Chapter 8

Hypothesis Testing for Population Proportions

Confidence Intervals

- What did we learn in the previous chapter about confidence intervals?
- Earlier we calculated a 95% confidence interval for the true population proportion of UCLA students who traveled 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 traveled abroad 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 found that 35 of them have at some point in their lives traveled 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=0.30 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 consistency 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 traveled outside the US. UCLA has recently implemented a new study abroad program and results of a new survey show that out of 100 randomly sample students 42 have traveled abroad. Is there significant evidence to suggest that the proportion of students at UCLA who have traveled abroad has increased after the implementation of the study abroad program?

Clicker!

  - What is the sample size?

    A. 35

    B. 42

    C. 100

Hypothesis Testing for One Proportion

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

    A. 35

    B. 42

    C. 100

Hypothesis Testing for One Proportion

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

Clicker!

A. 35%

B. 42%

C. 100%

D. It is not given

Hypothesis Testing for One Proportion

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

    A. 35%

    B. 42%

    C. 100%

    D. It is not given

Hypothesis Testing for One Proportion

- Research conducted a few years ago showed that 35% of UCLA students had traveled outside the US. UCLA has recently implemented a new study abroad program and results of a new survey show that out of 100 randomly sample students 42 have traveled abroad. Is there significant evidence to suggest that the proportion of students at UCLA who have traveled 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 that 0.35.
    - The new data indicate that it may be since 0.42 is clearly greater than 0.35.
    - But is the 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_0, specifies a population model parameter of interest and proposes a value for that parameter (p).

$$H_0: p=0.35$$

- **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_0 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

- If the evidence is not strong enough to reject the presumption of innocence, 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).

A Pair of Hypotheses

- Hypotheses are statements about population parameters.
- The null hypothesis is the neutral, status quo, skeptical statement about a population parameter.
    - It often represents “no change”, “no effect” or “no difference”.
    - It will always have an equal sign.
- The alternative hypothesis is the research hypothesis.
    - It is a statement about the value of a parameter that we intend to demonstrate is true.
    - It can have a less than, greater than or not equal to sign.

Alternative Hypothesis

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

$$H_a: p < p_0$$

$$H_a: p > p_0$$

$$H_a: p \neq p_0$$

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

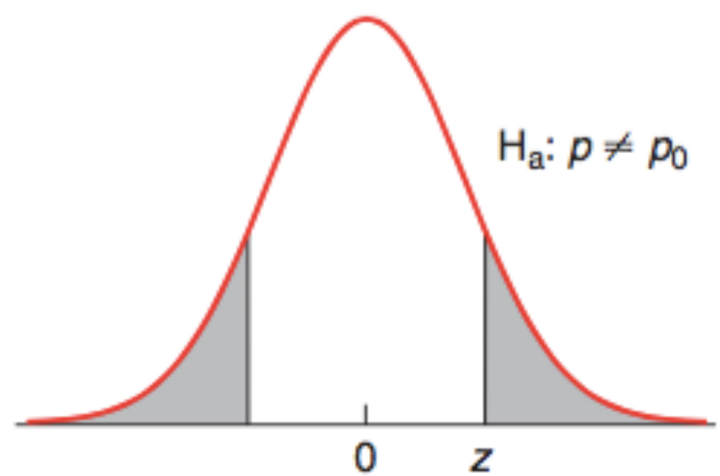
Alternative Hypothesis

$$\mathbf{H_a: p \neq p_0}$$

- This is known as the two-sided alternative hypothesis.
- We are equally interested in deviations on either side of the null hypothesis value.
- Example: A coin is flipped 20 times and comes up heads 18 times. You want to test the hypothesis that the coin does not come up 50% heads in the long run.

$$H_0: p = 0.50$$

$$H_a: p \neq 0.50$$



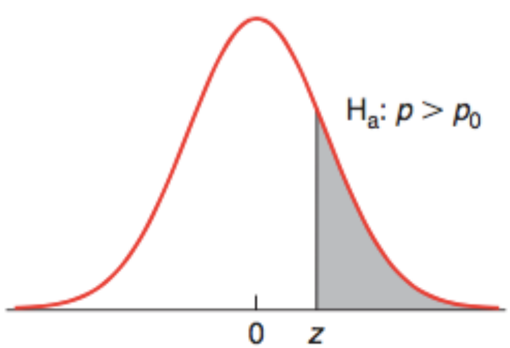
Alternative Hypothesis

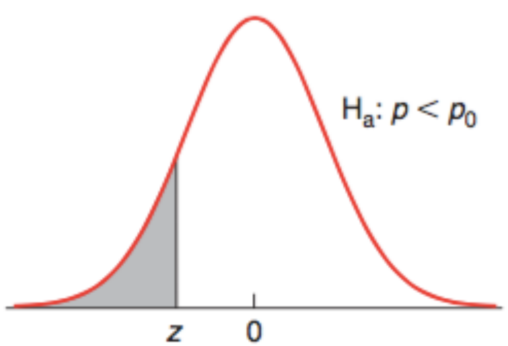
H_a: $p > p_0$ **H_a: $p < p_0$**

- These are known as the one-sided alternative hypothesis.
- A one-sided alternative focuses on deviations from the null hypothesis value in only one direction.
- Example: A coin is flipped 20 times and comes up heads 18 times. You want to test the hypothesis that the coin comes up more than 50% heads in the long run.

$$H_0: p = 0.50$$

$$H_a: p > 0.50$$





Example

Clicker!

