Chapter 10: Hypothesis Testing for Population Proportions

1 What is a hypothesis?

A hypothesis can be defined as a proposed explanation for a phenomenon. It is not the absolute truth, but a provisional, working assumption, perhaps best thought of as a potential suspect in a criminal case.

2 Null and alternate hypothesis in a non-"mathy" way

In scientific research, we often start with two competing claims: the *null hypothesis* and the *alternative hypothesis*. Let's break these down in simple terms.

The null hypothesis is what we are willing to assume is the case until proven otherwise. It is relentlessly negative, denying all progress and change. But this does not mean that we actually believe the null hypothesis is literally true. The alternative hypothesis is the suspected claim that contradicts or attempts to replace the null hypothesis.

Example: Legal System Analogy. There is a strong analogy with criminal trials in the legal system. It is said that *"innocent until proven guilty."* Here, the null hypothesis is that the person is not guilty and the alternative hypothesis is that the person is guilty. Unless there are compelling arguments to refute the null hypothesis, the court believes that the person is innocent.

Anothe example. Imagine you're a teacher who wants to determine whether a new teaching method is more effective than the traditional method. You decide to test this by comparing the test scores of two groups of students: one group taught using the new method and the other group taught using the traditional method. The null hypothesis is that the new teaching method is no more effective than the traditional method. This means there is no difference in the average test scores of the two groups. The alternate hypothesis that the new teaching method is more effective than the traditional method. This means the average test score of the group taught with the new method is higher than that of the group taught with the traditional method.

Example: Assume you are in the forest and you hear a rustling in the bushes. Now you have to decide whether you think there is a tiger behind the bushes or not. In this case, the null hypothesis is that there is no tiger behind the bushes. The alternate hypothesis is that there is a tiger behind the bushes.

The process of hypothesis testing in science is like a trial in court. We start with an assumption (null hypothesis) and gather evidence to see if we can reject it in favor of a new claim (alternative hypothesis). This systematic approach helps ensure that our conclusions are based on solid evidence rather than assumptions or biases.

3 Null and alternate hypothesis in a "mathy" way

1. Null Hypothesis (H_0) - Currently accepted or default assumption
2. Alternate Hypothesis (H_A) - The suspected claim that contradicts or attempts to replace the null hypothesis.

Example: Data from the Center for Disease Control estimates that about 30% of American teenagers were overweight in 2008. A professor in public health at a major university wants to determine whether the proportion has changed since 2008. Write down the null and alternate hypothesis.

$H_0 : p = 0.3$

$H_a : p \neq 0.3$

Example: An article recently published in a local newspaper states that 80% of all human population is right handed.You are skeptical of it and want to determine whether the proportion ios less than stated. Write down the null and alternate hypothesis.

$H_0 : p = 0.9$

$H_a : p < 0.9$

Note: Hypotheses are always statements about population parameters; they are never statements about sample statistics.

4 Hypothesis testing in a "non-mathy" way

To actually test a null hypothesis and establish evidence against it, we can perform what is known as a 'Test of Significance.' We do not ever try to "prove" an alternative hypothesis in hypothesis testing. We are trying to refute the null hypothesis; therefore we have two possible outcomes:

1. Reject the null hypothesis
2. Fail to reject the null hypothesis

Example: Height and Starting Salaries. Suppose your best friend believes there is a correlation between one's Height and their starting salary in their first full-time job. You, being a Statistics student, are skeptical of the claim and wish to test this claim.

$H_0 :$ There is some correlation between Height and Starting Salary (claim)

$H_a :$ There is no correlation between Height and Starting Salary (suspicion)

Once we've performed a Significance Test, depending on the result we may:

- Reject the claim that there is a correlation between height and starting salary
- Fail to reject the claim that there is a correlation between height and starting salary

Note: We never try to prove or establish either claim. Tests of significance are inherently tests of refutation.

Example: Suppose you and your friend like to flip a coin to decide who pays when you go out to eat. Your friend has a "lucky" coin that they love to use. If the coin lands on heads, they pay. If the coin lands on tails, you pay. The last 7 times you've gone out to eat, the coin has come up as tails every time. Your friend claims this is a perfectly fair coin, but you have your suspicions.

a) Identify H_0 and H_a :

H_0 : Your friend's cain is not biased (claim)

H_a : Your friend's coin is bias (suspicion)

b) Suppose a significance test finds that the past 7 outcomes being tails is too unlikely to occur purely by chance (by comparing its likelihood to a judgment threshold i.e. significance level). What, then, should the conclusion of your test be?

