of cadets who choose the candidate on the left (or right) is not equal to $\frac{1}{2}$

b) List the number of successes, number of failures, observed statistic as a proportion, and the sample size.

Here, we are defining the candidate on the left as a "success."
Number of "successes" = 59
Number of "failures" = 4
$\hat{p} = \frac{59}{63} = 0.9365$
$n = 63$

```
FacesResults %>%
  tabyl(Faces) %>%
  adorn_totals()
```

c) Do we meet the validity conditions to use theoretical methods?

No, we do not have at least 10 failures.

d) Calculate the p-value and standardized statistic. Justify why you used simulation or theoretical methods.

p-value = 0 (Note, with simulation, may have some variation)
$z = \frac{0.93650794 - 0.5}{0.068} = 6.613757$

```
phat <- 59/63
z <- (phat - 0.5)/0.066; z
    ##The 0.066 SD came from the applet simulation, answers may vary
```

e) Interpret your results.

With a p-value of computationally 0 and a standardized statistic of 6.062951, we have **very strong** evidence against the null hypothesis that cadets randomly choose which candidate appears more competent.

f) Have we proven that cadets chose the candidate on the left because he appeared more competent and did not randomly select him?

No, we have not proven this. However, we have deemed it is very unlikely - we have very strong evidence that our observed statistic would not occur if cadets were choosing randomly.

4) On the television show *Mythbusters*, the hosts Jamie and Adam wanted to investigate which side buttered toast prefers to land on when it falls through the air. To replicate a piece of toast falling through the air, they set up a specially designed rig on the roof shot buttered toast into the air. Their results are replicated in the Toast.csv file.

$H_0 : \pi = 0.5$. The true long-run proportion of times that buttered toast lands with buttered side down is 0.5.

$H_a : \pi \neq 0.5$. The true long-run proportion of times that buttered toast lands with buttered side down is not equal to 0.5.

b) List your p-value, standardized statistic, and if you used theoretical methods.

We have at least 10 successes and 10 failures, so we will use theoretical methods to calculate z and p-value.

$\hat{p} = \frac{19}{48} = 0.395833$

$z = \frac{\hat{p} - \pi_0}{\sqrt{\frac{\pi \times (1-\pi)}{n}}} = \frac{0.395833 - 0.5}{\sqrt{\frac{0.5 \times 0.5}{48}}} = -1.443376$

p-value $= 2 \times (1 - pnorm(abs(-1.443376))) = 0.1489147$

```
Toast <- read_csv("Toast.csv")

Toast %>%
  tabyl(Butter) %>%
  adorn_totals()



n = 48
phat <- 19/n
null <- 0.5
z <- (phat - null)/sqrt(null * (1 - null) / n); z
pvalue = 2 * (1 - pnorm(abs(z)))
```

c) Write your conclusions, ensuring you reference your strength of evidence and what parameter you are inferring about.

With a p-value of 0.1489147 and standardized statistic of -1.443376, we conclude that our observed statistic of 0.395833 provides weak evidence against our null hypothesis that the true long-run proportion of times that buttered toast lands with buttered side down is 50%. It is feasible that the side buttered toast lands on is purely random with no preference with regards to side.