Historically, about 70% of all US adults were married. A sociologist who believes that marriage rates in the United States have declined will take a random sample of US adults and record whether or not they are married.

- State the null hypothesis in words.

Answer:

A. The same proportion of adults are married now as in the past.

B. The proportion of adults who are married now is less than it was in the past.

C. The proportion of adults who are married now is greater than it was in the past.

D. The proportion of adults who are married now is different than it was in the past.

Example

Clicker!

Historically, about 70% of all US adults were married. A sociologist who believes that marriage rates in the United States have declined will take a random sample of US adults and record whether or not they are married.

- State the alternative hypothesis in words.

Answer:

A. The same proportion of adults are married now as in the past.

B. The proportion of adults who are married now is less than it was in the past.

C. The proportion of adults who are married now is greater than it was in the past.

D. The proportion of adults who are married now is different than it was in the past.

Example

Historically, about 70% of all US adults were married. A sociologist who believes that marriage rates in the United States have declined will take a random sample of US adults and record whether or not they are married.

- State the null hypothesis and alternative hypothesis in words.

  Null hypothesis: The same proportion of adults are married now as in the past.

  Alternative hypothesis: The proportion of adults who are married now is less than it was in the past.

Example

Clicker!

Historically, about 70% of all US adults were married. A sociologist who believes that marriage rates in the United States have declined will take a random sample of US adults and record whether or not they are married.

- State the null hypothesis in terms of the population parameter, p.

Answer:

A. H_0: $p = 0.70$

B. H_0: $p > 0.70$

C. H_0: $p < 0.70$

D. H_0: $p \neq 0.70$

Example

Clicker!

Historically, about 70% of all US adults were married. A sociologist who believes that marriage rates in the United States have declined will take a random sample of US adults and record whether or not they are married.

- State the alternative hypothesis in terms of the population parameter, p.

Answer:

A. H_a: $p = 0.70$

B. H_a: $p > 0.70$

C. H_a: $p < 0.70$

D. H_a: $p \neq 0.70$

Example

Historically, about 70% of all US adults were married. A sociologist who believes that marriage rates in the United States have declined will take a random sample of US adults and record whether or not they are married.

- State the null hypothesis and alternative hypothesis in terms of the population parameter, p.

  We are told historically p is 0.70 so,

$$H_0: p = 0.70$$

$$H_a: p < 0.70$$

Example

An Internet retail business is trying to decide whether to pay a search engine company to upgrade its advertising. In the past, 15% of customers who visited the company's webpage by clicking on the advertisement bought something (this is called a "click-through"). If the business decides to purchase premium advertising, then the search engine company will make that company's ad more prominent.

The search engine company offers to do an experiment: For one day, customers will see the retail business's ad in the more prominent position. The retail business can then decide whether the advertising improves the percentage of click-throughs. The retailer agrees to the experiment, and when it's over, 17% of the customers have bought something.

A marketing executive wrote the following hypotheses:

H_0: $\hat{p} = 0.15$

H_a: $\hat{p} = 0.17$

- What is wrong with they hypotheses? Rewrite them so they are correct.

Example

A marketing executive wrote the following hypotheses:

H_0: $\hat{p} = 0.15$

H_a: $\hat{p} = 0.17$

- What is wrong with they hypotheses? Rewrite them so they are correct.

First, these hypotheses are written about the ***sample proportion***. We know that 17% of the sample bought something, so there is no need to make a hypothesis about it. What we don't know is what proportion of the entire population of people who will click on the advertisement will purchase something. The hypotheses should be written in terms of p.

A second problem is with the ***alternative hypothesis***. The research question that the company wants to answer is not whether 17% of customers will purchase something. It wants to know whether the percentage of customers who do so has increased over what has happened in the past.

The correct hypotheses:

H_0: $p = 0.15$

H_a: $p > 0.15$

Making Mistakes

- Mistakes are an inevitable part of the hypothesis testing process. The trick is not to make them too often.
- The significance level is the name of a special probability.
    - It is the probability of making the mistake of rejecting the null hypothesis when, in fact, the null hypothesis is true.
    - The significance level is known as alpha, α.
    - We want a small significance level. Most commonly $\alpha=0.05$ is used, but also sometimes $\alpha=0.01$ and $\alpha=0.10$.
- Examples:
    - In flipping a coin, the significance level is the probability that we will conclude that flipping a coin is not fair when, in fact, it really is fair.
    - In the criminal justice setting, the significance level is the probability that we conclude that the suspect is guilty when he is actually innocent.

