Day 16 - Non-parametric inference for one-proportion

Learning outcomes and terms to know

In this section, we explore how to leverage the null hypothesis significance testing framework in the context of a particular type of problem: one in which we have a single categorical response variable.

Simulation-based inference for a single proportion (hypothesis testing)

Simulation-based inference

_____ may be used to generate the variability we would expect to see under the null hypothesis by _____ sample proportions we would expect to see if the true population proportion was equal to the null hypothesized value.

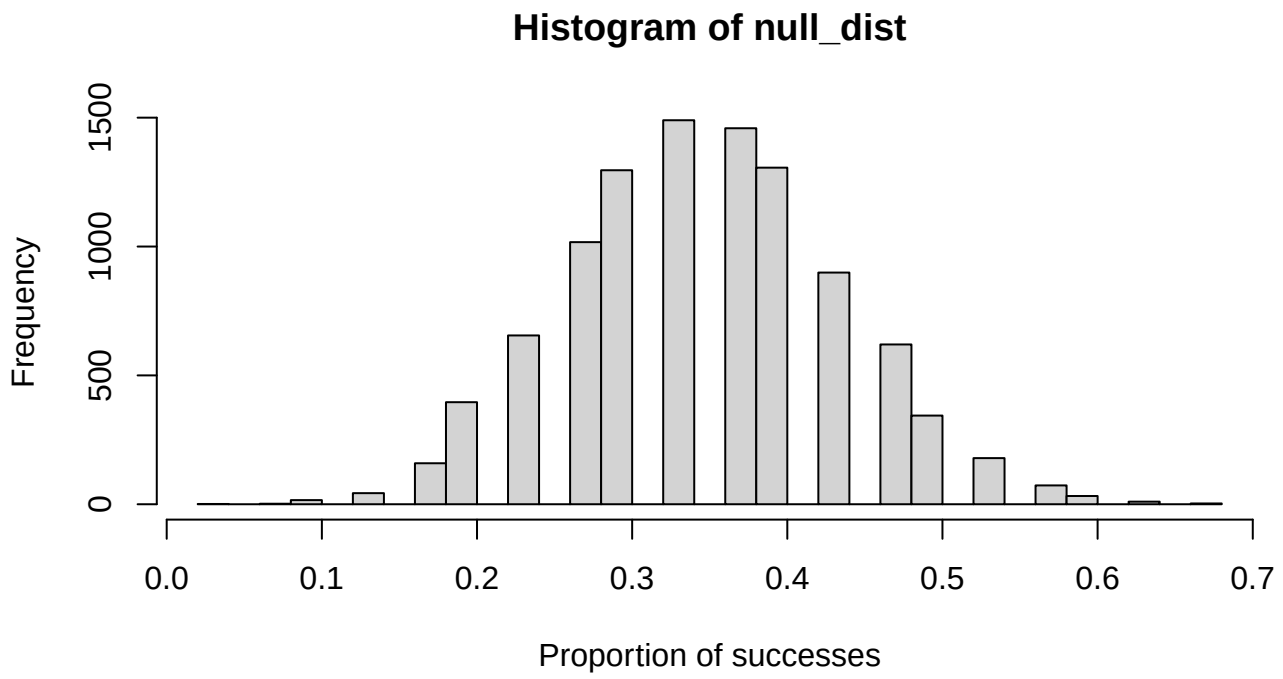
Assumptions

We require that observational units are independent of one another.

rbinom

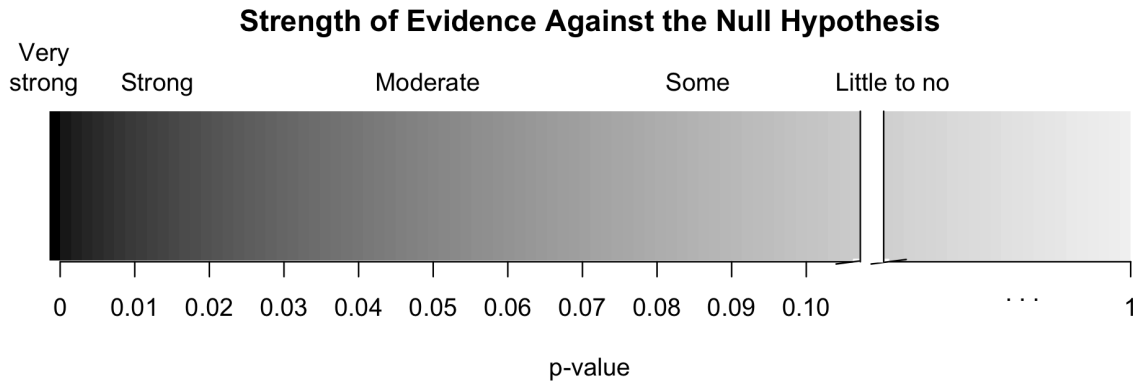
We can generate this distribution in `R` using the `rbinom` function.

```
# generate samples under the null hypothesis
null_dist <- rbinom(10000, 30, .35)/30 ## 10000 simulations of n = 30 with pi = .35
hist(null_dist, breaks = 24, xlab = "Proportion of successes")
```



