ANNOUNCEMENTS

- RECAP OF CHAPTER 7 TODAY
- PLEASE ASK QUESTIONS (how? why?)
- INTRODUCTION TO HYPOTHESIS TESTING BY END OF CLASS
- COLTON STARTS WORKING AT NCAR ON TUES/THURS
- QUIZZES EVERY FRIDAY UNTIL THE END OF SEMESTER
- I CANNOT ACCEPT ANY MORE RETAKES — PLEASE COME TO CLASS

SCHEDULE

week 11	Mon	Intro to Hypothesis Testing
	Wed	Testing Means and Proportions
	Fri	Quiz on Hypothesis Testing, finish means and proportions
week 12	Mon	Dependent Samples
	Wed	Independent Samples
	Fri	Quiz on "paired" hypotheses begin midterm review
week 13	Mon	Review
	Wed	Midterm 2

Point Estimates and Sampling Variability

$\bar{x}$ is a point estimate for μ, the popn mean

$\hat{p}$ is a point est. for p, the popn proportion

Parameter estimation

- We are often interested in *population parameters*. (e.g. μ and p)
- Since complete populations are difficult (or impossible) to collect data on, we use *sample statistics* as *point estimates* for the unknown population parameters of interest.
- Sample statistics vary from sample to sample.
- Quantifying how sample statistics vary provides a way to estimate the *margin of error* associated with our point estimate.
- But before we get to quantifying the variability among samples, let's try to understand how and why point estimates vary from sample to sample.

Suppose we randomly sample 1,000 adults from each state in the US. Would you expect the sample means of their heights to be the same, somewhat different, or very different?

Suppose we randomly sample 1,000 adults from each state in the US. Would you expect the sample means of their heights to be the same, somewhat different, or very different?

Not the same, but only somewhat different.

(1) Known

- $\bar{x} = \dfrac{x_1 + \dots + x_n}{n}$
- $n = 1000$
- $\bar{x}$ should be "close" to the pop$^{\underline{n}}$ mean μ.

(2) Unknown

- $\mu = ?$

Suppose the proportion of American adults who support the expansion of solar energy is $p = 0.88$, which is our parameter of interest. Is a randomly selected American adult more or less likely to support the expansion of solar energy?

Suppose the proportion of American adults who support the expansion of solar energy is $p = 0.88$, which is our parameter of interest. Is a randomly selected American adult more or less likely to support the expansion of solar energy?

More likely.

Suppose that you don't have access to the $\boxed{?}$ of all Ameri- can adults, which is a quite likely scenario. In order to estimate the proportion of American adults who support solar power expansion, you might sample from the population and use your sample propor- tion as the best guess for the unknown population proportion.

1. Sample, with replacement, 1000 American adults from the population, and record whether they support solar power or not expansion.
2. Find the sample proportion.

loop

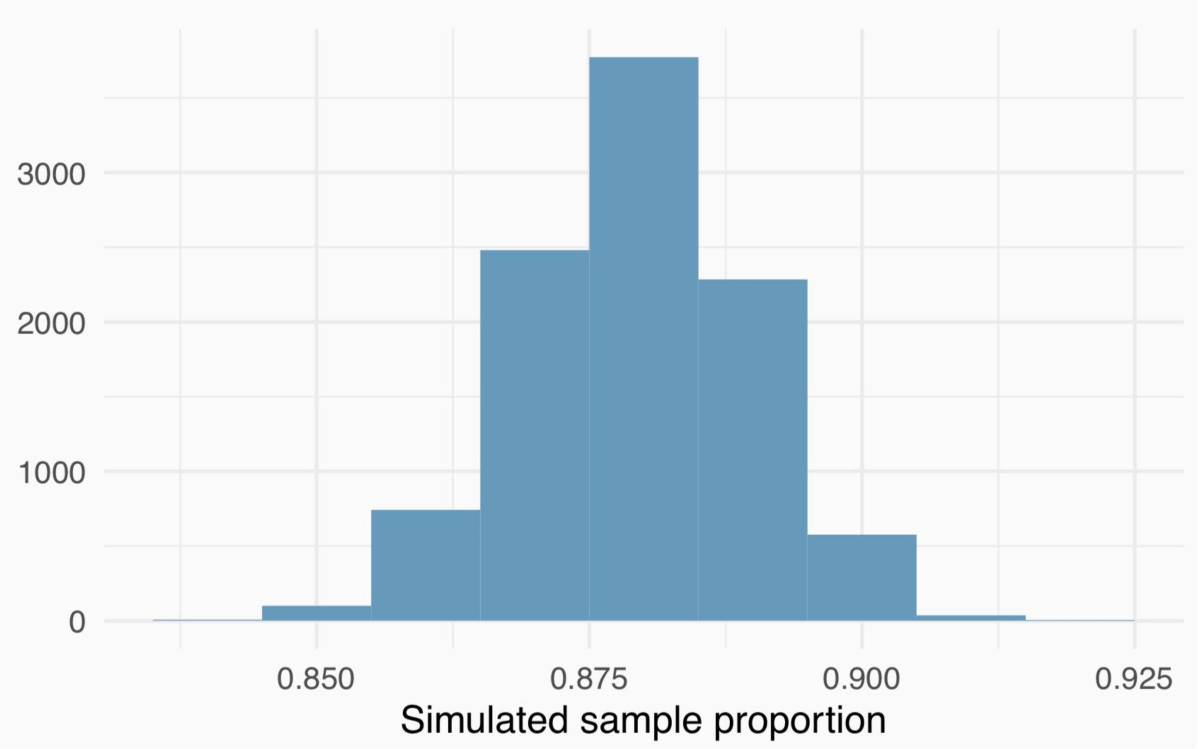
stop

Plot the distribution of the sample proportions obtained by ...

Sampling distribution

on Quiz!

