

Cancer Clusters – ID: 9996

By Pat Flynn

Time required

45 minutes

Activity Overview

Students will investigate cancer incidence rates in a number of states. Hypothesis testing is introduced and used along with a two-proportion z-test to compare cancer rates. This activity helps students to determine when a difference in data is actually statistically significant. This should encourage students to think carefully about claims made (for instance the presence of cancer clusters) and give them the tools to scrutinize these claims using statistical methods.

Concepts

- Hypothesis testing
- Two-proportion z-test

Teacher Preparation

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 z-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 in a criminal 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.
- The screenshots on pages 2–4 (top) demonstrate expected student results. Refer to the screenshots on pages 4 (bottom) and 5 for a preview of the student TI-Nspire document (.tns file).
- **To download the student .tns file and student worksheet, go to education.ti.com/exchange and enter “9996” in the quick search box.**

Classroom Management

- This activity is intended to be **teacher-led**. You may use the following pages to present the material to the class and encourage discussion. Students will follow along using their handhelds, although the majority of the ideas and concepts are only presented in this document; be sure to cover all the material necessary for students’ total comprehension.
- The student worksheet *StatAct06_CancerClusters_worksheet_EN* serves as a place for students to record their answers. Alternatively, you may wish to have the class record their answers on separate sheets of paper, or just use the questions posed to engage a class discussion.

TI-Nspire™ Applications
Calculator, Notes

Cancer Clusters

This activity gives students a basic overview of hypothesis testing. The questions in the student worksheet should be augmented with a more thorough look at hypothesis testing and z-tests.

Students should start by looking at the data and determining which states have the highest percentage of cancer cases. Students may question how the population size of each state will affect the outcomes. Assure the student that for the purposes of this activity the variation can be ignored and assumed to be addressed by the random sample.

On page 1.4, students should calculate the following percentages:

Maine – 0.582% New Hampshire – 0.502%

Opinions will vary about whether the difference of the rates is considerable.

Discuss with students the general concept of hypothesis testing. Make sure students understand that the null hypothesis and alternate hypothesis must cover all possibilities, so that if the null hypothesis is proved false the alternate hypothesis must be true. Note that if you are trying to support your claim, you will form your claim as the alternate hypothesis.

Discuss with students how the notation on page 1.7 is used to formally describe the hypothesis test described on page 1.6.

Students should have some familiarity with statistical tests and the two-proportion z-test can be linked back to their study of the z- and t-tests.

To perform this test on the handheld, select **MENU > Statistics > Stat Tests... > 2-Prop z Test**.

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 **p1 > p2**.

Students are told informally that *p*-values lower than 0.05 should lead us to reject the null hypothesis. This is a reference to the **critical region** which can now be defined more formally. Although not covered in this activity, discussions of **significance level** (the probability that based on the test statistic the null hypothesis will be rejected when it is in fact true) and **critical value** (the value that defines the edge of the critical region, in this example a *p*-value of 0.05) are appropriate at this time.

1.4 1.5 1.6 1.7 RAD AUTO REAL

p_1 = rate of cancer in Maine
 p_2 = rate of cancer in New Hampshire
 $H_0: p_1 \leq p_2$
 $H_1: p_1 > p_2$

1.4 2-Prop z Test

Successes, x1: 291
n1: 50000
Successes, x2: 251
n2: 50000
Alternate Hyp: Ha: p1 > p2

OK Cancel

1/1

An important part of hypothesis testing is in drawing a conclusion from the final data. When the test statistic falls within the critical region the claim made in the alternate hypothesis is supported. In this example, students should calculate a p -value of 0.04246 and conclude that their alternate hypothesis is supported. Caution students that this is not in itself proof, but it does help to confirm the claim.

"Title"	"2-Prop z Test"
"Alternate Hyp"	"p1 > p2"
"z"	1.72282
"PVal"	.04246
"p̂1"	.00582
"p̂2"	.00502
"p̂"	.00542
"n1"	50000.
"n2"	50000.

