

7.2: One-Sample Proportion Test

There are many different parameters that you can test. There is a test for the mean, such as was introduced with the z-test. There is also a test for the population proportion, p . This is where you might be curious if the proportion of students who smoke at your school is lower than the proportion in your area. Or you could question if the proportion of accidents caused by teenage drivers who do not have a drivers' education class is more than the national proportion.

To test a population proportion, there are a few things that need to be defined first. Usually, Greek letters are used for parameters and Latin letters for statistics. When talking about proportions, it makes sense to use p for proportion. The Greek letter for p is π , but that is too confusing to use. Instead, it is best to use p for the population proportion. That means that a different symbol is needed for the sample proportion. The convention is to use, \hat{p} , known as p-hat. This way you know that p is the population proportion, and that \hat{p} is the sample proportion related to it.

Now proportion tests are about looking for the percentage of individuals who have a particular attribute. You are really looking for the number of successes that happen. Thus, a proportion test involves a binomial distribution.

Hypothesis Test for One Population Proportion (1-Prop Test)

1. State the random variable and the parameter in words.

x = number of successes

I = proportion of successes

2. State the null and alternative hypotheses and the level of significance

$H_o : p = p_o$, where p_o is the known proportion

$H_A : p < p_o$

$H_A : p > p_o$, use the appropriate one for your problem

$H_A : p \neq p_o$

Also, state your α level here.

3. State and check the assumptions for a hypothesis test

a. A simple random sample of size n is taken.

b. The conditions for the binomial distribution are satisfied

c. To determine the sampling distribution of \hat{p} , you need to show that $np \geq 5$ and $nq \geq 5$, where $q = 1 - p$. If this requirement is true, then the sampling distribution of \hat{p} is well approximated by a normal curve.

4. Find the sample statistic, test statistic, and p-value

Sample Proportion:

$$\hat{p} = \frac{x}{n} = \frac{\# \text{ of successes}}{\# \text{ of trials}}$$

Test Statistic:

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

p-value:

TI-83/84: Use normalcdf(lower limit, upper limit, 0, 1)

Note

if $H_A : p < p_o$, then lower limit is $-1E99$ and upper limit is your test statistic. If $H_A : p > p_o$, then lower limit is your test statistic and the upper limit is $1E99$. If $H_A : p \neq p_o$, then find the p-value for $H_A : p < p_o$, and multiply by 2.

R: Use `pnorm(z, 0, 1)`

Note

If $H_A : p < p_o$, then you can use `pnorm`. If $H_A : p > p_o$, then you have to find `pnorm` and then subtract from 1. If $H_A : p \neq p_o$, then find the p-value for $H_A : p < p_o$, and multiply by 2.

5. Conclusion

This is where you write reject H_o or fail to reject H_o . The rule is: if the p-value $< \alpha$, then reject H_o . If the p-value $\geq \alpha$, then fail to reject H_o .

6. Interpretation

This is where you interpret in real world terms the conclusion to the test. The conclusion for a hypothesis test is that you either have enough evidence to show H_A is true, or you do not have enough evidence to show H_A is true.

Example 7.2.1 hypothesis test for one proportion using formula

A concern was raised in Australia that the percentage of deaths of Aboriginal prisoners was higher than the percent of deaths of non-Aboriginal prisoners, which is 0.27%. A sample of six years (1990-1995) of data was collected, and it was found that out of 14,495 Aboriginal prisoners, 51 died ("Indigenous deaths in," 1996). Do the data provide enough evidence to show that the proportion of deaths of Aboriginal prisoners is more than 0.27%?

1. State the random variable and the parameter in words.
2. State the null and alternative hypotheses and the level of significance.
3. State and check the assumptions for a hypothesis test.
4. Find the sample statistic, test statistic, and p-value.
5. Conclusion
6. Interpretation

Solution

1. x = number of Aboriginal prisoners who die

p = proportion of Aboriginal prisoners who die

2. $H_o : p = 0.0027$
 $H_A : p > 0.0027$

Example 7.2.5b argued that the $\alpha = 0.05$.

3.