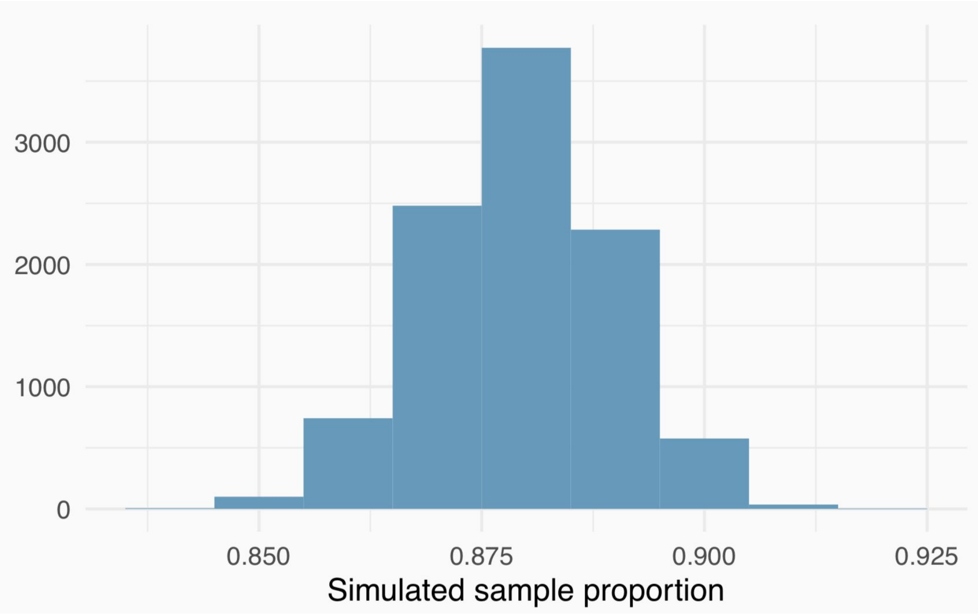
Suppose you were to repeat this process many times and plot the results. What you just constructed is called a sampling distribution.



Sampling distribution

What is the shape and center of this distribution?

The distribution looks symmetric and somewhat bell-shaped.

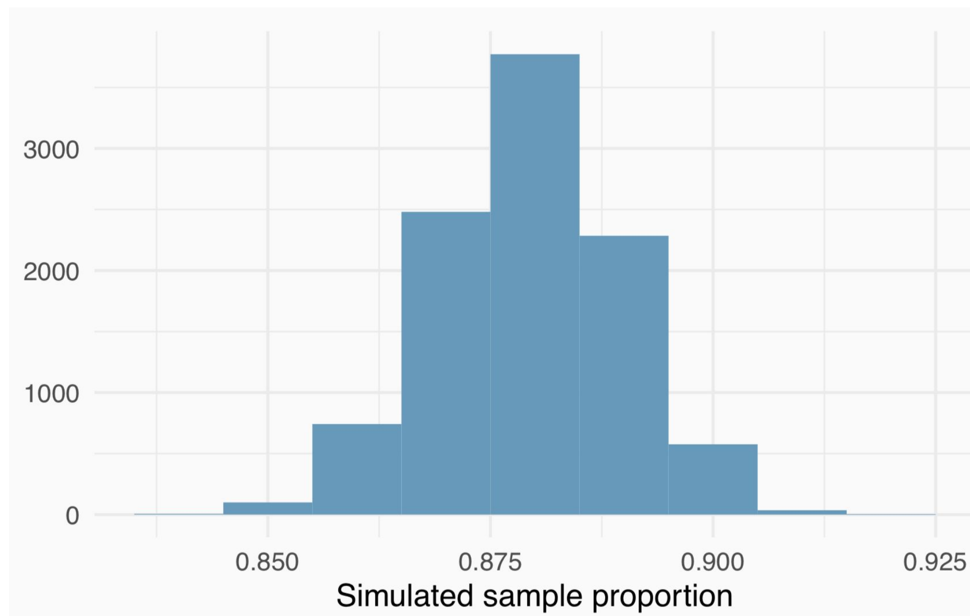


Sampling distribution

Based on this distribution, what do you think is the true population proportion?

The center of the distribution: about 0.88.

almost p,
the population proportion



Sampling distributions are never observed

- In real-world applications, we never actually observe the sampling distribution, yet it is useful to always think of a point estimate as coming from such a hypothetical distribution.

- Understanding the sampling distribution will help us characterize and make sense of the point estimates that we do observe.

Central Limit Theorem

Sample proportions will be nearly normally distributed with mean equal to the population proportion, *p*, and standard error equal to $\sqrt{\frac{p\,(1-p)}{n}}$

$$\hat{p} \sim N\left(mean = p, SE = \sqrt{\frac{p\,(1-p)}{n}}\right)$$

It wasn't a coincidence that the sampling distribution we saw earlier was symmetric, and centered at the true population population.

We won't go through a detailed proof of why $SE = \sqrt{\frac{p\,(1-p)}{n}}$ ← last week

but note that as n increases SE decreases.

- As n increases samples will yield more consistent $\hat{p}$s, i.e. variability among $\hat{p}$s will be lower.

CLT - conditions

Certain conditions must be met for the CLT to apply:

Independence

Sampled observations must be independent. This is difficult to verify, but is more likely if

- random sampling/assignment is used, and
- if sampling without replacement, n < 10% of the population.

Sample size

There should be at least 10 expected successes and 10 expected failures in the observed sample.

This is difficult to verify if you don't know the population proportion (or can't assume a value for it). In those cases we look for the number of observed successes and failures to be at least 10.

