Chapter 8: Hypothesis Testing for Population Proportions

Testing a claim

- Earlier we calculated a 95% confidence interval for the true population proportion of UCLA students who travelled outside the US.

$$(0.26, 0.44)$$

- A 95% confidence interval of 26% to 44% means that

  - We are 95% confident that the true population proportion of UCLA students who travelled outside the US is between 26% and 44%.

  - 95% of random samples of size n = 100 will produce confidence intervals that contain the true population proportion.

Testing a claim

- The true population proportion, p, may be outside the interval, but we would expect it to be somewhat close to $\hat{p}$.

- In our random sample of 100 students we had found that 35 of them have at some point in their lives travelled outside the US, $\hat{p}$ = 0.35.

- Do you think it's possible that:

  - p = 0.90?

  - p = 0.05?

  - p = 0.34 or p = 0.36?

  - p = 0.25 or p = 0.45?

- It is difficult to decide how close is close enough, or how far is too far, and this decision should not be made subjectively.

Hypothesis Testing

- In Statistics when testing claims we use an objective method called hypothesis testing

- Given a sample proportion, $\hat{p}$, and sample size, n, we can test claims about the population proportion, p.

- We call these claims hypotheses

- Our starting point, the status quo, is called the null hypothesis and the alternative claim is called the alternative hypothesis

Hypothesis Testing

- If our null hypothesis was that p = 0.35 and our sample yields $\hat{p}$ = 0.35, then the data are consistent with the null hypothesis, and we have no reason to not believe this hypothesis.

  - This doesn't prove the hypothesis but we can say that the data support it.

- If our null hypothesis was different than p = 0.35, let's say p = 30 and our sample yields $\hat{p}$ = 0.35, then the data are not consistent with the hypothesis and we need to make choices as to whether this inconsistency is large enough to not believe the hypothesis.

  - If the inconsistency is significant, we reject the null hypothesis

Hypothesis Testing for One Proportion

Research conducted a few years ago showed that 35% of UCLA students had travelled outside the US. UCLA has recently implemented a new study abroad program and results of a new survey show that out of the 100 randomly sampled students 42 have travelled abroad. Is there significant evidence to suggest that the proportion of students at UCLA who have travelled abroad has increased after the implementation of the study abroad program?

- Population proportion used to be 0.35.

- New sample proportion is 0.42.

- We are testing the claim that the population proportion is now greater than 0.35.

- The new data indicate that it may be since 0.42 is clearly grater than 0.35.

- But is this difference statistically significant, i.e. are the data inconsistent enough?

- We do a formal hypothesis test to answer this question.

Setting Up Hypotheses

- Null hypothesis, denoted by H_o, specifies a population model parameter of interest and proposes a value for that parameter (p).

$$H_o: p = 0.35$$

- Alternative hypothesis, denoted by H_A, is the claim we are testing for.

$$H_A: p > 0.35$$

- Even though we are testing for the alternative hypothesis, we check to see whether or not the null hypothesis is plausible.

- If the null hypothesis is not plausible, we reject the null hypothesis and conclude that there is sufficient evidence to support the alternative. If the null hypothesis is plausible, we fail to reject the null hypothesis and conclude that there isn't sufficient evidence to support the alternative.

Why do we check if H_o is plausible and not if H_A is plausible?

- Think about the logic of jury trials:

    - To prove someone is guilty, we start by assuming they are innocent.

    - We retain that hypothesis until the facts make it unlikely beyond a reasonable doubt.

    - Then, and only then, we reject the hypothesis of innocence and declare the person guilty.

- The same logic used in jury trials is used in statistical tests of hypotheses:

    - We begin by assuming that the null hypothesis is true.

    - Next we consider whether the data are consistent with this hypothesis.

    - If they are, all we can do is retain the hypothesis we started with. If they are not, then like a jury, we ask whether they are unlikely beyond a reasonable doubt.

A Trial as a Hypothesis Test

- Hypothesis testing is very much like a court trial.

  - This doesn't prove the hypothesis but we can say that the data support it.

- H_o: Defendant is innocent

  H_A: Defendant is guilty

- We then present the evidence - collect data.

- Then we judge the evidence - "Could these data plausibly have happened by chance if the null hypothesis were true?"