Example

Referring to the internet retail business example. Recall that in the past, the proportion of customers who brought the product was 0.15, and the company hopes this proportion has increased. It intends to test these hypotheses with a significance level of 5%. In other words, α=0.05.

$$H_0: p = 0.15$$

$$H_a: p > 0.15$$

- Describe the significance level in context.

Example

Referring to the internet retail business example. Recall that in the past, the proportion of customers who bought the product was 0.15, and the company hopes this proportion has increased. It intends to test these hypotheses with a significance level of 5%. In other words, $\alpha=0.05$.

$$H_0: p = 0.15$$

$$H_a: p > 0.15$$

- Describe the significance level in context.

The significance level is the probability of rejecting H_0 when in fact it is true. In this context, this means that the probability is 5% that the company will conclude that the proportion of customers who will buy its product is bigger than 0.15 when in fact, it is 0.15.

The Test Statistic

- How unlikely is it to get a random sample of 100 students where 42 have traveled abroad if in fact the true population proportion is 35%?
- A test statistic compares the real world with the null hypothesis world.
- It compares our observed outcome with the outcome the null hypothesis says we should see.
- We use the following formula to find the test statistic:

$$z = \frac{\text{observed value - null value}}{SE}$$

- For this example,

$$z = \frac{0.42 - 0.35}{\sqrt{\frac{0.35(1-0.35)}{100}}} = 1.47$$

- A positive value means the outcome was greater than what was expected, and a negative value means it was smaller than what was expected.

How do we determine if H_0 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 traveled 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 | H_0 true)
    =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.

P-values

- 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 (closer to 0) suggests that a surprising outcome has occurred and discredits the null hypothesis.
- A large p-value (closer to 1) means there is no surprise and the outcome occurs fairly often.

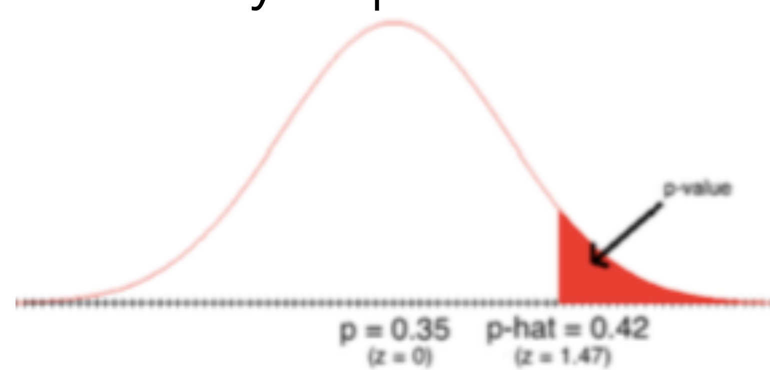
Example

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

  p-value = Prob($\hat{p}$> 0.42 | p=0.35) = Prob(z > 1.47)

  = 1 - 0.9293 = 0.0708

- Here we get a p-value that is very small (close to 0). This tells that if it is true that 35% of students have traveled abroad, then getting a test statistic as extreme as or more extreme as 1.47 is very improbable. If you believed that the null hypothesis was true and p=0.35, then you should be very surprised, because what you saw is nearly impossible.



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 $< \alpha$, we reject the H_0.
    - There is sufficient evidence to suggest that H_a is plausible.
  - If p-value $> \alpha$, we fail to reject the H_0.
    - 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 or even a more extreme outcome is highly unlikely under the assumption that H_0 is true, therefore we reject that assumption. α level is the complement of the confidence level.
- Note: 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 $\alpha=0.05$.

Steps for a Hypothesis Test

1. Hypothesize: state the null and alternative hypotheses
2. Prepare. Type of test: one-proportion z-test. Check conditions: check to see if conditions are met 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.

$$z = \frac{\text{observed value - null value}}{SE} = \frac{\hat{p} - p}{\sqrt{\frac{p(1-p)}{n}}}$$

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 ≠ 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 < α, reject the H_0, there is sufficient evidence to suggest that the alternative hypothesis is plausible.
   - If p-value > α, 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.

Hypothesis Test Example

In one Florida election, 47% of all registered voters voted. A researcher studying the behavior of academics was curious about whether political scientists voted in the same proportion as the rest of the people in the state. To answer this question, the researcher took a random sample, without replacement, of 54 political scientists who live in Florida and interviewed them to determine whether they voted in this election. As it turned out, 40 out of 54 political scientists in the sample voted, which is a sample proportion of 0.74.

We wish to carry out a hypothesis test to determine whether the proportion of all political scientists who voted in this election differed from the proportion of the general public.