We reject the null hypothesis that the coin is unbiased.

5 Hypothesis testing in a "mathy" way

As have written our pair of hypotheses, we need to perform the test of significance. The idea behind test of significance is that there are two possible conclusions: We either reject H_0 or we fail to reject H_0 depending on the test result. We reject the null hypothesis if the observed sample is very unlikely to have occurred when H_0 is true.

In this chapter, we consider testing hypotheses about a population proportion when the sample size n is large. We know that p denotes the population parameter. A random sample of n individuals or objects is selected from the population. The sample proportion $\hat{p}$ is the natural statistic for making inferences about p.

Firstly, we check the conditions of the **Central Limit Theorem** (CLT) that applies to estimating proportions in a population to make sure that the the sampling distribution of the sample proportion is close to the Normal distribution. The standard error is calcuted as

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

Now we calculate the test statistic to perform the test of significance as:

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

We call the z here $z-$statistic. The $z-$test statistic has the same structure as the $z-$score introduced in earlier chapter, and it serves the same purpose. By subtracting the value the null hypothesis expects from the observed value, $\hat{p} - p_0$, we learn how far away the actual sample value was from the expected value. A positive value means the outcome was greater than what was expected, and a negative value means it was smaller than what was expected.

If the test statistic value is 0, then the observed value and the expected value are the same. This means we have little reason to doubt the null hypothesis. The null hypothesis tells us that the test statistic should be 0, give or take some amount. If the value is far from 0, then we doubt the null hypothesis.

By dividing this distance by the standard error, we convert the distance into "standard error" units, and we learn how many standard errors away our outcome lies from what was expected. For example, we have a $z-$statistic of -1.34. This tells us that our sample proportion was 1.34 standard errors below the null hypothesis proportion.

6 What if we are wrong?

A Hypothesis Test or Significance Test is ultimately founded upon likelihoods, not definitive answers about reality. Because of that, there is a small chance of arriving at the incorrect conclusion. Hence, we use significance level to compare our $z-$statistic. The significance level is the probability of making the mistake of rejecting the null hypothesis when, in fact, the null hypothesis is true. The significance level is such an important probability that it has its own symbol, α.

6.1 $p-$value

The $p-$value (also sometimes called the observed significance level) is a measure of inconsistency between the hypothesized value for a population characteristic and the observed sample. The $p-$ value is a probability. Assuming that the null hypothesis is true, the $p-$ value is the probability that if the experiment were repeated, you would get a test statistic as extreme as or more extreme than the one you actually got. A small p-value suggests that a surprising outcome has occurred and discredits the null hypothesis.

A decision about whether to reject or to fail to reject H_0 results from comparing the $p-$value to the chosen significance level α:

$$\text{We reject } H_0 \text{ if } p\text{−value} \leq \alpha.$$

$$\text{We fail to reject } H_0 \text{ if } p\text{−value} > \alpha.$$

7 Summary

- Remember hypothesis testing is about interpreting or drawing conclusions about the population using sample data.
- Figure out n, $\widehat{p}$, and p_0 from the problem.
- Write down the pair of hypotheses.

- Step 3: Find the corresponding one-proportion $z-$test statistics. $z = \frac{\widehat{p}-p_0}{\sqrt{\frac{p_O(1-p_O)}{n}}}$
- Step 4: Find the corresponding $p-$value of your $z-$test statistics.
- Step 5: Compare your $p-$value with significance level α

8 Examples

Example 1: A study at a popular gym across the United States, Anytime Fitness, introduced a new fitness program aimed at reducing stress among participants. The researchers surveyed 150 randomly selected participants enrolled in the program. In the survey, 65 of the 150 participants indicated that they believed the fitness program significantly reduced their stress levels. Does the sample provide convincing evidence that more than 40% of the participants believe the fitness program significantly reduces stress levels? Perform hypothesis testing with a significance level of 0.05.

Write down the null and alternative hypothesis:

$$H_0 : p = 0.40$$
$$H_a : p > 0.40$$

Check the conditions of the Central Limit Theorem:

The three necessary conditions of the Central Limit Theorem are met as:

- The sample is random and independent.
- We have enough success and failures within the sample.

  $np = 150(0.4) \geq 10$ and $n(1-p) = 150(1-0.4) \geq 10$
- The population size is at least 10 times the sample size.

Calculate the value of the z-statistic.