  - If they were very unlikely to have occurred, then the evidence raises more than a reasonable doubt in our minds about the null hypothesis.

- Ultimately we must make a decision. How unlikely is unlikely?

A Trial as a Hypothesis Test

- If the evidence is not strong enough to reject the presumption of innocent, the jury returns with a verdict of "not guilty"

  - The jury does not say that the defendant is innocent.

  - All it says is that there is not enough evidence to convict, to reject innocence.

  - The defendant may, in fact, be innocent, but the jury has no way to be sure.

- Said statistically, we fail to reject the null hypothesis.

  - We never declare the null hypothesis to be true, because we simply do not know whether it's true or not.

  - Therefore we never "accept the null hypothesis"

A Trial as a Hypothesis Test

- In a trial, the burden of proof is on the prosecution.

- In a hypothesis test, the burden of proof is on the unusual claim.

- The null hypothesis is the ordinary state of affairs (the status quo), so it's the alternative hypothesis that we consider unusual (and for which we must gather evidence).

How do we determine if H₀ is plausible?

- In Statistics we can quantify our level of doubt.

  - How unlikely is it to get a random sample of 100 students where 42 have travelled abroad if in fact the true population proportion is 35%?

  - To answer this question we use the model proposed by the null hypothesis as a given and calculate the probability that the event we have witnessed could happen.

    Prob(observed or more extreme outcome | H₀ true)

    $$= \text{Prob}(\ \hat{p}\ > 0.42\ |\ p = 0.35)$$

  - This probability quantifies exactly how surprised we are to see our results and is called the p-value

Calculating the p-value.

- First, we calculate the test statistic (z-score)

- The test statistic used for hypothesis testing for proportions is a z-score.

$$z = \frac{obs - mean}{SD}$$

- Remember the CLT: In calculation of the z-score, we use the mean and the SD suggested by the CLT, and the observed value is the $\hat{p}$

$$\hat{p} \sim N\left(mean = p, SD = \sqrt{\frac{pq}{n}}\right)$$

Note: The p comes from the null hypothesis, i.e. it is value proposed for the parameter in the null hypothesis.

Check the Conditions for the CLT

- **Random and Independent:** The sample is collected randomly and the trials are independent of each other.

- **Large Sample:**

  - The sample has at least 10 successes, $np \geq 10$, and

  - at least 10 failures $n(1 - p) \geq 10$.

- **Large Population:** If the sample is collected without replacement, then the population size is at least 10 times the sample size.

Calculating the p-value.

$H_0 : p = 0.35$

$H_A : p > 0.35$

$$z = \frac{\hat{p} - p}{\sqrt{\frac{pq}{n}}}$$

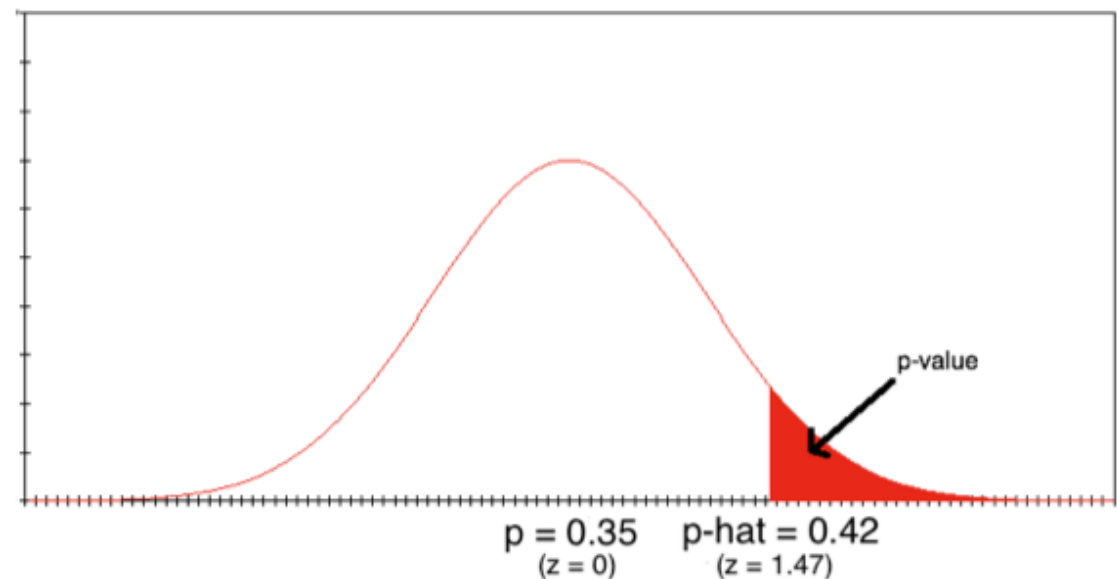$$= \frac{0.42 - 0.35}{\sqrt{\frac{0.35 \times 0.65}{100}}} = 1.47$$