When *p* is unknown

The CLT states

$$SE = \sqrt{\frac{p\,(1-p)}{n}}$$

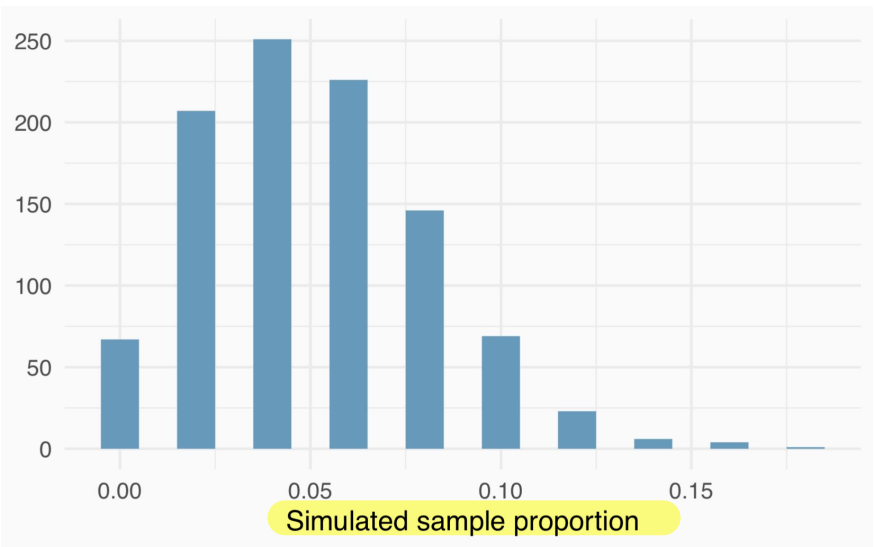
with the condition that *np* and *n(1 − p)* are at least 10. (webassign ≥ 5, ok?)

However, we often don't know the value of *p*, the population proportion.
In these cases we substitute $\hat{p}$ for *p*.

When *np* or *n(1 - p)* is small

Suppose we have a population where the true population proportion is $p = 0.05$, and we take random samples of size $n = 50$ from this population. We calculate the sample proportion in each sample and plot these proportions. Would you expect this distribution to be nearly normal? Why, or why not?

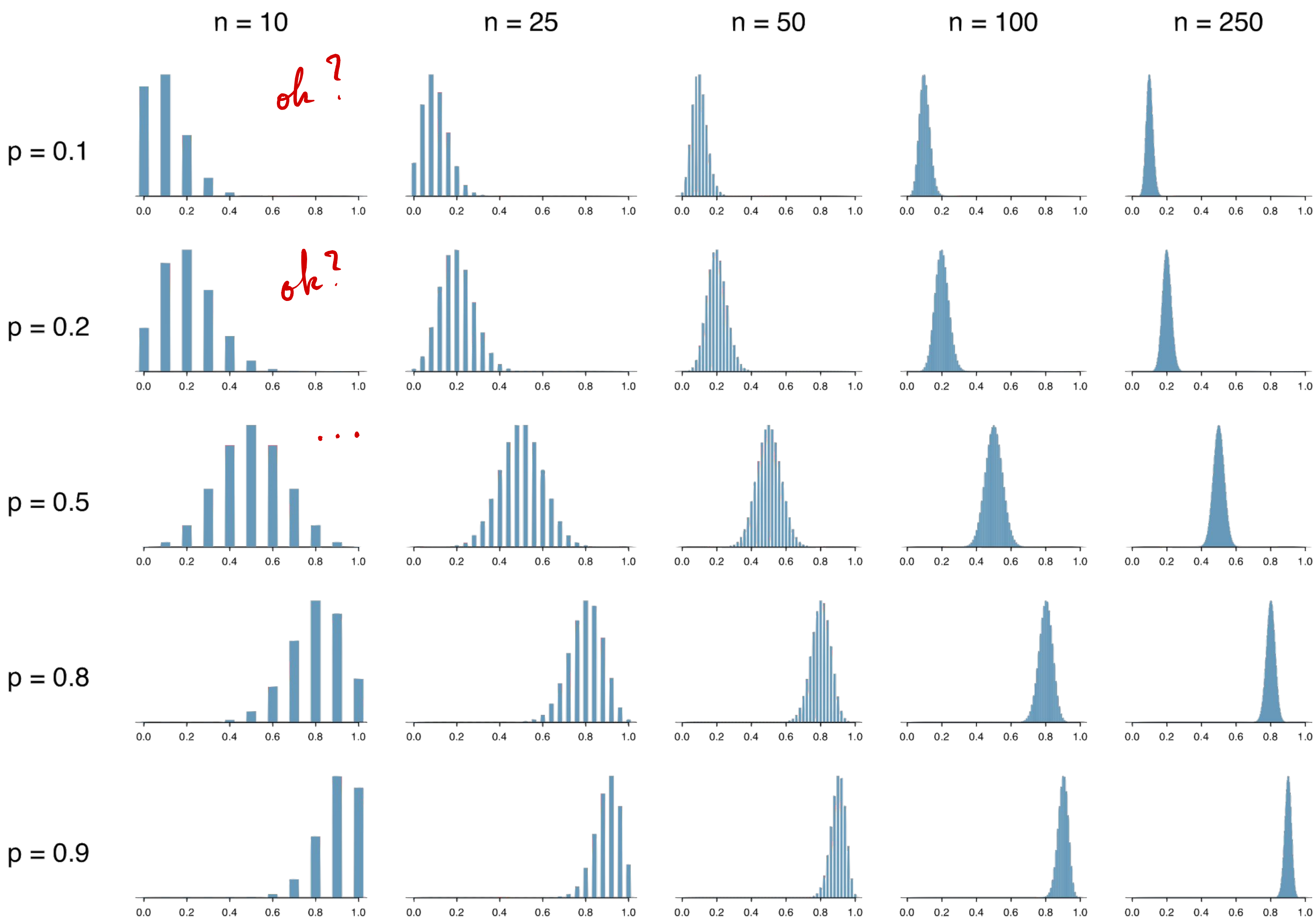
No, the success-failure condition is not met ($50 \times 0.05 = 2.5$), so we would not expect the sampling distribution to be nearly normal.