Hypothesis Test Example

1. H_0: Political scientists vote in the same proportion as the public, 0.47.

   H_a: Political scientists do not vote in the same proportion as the public.

$$H_0: p = 0.47$$

$$H_a: p \neq 0.47$$

2. Type of test: one-proportion z-test. Check conditions:

Random and Independent: We are told the data come from a random sample of 54 political scientists. The researcher used a random sample of political scientists, so we assume their responses were independent of one another. Condition is satisfied.

Large Sample: Condition is satisfied.

$np = 54 \times 0.47 = 25.38 > 10$

$n(1-p) = 54 \times (1-0.47) = 28.62 > 10$

Big Population: The population size needs to be greater than $10 \times 54 = 540$. Condition is satisfied.

Hypothesis Test Example

3. Test statistic:

Since the problem states 40 out of 54 political scientists voted then our sample proportion is 40/54=0.74.

$$z = \frac{\text{observed value - null value}}{SE} = \frac{\hat{p} - p}{\sqrt{\frac{p(1-p)}{n}}}$$

$$= \frac{0.74 - 0.47}{\sqrt{\frac{0.47(1-0.47)}{54}}} = 3.98$$

Hypothesis Test Example

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

$Prob(z<|z_{obs}|) = Prob(z<z_{obs})+Prob(z>z_{obs})$

The p-value is $Prob(z<-3.98)+Prob(z>3.98)$ = $0.00003+(1-.9999683) = 0.00006$

5. Conclusion:

We choose the significance level is $\alpha=0.05$.

The p-value = $0.00006 < 0.05$, so we reject the null hypothesis. We conclude that political scientists did not vote in the same proportion as the rest of the population.

Example

Health professionals are often concerned about our lifestyles and how they affect our well-being. A group of medical researchers knew from previous studies that in the past, about 39% of all men between the ages of 45 and 59 were regularly active. Because regular activity is good for our health, researchers were concerned that this percentage had declined over time. For this reason they selected a random sample, without replacement, of 1927 men in this age group and interviewed them. Out of this sample, 680 said that they were regularly active.

Carry out the first two steps of a hypothesis test that will test whether the proportion of regularly active men in this age group has declined. Use a significance level of 5%.

Hypothesis Test Example

1. In the past, $p=0.39$. The researchers wish to know whether this proportion has declined (or decreased), which means we have a left-sided alternative.

$$H_0: p = 0.39$$

$$H_a: p < 0.39$$

2. Type of test: one-proportion z-test. Check conditions:

Random and Independent: We are told the sample is random. As long as the sample was random and the men were interviewed independently, this condition is satisfied.

Large Sample: Condition is satisfied.

$np = 1927 \times 0.39 = 752 > 10$

$n(1-p) = 1927 \times (1-0.39) = 1175 > 10$

Big Population: The population size needs to be greater than $10 \times 1927 = 19270$. Condition is satisfied.

Example

Health professionals are often concerned about our lifestyles and how they affect our well-being. A group of medical researchers knew from previous studies that in the past, about 39% of all men between the ages of 45 and 59 were regularly active. Because regular activity is good for our health, researchers were concerned that this percentage had declined over time. For this reason they selected a random sample, without replacement, of 1927 men in this age group and interviewed them. Out of this sample, 680 said that they were regularly active.

Carry out the final steps of a hypothesis test that will test whether the proportion of regularly active men in this age group has declined. Use a significance level of 5%.

Hypothesis Test Example

3. Test statistic:

Since the problem states 680 out of 1927 men are regularly active then our sample proportion is 680/1927=0.353.

$$z = \frac{\text{observed value - null value}}{SE} = \frac{\hat{p} - p}{\sqrt{\frac{p(1-p)}{n}}}$$

$$= \frac{0.353 - 0.39}{\sqrt{\frac{0.39(1-0.39)}{1927}}} = -3.33$$

Hypothesis Test Example

4. p-value: If H_a has a < sign, p-value = Prob($z<z_{obs}$).