$p - value$

$= Prob(\hat{p} > 0.42 | p = 0.35)$

$= Prob(z > 1.47)$

$= 1 - 0.9292$

$= 0.0708$

p-value

p = 0.35
(z = 0)

p-hat = 0.42
(z = 1.47)

Decision based on the p-value

- When the data are consistent with the model from the null hypothesis, the p-value is high and we are unable to reject the null hypothesis.

  - In that case, we have to "retain" the null hypothesis.

  - We can't claim to have proved it; instead we fail to reject the null hypothesis and conclude that the difference we observed between the null hypothesis (p = 0.35) and the observed outcome ($\hat{p}$ = 0.42) is due to natural sampling variability (or chance).

- If the p-value is low enough, we reject the null hypothesis since what we observed would be very unlikely if in fact the null model was true.

  - We call such results statistically significant

How low a p-value is low enough?

- We compare the p-value to a given α:

  If p-value < α → Reject H₀

  There is sufficient evidence to suggest that H_A is plausible.

  If p-value > α → Fail to reject H₀

  There is not sufficient evidence to suggest that H_A is plausible. The difference we are seeing between the null model and the observed outcome is due to natural sampling variability.

- When p-value is low, it indicates that obtaining the observed $\hat{p}$ or even a more extreme outcome is highly unlikely under the assumption that H₀ is true, therefore we reject that assumption. α level is the complement of the confidence level.

  Remember: When constructing confidence intervals if a confidence level is not specified use 95% confidence.

- Since $1 - 0.95 = 0.05$, if a α level is not specified use $\alpha = 0.05$.

Decision based on the p-value

With a p-value of 0.0708, do we reject or fail to reject H_o?

a) Reject H_o

b) Fail to reject H_o

Decision based on the p-value

What does the conclusion of the hypothesis mean in context of the research question?

a) The data provide convincing evidence to suggest that the true population proportion of UCLA students who have travelled outside the US is 0.35.

b) The data provide convincing evidence to suggest that the true population proportion of UCLA students who have travelled outside the US is 0.42.

c) The data do not provide convincing evidence to suggest that the true population proportion of UCLA students who have travelled outside the US has increased.

d) The data do not provide convincing evidence to suggest that the true population proportion of UCLA students who have travelled outside the US has not changed.

Interpreting the p-value

- A p-value is a conditional probability - the probability of the observed or a more extreme outcome given that the null hypothesis is true.

$$\text{Prob(observed or more extreme outcome} \mid H_0 \text{ true)}$$

- The p-value is NOT the probability that the null hypothesis is true.

- It's not even the conditional probability that the null hypothesis is true given the data.

Interpreting the p-value

- Which of the below is the correct interpretation of the p-value?

    a) The probability that 35% of UCLA students have travelled outside the US is 0.0708.

    b) The probability that more than 35% of UCLA students have travelled outside the US is 0.0708.

    c) If in fact the true population proportion of UCLA students who have travelled outside the US is greater than 0.35, the probability of getting a random sample where the sample proportion is 0.42 or higher would be 0.0708.

    d) If in fact the true population proportion of UCLA students who have travelled outside the US is 0.35, the probability of getting a random sample where the sample proportion is 0.42 or higher would be 0.0708.