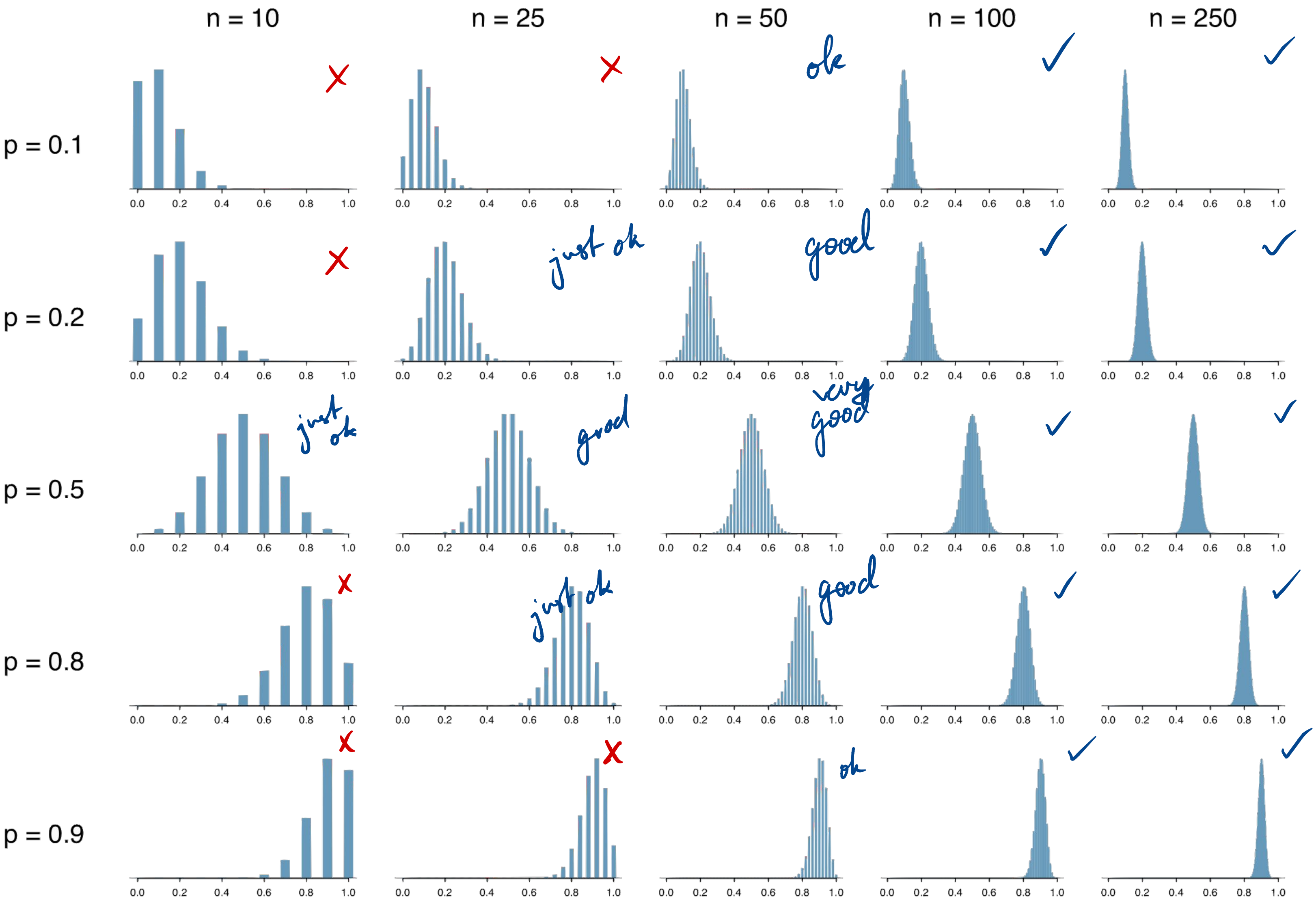
EXERCISE

which distributions satisfy the CLT?

What happens when np and/or $n(1-p) < 10$



What happens when *np* and/or *n(1 − p)* < 10



Confidence Intervals for a Proportion

EXERCISE

Facebook's categorization of user interests

Most commercial websites (e.g. social media platforms, news out- lets, online retailers) collect a data about their users' behaviors and use these data to deliver targeted content, recommendations, and ads. To understand whether Americans think their lives line up with how the algorithm-driven classification systems categorizes them, Pew Research asked a representative sample of 850 American Facebook users how accurately they feel the list of categories Facebook has listed for them on the page of their supposed interests actually represents them and their interests. 67% of the respondents said that the listed categories were accurate. Estimate the true proportion of American Facebook users who think the Facebook categorizes their interests accurately.

https://www.pewinternet.org/2019/01/16/facebook-algorithms-and-personal-data/

Facebook's categorization of user interests

① Known?

(sample proportion) $\hat{p} = 0.67$ (sample size) $n = 850$

② Unknown? p?

Want To Find

W T F

③ Relevant mathematical relationships?

The approximate 95% confidence interval is defined as

$$\text{point estimate} \pm 1.96 \times SE$$

④ Compute!

$$SE = \sqrt{\frac{p(1-p)}{n}} = \sqrt{\frac{0.67 \times 0.33}{850}} \approx 0.016$$

$$\begin{aligned} \hat{p} \pm 1.96 \times SE &= 0.67 \pm 1.96 \times 0.016 \\ &= (0.67 - 0.03, 0.67 + 0.03) \\ &= (0.64, 0.70) \end{aligned}$$

$\boxed{\text{1 Prop Z Int}}$!

EXERCISE

Facebook's categorization of user interests

Which of the following is the correct interpretation of this confidence interval? We are 95% confident that...

(a) 64% to ~~67%~~ 70% of American Facebook users in this sample think Facebook categorizes their interests accurately.

(b) 64% to ~~67%~~ 70% of all American Facebook users think Facebook categorizes their interests accurately

(c) there is a 64% to ~~67%~~ 70% chance that a randomly chosen American Facebook user's interests are categorized accurately.

(d) there is a 64% to ~~67%~~ 70% chance that 95% of American Facebook users' interests are categorized accurately.

Facebook's categorization of user interests

Which of the following is the correct interpretation of this confidence interval? We are 95% confident that...

(a) 64% to 67% of American Facebook users in this sample think Facebook categorizes their interests accurately.