Strength of evidence

A _____ represents the strength of evidence **against** the null hypothesis. This is the preferred way to think about a p-value.



Fixed-level testing

A result is considered _____ if the p-value for the test is less than some predefined threshold, called a significance level (α). This is a common way to think about p-values, but is prone to some errors.

Making decisions

You may **never** accept the null the hypothesis! When conducting a hypothesis test, you have two options:

- to reject the null hypothesis and conclude the alternative hypothesis is true
- to fail to reject the null hypothesis and conclude that there is insufficient evidence to suggest the alternative hypothesis is true

Recall the first day of class, **when we sought to determine whether we were more likely to correctly identify Coke in the blind taste test.** Of the 32 students in class that day, 23 correctly identified Coke.

1. State the parameter of interest for this analysis using both words and notation.

   > π, the true probability of correctly identifying Coke in a blind taste test of Coke vs. Pepsi.

2. State the null and alternative hypotheses for this analysis using both words and notation.

   > $H_0 : \pi = .5$, the true probability of correctly identifying Coke is 0.5 (i.e., we are just guessing).
   > $H_a : \pi > .5$, the true probability of correctly identifying Coke is greater than 0.5.

3. State the statistic for this study using both words and notation.

   > $\hat{p} = \frac{23}{32} \approx 0.719$, the sample proportion of students who correctly identified Coke.

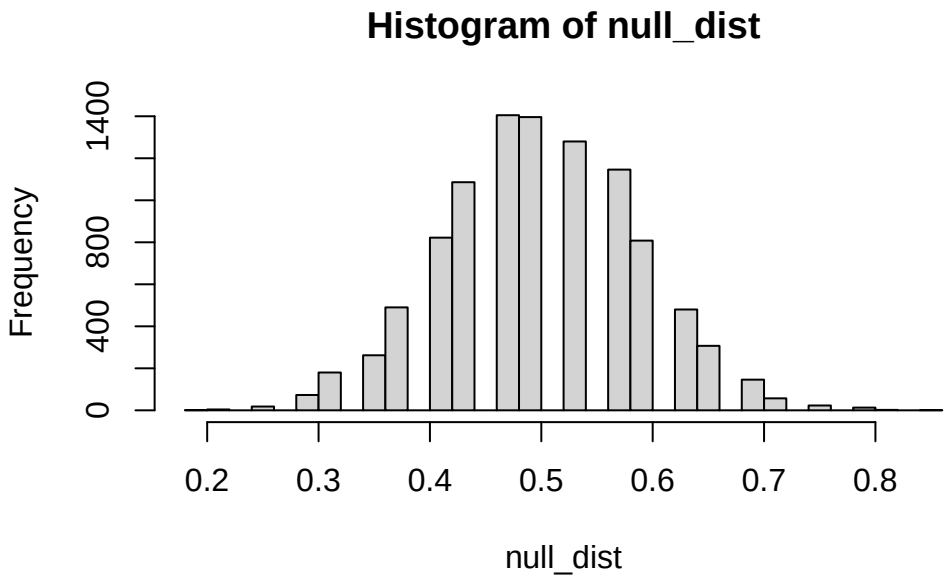
Creating the null distribution by hand

4. Describe how we could use a spinner to simulate the sampling distribution of the sample proportion under the null hypothesis. Be sure to describe: how you would label the spinner, how many times you would flick the spinner, and what you would record on each flick.

   > We could label half of a spinner "correct" and the other half "incorrect" (a .5 probability of each, since we are assuming everyone is just guessing). Then flick the spinner 32 times (once per student) and record the proportion of flicks that landed on "correct". Repeat this a couple of thousand times to create the null distribution.