Alternative Hypotheses

- When testing for population proportions, there are three possible alternative hypotheses:

  - H_A: $p < p_o$

  - H_A: $p > p_o$

  - H_A: $p \neq p_o$

- We decide on which alternative hypothesis to use based on we hypothesize (or what the wording of the question instructs as to hypothesize):

  - Smaller, less, decreased, fewer

  - Larger, greater, more, increased

  - Different, not equal to, changed

- We DO NOT decide on which alternative hypothesis to use based on what the data suggest.
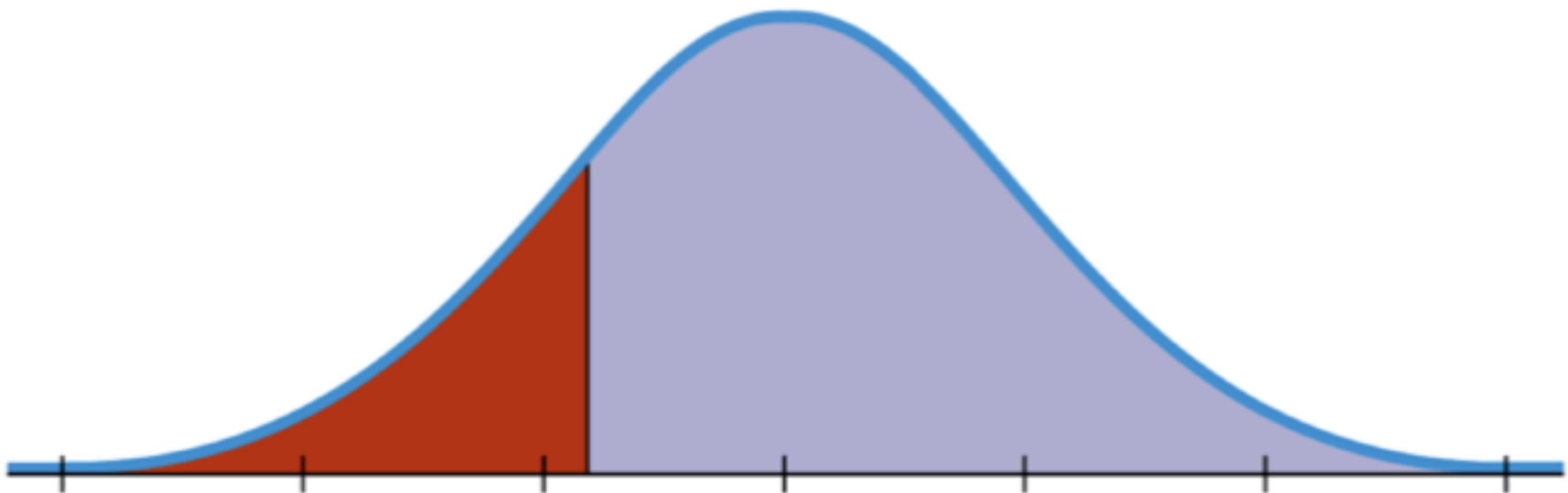
Alternative Hypotheses

- H_A: parameter $\neq$ hypothesized value is known as a two-sided alternative because we are equally interested in deviations on either side of the null hypothesis value.

- For two-sided alternatives, the p-value is the probability of deviating in either direction from the null hypothesis value.

- p-value is the sum of the shaded (red) areas.
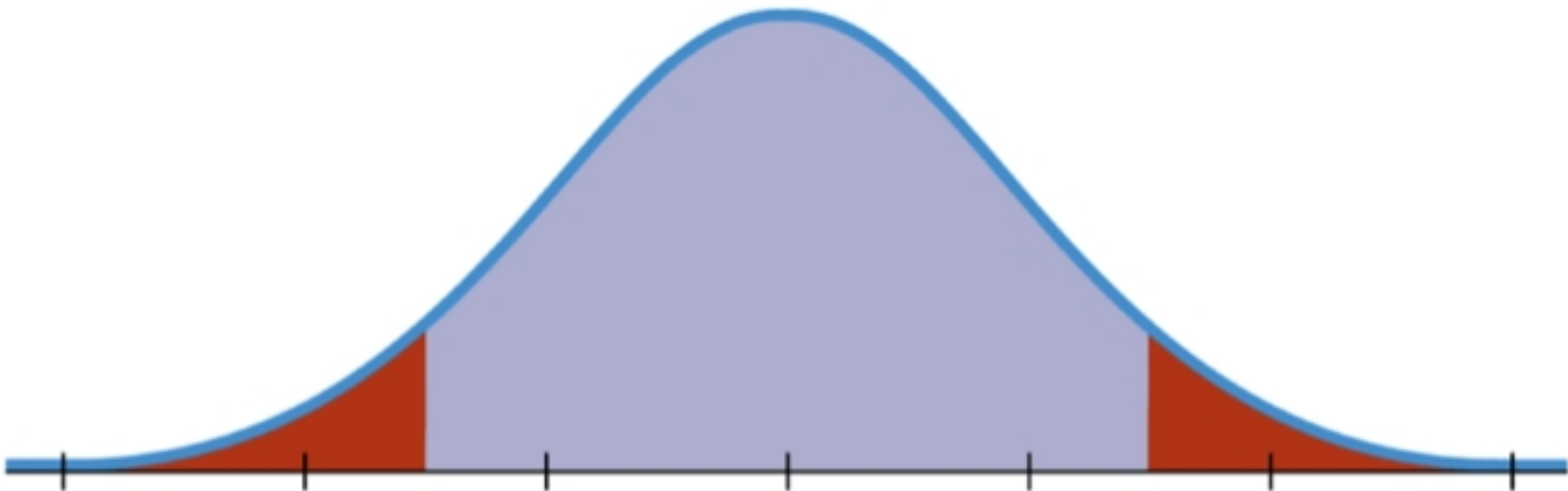
Alternative Hypotheses