(b) 64% to 67% of all American Facebook users think Facebook categorizes their interests accurately

(c) there is a 64% to 67% chance that a randomly chosen American Facebook user's interests are categorized accurately.

(d) there is a 64% to 67% chance that 95% of American Facebook users' interests are categorized accurately.

What does 95% confident mean?

Quiz!

Suppose we took many samples and built a confidence interval from each sample using the equation

why?

point estimate ± 1.96 × SE

Then about 95% of those intervals would contain the true population proportion (p).

critical (z) values always come from $Z \sim N(0,1)$

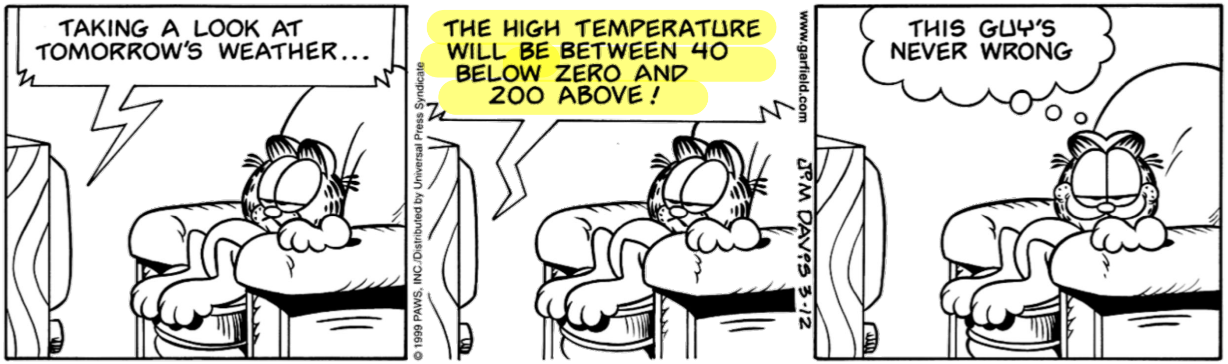
$$z_{0.95} = \text{invNorm}(\mu = 0, \sigma = 1, 0.975)$$

Width of an interval

If we want to be more certain that we capture the population parameter, i.e. increase our confidence level, should we use a wider interval or a smaller interval?

A wider interval.

Can you see any drawbacks to using a wider interval?



If the interval is too wide it may not be very informative.

Changing the confidence level

$\boxed{t^* \times SE}$

also true for Student's t distr. ok?

$$\text{point estimate} \pm z^\star \times SE$$

- In a confidence interval, $z^\star \times SE$ is called the **margin of error**, and for a given sample, the margin of error changes as the confidence level changes.
- In order to change the confidence level we need to adjust $z^\star$ in the above formula.
- Commonly used confidence levels in practice are 90%, 95%, 98%, and 99%.
- For a 95% confidence interval, $z^\star = 1.96$. (memorize)
- However, using the standard normal (z) distribution, it is possible to find the appropriate $z^\star$ for any confidence level. (invNorm)

Hypothesis Testing for a Proportion

(Recall Conditional Probabilities)

Gender discrimination experiment:

		Promotion		
		Promoted	Not Promoted	Total
Gender	Male	21	3	24
	Female	14	10	24
	Total	35	13	48

EXERCISE

$\mathbb{P}(\text{Promoted} \mid \text{Male}) = ?$

$\mathbb{P}(\text{Promoted} \mid \text{Female}) = ?$

Gender discrimination experiment:

		Promotion		
		Promoted	Not Promoted	Total
Gender	Male	21	3	24
	Female	14	10	24
	Total	35	13	48

