Statistics Package Exercise: Carrying Out the z-test for the Population Proportion (p)

Learning Objective: In a given context, specify the null and alternative hypotheses for the population proportion and mean.

Learning Objective: Carry out hypothesis testing for the population proportion and mean (when appropriate), and draw conclusions in context.

The objectives of this activity are:

1. To give you guided practice in carrying out the z-test for the population proportion (p).

2. To learn how to use statistical software to help you carry out the test.

Background: This activity is based on the results of a recent study on the safety of airplane drinking water that was conducted by the U.S. Environmental Protection Agency (EPA). A study found that out of a random sample of 316 airplanes tested, 40 had coliform bacteria in the drinking water drawn from restrooms and kitchens. As a benchmark comparison, in 2003 the EPA found that about 3.5% of the U.S. population have coliform bacteria-infected drinking water. The question of interest is whether, based on the results of this study, we can conclude that drinking water on airplanes is more contaminated than drinking water in general.

Learn By Doing (1/1 point)

Let p be the proportion of contaminated drinking water in airplanes. Write down the appropriate null and alternative hypotheses.

Your Answer:

Ho: p = 0.035
Ha: p > 0.035

Our Answer:

Ho: p = 0.035 Ha: p > 0.035 As usual, Ho claims that "nothing special is going on" with the drinking water in airplanes—the contamination rate is the same as the contamination rate in drinking water in general. Ha represents what we suspect, or what we want to check. In this case, we want to check whether drinking water on airplanes is more contaminated than drinking water in general.

Resubmit Reset

Learn By Doing (1/1 point)

Based on the collected data, is it safe to use the z-test for p in this scenario? Explain.

Your Answer:

Why a repeat?

Anyway, yes it is, because n*p = 316*0.035 = 11.06 which is greater than 10. n*(1-p) = 305, which is also greater than 10. It was also randomly sampled.

Our Answer:

Let's check the conditions. • The sample of airplanes is random. • n * po = 316 * 0.035 = 11.06 > 10. • n * (1 - po) = 316 * .965 = 304.94 > 10. Yes, it is safe to use the test.

Resubmit Reset

Use the following instructions to conduct the z-test for the population proportion:

- R
- StatCrunch
- TI Calculator
- Minitab
- Excel

R Instructions

In R, the default command for inference for proportions,

```
prop.test()
```

does not use the traditional z-test, but instead uses a related test called a chi-square (χ^2) test. Therefore, to conduct the z-test for the population proportion (

```
p
```

) using R, we must modify the output to acquire the z-test statistic.

From the background, we know that there are

```
n=316
```

total airplanes,

```
x=40
```

contaminate samples, and the baseline comparison (null value) is

```
p=0.035
```

. Here are the basic commands:

- ```
  p =
  prop.test(x=40,n=316,p=0.035,alternative="greater",conf.
  level=0.95, correct=FALSE);p
  ```

The parameter

```
alternative=
```

may take on the values

```
"greater"
```

,

```
"less"
```

, or
```

```
"two.sided"
```

depending on the alternative hypothesis.

Notice that

```
X-squared
```

can be pulled from the output by entering the command

```
p$statistic
```

, referred to as the chi-square (χ^2) test statistic. The χ^2 test statistic is equivalent to z^2, so we can determine the z-test statistic by calculating $z = \sqrt{X^2}$).

- ```
  z = sqrt(p$statistic);z
  ```

The provided p-value is equivalent to the p-value we might find from the z-test we hand calculate for proportions.

Note: The χ^2-test statistic will always be positive so its square root will be positive in calculation, but that does not mean that the z-test statistic is positive. If the sample proportion is greater than the null proportion then the z-test statistic is positive. If the sample proportion is less than the null proportion then the z-test statistic is negative.

Learn By Doing (1/1 point)

Now that we have established that it is safe to use the z-test for p for our problem, go ahead and carry out the test. Paste the output below.

Your Answer:

z = 8.858441 via the code provided above. Since the smaple proportion 40/316 (0.1266) is greater than the null proportion (0.035), this value is positive.

We also found that p-value < 2.2e-16. That means we vehemently reject Ho and accept Ha.

Our Answer:
  ```