- The other two alternative hypotheses are called

- A one-sided alternative focuses on deviations from the null hypothesis value in only one direction.

- Thus, the p-value for one-sided alternatives is the probability of deviating only in the direction of the alternative away from the null hypothesis value.

- p-value is the shaded (red) area.

Steps for a Hypothesis Test

1 *Hypotheses:* state the null and alternative hypotheses

$$H_0 : p = p$$

$$H_A : p < p_0 \text{ OR } p > p_0 \text{ OR } p \neq p_0$$

2 *Assumptions/Conditions:* check to see if assumptions are satisfied and state "Because the conditions are (not) satisfied, I can (not) model the sampling distribution of the sample proportion with the Normal model."

3 *test statistic:* calculate the appropriate test statistic.

$$z = \frac{obs - mean}{SD} = \frac{\hat{p} - p}{\sqrt{\frac{pq}{n}}}$$

Steps for a Hypothesis Test

4. *p-value:* Find the p-value - the probability that the observed statistic value (or an even more extreme value) could occur if the null model were correct.

   - If H_A has a $<$ sign, p-value $= Prob(z < z_{obs})$.
   - If H_A has a $>$ sign, p-value $= Prob(z > z_{obs})$.
   - If H_A has a $\neq$ sign, p-value $= Prob(z > |z_{obs}|)$.

5. *Conclusion:* Either reject or fail to reject the null hypothesis and put your conclusion in context.

   - If p-value $< \alpha$, reject H_0, there is sufficient evidence to suggest that the alternative hypothesis is plausible.
   - If p-value $> \alpha$, fail to reject H_0, there isn't sufficient evidence to suggest that the alternative hypothesis is plausible, the difference between the null hypothesized population proportion and the sample proportion is due to natural sampling variability.

Write the appropriate hypotheses

A company with a fleet of 150 cars found that the emissions systems of 7 out of the 22 they randomly chose and tested failed to meet pollution control guidelines. We would like to test if this is strong evidence that more than 20% of the fleet might be out of compliance. What are the appropriate hypotheses?

Do a Hypothesis Test

A survey conducted five years ago by the health center at a university showed that 18% of the students smoked at the time. This year a new survey was conducted on a random sample of 200 students from this university, and it was found that 40 of them smoke. Do these data provide convincing evidence to suggest that the percentage of students who smoke has changed over the last five years?

Comparing 2 proportions

- Compare 2 sample proportions each from a different group

  - Are the two groups the same? Are they different?

- Same ideas as for 1 proportion, but the formulas change

  - Confidence Interval for difference in proportions

  - Hypothesis test for difference in proportions

Remember: CI for 1 proportion.

Depending on the confidence level, we choose how many SEs to add/subtract to/from $\hat{p}$. So, a more generic formula for the CI can be written as

$$\hat{p} \pm \underbrace{z^* SE(\hat{p})}_{ME}$$

where

$$SE(\hat{p}) = \sqrt{\frac{\hat{p}\hat{q}}{n}}$$

and z* is a critical z-score that depends on the confidence level.