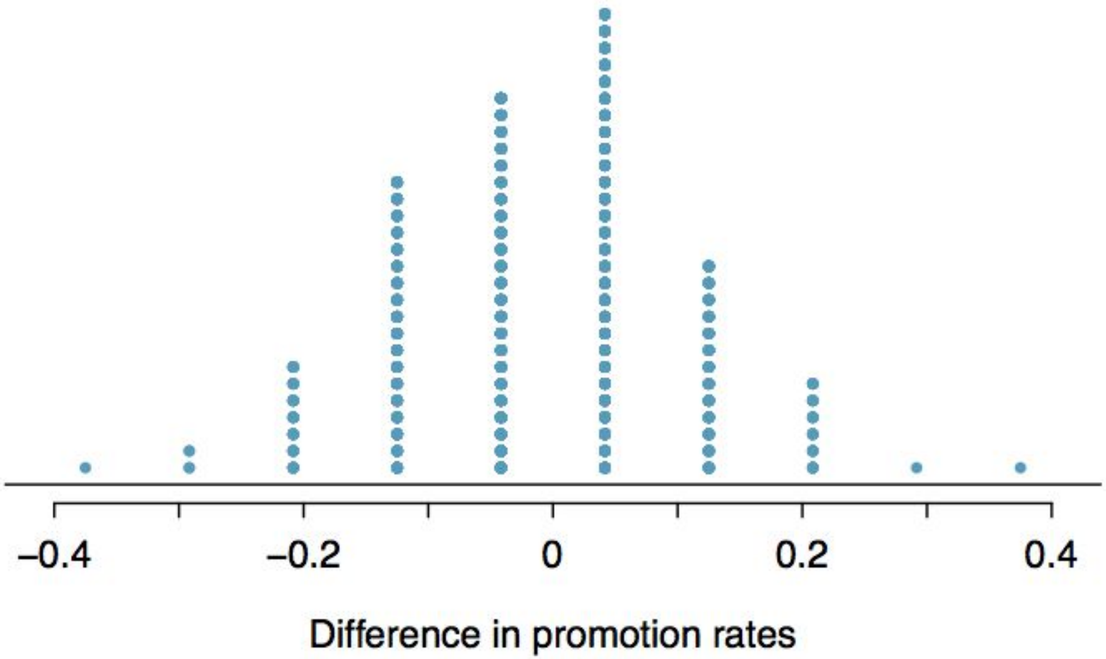
$\hat{p}_{males} = 21 / 24 = 0.88$

$\hat{p}_{females} = 14 / 24 = 0.58$

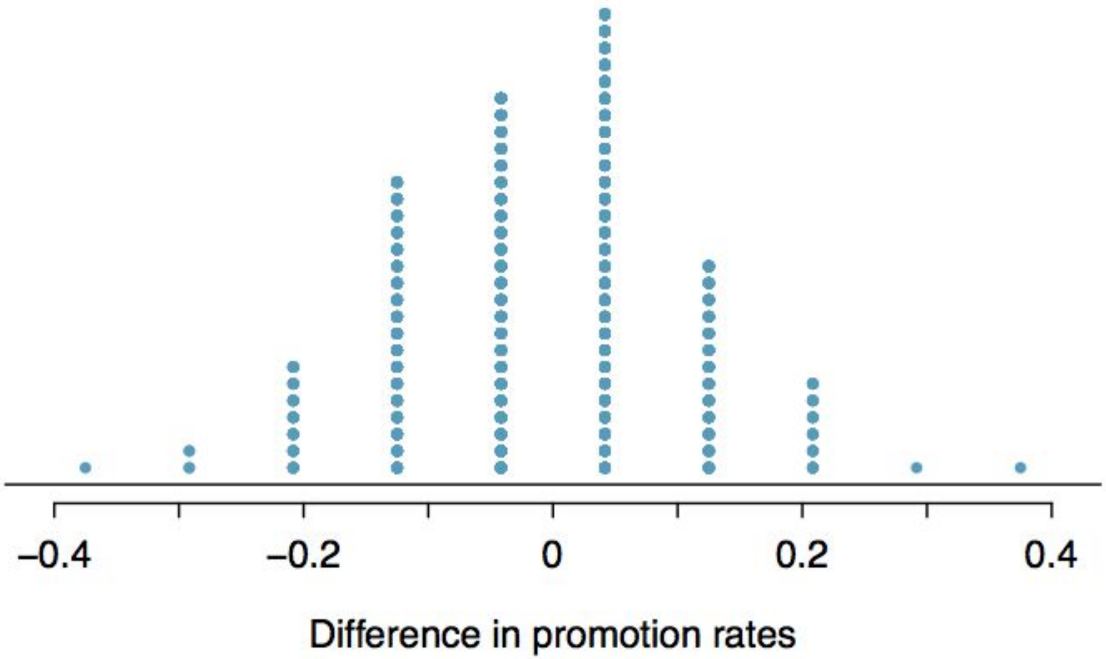
Possible explanations:

- Promotion and gender are *independent*, no gender discrimination, observed difference in proportions is simply due to chance.
  → null (nothing is going on)
- Promotion and gender are *dependent*, there is gender discrimination, observed difference in proportions is not due to chance.
  → alternative (something is going on)

Result (Simulated)



Result



Since it was quite unlikely to obtain results like the actual data or something more extreme in the simulations (male promotions being 30% or more higher than female promotions), we decided to reject the null hypothesis in favor of the alternative.

Recap: hypothesis testing framework

- We start with a *null hypothesis* (H_0) that represents the status quo.

Recap: hypothesis testing framework

- We start with a *null hypothesis* (H_0) that represents the status quo.
- We also have an *alternative hypothesis* (H_A) that represents our research question, i.e. what we're testing for.

Recap: hypothesis testing framework

- We start with a *null hypothesis* (H_0) that represents the status quo.
- We also have an *alternative hypothesis* (H_A) that represents our research question, i.e. what we're testing for.
- We conduct a hypothesis test under the assumption that the null hypothesis is true, either via simulation or traditional methods based on the central limit theorem (coming up next...).

Recap: hypothesis testing framework

- We start with a *null hypothesis* (H_0) that represents the status quo.
- We also have an *alternative hypothesis* (H_A) that represents our research question, i.e. what we're testing for.
- We conduct a hypothesis test under the assumption that the null hypothesis is true, either via simulation or traditional methods based on the central limit theorem (coming up next...).
- If the test results suggest that the data do not provide convincing evidence for the alternative hypothesis, we stick with the null hypothesis. If they do, then we reject the null hypothesis in favor of the alternative.

Recap: hypothesis testing framework

- We start with a *null hypothesis* (H_0) that represents the status quo.
- We also have an *alternative hypothesis* (H_A) that represents our research question, i.e. what we're testing for.
- We conduct a hypothesis test under the assumption that the null hypothesis is true, either via simulation or traditional methods based on the central limit theorem (coming up next...).
- If the test results suggest that the data do not provide convincing evidence for the alternative hypothesis, we stick with the null hypothesis. If they do, then we reject the null hypothesis in favor of the alternative.

We'll formally introduce the hypothesis testing framework using an example on testing a claim about a population

Testing hypotheses using confidence intervals

Earlier we calculated a 95% confidence interval for the $\boxed{?}$ of American Facebook users who think Facebook categorizes their interests accurately as 64% to 67%. Based on this confidence interval, do the data support the hypothesis that majority of American Facebook users think Facebook categorizes their interests accurately $\boxed{?}$

The associated hypotheses are:

H_0: $p = 0.50$: 50% of American Facebook users think Facebook categorizes their interests accurately

H_A: $p > 0.50$: More than 50% of American Facebook users think Facebook categorizes their interests accurately

Conclusion ?

Null value is not included in the interval $\xrightarrow{\text{therefore}}$ reject the null hypothesis. (reject H_0)

This is a "quick-and-dirty approach" for hypothesis testing, but it doesn't tell us the likelihood of certain outcomes under the null hypothesis (p-value)

Define in your reading!