- a. A simple random sample of 14,495 Aboriginal prisoners was taken. However, the sample was not a random sample, since it was data from six years. It is the numbers for all prisoners in these six years, but the six years were not picked at random. Unless there was something special about the six years that were chosen, the sample is probably a representative sample. This assumption is probably met.
- b. There are 14,495 prisoners in this case. The prisoners are all Aboriginals, so you are not mixing Aboriginal with non-Aboriginal prisoners. There are only two outcomes, either the prisoner dies or doesn't. The chance that one prisoner dies over another may not be constant, but if you consider all prisoners the same, then it may be close to the same probability. Thus the conditions for the binomial distribution are satisfied
- c. In this case $p = 0.0027$ and $n = 14,495$. $np = 14495 * 0.0027 \approx 39 \geq 5$ and $nq = 14495 * (1 - 0.0027) \approx 14456 \geq 5$. So, the sampling distribution for \hat{p} is a normal distribution.

4. Sample Proportion:

$$x = 51$$

$$n = 14495$$

$$\hat{p} = \frac{x}{n} = \frac{51}{14495} \approx 0.003518$$

Test Statistic:

$$z = \frac{\hat{p} - p}{\sqrt{\frac{pq}{n}}} = \frac{0.003518 - 0.0027}{\sqrt{\frac{0.0027(1 - 0.0027)}{14495}}} \approx 1.8979$$

p-value:

$$TI-83/84: \text{p-value} = P(z > 1.8979) = \text{normalcdf}(1.8979, 1E99, 0, 1) \approx 0.029$$

R: p-value = $P(z > 1.8979) = 1 - \text{pnorm}(1.8979, 0, 1) \approx 0.029$

5. Since the p-value < 0.05 , then reject H_0 .

6. There is enough evidence to show that the proportion of deaths of Aboriginal prisoners is more than for non-Aboriginal prisoners.

Example 7.2.2 hypothesis test for one proportion using technology

A researcher who is studying the effects of income levels on breastfeeding of infants hypothesizes that countries where the income level is lower have a higher rate of infant breastfeeding than higher income countries. It is known that in Germany, considered a high-income country by the World Bank, 22% of all babies are breastfed. In Tajikistan, considered a low-income country by the World Bank, researchers found that in a random sample of 500 new mothers that 125 were breastfeeding their infant. At the 5% level of significance, does this show that low-income countries have a higher incident of breastfeeding?

1. State your random variable and the parameter in words.
2. State the null and alternative hypotheses and the level of significance.
3. State and check the assumptions for a hypothesis test.
4. Find the sample statistic, test statistic, and p-value.
5. Conclusion
6. Interpretation

Solution

1. x = number of woman who breastfeed in a low-income country

p = proportion of woman who breastfeed in a low-income country

$$H_0 : p = 0.22$$

2. $H_A : p > 0.22$

$$\alpha = 0.05$$

3.

- a. A simple random sample of 500 breastfeeding habits of woman in a low-income country was taken as was stated in the problem.
- b. There were 500 women in the study. The women are considered identical, though they probably have some differences. There are only two outcomes, either the woman breastfeeds or she doesn't. The probability of a woman breastfeeding is probably not the same for each woman, but it is probably not very different for each woman. The conditions for the binomial distribution are satisfied
- c. In this case, $n = 500$ and $p = 0.22$. $np = 500(0.22) = 110 \geq 5$ and $nq = 500(1 - 0.22) = 390 \geq 5$, so the sampling distribution of \hat{p} is well approximated by a normal curve.

4. This time, all calculations will be done with technology. On the TI-83/84 calculator. Go into the STAT menu, then arrow over to TESTS. This test is a 1-propZTest. Then type in the information just as shown in Figure 7.2.1.

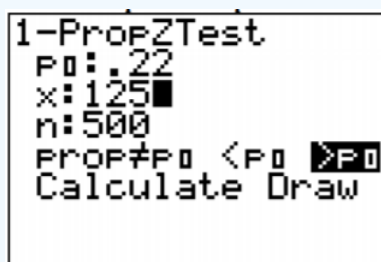


Figure 7.2.1: Setup for 1-Proportion Test

Once you press Calculate, you will see the results as in Figure 7.2.2.

```
1-PropZTest
PROP>.22
z=1.619375689
P=.052683219
p=.25
n=500
```

Figure 7.2.2: Results for 1-Proportion Test

The z in the results is the test statistic. The $p = 0.052683219$ is the p-value, and the $\hat{p} = 0.25$ is the sample proportion.

The p-value is approximately 0.053.

