

NUMB3RS Activity: Different or Not? Episode: "Waste Not"

Topic: Two-proportion Z-test

Grade Level: 11 - 12

Objective: Use the two-proportion Z-test to determine if there is a difference between two proportions.

Materials: TI-83 Plus/TI-84 Plus graphing calculator

Time: 15 - 25 minutes

Introduction

In "Waste Not," the FBI discovers that children in the area of a sinkhole seem to have an unusually high occurrence of cancer. FBI agent Megan Reeves believes that this might represent a "cancer cluster." Charlie warns her not to jump to this conclusion too quickly.

The FBI may have used a statistical technique called a *two-proportion test* to determine whether the number of cancer cases is unusually high. This activity contains an introduction to this test.

Discuss with Students

This activity is designed to introduce the concept of hypothesis testing. Students are given information (based on data from the Centers for Disease Control) on cancer rates in a number of states. They use a two-proportion test to determine whether the rates are appreciably different.

Remind students that this activity gives only an introduction to the concept of hypothesis tests, a major topic in statistical inference. Hypothesis tests work in a manner similar to a jury at a trial. In a jury trial, a defendant is presumed innocent. Only if the evidence convinces the jury beyond any reasonable doubt, does the jury decide that the person is guilty. Statisticians use hypothesis tests to make inferences about a population based on random samples.

In courses like AP Statistics, students learn that three assumptions need to be checked before using a two-proportion test. The samples need to be chosen randomly, the samples need to be independent, and a sample's size must be less than 10% of the population.

Student Page Answers:

1. *Maine: 0.582%; New Hampshire: 0.502%. Opinions will vary about whether the difference of the rates is considerable.* **2.** *p-value = 0.289 If there were no difference between the rates, we would get results like this by chance about 28.9% of the time. There is not enough information to conclude that one rate is bigger than the other. (Note: A student who gets a p-value of 0.711 made the wrong choice for the inequality.)* **3.** *No, just because the two sample proportions are the same, this does not mean that the actual rates of incidence of cancer in the two states are the same. With either inequality the p-value = 0.5. If there were no difference between the rates, we would get results like this about half of the time, just by chance.* **4.** *The p-value for the \neq p2 test is twice the smaller of the two p-values from the other tests.*

Name: _____

Date: _____

NUMB3RS Activity: Different or Not?

In "Waste Not," the FBI discovers that the children in the area of a sinkhole seem to have an unusually high occurrence of cancer. FBI agent Megan Reeves believes that this might represent a "cancer cluster." Charlie warns her not to jump to this conclusion too quickly. The FBI may have used a statistical technique called a *two-proportion test* to decide whether the rate of cancer seems to be unusually high in this location. A two-proportion test is a statistical technique that can be used to decide if the rate of cancer seems to be different in different areas of the country. This activity contains an introduction to this test.

Consider the following data on cases of cancer, based on a hypothetical random sample of 50,000 people from each of six states in New England. A large sample is used here because the rate of incidence of cancer is quite small.

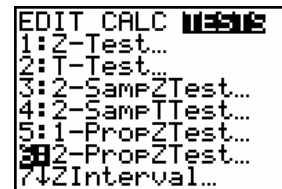
State	Cancer Cases
Connecticut	268
Maine	291
Massachusetts	268
New Hampshire	251
Rhode Island	281
Vermont	249

Based on these data, does there seem to be an association between location and incidence of cancer? One way of attempting to answer this question is to perform a statistical procedure called a two-proportion test. This is one of many procedures for testing hypotheses that are studied in statistics classes.

1. Maine and New Hampshire are neighboring states. What percent of the Maine sample had cancer? What percent of the New Hampshire sample had cancer? Do you think that the difference between these two percentages is considerable?

The data shows that the proportion of cancer cases in the sample from Maine is higher than the proportion of cancer cases in the sample from New Hampshire. If the samples involved are randomly chosen, a two-proportion test can determine if it is likely that the proportion of cases for the entire state's population is higher in Maine than New Hampshire. A hypothesis test assumes that there is no difference and inspects the data to see if they in fact do not support this assumption. In this case, the hypothesis is that there is no difference in the cancer rates for the states. The alternative hypothesis is that the cancer rate is higher in Maine than in New Hampshire. A two-proportion test can indicate that the initial hypothesis should be questioned.