Practice (in class, on wrksheet)

- ☆ pg. 3 prob. 2 parts (a), (b), (c), (d)
- ☆ pg. 6 prob 1 parts (a), (b)
  ↳ guess part (d)
- ☆ pg. 6 prob 2 parts (a), (b)
  ↳ guess part (d)
- pg. 4 prob. 3 part (a)
- pg. 4 prob 4 part (a)
- pg. 5 prob 5 part (a)
- pg. 7 prob 3
  ↳ guess

Homework (for next class)

- define p-value
- define α, the level of significance
- practice using Z-test and T-test
- finish Quiz Week 10 !
  → turn in to Colton at 8am
  → or turn in to Maya at 5pm
  @ MARC Math 175

Decision errors

- Hypothesis tests are not flawless.
- In the court system innocent people are sometimes wrongly convicted, and the guilty sometimes walk free.
- Similarly, we can make a wrong decision in statistical hypothesis tests as well.
- The difference is that we have the tools necessary to quantify how often we make errors in statistics.

Decision errors (cont.)

There are two competing hypotheses: the null and the alternative. In a hypothesis test, we make a decision about which might be true, but our choice might be incorrect.

Decision errors (cont.)

There are two competing hypotheses: the null and the alternative. In a hypothesis test, we make a decision about which might be true, but our choice might be incorrect.

		Decision	
		fail to reject H_0	reject H_0
Truth	H_0 true		
	H_A true		

Decision errors (cont.)

There are two competing hypotheses: the null and the alternative. In a hypothesis test, we make a decision about which might be true, but our choice might be incorrect.

		Decision	
		fail to reject H_0	reject H_0
Truth	H_0 true	✓	
	H_A true		

Decision errors (cont.)

There are two competing hypotheses: the null and the alternative. In a hypothesis test, we make a decision about which might be true, but our choice might be incorrect.

		Decision	
		fail to reject H_0	reject H_0
Truth	H_0 true	✓	
	H_A true		✓

Decision errors (cont.)

There are two competing hypotheses: the null and the alternative. In a hypothesis test, we make a decision about which might be true, but our choice might be incorrect.

		Decision	
		fail to reject H_0	reject H_0
Truth	H_0 true	✓	*Type 1 Error*
	H_A true		✓

- A *Type 1 Error* is rejecting the null hypothesis when H_0 is true.

Decision errors (cont.)

There are two competing hypotheses: the null and the alternative. In a hypothesis test, we make a decision about which might be true, but our choice might be incorrect.

		Decision	
		fail to reject H_0	reject H_0
Truth	H_0 true	✓	*Type 1 Error*
	H_A true	*Type 2 Error*	✓

- A *Type 1 Error* is rejecting the null hypothesis when H_0 is true.
- A *Type 2 Error* is failing to reject the null hypothesis when H_A is true.

Decision errors (cont.)

There are two competing hypotheses: the null and the alternative. In a hypothesis test, we make a decision about which might be true, but our choice might be incorrect.

		Decision	
		fail to reject H_0	reject H_0
Truth	H_0 true	✓	*Type 1 Error*
	H_A true	*Type 2 Error*	✓

- A *Type 1 Error* is rejecting the null hypothesis when H_0 is true.
- A *Type 2 Error* is failing to reject the null hypothesis when H_A is true.

We (almost) never know if H_0 or H_A is true, but we need to consider all possibilities.

Hypothesis Test as a trial

If we again think of a hypothesis test as a criminal trial then it makes sense to frame the verdict in terms of the null and alternative hypotheses:

H_0: Defendant is innocent

H_A: Defendant is guilty

Which type of error is being committed in the following circumstances?

- Declaring the defendant innocent when they are actually guilty
- Declaring the defendant guilty when they are actually innocent

Hypothesis Test as a trial

If we again think of a hypothesis test as a criminal trial then it makes sense to frame the verdict in terms of the null and alternative hypotheses:

H_0: Defendant is innocent

H_A: Defendant is guilty

Which type of error is being committed in the following circumstances?

- Declaring the defendant innocent when they are actually guilty

Type 2 error

- Declaring the defendant guilty when they are actually innocent

Hypothesis Test as a trial

If we again think of a hypothesis test as a criminal trial then it makes sense to frame the verdict in terms of the null and alternative hypotheses:

H_0: Defendant is innocent

H_A: Defendant is guilty

Which type of error is being committed in the following circumstances?

- Declaring the defendant innocent when they are actually guilty

  Type 2 error

- Declaring the defendant guilty when they are actually innocent

  Type 1 error

Which error do you think is the worse error to make?

Hypothesis Test as a trial

If we again think of a hypothesis test as a criminal trial then it makes sense to frame the verdict in terms of the null and alternative hypotheses:

H_0: Defendant is innocent

H_A: Defendant is guilty

Which type of error is being committed in the following circumstances?

- Declaring the defendant innocent when they are actually guilty

Type 2 error

- Declaring the defendant guilty when they are actually innocent

Type 1 error

Which error do you think is the worse error to make?

“better that ten guilty persons escape than that one innocent suffer”
- William Blackstone

Type 1 error rate

- As a general rule we reject H_0 when the p-value is less than 0.05, i.e. we use a *significance level* of 0.05, $\alpha = 0.05$.

Type 1 error rate

- As a general rule we reject H_0 when the p-value is less than 0.05, i.e. we use a *significance level* of 0.05, $\alpha = 0.05$.
- This means that, for those cases where H_0 is actually true, we do not want to incorrectly reject it more than 5% of those times.

Type 1 error rate

- As a general rule we reject H_0 when the p-value is less than 0.05, i.e. we use a *significance level* of 0.05, $\alpha = 0.05$.
- This means that, for those cases where H_0 is actually true, we do not want to incorrectly reject it more than 5% of those times.
- In other words, when using a 5% significance level there is about 5% chance of making a Type 1 error if the null hypothesis is true.

$$P(\text{Type 1 error} \mid H_0 \text{ true}) = \alpha$$

Type 1 error rate

- As a general rule we reject H_0 when the p-value is less than 0.05, i.e. we use a *significance level* of 0.05, $\alpha = 0.05$.
- This means that, for those cases where H_0 is actually true, we do not want to incorrectly reject it more than 5% of those times.
- In other words, when using a 5% significance level there is about 5% chance of making a Type 1 error if the null hypothesis is true.

$$P(\text{Type 1 error} \mid H_0 \text{ true}) = \alpha$$

- This is why we prefer small values of α -- increasing α increases the Type 1 error rate.