The z-statistic is calculated as:

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

Thus,

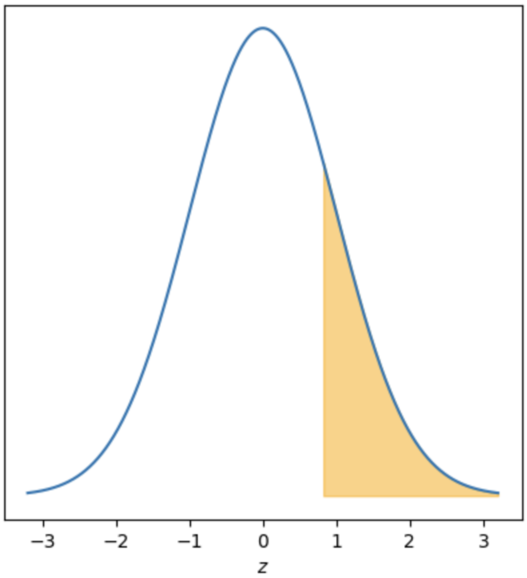
$$z = \frac{0.433-0.40}{\sqrt{\frac{(0.40)(1-0.40)}{150}}} = \frac{0.033}{0.040} = 0.825$$

Calculate the p-value for the corresponding z-statistic.

$$p-\text{value} = \text{area under the } z \text{ curve to the right of } 0.825 = 1 - 0.7954 = 0.2046$$

Interpret the meaning of $p-$value.

This $p-$value is shown in figure below.



If H_0 is true, in the long run, 2046 out of 10,000 samples would result in a $z-$value as or more extreme than (or "greater than or equal to the observed value" than) what actually resulted. This is not a surprising outcome and we can't discredit the null hypothesis.

Compare the p-value with the significance level. Do you 'reject' or 'fail to reject' the null hypothesis?

Since the p-value is 0.2046 and our significance level is 0.05, we have p-value $> \alpha$. Hence, we fail to reject the null hypothesis.

Is there sufficient evidence to suggest that the proportion of participants who believe that the fitness program significantly reduces stress levels is greater than 40%?

There is not strong evidence that the proportion of participants who believe that the fitness program significantly reduces stress levels is greater than 40%. We failed to reject the null hypothesis.

Example 2: In December 2009, a county-wide water conservation campaign was conducted in a particular county. In January 2010, a random and independent sample of 500 homes was selected, and water usage was recorded for each home in the sample. It was found that 220 of the houses reduced their water usage. The county supervisors wanted to know whether their data supported the claim that fewer than half the households in the county reduced water consumption. Perform a hypothesis testing with significance level of 0.01.

Write down the null and alternate hypothesis:

$$H_0 : p = 0.5$$
$$H_a : p < 0.5$$

Check the conditions of Central Limit Theorem:

The three necessary conditions of Central Limit Theorem are met as:

- The sample is random and independent.
- We have enough success and failures within the sample.

  $np = 500(0.5) \geq 10$ and $n(1-p) = 500(1-0.5) \geq 10$

- The population size is at least 10 times the sample size.

Calculate the value of $z-$stat.

$z-$stat is calculated as:

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

Thus,

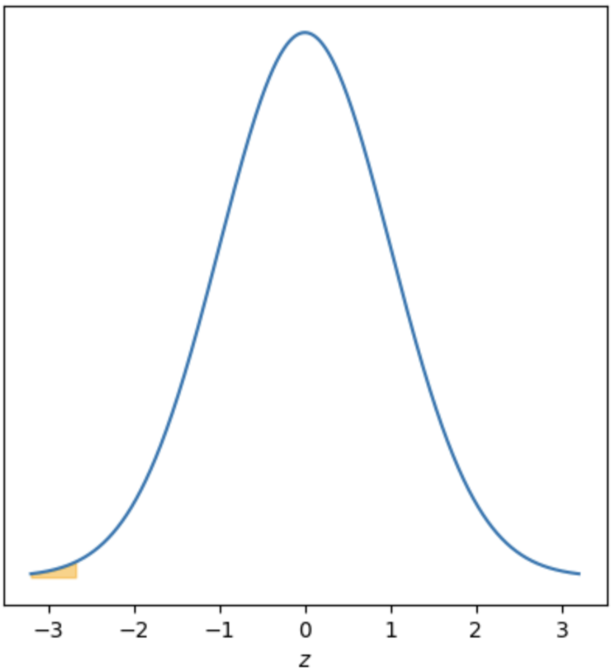
$$z = \frac{0.44 - 0.50}{\sqrt{\frac{(0.50)(1-0.50)}{500}}} = -\frac{0.06}{0.0224} = -2.68$$

Calculate p value for the corresponding $z-$stat.

$$\begin{aligned} P\text{-value} &= \text{ area under the } z \text{ curve to the right of } -2.68 \\ &= 0.0037 \end{aligned}$$

Interpret the meaning of $p-$value.