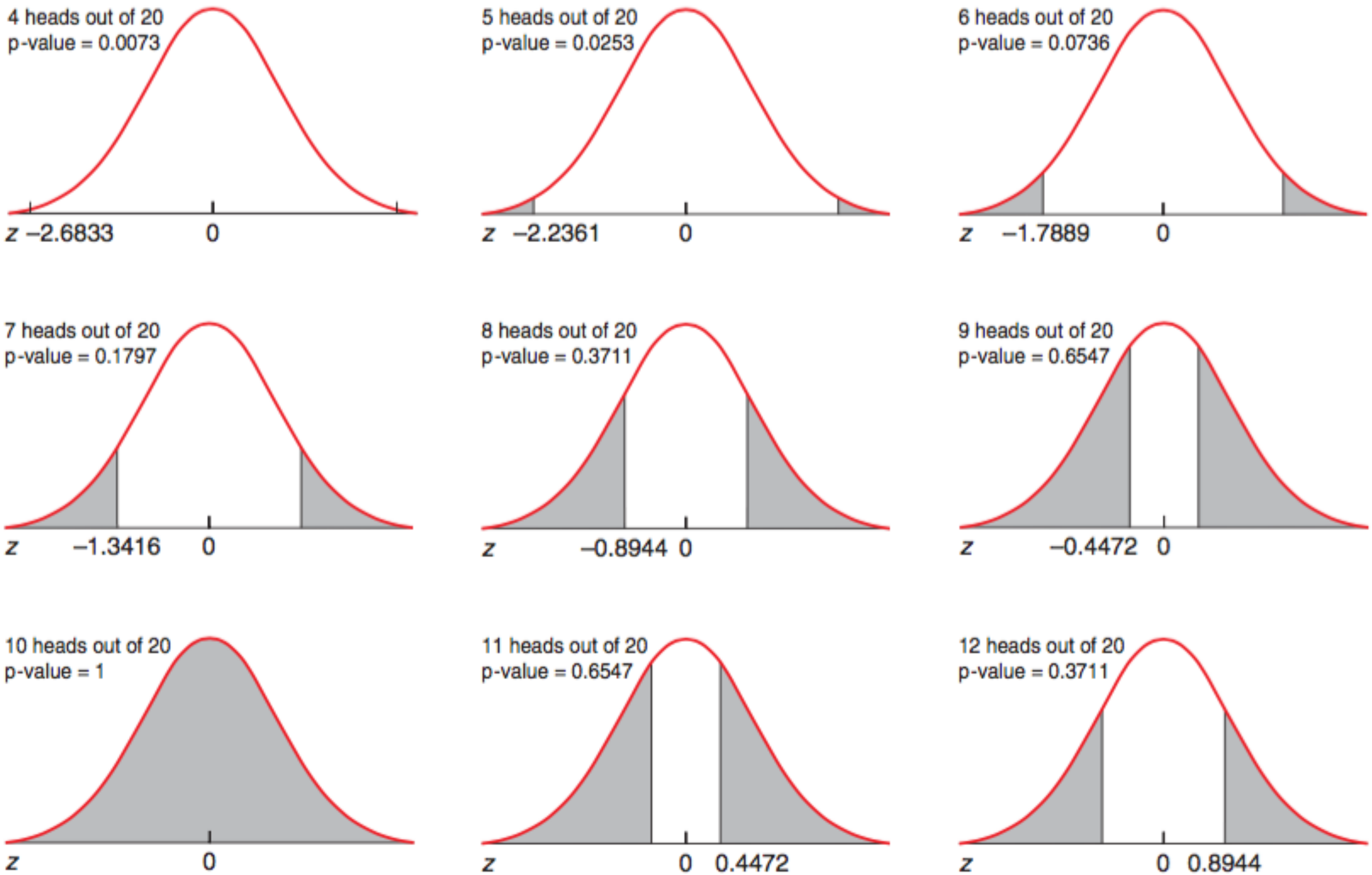
The p-value is Prob($z<-3.33$) = 0.0004

5. Conclusion:

The significance level is $\alpha=0.05$.

The p-value = 0.0004 < 0.05, so we reject the null hypothesis. We conclude that the proportion of regularly active men in this age group has declined.

Test Statistics and P-values



Each graph shows the p-value (shaded) for a different number of heads out of 20 flips of a coin. The closer the number of heads is to 10, the closer the observed test statistic is to 0 and the larger the p-value is. If the null hypothesis is correct, then the test statistic should be close to 0.

Types of Mistakes

- The first type of mistake is to reject the null hypothesis when it is true. This is also known as a **Type I Error**.
- The second type of mistake is to fail to reject the null hypothesis when it is false. This is also known as a **Type II Error.**

Type I and Type II Errors

- In medical tests:
    - A type I error is a false positive. (They conclude someone has a disease when they don’t.)
    - A type II error is a false negative. (They conclude someone does not have a disease then they actually do.)
- As you can see, these errors can have very different consequences.

Type I and Type II Errors

TABLE 2.9 **A summary of Type I and Type II errors**

		What is true (unknown to us)	
		Null hypothesis is true	Null hypothesis is false
What we decide (based on data)	Reject null hypothesis	Type I error (false alarm)	Correct decision
	Do not reject null hypothesis	Correct decision	Type II error (missed opportunity)

Type I and Type II Errors

- In medical tests:
    - A type I error is a false positive. (They conclude someone has a disease when they don't.)
    - A type II error is a false negative. (They conclude someone does not have a disease then they actually do.)
- As you can see, these errors can have very different consequences.

Example

Clicker!

- Lie detector tests are similar to hypothesis tests in that there are two possible decisions and two possible realities and therefore two kinds of errors can be made. Which of the following is a Type I error?