To perform this test on your calculator, press **STAT** and go to the **TESTS** menu. Then select **6:2-PropZTest....**



Input the data as follows:

x1 represents the number of cases of cancer in the sample from Maine. **x2** represents the number from New Hampshire. **n1** and **n2**, represent the number of people in each sample (50,000). Because we are interested in whether the proportion of cases of cancer is higher in Maine than in New Hampshire, choose **>p2**.

```
2-PropZTest
x1:291
n1:50000
x2:251
n2:50000
P1:#P2 <P2  
Calculate Draw
```

Then select **Calculate** to produce the results shown at the right.

```
2-PropZTest
P1>P2
z=1.72282231
P=.0424602946
P1=.00582
P2=.00502
↓P=.00542
```

If the samples were truly random, the set of all possible differences between the sample proportions would form a normal distribution. The informal description of a normal distribution is a "bell shaped curve". The results of the test show whether the difference observed in the data is typical of most of the differences or if the difference is in fact considerable.

The screen shows a p -value of about 0.042. This value means that if there were no actual difference between the cancer rates of the two states, the difference observed in these samples would happen by chance about 4.2% of the time. Many statisticians use 5% (0.05) as a maximum value to decide if an observation is rare, so they would conclude in this example that it is unlikely to get such results as these by chance. Statisticians use sampling to make predictions about the whole population. In this example, it is likely that the actual cancer rate in Maine is higher than that in New Hampshire, based on the sample data. For a p -value higher than 5%, statisticians would state that there is not enough information to conclude that one rate was actually higher than the other. They would not conclude, however, that there is no difference between the rates.

2. Connecticut and Rhode Island are also neighboring states. Compare their cancer data using a two-proportion test. Choose **<p2** or **>p2**, as appropriate. What is the p -value for this test? Explain the meaning of this number in the context of this example. Do you have enough information to conclude that one rate is larger than the other?
3. The sample data in this activity indicate that 268 out of 50,000 people in both the Connecticut and Massachusetts samples were diagnosed with cancer. Does this mean that the *actual* rates of incidence of cancer in the total populations of the two states were the same? Compare their cancer data using a two-proportion test. Choose either **<p2** or **>p2**. What is the p -value for this test? Explain the meaning of this number in the context of this example.
4. What is the relationship between the p -value obtained for the **< p2** or **> p2** test and the value for the **≠ p2** test?

The goal of this activity is to give your students a short and simple snapshot into a very extensive mathematical topic. TI and NCTM encourage you and your students to learn more about this topic using the extensions provided below and through your own independent research.

Extensions

For the Student

1. Compare the incidence of cancer in your state with that of neighboring states using data from the Centers for Disease Control. The Web site for United States Cancer Statistics includes information based on gender and ethnicity as well as different types of cancer. See: <http://apps.nccd.cdc.gov/uscs>
2. Complete the activity of identifying the source of an outbreak of Legionnaire's disease in Louisiana on the Centers for Disease Control Web site at: <http://www.cdc.gov/excite/classroom/legionnairesQ.htm>
3. Conduct a survey in your school to learn if there is a significant difference between the views of boys and girls about a particular issue and analyze the data using the techniques from this activity.

Additional Resources

The Robert Wood Johnson Foundation's Young Epidemiology Scholars project has an annual scholarship competition. The Web site features a considerable number of curriculum units on epidemiology. One activity, "Disease Outbreak Investigation" focuses on a leukemia cluster in Massachusetts that was the inspiration for the film "A Civil Action" starring John Travolta. See: <http://www.collegeboard.com/yes>

For more information about cancer rates and risk factors by state, see: http://www.cancer.org/docroot/STT/STT_2.asp

Sometimes people suspect that cancer is being caused by something in the environment because there is a cluster of cases in one specific geographic area. For more about cancer clusters, see: <http://www.cdc.gov/nceh/clusters/about.htm>

Burrill, Gail, Christine A. Frankin, Landy Godbold, and Linda J. Young; *Navigating through Data Analysis in Grades 9–12*, [NCTM, 2003.] The activities in this book introduce students to simple random sampling, sampling techniques, and simulation as a tool for analyzing both categorical and numerical data.