MA256 Lesson 4 - Significance - Strength of Evidence (1.3, 1.4, 1.5)

Define and describe the standardized statistic for a proportion:

z = $\frac{statistic - mean_{null}}{SD_{null}} = \frac{\hat{p} - \pi}{SD_{null}}$

When can we use a theory-based approach to calculate a p-value? (for a one-proportion z-test)

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? Why?

SD = $\sqrt{\frac{\pi(1-\pi)}{n}}$. This is due to the central limit theorem.

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

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

Guidelines for using p-value/standardized test statistic for strength of evidence.

	p-value	standardized test-stat
Weak Evidence against the null	$0.1 < p$	$\lvert z\rvert \leq 1.5$
Moderate Evidence against the null	$0.05 < p \leq 0.1$	$1.5 < \lvert z\rvert \leq 2$
Strong Evidence against the null	$0.01 < p \leq 0.05$	$2 < \lvert z\rvert \leq 3$
Very Strong Evidence against the null	$p \leq 0.01$	$\lvert z\rvert > 3$

What factors impact the strength of evidence?

Distance from the Null

Sample Size

1 Tail vs 2 Tail test

Using a theoretical z-score, how do we calculate the p-value? (Hint: Take a look at the Course Guide - Block 2/Lesson 3-4)

(two-sided) $H_0 : z = 0$		
(two-sided) $H_a : z \neq 0$	2*(1 - pnorm(abs(z)))	$p - val = \int_{-\infty}^{-\lvert z\rvert} f(x)dx + \int_{\lvert z\rvert}^{\infty} f(x)dx = 2\int_{\lvert z\rvert}^{\infty} f(x)dx$
(lower-tail) $H_a : z < 0$	pnorm(z)	$p - val = \int_{-\infty}^{z} f(x)dx$
(upper-tail) $H_a : z > 0$	1 - pnorm(z)	$p - val = \int_{z}^{\infty} f(x)dx$

1) Twenty-eight firsties miss recall formation because they partied hard over the weekend, but blamed their lateness on a flat tire. The TAC team brings them into their office and asks them 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. The results of the cadets' responses 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

```
pi <- 0.25
```

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

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

```
n <- 28
phat <- 14 / n
c(n, phat)
```

```
## [1] 28.0  0.5
```

f) 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

g) Assume that validity conditions are met. What is the theory-based standardized statistic and p-value?

z = $\frac{0.5-0.25}{\sqrt{\frac{0.25\times(1-0.25)}{28}}} = 3.05505$; p-value = 1 - pnorm(3.05505) = 0.001125

```
z <- (phat-pi) / sqrt(pi *(1-pi) / n)
z
```

```
## [1] 3.05505
```

```r
1 - pnorm(z)
```

```
## [1] 0.001125113
```

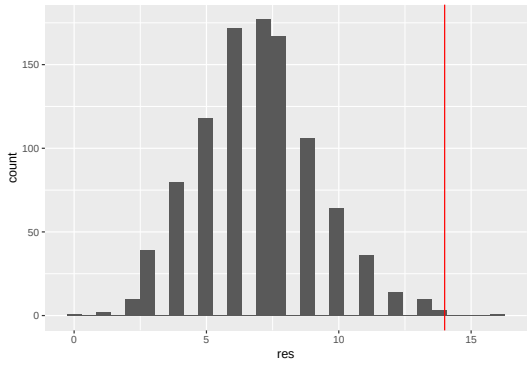
h) Summarize the conclusion that you draw from this study and your analysis. Explain your reasoning.

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 0.1% of the time if the true long-run proportion of this process was 0.25.

i) Simulate the results of the cadets answering the TAC team. Use 1000 replications. List the simulated p-value and interpret the strength of evidence (as compared to your answer above).

```r
set.seed(256)
M <- 1000
pi <- 0.25
ncadets <- 28
n.succ <- 14
phat <- n.succ / ncadets
RES <- data.frame(res = rep(NA, M))  # create a data frame to hold the results of the simulation

for(i in 1:M){
  myobs <- rbinom(1, ncadets, pi)
  RES$res[i] <- myobs
}
RES %>% ggplot(aes(x = res)) + geom_histogram() +
  geom_vline(xintercept = n.succ, color = "red")
```




```r
sum(RES$res >= n.succ) / M
```

```
## [1] 0.004
```

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 $\frac{1}{4}$.