This $p-$value is shown in figure below.



If H_0 is true, in the long run, only 37 out of 10,000 samples would result in a $z-$value as or more extreme than (or "less than or equal to the observed value" than) what actually resulted. This is a surprising outcome and we will discredit the null hypothesis.

Compare $p-$value with the significance level. Do you 'reject' or 'fail to reject' the null hypothesis?

Since $p-$value is 0.0037 and our significance level is 0.01. Thus, p-value $< \alpha$. Hence, we reject the null hypothesis.

Is there sufficient evidence to suggest that fewer than half the households in the county reduced water consumption.

Using a 0.01 significance level, we rejected null hypothesis, leading us to conclude that there is convincing evidence that the proportion with reduced water usage was less than half.

Example 3: "Around half the Americans Like it Hot" is the title of a press release issued by the Pew Research Center. The press release states that "around half the Americans, Americans want to live in a sunny place." This statement is based on data from a nationally representative sample of 2260 adult Americans. Of those surveyed, 1200 indicated that they would prefer to live in a hot climate rather than a cold climate. Do the sample data provide convincing evidence that the proportion of all adult Americans who prefer a hot climate over a cold climate is different from 50%? Test the hypothesis with a significance level of 0.05 to answer this question.

Write down the null and alternative hypothesis:

$$H_0 : p = 0.50$$
$$H_a : p \neq 0.50$$

Check the conditions of the Central Limit Theorem:

The three necessary conditions of the Central Limit Theorem are met as:

- The sample is random and independent.
- We have enough successes and failures within the sample.

  $np = 2260 \times 0.50 = 1130 \geq 10$ and $n(1-p) = 2260 \times 0.50 = 1130 \geq 10$

- The population size is at least 10 times the sample size.

Calculate the value of the z-statistic.

The z-statistic is calculated as:

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

Thus,

$$z = \frac{0.509 - 0.50}{\sqrt{\frac{(0.50)(1-0.50)}{2260}}} = \frac{0.009}{0.0105} \approx 0.86$$

Calculate the p-value for the corresponding z-statistic.

$$p\text{-value} = 2 \times \text{area under the } z \text{ curve to the right of } 0.86 = 2 \times (1 - 0.8051) = 0.3898$$

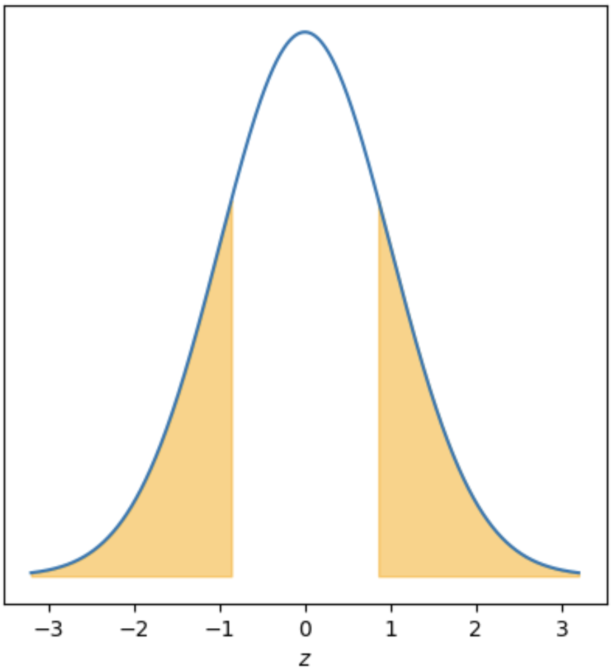
Or,

$$p\text{-value} = 2 \times \text{area under the } z \text{ curve to the left of } -0.86 = 2 \times (0.1949) = 0.3898$$

We will get the same $p-$value. So you can pick one you're more comfortable with.

Interpret the meaning of the p-value.

This $p-$value is shown in figure below.



If H_0 is true, in the long run, 3898 out of 10,000 samples would result in a $z-$value as or more extreme than what actually resulted. This is not a surprising outcome and we won't discredit the null hypothesis.
Compare the p-value with the significance level. Do you 'reject' or 'fail to reject' the null hypothesis?