Confidence Intervals for the difference in proportions

When we have two populations to compare, we can also find a confidence interval for the difference between the two population proportions:

$$(\hat{p}_1 - \hat{p}_2) \pm z^* \times \sqrt{\frac{\hat{p}_1 \hat{q}_1}{n_1} + \frac{\hat{p}_2 \hat{q}_2}{n_2}}$$

where z* is still the critical z-score that depends on the confidence level. The numbers 1 and 2 indicate the two different groups.

Confidence Intervals for the difference in proportions

A recent survey asked San Diego residents the following question: "If you had been aware that before Schwarzenegger ran for Governor, he had fathered a child outside of his marriage, would you still have voted or him?" Out of the 136 males surveyed who had voted for Schwarzenegger at least once 78 said they would still vote for him, and out of the 128 females 48 said they would. Calculate a 95% confidence interval for the difference between the proportions of male and female SD residents who have would have voted for Schwarzenegger even if they knew that he had fathered a child outside of his marriage.

Confidence Intervals for the difference in proportions

Calculate a 95% confidence interval for the difference between the proportions of male and female SD residents who have would have voted for Schwarzenegger even if they knew that he had fathered a child outside of his marriage.

	Males Group 1	Females Group 2
n	136	128
success	78	48
$\hat{p}$	$\frac{78}{136} = 0.57$	$\frac{48}{128} = 0.38$
$\hat{q}$	0.43	0.62

Confidence Intervals for the difference in proportions

Randomization condition: We are told that the sample is random.

Large Sample: There are 78 successes and 58 failures in the male group (both ≥ 10) and 48 successes and 80 failures in the female group (also both ≥ 10). Since the success/failure condition is met, we can assume that the distributions of males and females are nearly normal.

Big Population (10% Condition): 136 < 10% of all male SD residents who voted for Schwarzenegger before and 128 < 10% of all female SD residents who voted for Schwarzenegger before.

Independent groups: The two groups should be independent of one another. Males are independent of females, i.e., you can't be in both groups.

Confidence Intervals for the difference in proportions

$$(\hat{p}_M - \hat{p}_F) \pm z^* \times \sqrt{\frac{\hat{p}_M \hat{q}_M}{n_M} + \frac{\hat{p}_F \hat{q}_F}{n_F}}$$

$$= (0.57 - 0.38) \pm 1.96 \times \sqrt{\frac{0.57 \times 0.43}{136} + \frac{0.38 \times 0.62}{128}}$$

$$= 0.19 \pm 0.06$$

$$= (0.13, 0.25)$$

We are 95% confident that the proportion of male SD residents who would have voted for Schwarzenegger even if they knew that he had fathered a child outside of his marriage is 13% to 25% higher than the proportion of females.

Determining if there is a difference in proportions between groups

If the confidence interval for the difference between two population proportions includes 0, then it is possible that the difference between the two population proportions equals 0, i.e. that there is no significant difference between the two population proportions.

$$95\% \text{ CI} : (13\%, 25\%)$$

The confidence interval does not include 0, which means that we are 95% confident that the two population proportions are not equal; that proportion of males that would have voted for Schwarzenegger is higher than the proportion of females.

CI for the difference in proportions

We buy a bag of Skittles and M&M's. The bag of Skittles has 24% red out of a total 65 Skittles. The bag of M&M's has 19% red out of a total of 55 M&M's. Construct a 90% confidence interval for the difference in percent red for Skittles versus M&M's. Interpret the interval. Is there evidence for a difference in proportions between the two candies?

CI for the difference in proportions

The number of Prius owners in a random sample of 500 Californians is 50. The number of Prius owners in random sample of 1000 Iowans is 25. Create a 95% CI for the difference in proportions of Californians and Iowans who own Priuses. Is this evidence of a difference that proportion of Prius owners is different in CA than in IA?

Hypothesis Testing for Two Proportions

Hypotheses: state the null and alternative hypotheses

H_o: p1 = p2
H_A: p1 < p2   OR H_A: p1 > p2   OR H_A: p1 ≠ p2

Note:
If p1 = p2, then p1 − p2 = 0. This is always what goes in the null hypothesis since we start by assuming that the two population means are equal, i.e. the two samples come from the same population.

If p1 > p2, then p1 − p2 > 0.
If p1 < p2, then p1 − p2 < 0.
These are one-sided alternative hypotheses.