Exercise 1

Remind students to compare percentage of cancer incidence to formulate the alternate hypothesis. If needed, provide students with the options ($p_1 < p_2$, $p_1 > p_2$, or $p_1 \neq p_2$).

Students should provide the following answers:

The p -value = 0.289 means that the data in our sample would be expected 28.9% of the time, if the null hypothesis is true.

2-Prop z Test

Successes, x1: 268

n1: 50000

Successes, x2: 281

n2: 50000

Alternate Hyp: Ha: p1 < p2

OK Cancel

"Title"	"2-Prop z Test"
"Alternate Hyp"	"p1 < p2"
"z"	-1.556356
"PVal"	.288984
"p̂1"	.00536
"p̂2"	.00562
"p̂"	.00549
"n1"	50000.
"n2"	50000.

Exercise 2

Remind students that their p -value does not prove that the states have the same cancer rates. The conclusion is that the data from this study does not support the alternate hypothesis.

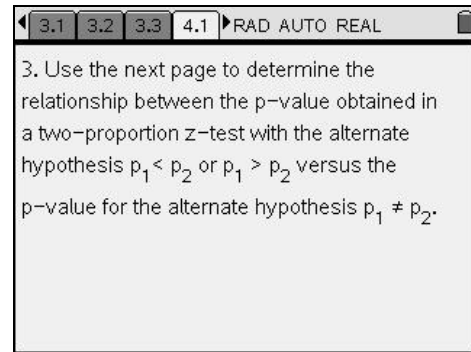
Students should provide answers similar to:

No, just because the two sample proportions are the same, does not mean that the actual incidence rates 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.

"Title"	"2-Prop z Test"
"Alternate Hyp"	"p1 > p2"
"z"	0.
"PVal"	.5
"p̂1"	.00536
"p̂2"	.00536
"p̂"	.00536
"n1"	50000.
"n2"	50000.

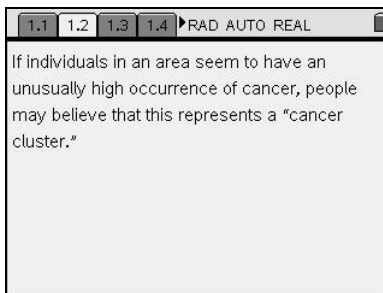
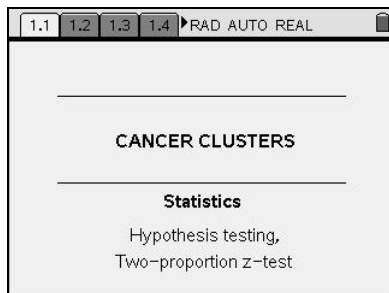
Exercise 3

Encourage students to work together in groups of 3 and calculate each of the three alternate hypotheses for a unique set of data. Have students make observations and share them with the class. Students should notice that the sum of the p -values for the $p_1 < p_2$ and $p_1 > p_2$ tests is 1. Students should also discover that the p -value for the $p_1 \neq p_2$ test is twice the smaller of the two p -values calculated in the $p_1 < p_2$ and $p_1 > p_2$ tests.