RStatCrunch TI CalculatorMinitabExcel R Here is the R output: A lot of information is returned; let's review it item by item. The data we entered is echoed back: we have a random sample of 316 airplanes, out of which 40 were found to have contaminated water. The null hypothesis was that the proportion of planes with contaminated water is 0.035. The chi-squared statistic for the test was 78.472, with one degree of freedom. The most important result is next: the p-value of the test was 2.2e - 16, which is essentially 0. The alternate hypothesis for our test was that the proportion of planes with contaminated water was greater than 0.035. StatCrunch In our case, the data are summarized. We have a random sample of 316 airplanes, out of which 40 were found to have contaminated water. When you enter these two values into StatCrunch, set the null value to 0.035, and choose the "greater than" alternative, you get: TI Calculator Here is the TI output: A lot of information is returned; let's review it item by item. The alternate hypothesis for our test was that the proportion of planes with contaminated water was greater than 0.035. The z-score calculated was 8.86. The most important result is next: the p-value of the test was 4.1e - 19, which is essentially 0. We then see the sample proportion (p-hat) of .126, which was calculated by p-hat = 40 / 316 (the number of airplanes in our sample with contaminated water divided by the total number of airplanes). Finally, our n-value of 316 is echoed back. Minitab In our case, the data are summarized. We have a random sample of 316 airplanes, out of which 40 were found to have contaminated water. When you enter these two values into Minitab, set the null value to .035, and choose the "greater than" alternative, you get: Excel We have a random sample of 316 airplanes, out of which 40 were found to have contaminated water. When we calculate p-hat, we get a value of 0.1266. When we use p-hat to calculate the test statistic (z), we get a value of 8.858. Using the NORMSDIST function gives a value of 1, which is the probability that p is less than or equal to the test statistic z. Since we want the probability that p is greater than the test statistic, we subtract this value from 1, giving us a p-value of 0.

Resubmit Reset

Learn By Doing (1/1 point)

Note that, according to the output, the test statistic for this test is 8.86. Make sure you understand how this was calculated, and give an interpretation of its value.

Your Answer:

The sample proportion we observed was 8.86 standard deviations above the null proportion. (edited after seeing the answer) Assuming it's true that the population of drinking water is 0.035, our finding was 8.9 sd above that. Since 99.7% of values fall within 3sd of the mean, this is very very unlikely.

Our Answer:

In our case: • n = 316 • $\hat{p}$ = 40/316 ≈ .127 • po = .035 And therefore, z = 8.9 means that assuming that the Ho is true (i.e., that the proportion of contaminated drinking water on airplanes is indeed .035, the same as drinking water in general), the results of our study provided a sample proportion that is 8.9 standard deviations above that proportion. Recall that the standard deviation rule for normal distributions tells us that 99.7% of normal values fall within 3 standard deviations of the mean. A sample proportion that falls 8.9 standard deviations above the true proportion is, therefore, extremely unlikely. As you'll see, this fact will also be expressed by the p-value.

Resubmit Reset

Learn By Doing (1/1 point)

We calculated a p-value of 0 in this test. Interpret what that means, and draw your conclusions.

Your Answer:

Whatever acceptance level we use, there's should be practically a 0% chance to observe 8.9 standard deviations above the null proportion, if Ho was true.

Our Answer:

A p-value that is so close to 0 tells us that it would be almost impossible to get a sample proportion of 12.5% (or larger) of contaminated drinking water had the true proportion been 3.5%. In other words, the airline industry cannot claim that this just happened to be a "bad" sample that occurred by chance. A p-value that is essentially 0 tells us that it is highly unlikely that such a sample happened just by chance. Our conclusion is therefore that we have an extremely strong reason to reject Ho and conclude that the proportion of contaminated drinking water on airplanes is higher than the proportion in general. On a technical level, the p-value is smaller than any significance level that we are going to set, so Ho can be rejected. Comment: In the original study, there were 158 randomly chosen airplanes, and in 20 of them the drinking water was found to be contaminated. As we mentioned, we based the context of this activity on these results, and we simply doubled the counts. Instead of 158 airplanes, we had 316; instead of 20 airplanes with contaminated drinking water, we had 40). We did that because technically, otherwise the conditions under which this test can be used would not have been met. Practically, the results of this study are so extreme that the fact that not all the conditions were met has no effect on the actual conclusion.

Resubmit Reset