Facebook interest categories

The same survey asked the 850 respondents how comfortable they are with Facebook creating a list of categories for them. 41% of the respondents said they are comfortable. Do these data provide convincing evidence that the proportion of American Facebook users are comfortable with Facebook creating a list of interest categories for them is different than 50%?

https://www.pewinternet.org/2019/01/16/facebook-algorithms-and-personal-data/

Facebook interest categories

The same survey asked the 850 respondents how comfortable they are with Facebook creating a list of categories for them. 41% of the respondents said they are comfortable. Do these data provide convincing evidence that the proportion of American Facebook users are comfortable with Facebook creating a list of interest categories for them is different than 50%?

Setting the hypotheses

The *parameter of interest* is the proportion of <u>all</u> American Facebook users who are comfortable with Facebook creating categories of interests for them.

There may be two explanations why our sample proportion is lower than 0.50 (minority).

- The true population proportion is different than 0.50.
- The true population mean is 0.50, and the difference between the true population proportion and the sample proportion is simply due to natural sampling variability.

Facebook interest categories

The same survey asked the 850 respondents how comfortable they are with Facebook creating a list of categories for them. 41% of the respondents said they are comfortable. Do these data provide convincing evidence that the proportion of American Facebook users are comfortable with Facebook creating a list of interest categories for them is different than 50%?

Setting the hypotheses

We start with the assumption that 50% of American Facebook users are comfortable with Facebook creating categories of interests for them

H_0: $p = 0.50$

We test the claim that the proportion of American Facebook users who are comfortable with Facebook creating categories of interests for them is different than 50%.

H_A: $p \neq 0.50$

Facebook interest categories - conditions

Which of the following is not a condition that needs to be met to proceed with this hypothesis test?

(a) Respondents in the sample should be independent of each other with respect to whether or not they feel comfortable with their interests being categorized by Facebook.

(b) Sampling should have been done randomly.

(c) The sample size should be less than 10% of the population of all American Facebook users.

(d) There should be at least 30 respondents in the sample.

(e) There should be at least 10 expected successes and 10 expected failure.

Facebook interest categories - conditions

Which of the following is not a condition that needs to be met to proceed with this hypothesis test?

(a) Respondents in the sample should be independent of each other with respect to whether or not they feel comfortable with their interests being categorized by Facebook.

(b) Sampling should have been done randomly.

(c) The sample size should be less than 10% of the population of all American Facebook users.

(d) There should be at least 30 respondents in the sample.

(e) There should be at least 10 expected successes and 10 expected failure.

Test statistic

In order to evaluate if the observed sample mean is unusual for the hypothesized sampling distribution, we determine how many standard errors away from the null it is, which is also called the *test statistic*.

$$\hat{p} \sim N\left(\mu = 0.50, SE = \sqrt{\frac{0.50 \times 0.50}{850}}\right)$$

$$Z = \frac{0.41 - 0.50}{0.0171} = -5.26$$

The sample proportion is 5.26 standard errors away from the hypothesized value. Is this considered unusually low? That is, is the result *statistically significant*?

Yes, and we can quantify how unusual it is using a p-value.

p-values

We then use this test statistic to calculate the *p-value*, the probability of observing data at least as favorable to the alternative hypothesis as our current data set, if the null hypothesis were true.

If the p-value is *low* (lower than the significance level, α, which is usually 5%) we say that it would be very unlikely to observe the data if the null hypothesis were true, and hence *reject* H_0.

If the p-value is *high* (higher than α) we say that it is likely to observe the data even if the null hypothesis were true, and hence *do not reject* H_0.

Facebook interest categories - p-value

p-value: probability of observing data at least as favorable to H_A as our current data set (a sample proportion lower than 0.41), if in fact H_0 were true (the true population proportion was 0.50).

$$P(\hat{p} < 0.41 \; or \; \hat{p} > 0.59 \mid p = 0.50) = P(|Z| > 5.26) < 0.0001$$

Facebook interest categories - Making a decision

p-value < 0.0001

- If 50% of all American Facebook users are comfortable with Facebook creating these interest categories, there is less than a 0.01% chance of observing a random sample of 850 American Facebook users where 41% or fewer or 59% of higher feel comfortable with it.
- This is a pretty low probability for us to think that the observed sample proportion, or something more extreme, is likely to happen simply by chance.

Since p-value is *low* (lower than 5%) we *reject* H_0.

The data provide convincing evidence that the proportion of American Facebook users who are comfortable with Facebook creating a list of interest categories for them is different than 50%.

The difference between the null value of 0.50 and observed sample proportion of 0.41 is *not due to chance* or sampling variability.

Choosing a significance level

Choosing a significance level for a test is important in many contexts, and the traditional level is 0.05. However, it is often helpful to adjust the significance level based on the application.

We may select a level that is smaller or larger than 0.05 depending on the consequences of any conclusions reached from the test.

If making a Type 1 Error is dangerous or especially costly, we should choose a small significance level (e.g. 0.01). Under this scenario we want to be very cautious about rejecting the null hypothesis, so we demand very strong evidence favoring H_A before we would reject H_0.

If a Type 2 Error is relatively more dangerous or much more costly than a Type 1 Error, then we should choose a higher significance level (e.g. 0.10). Here we want to be cautious about failing to reject H_0 when the null is actually false.

One vs. two sided hypothesis tests

In two sided hypothesis tests we are interested in whether p is either above or below some null value p_0: $H_A : p \neq p_0$.

In one sided hypothesis tests we are interested in p differing from the null value p_0 in one direction (and not the other):

If there is only value in detecting if population parameter is less than p_0, then H_A: $p < p_0$.

If there is only value in detecting if population parameter is greater than p_0, then $H_A : p > p_0$.

Two-sided tests are often more appropriate as we often want to detect if the data goes clearly in the opposite direction of a hypothesis direction as well.