Answer:

A. We have strong evidence the subject is not telling the truth, when in fact they are telling the truth.

B. We don’t have strong evidence the subject is not telling the truth (it’s plausible they are telling the truth), when in fact they are lying.

Example

Clicker!

- Lie detector tests are similar to hypothesis tests in that there are two possible decisions and two possible realities and therefore two kinds of errors can be made. Which of the following is a Type II error?

Answer:

A. We have strong evidence the subject is not telling the truth, when in fact they are telling the truth.

B. We don't have strong evidence the subject is not telling the truth (it's plausible they are telling the truth), when in fact they are lying.

Example

- Lie detector tests are similar to hypothesis tests in that there are two possible decisions and two possible realities and therefore two kinds of errors can be made.

Type I error: We have strong evidence the subject is not telling the truth, when in fact they are telling the truth.

Type II error: We don't have strong evidence the subject is not telling the truth (it's plausible they are telling the truth), when in fact they are lying.

CI and Hypothesis Testing

- Confidence Intervals and hypothesis tests are closely related, even though they are used to answer (slightly) different questions.
    - CI: “What is the value of this parameter?”
    - Hypothesis test: “Are the data consistent with the parameter being one particular value, or might the parameter be something else?”
- They are similar enough that you can often use a confidence interval to reach the same types of conclusions you would reach with a hypothesis test using a two-sided alternative hypothesis.

Comparing Proportions from 2 Populations

- Compare 2 sample proportions each from a different group
    - Are the two groups the same? Are they different?
- Same as the ideas for the one proportion, but the formulas are different.
    - Confidence interval for difference in proportions
    - Hypothesis test for difference in proportions

Hypothesis Testing for Two Proportions

- Hypotheses: state the null and alternative

Hypothesis	Symbols	The Alternative in Words
Two-tailed	$H_0: p_1 = p_2$ $H_a: p_1 \neq p_2$	The proportions are different in the two populations.
One-tailed (left)	$H_0: p_1 = p_2$ $H_a: p_1 < p_2$	The proportion in population 1 is less than the proportion in population 2.
One-tailed (right)	$H_0: p_1 = p_2$ $H_a: p_1 > p_2$	The proportion in population 1 is greater than the proportion in population 2.

Note: If p1=p2, then p1-p2=0. This is always the null hypothesis since we start by assuming that the two population proportions are the same.

The Test Statistic

- We are interested in how p_1 and p_2 differ, so our test statistic is based on the difference between our sample proportions from the two populations.
- The two-proportion z-test statistic is:

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

$$\text{where } \hat{p} = \frac{\text{number of successes in sample 1} + \text{number of successes in sample 2}}{n_1 + n_2}$$

Example

The researcher from the Pew study interviewed two random samples. Both samples, one taken in 2002 and one taken in 2009, had 1500 people. In 2002, 645 people expressed support for stem cell research. In 2009, 870 expressed support. The researcher wants to know if the proportion of Americans who support stem cell research has changed from 2002 to 2009. Use significance level of 5%.

Test the hypothesis, is there a difference in proportions between the two populations?

Example

Clicker!

The researcher from the Pew study interviewed two random samples. Both samples, one taken in 2002 and one taken in 2009, had 1500 people. In 2002, 645 people expressed support for stem cell research. In 2009, 870 expressed support. The researcher wants to know if the proportion of Americans who support stem cell research has changed from 2002 to 2009.

State the null hypothesis in symbols.

Answer:

A. H_0: $p_1 = p_2$

B. H_0: $p_1 < p_2$

C. H_0: $p_1 > p_2$

D. H_0: $p_1 \neq p_2$

Example

Clicker!

The researcher from the Pew study interviewed two random samples. Both samples, one taken in 2002 and one taken in 2009, had 1500 people. In 2002, 645 people expressed support for stem cell research. In 2009, 870 expressed support. The researcher wants to know if the proportion of Americans who support stem cell research has changed from 2002 to 2009.

State the alternative hypothesis in symbols.

Answer:

A. H_a: $p_1 = p_2$

B. H_a: $p_1 < p_2$

C. H_a: $p_1 > p_2$

D. H_a: $p_1 \neq p_2$

Example

The researcher from the Pew study interviewed two random samples. Both samples, one taken in 2002 and one taken in 2009, had 1500 people. In 2002, 645 people expressed support for stem cell research. In 2009, 870 expressed support. The researcher wants to know if the proportion of Americans who support stem cell research has changed from 2002 to 2009. Use significance level of 5%.

Write the null and alternative hypotheses:

$$H_0: p_1 = p_2$$

$$H_a: p_1 \neq p_2$$

Example

Clicker!

The researcher from the Pew study interviewed two random samples. Both samples, one taken in 2002 and one taken in 2009, had 1500 people. In 2002, 645 people expressed support for stem cell research. In 2009, 870 expressed support. The researcher wants to know if the proportion of Americans who support stem cell research has changed from 2002 to 2009.