On R, the command is `prop.test(x, n, po, alternative = "less" or "greater")`, where p_0 is what H_0 says p equals, and you use less if your H_A is less and greater if your H_A is greater. If your H_A is not equal to, then leave off the alternative statement. So for this example, the command would be `prop.test(125, 500, .22, alternative = "greater")`

1-sample proportions test with continuity correction

data: 125 out of 500, null probability 0.22

X-squared = 2.4505, df = 1, p-value = 0.05874

alternative hypothesis: true p is greater than 0.22

95 percent confidence interval:

0.218598 1.000000

sample estimates:

p

0.25

Note

R does a continuity correction that the formula and the TI-83/84 calculator do not do. You can put in a command that says not to use the continuity correction, but it is correct to use it. Also, R doesn't give the z test statistic, so you don't need to worry about this. It does give a p-value that is slightly off from the formula and the calculator due to the continuity correction.

p-value = 0.05874

5. Since the p-value is more than 0.05, you fail to reject H_0 .

6. There is not enough evidence to show that the proportion of women who breastfeed in low-income countries is more than in high-income countries.

Notice, the conclusion is that there wasn't enough evidence to show what H_1 said. The conclusion was not that you proved H_0 true. There are many reasons why you can't say that H_0 is true. It could be that the countries you chose were not very representative of what truly happens. If you instead looked at all high-income countries and compared them to low-income countries, you might have different results. It could also be that the sample you collected in the low-income country was not representative. It could also be that income level is not an indication of breastfeeding habits. There could be other factors involved. This is why you can't say that you have proven H_0 is true. There are too many other factors that could be the reason that you failed to reject H_0 .

Homework

Exercise 7.2.1

In each problem show all steps of the hypothesis test. If some of the assumptions are not met, note that the results of the test may not be correct and then continue the process of the hypothesis test.

1. Eyeglassomatic manufactures eyeglasses for different retailers. They test to see how many defective lenses they made in a given time period and found that 11% of all lenses had defects of some type. Looking at the type of defects, they found in a three-month time period that out of 34,641 defective lenses, 5865 were due to scratches. Are there more defects from scratches than from all other causes? Use a 1% level of significance.
2. In July of 1997, Australians were asked if they thought unemployment would increase, and 47% thought that it would increase. In November of 1997, they were asked again. At that time 284 out of 631 said that they thought unemployment would increase ("Morgan gallup poll," 2013). At the 5% level, is there enough evidence to show that the proportion of Australians in November 1997 who believe unemployment would increase is less than the proportion who felt it would increase in July 1997?
3. According to the February 2008 Federal Trade Commission report on consumer fraud and identity theft, 23% of all complaints in 2007 were for identity theft. In that year, Arkansas had 1,601 complaints of identity theft out of 3,482 consumer complaints ("Consumer fraud and," 2008). Does this data provide enough evidence to show that Arkansas had a higher proportion of identity theft than 23%? Test at the 5% level.
4. According to the February 2008 Federal Trade Commission report on consumer fraud and identity theft, 23% of all complaints in 2007 were for identity theft. In that year, Alaska had 321 complaints of identity theft out of 1,432 consumer complaints ("Consumer fraud and," 2008). Does this data provide enough evidence to show that Alaska had a lower proportion of identity theft than 23%? Test at the 5% level.
5. In 2001, the Gallup poll found that 81% of American adults believed that there was a conspiracy in the death of President Kennedy. In 2013, the Gallup poll asked 1,039 American adults if they believe there was a conspiracy in the assassination, and found that 634 believe there was a conspiracy ("Gallup news service," 2013). Do the data show that the proportion of Americans who believe in this conspiracy has decreased? Test at the 1% level.
6. In 2008, there were 507 children in Arizona out of 32,601 who were diagnosed with Autism Spectrum Disorder (ASD) ("Autism and developmental," 2008). Nationally 1 in 88 children are diagnosed with ASD ("CDC features -," 2013). Is there sufficient data to show that the incident of ASD is more in Arizona than nationally? Test at the 1% level.

Answer

For all hypothesis tests, just the conclusion is given. See solutions for the entire answer.

1. Reject H_0 .
3. Reject H_0 .
5. Reject H_0 .

This page titled [7.2: One-Sample Proportion Test](#) is shared under a [CC BY-SA 4.0](#) license and was authored, remixed, and/or curated by [Kathryn Kozak](#) via [source content](#) that was edited to the style and standards of the LibreTexts platform.