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*.

j) What would you expect to happen to the strength of evidence against the null hypothesis in this case?

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

l) Using our reduced sample, calculate the new p-value and standardized statistic. Specify if you simulated or used theoretical methods.

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

```
z <- (0.5-0.25) / sqrt(0.25 *(1-0.25) / 14); z
```

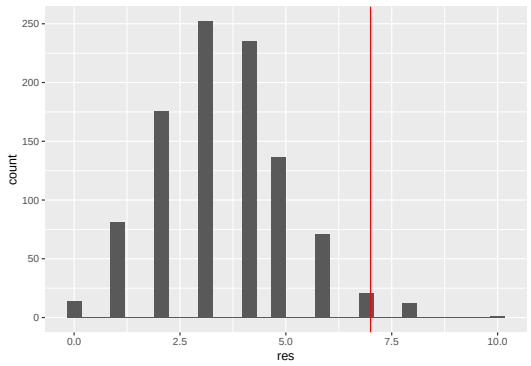
```
## [1] 2.160247
```

```
1 - pnorm(z)
```

```
## [1] 0.01537678
```

```r
set.seed(256)
M <- 1000
pi <- 0.25
ncadets <- 14
n.succ <- 7
phat <- n.succ / ncadets
RES <- data.frame(res = rep(NA, M))  # create a data frame to hold the results of the simulation

for(i in 1:M){
  myobs <- rbinom(1, ncadets, pi)
  RES$res[i] <- myobs
}
RES %>% ggplot(aes(x = res)) + geom_histogram() +
  geom_vline(xintercept = n.succ, color = "red")
```




```r
sum(RES$res >= n.succ) / M
```

```
## [1] 0.034
```

2) An article published in *College Mathematics Journal* (Eyler, Shalla, Doumaux, and McDevitt, 2009) found that players tend to not prefer scissors when playing Rock-Paper-Scissors. You want to test if people really choose scissors less, and conduct a test. You played 120 games and your friend chose scissors 31 times.

a) List the null and alternate hypothesis in words and symbols.

$H_0 : \pi = 0.3333$. The true proportion of times that a person choose scissors in Rock-Paper-Scissors is $\frac{1}{3}$.

$H_a : \pi < 0.333$. The true proportion fo times that a person chooses scissors in Rock-Paper-Scissors is less than $\frac{1}{3}$.

b) Calculate the mean and standard deviation associated with your null distribution. Calculate the observed proportion that chose scissors.

Mean $\approx 0.333, SD \approx 0.043$.

```r
pi <- 1/3
n <- 120
sd.RPS <- sqrt(pi * (1-pi) / n)
phat <- 31/n
c(pi, sd.RPS, phat)
```

```
## [1] 0.33333333 0.04303315 0.25833333
```

c) What is the standardized statistic (z) for your test? p-value? Comment on the strength of evidence.

```
z <- (phat - pi) / sd.RPS
pval <- pnorm(z)
c(z, pval)
```

```
## [1] -1.74284251  0.04068056
```

The z-score should be close to -1.74, which is moderate evidence against the null hypothesis that scissors is chosen randomly at a rate of $\frac{1}{3}$. The p-value is less than 0.05, indicating strong evidence against the null hypothesis.

d) If you repeated the test another 240 times and your friend chose scissors the same proportion of times ($\hat{p} = 0.258333$), would you expect your strength of evidence to increase, decrease, or stay the same?

We would expect our strength of evidence to increase if the sample size is larger but the observed proportion is the same.

e) If we repeated the experiment with a different friend and our sample size stayed the same (120), but the number of times he chose scissors was 38, would the strength of evidence increase, decrease, or stay the same?

We would expect the strength of evidence to decrease if the observed statistic is closer to the null (less distance) and the sample size stayed the same.

f) What if we used our original experimental data ($\frac{31}{120}$ scissors), but instead we wanted to do a two-sided test instead of a one-sided test. Would our strength of evidence increase, decrease, or stay the same? Use R to verify.

The strength of evidence would decrease if we go from a one-sided to a two-sided test. Of note, our p-value will roughly double (higher p-value is less strength of evidence), but our z-statistic is still the same.

```
2 * (1 - pnorm(abs(z)))
```

```
## [1] 0.08136113
```