5. Use `R` to generate the sampling distribution of the sample proportion under the null hypothesis (i.e. the *null distribution*). Plot the null distribution. Where is the distribution centered?

```
null_dist <- rbinom(10000, size = 32, prob = .5)/32
hist(null_dist, breaks = 30)
```



It is centered at 0.5. This makes sense, as we assumed $\pi = 0.5$ to create the distribution!

6. Compute the standard deviation of the null distribution. What does this quantity represent?

```
sd(null_dist)
```

```
[1] 0.08759132
```

The standard deviation is 0.0876. On average, sample proportions from samples of 32 students fall that far from the hypothesized probability of 0.5, **if** the true probability of correctly identifying Coke is 0.5.

7. Use the null distribution to calculate a p-value for your test. Interpret this p-value as a conditional probability in the context of the problem.

```
phat <- 23/32
mean(null_dist >= phat)
```

```
[1] 0.0096
```

If the probability of correctly identifying Coke is actually 0.5 (i.e., everyone is just guessing), there is a 0.0096 probability of observing a sample proportion of 23/32 correctly identifying Coke, or even more than that.

8. Interpret the p-value as strength of evidence against the null hypothesis and state a conclusion to this test.

A p-value of 0.0096 provides strong evidence against the null hypothesis, suggesting that the probability of correctly identifying Coke is greater than 0.5.

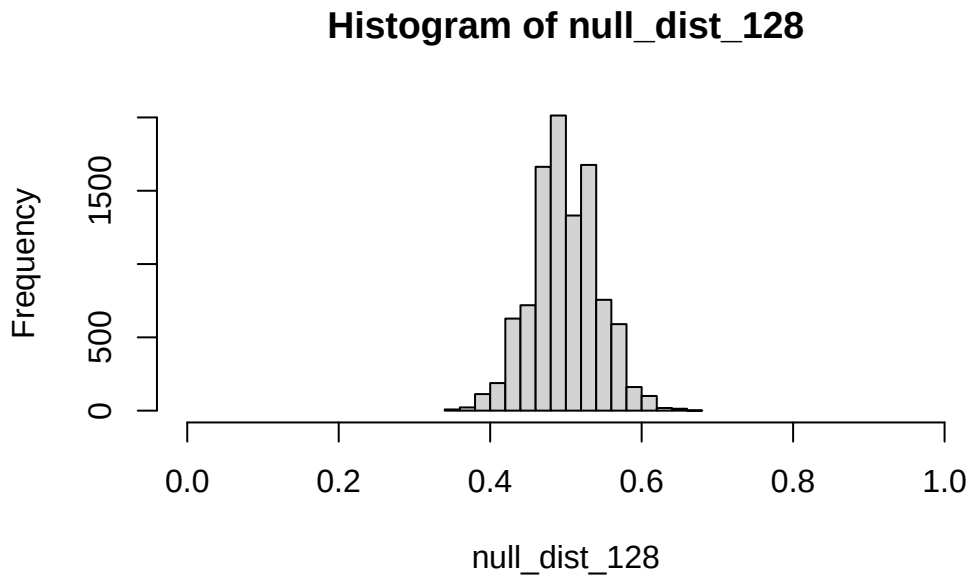
9. Make a decision at the $\alpha = 0.05$ level.

> Since the p-value is less than 0.05, we reject the null hypothesis.

Effect of n on p-values

10. Suppose we had instead observed 92/128 (same proportion, but four times as many students) correctly identifying Coke. Compute a p-value for these new data, plot the null distribution, and comment on how the new p-value relates to your previous p-value. Does this change make sense?

```
null_dist_128 <- rbinom(10000, size = 128, prob = .5)/128
hist(null_dist_128, breaks = 20, xlim = c(0, 1))
```




```
mean(null_dist_128 >= 92/128)
```

```
[1] 0
```

> The simulated p-value is 0, because none of the 10,000 simulated sample proportions were as large as 92/128. This is much smaller than our previous p-value of 0.0096, even though the sample proportion is identical. This makes sense: with four times as many students, the null distribution is much less variable (its standard deviation is 0.0443, about half of what it was with 32 students), so a sample proportion of 0.72 is much farther out in the tail of the null distribution. The same observed proportion provides much stronger evidence against H_0 when it is based on more data.