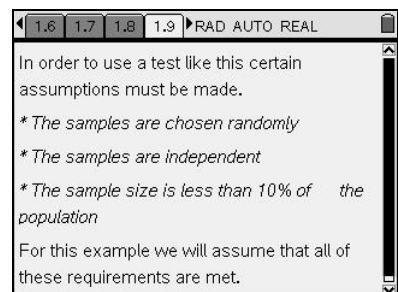
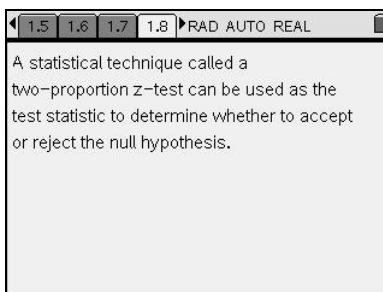
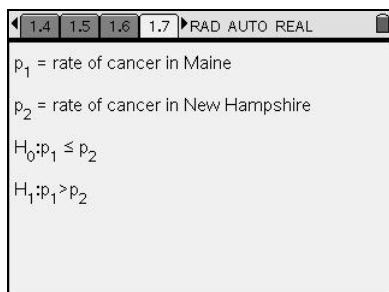
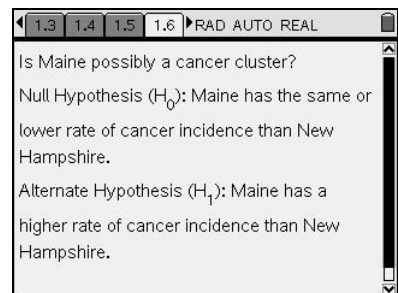
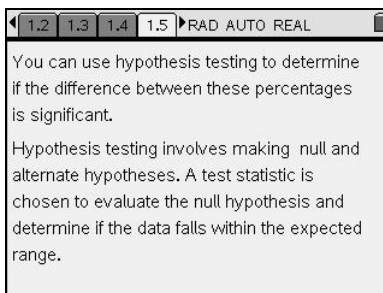
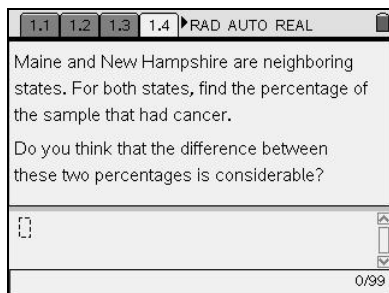


Cancer Clusters – ID: 9996

(Student)TI-Nspire File: *StatAct06_CancerClusters_EN.tns*



State	Cancer Cases
CT	268
ME	291
MA	268
NH	251
RI	281
VT	249



1.7 1.8 1.9 1.10 ▸RAD AUTO REAL

On the next page calculate the 2-prop z test. Use the p-value to determine whether to reject the null hypothesis. A p-value of 0,05 or lower is considered unusual.

1.8 1.9 1.10 1.11 ▸RAD AUTO REAL

0/99

1.10 1.11 1.12 1.13 ▸RAD AUTO REAL

The test calculated a p-value of about 0,042. This value means that if Maine did not have a higher cancer rate, the difference observed in these samples would happen by chance about 4.2% of the time, which is highly unlikely.

1.10 1.11 1.12 1.13 ▸RAD AUTO REAL

Based on the p-value, the null hypothesis is rejected and statisticians can conclude that the sample data supports the claim that the cancer incidence in Maine is higher than in New Hampshire.

1.12 1.13 2.1 2.2 ▸RAD AUTO REAL

Exercises

1.12 1.13 2.1 2.2 ▸RAD AUTO REAL

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.

State	Cancer Cases
CT	268
ME	291
MA	268
NH	251
RI	281
VT	249

1.13 2.1 2.2 2.3 ▸RAD AUTO REAL

1. Compare the cancer data between Connecticut and Rhode Island using a two-proportion z-test. Choose an appropriate alternate hypothesis. 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

2.1 2.2 2.3 2.4 ▸RAD AUTO REAL

0/99

2.2 2.3 2.4 3.1 ▸RAD AUTO REAL

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.

State	Cancer Cases
CT	268
ME	291
MA	268
NH	251
RI	281
VT	249

2.3 2.4 3.1 3.2 ▸RAD AUTO REAL

2. 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.

2.4 3.1 3.2 3.3 ▸RAD AUTO REAL

0/99

3.1 3.2 3.3 4.1 ▸RAD AUTO REAL

3. Use the next page to determine the relationship between the p-value obtained in a two-proportion z-test with the alternate hypothesis $p_1 < p_2$ or $p_1 > p_2$ versus the p-value for the alternate hypothesis $p_1 \neq p_2$.

3.2 3.3 4.1 4.2 ▸RAD AUTO REAL

0/99