If p1 ≠ p2, then p1 − p2 < 0. This is a two-sided alternative hypothesis.

Hypothesis Test for the difference in proportions

A recent survey asked San Diego residents the following question: "If you had been aware that before Schwarzenegger ran for Governor, he had fathered a child outside of his marriage, would you still have voted or him?" Out of the 136 males surveyed who had voted for Schwarzenegger at least once 78 said they would still vote for him, and out of the 128 females 48 said they would. Do these data provide convincing evidence to suggest that there is a difference between the proportion of male and female SD residents who would have voted for Schwarzenegger had they known he had fathered a child outside of his marriage?

Hypothesis Test for the difference in proportions

Out of the 136 males surveyed who had voted for Schwarzenegger at least once 78 said they would still vote for him, and out of the 128 females 48 said they would. Do these data provide convincing evidence to suggest that there is a difference between the proportion of male and female SD residents who would have voted for Schwarzenegger had they known he had fathered a child outside of his marriage?

Hypotheses: H_0: $p_M = p_F$

H_A: $p_M \neq p_F$

alternatively... H_0: $p_M - p_F = 0$

H_A: $p_M - p_F \neq 0$

Check the Conditions

The conditions are the same as in the CI for two proportions. We have already checked these and found that the sampling distributions for male and female voters follow a normal distribution.

We can proceed with the hypothesis test.

Combine the sample proportions

Calculate $\hat{p}_{pooled}$: "Pool" the counts from the two groups to calculate a *pooled proportion* ($\hat{p}_{pooled}$) to be used in calculation of the standard error of the difference between the two sample means, $SE(\hat{p}_1 - \hat{p}_2)$.

$$\hat{p}_{pooled} = \frac{\text{\# of successes}_1 + \text{\# of successes}_2}{n_1 + n_2}$$

Remember: The definition of p-value is:

p-value $= P(\text{observed or more extreme outcome} \mid H_0 \text{ true})$

The p-value of a hypothesis test is calculated assuming H_0 is true, since H_0 of a two proportions test sets the population proportions equal to each other ($H_0 : p_1 = p_2$), standard error should be calculated using a single proportion value, not different values for each sample.

Combine the sample proportions

	Males	Females
	Group 1	Group 2
n	136	128
success	78	48
$\hat{p}$	$\frac{78}{136} = 0.57$	$\frac{48}{128} = 0.38$
$\hat{q}$	0.43	0.62

$\hat{p}_{pooled}$:

$$\hat{p}_{pooled} = \frac{78 + 48}{136 + 128} = \frac{126}{264} = 0.48$$

$$\hat{q}_{pooled} = 1 - 0.48 = 0.52$$

Calculate the test statistic (z-score)

Test-statistic: calculate the appropriate test statistic.

$$z = \frac{obs - mean}{SD} = \frac{(\hat{p}_1 - \hat{p}_2) - (p_1 - p_2)}{SE(\hat{p}_1 - \hat{p}_2)},$$

where $SE(\hat{p}_1 - \hat{p}_2) = \sqrt{\dfrac{\hat{p}_{pooled}\,\hat{q}_{pooled}}{n_1} + \dfrac{\hat{p}_{pooled}\,\hat{q}_{pooled}}{n_2}}$

Note: $(p_1 - p_2)$ *is always 0 since we set the two population proportions equal to each other in the null hypothesis.*

Therefore, the calculation of the test statistic simplifies to:

$$z = \frac{(\hat{p}_1 - \hat{p}_2)}{\sqrt{\dfrac{\hat{p}_{pooled}\,\hat{q}_{pooled}}{n_1} + \dfrac{\hat{p}_{pooled}\,\hat{q}_{pooled}}{n_2}}}$$
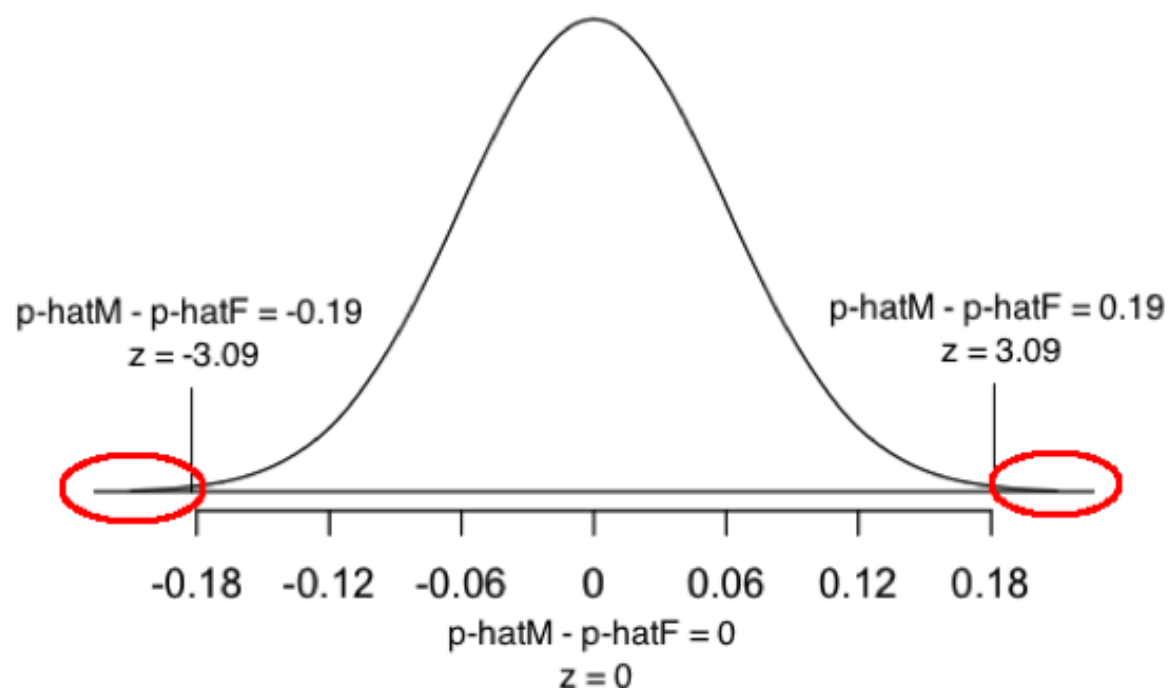
Calculate the test statistic (z-score)

Test-statistic: calculate the appropriate test statistic.

$$z = \frac{(\hat{p}_M - \hat{p}_F)}{\sqrt{\frac{\hat{p}_{pooled}\,\hat{q}_{pooled}}{n_1} + \frac{\hat{p}_{pooled}\,\hat{q}_{pooled}}{n_2}}}$$

$$= \frac{(0.57 - 0.38)}{\sqrt{\frac{0.48*0.52}{138} + \frac{0.48*0.52}{126}}}$$

$$= \frac{0.19}{0.0615}$$

$$= 3.09$$

Find the p-value

p-value:

$$p - value = Prob(z > 3.09) + Prob(z < -3.09)$$
$$= 0.0010 + 0.0010 = 0.0020$$

p-hatM - p-hatF = -0.19
z = -3.09

p-hatM - p-hatF = 0.19
z = 3.09

-0.18   -0.12   -0.06   0   0.06   0.12   0.18

p-hatM - p-hatF = 0
z = 0

Make a conclusion

Based on the p-value of 0.0020, which of the below is the most appropriate conclusion for this hypothesis test?

(a) We can Reject H$_o$. There is evidence to suggest that males are more likely ...

(b) Reject H$_o$. There is evidence to suggest that males and females are not equally likely ...

(c) Fail to reject H$_o$. There males and females are equally likely ...

(d) Fail to reject H$_o$. There is evidence to suggest that males and females are not equally likely ...

to have voted for Schwarzenegger even if they knew that he had fathered a child outside of his marriage.

2 proportion hypothesis test example

A USA Today/Gallup poll surveyed 1,145 unemployed and 675 underemployed randomly sampled people. 27% of the unemployed and 25% of the underemployed people said they had major problems in relationships as a result of their employment status. Do the data provide convincing evidence to suggest a difference between the true population proportions of unemployed and underemployed people who had such problems. Use a 5% significance level.

2 proportion hypothesis test example

Do the data provide convincing evidence to suggest a difference between the true population proportions of unemployed and underemployed people who had such problems. Use a 5% significance level.

	Unemployed Group 1	Underemployed Group 2
n	1,145	675
$\hat{p}$	0.27	0.25
$\hat{q}$	0.73	0.75