What's the sample statistic for the 2002 group?

Answer:

A. 870/1500=0.58

B. 645/1500=0.43

C. 645/870=0.74

D. 870/645=1.35

Example

Clicker!

The researcher from the Pew study interviewed two random samples. Both samples, one taken in 2002 and one taken in 2009, had 1500 people. In 2002, 645 people expressed support for stem cell research. In 2009, 870 expressed support. The researcher wants to know if the proportion of Americans who support stem cell research has changed from 2002 to 2009.

What's the sample statistic for the 2009 group?

Answer:

A. 870/1500=0.58

B. 645/1500=0.43

C. 645/870=0.74

D. 870/645=1.35

Example

The researcher from the Pew study interviewed two random samples. Both samples, one taken in 2002 and one taken in 2009, had 1500 people. In 2002, 645 people expressed support for stem cell research. In 2009, 870 expressed support. The researcher wants to know if the proportion of Americans who support stem cell research has changed from 2002 to 2009. Use significance level of 5%.

Establish samples:

Sample 1: 2009 data, n=1500, $\hat{p}_1$=870/1500=0.58

Sample 2: 2002 data, n=1500, $\hat{p}_2$=645/1500=0.43

Example

The researcher from the Pew study interviewed two random samples. Both samples, one taken in 2002 and one taken in 2009, had 1500 people. In 2002, 645 people expressed support for stem cell research. In 2009, 870 expressed support. The researcher wants to know if the proportion of Americans who support stem cell research has changed from 2002 to 2009.

2. Type of test: two-proportion z-test. Check conditions:

Random and Independent: We are told the data come from two random samples. The samples are independent of each other.

Large Sample: Condition is satisfied.

$$\hat{p} = \frac{870 + 645}{1500 + 1500} = 0.505$$

$$n_1\hat{p} \geq 10 \text{ and } n_1(1 - \hat{p}) \geq 10$$
$$n_2\hat{p} \geq 10 \text{ and } n_2(1 - \hat{p}) \geq 10$$

1500 x 0.505 = 757.5 > 10

1500 x (1-0.505) = 742.5 > 10

1500 x 0.505 = 757.5 > 10

1500 x (1-0.505) = 742.5 > 10

Big Population: The population size needs to be greater than 10 x 1500 = 15,000. Condition is satisfied.

Example

The researcher from the Pew study interviewed two random samples. Both samples, one taken in 2002 and one taken in 2009, had 1500 people. In 2002, 645 people expressed support for stem cell research. In 2009, 870 expressed support. The researcher wants to know if the proportion of Americans who support stem cell research has changed from 2002 to 2009.

3. Test statistic:

$$\hat{p} = \frac{870 + 645}{1500 + 1500} = 0.505$$

$$z = \frac{\text{estimator - null value}}{SE} = \frac{\hat{p}_1 - \hat{p}_2 - 0}{\sqrt{\hat{p}(1-\hat{p})\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}} = \frac{0.58 - 0.43 - 0}{\sqrt{0.505(1-0.505)\left(\frac{1}{1500} + \frac{1}{1500}\right)}} = 8.22$$

Example

The researcher from the Pew study interviewed two random samples. Both samples, one taken in 2002 and one taken in 2009, had 1500 people. In 2002, 645 people expressed support for stem cell research. In 2009, 870 expressed support. The researcher wants to know if the proportion of Americans who support stem cell research has changed from 2002 to 2009.

4. p-value:

$z = \text{Prob}(z < |z_{obs}|) = \text{Prob}(z < -z_{obs}) + \text{Prob}(z > z_{obs})$

The p-value is $\text{Prob}(z < -8.22) + \text{Prob}(z > 8.22) \approx 0.00$

5. Conclusion:

The p-value = $0.00 < 0.05$, so we reject the null hypothesis. We conclude that the proportions of Americans who support stem cell research in 2002 and 2009 are not equally likely, i.e. the proportion has changed from 2002 to 2009.

Example

Shankaran and colleagues (2012) reported the results of a randomized trial of whole-body hypothermia (cooling) for brain problems in babies due to lack of oxygen. 27 of the 97 infants randomly assigned to hypothermia died and 41 of the 93 infants in the control group died.

Test the hypothesis that the death rate was less for those treated with hypothermia at a 0.05 significance level.

Example

Clicker!

Shankaran and colleagues (2012) reported the results of a randomized trial of whole-body hypothermia (cooling) for brain problems in babies due to lack of oxygen. 27 of the 97 infants randomly assigned to hypothermia died and 41 of the 93 infants in the control group died.

Test the hypothesis that the death rate was less for those treated with hypothermia at a 0.05 significance level.

State the null hypothesis in symbols.

Answer:

A. H_0: $p_1 = p_2$

B. H_0: $p_1 < p_2$

C. H_0: $p_1 > p_2$

D. H_0: $p_1 \neq p_2$

Example

Clicker!