Since the p-value is 0.3898 and our significance level is 0.01, we have p-value $> \alpha$. Hence, we fail to reject the null hypothesis.

Is there sufficient evidence to suggest that the proportion of Americans who prefer a hot climate over a cold climate is different from 50%?

There is not strong evidence that the proportion of Americans who prefer a hot climate over a cold climate is different from 50%. We fail to reject the null hypothesis.

9 Comapring two population proportions

Just like we computed confidence intervals for two population proportions, we can use hypothesis testing to compare two population proportions. The underlying theory is the same as in hypothesis testing for one sample population proportions.

Sampling Distribution for Difference in Proportion: Firstly, we ensure the sampling distribution is normally distributed for which we need to check the following conditions for Central Limit Theorem:

1. Random and Independent. Both samples are *randomly* drawn from their populations, and observations are independent of each other.

2. Large Samples. Both sample sizes are large enough that at least 10 successes and 10

failures can be expected in both samples. Verify yourself!

3. Big Populations. If the samples are collected without replacement, then both population sizes must be at least 10 times bigger than their samples.

4. Independent Samples. The samples must be independent of each other.

We write the pair of hypotheses as:

Two-sided	One-sided (Left)	One-sided (Right)
$H_0 : p_1 = p_2$ $H_a : p_1 \neq p_2$	$H_0 : p_1 = p_2$ $H_a : p_1 < p_2$	$H_0 : p_1 = p_2$ $H_a : p_1 > p_2$

Now the $z-$statistic is calculated as:

$$z = \frac{\hat{p}_1 - \hat{p}_2 - 0}{\sqrt{\hat{p}(1-\hat{p})\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}}$$

We then compute the corresponding $p-$value and perform the test of significance.

Example: Two types of medicines are being tested in their effectiveness at reducing migraines. Two random samples of 200 patients were each given either Medicine A or Medicine B. In the group given Medicine A, 36 reported no headache after 30 minutes and in the group given Medicine B, 49 reported no headache after 30 minutes. Complete the setup for a hypothesis test at 5% significance level to determine if there is a difference in the proportion of relieved headaches.

Step 1: State the hypotheses We need to test if there is a difference in the proportions of patients who reported no headache after 30 minutes between Medicine A and Medicine B. We will use a two-sided test. Let p_1 be the proportion of patients who reported no headache after 30 minutes for Medicine A, and p_2 be the proportion for Medicine B. We need to test if there is a difference in the proportions of patients who reported no headache after 30 minutes between Medicine A and Medicine B. We will use a two-sided test.

- **Null Hypothesis:**
$$H_0 : p_1 = p_2$$

- **Alternative Hypothesis:**
$$H_a : p_1 \neq p_2$$

Step 2: Calculate the sample proportions

$$\hat{p}_1 = \frac{36}{200} = 0.18$$
$$\hat{p}_2 = \frac{49}{200} = 0.245$$

Step 3: Calculate the Combined Proportion

$$\hat{p} = \frac{36 + 49}{200 + 200} = \frac{85}{400} = 0.2125$$

Step 4: Calculate the standard error

$$SE = \sqrt{\hat{p}(1-\hat{p})\left(\frac{1}{n_1}+\frac{1}{n_2}\right)}$$
$$SE = \sqrt{0.2125 \times (1-0.2125)\left(\frac{1}{200}+\frac{1}{200}\right)}$$
$$SE = \sqrt{0.2125 \times 0.7875\left(\frac{1}{100}\right)}$$
$$SE = \sqrt{0.16734375 \times 0.01}$$
$$SE = \sqrt{0.0016734375}$$
$$SE = 0.04091$$

Step 5: Calculate the test statistic

$$z = \frac{\hat{p}_1 - \hat{p}_2}{SE}$$
$$z = \frac{0.18 - 0.245}{0.04091}$$
$$z = \frac{-0.065}{0.04091}$$
$$z = -1.59$$

Step 6: Calculate the $p-$value

$$p\text{-value} = 2 \times \text{area under the } z \text{ curve to the left of } 1.59 = 2 \times (0.0559) = 0.118$$

Step 7: Make the Decision

Since the $p-$value (0.1118) is greater than the significance level (0.05), we fail to reject reject H_0.

Conclusion

At the 5% level of significance, we do not have sufficient evidence to conclude that there is a difference in the proportions of patients who reported no headache after 30 minutes between Medicine A and Medicine B.