Shankaran and colleagues (2012) reported the results of a randomized trial of whole-body hypothermia (cooling) for brain problems in babies due to lack of oxygen. 27 of the 97 infants randomly assigned to hypothermia died and 41 of the 93 infants in the control group died.

Test the hypothesis that the death rate was less for those treated with hypothermia at a 0.05 significance level.

State the alternative hypothesis in symbols.

Answer:

A. H_a: $p_1 = p_2$

B. H_a: $p_1 < p_2$

C. H_a: $p_1 > p_2$

D. H_a: $p_1 \neq p_2$

Example

Shankaran and colleagues (2012) reported the results of a randomized trial of whole-body hypothermia (cooling) for brain problems in babies due to lack of oxygen. 27 of the 97 infants randomly assigned to hypothermia died and 41 of the 93 infants in the control group died.

Write the null and alternative hypotheses:

$$H_0: p_1 = p_2$$

$$H_a: p_1 < p_2$$

Example

Clicker!

Shankaran and colleagues (2012) reported the results of a randomized trial of whole-body hypothermia (cooling) for brain problems in babies due to lack of oxygen. 27 of the 97 infants randomly assigned to hypothermia died and 41 of the 93 infants in the control group died.

What's the sample statistic for the Hypothermia group?

Answer:

A. 27/93=0.290

B. 27/97=0.278

C. 41/93=0.440

D. 41/97=0.423

Example

Clicker!

Shankaran and colleagues (2012) reported the results of a randomized trial of whole-body hypothermia (cooling) for brain problems in babies due to lack of oxygen. 27 of the 97 infants randomly assigned to hypothermia died and 41 of the 93 infants in the control group died.

What's the sample statistic for the Control group?

Answer:

A. 27/93=0.290

B. 27/97=0.278

C. 41/93=0.440

D. 41/97=0.423

Example

Shankaran and colleagues (2012) reported the results of a randomized trial of whole-body hypothermia (cooling) for brain problems in babies due to lack of oxygen. 27 of the 97 infants randomly assigned to hypothermia died and 41 of the 93 infants in the control group died.

Establish samples:

Sample 1: Hypothermia group, n=97, $\hat{p}_1$=27/97=0.278

Sample 2: Control group, n=93, $\hat{p}_2$=41/93=0.44

Example

Shankaran and colleagues (2012) reported the results of a randomized trial of whole-body hypothermia (cooling) for brain problems in babies due to lack of oxygen. 27 of the 97 infants randomly assigned to hypothermia died and 41 of the 93 infants in the control group died.

2. Type of test: two-proportion z-test. Check conditions:

Random and Independent: We are told the infants are randomly assigned. Even though these are not random samples since all other conditions hold then it's acceptable. The infants are randomly assigned hence they are independent of each other.

Large Sample: Condition is satisfied.

$$\hat{p} = \frac{27+41}{97+93} = 0.36$$

97 x 0.36 = 34.92 > 10

97 x (1-0.36) = 62.08 > 10

93 x 0.36 = 33.48 > 10

93 x (1-0.36) = 59.52 > 10

$$n_1\hat{p} \geq 10 \text{ and } n_1(1-\hat{p}) \geq 10$$
$$n_2\hat{p} \geq 10 \text{ and } n_2(1-\hat{p}) \geq 10$$

Big Population: The population size needs to be greater than 10 x 97 = 970 and 10 x 93 = 930. Condition is satisfied.

Example

Shankaran and colleagues (2012) reported the results of a randomized trial of whole-body hypothermia (cooling) for brain problems in babies due to lack of oxygen. 27 of the 97 infants randomly assigned to hypothermia died and 41 of the 93 infants in the control group died.

3. Test statistic:

$$\hat{p} = \frac{27+41}{97+93} = \frac{68}{190} = 0.36$$

$$z = \frac{\text{estimator - null value}}{SE} = \frac{\hat{p}_1 - \hat{p}_2 - 0}{\sqrt{\hat{p}(1-\hat{p})\left(\frac{1}{n_1}+\frac{1}{n_2}\right)}} = \frac{0.278 - 0.44 - 0}{\sqrt{0.36(1-0.36)\left(\frac{1}{97}+\frac{1}{93}\right)}} = -2.34$$

Example

Shankaran and colleagues (2012) reported the results of a randomized trial of whole-body hypothermia (cooling) for brain problems in babies due to lack of oxygen. 27 of the 97 infants randomly assigned to hypothermia died and 41 of the 93 infants in the control group died.

4. p-value: $z=\text{Prob}(z<-z_{obs})=\text{Prob}(z<-2.34)=0.0096$

5. Conclusion:

The p-value = 0.0096 < 0.05, so we reject the null hypothesis. We conclude that the death rate is less for babies treated with hypothermia .