Candy Pieces

ID: 10039

Time required
45 minutes

Activity Overview

This activity introduces the statistics concept of hypothesis testing. Students are given information on the number of pieces of candy in a bag according to color. They are asked to consider whether the bag came from a manufacturing process designed to produce equal portions of each color. Students will then use a chi-square test for goodness-of-fit to determine if there is a significant difference between the proportions they find in the sample, and the proportions they would expect if the manufacturer produced equal proportions of each color.

Topic: Hypothesis Testing

- *Use a X^2 goodness-of-fit test to test the hypothesis that an observed frequency distribution fits an expected frequency distribution.*

Teacher Preparation and Notes

Tell students that this activity is only an informal introduction to the concept of hypothesis testing, a major topic in statistical inference. Hypothesis tests work in a manner similar to a jury in a criminal trial. In the American system of justice, a defendant is presumed innocent. If the evidence is convincing enough, the jury finds the defendant guilty. A jury never finds a defendant innocent—just guilty or not guilty. In a similar way, at the outset, a statistician assumes that there is no statistical difference between the observed data and the expected results. Statisticians use hypothesis tests to make inferences about a population based on random samples. With hypothesis tests, statisticians determine whether there is enough evidence to reject the hypothesis that the difference can be due to chance, or decide there is not enough evidence to reject it.

- *Actual bags of candy can be used in the exercise questions instead of using the supplied data.*
- *This activity is designed to be **teacher-led** with breaks for individual student work. Lead an interactive class discussion by using the questions found on the following pages.*
- *Each student should have a calculator and the worksheets to follow along with the teacher and to record their answers.*
- *The student worksheet is intended to guide students through the main ideas of the activity and provide a place to record their answers. Alternately, 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.*
- ***To download the student worksheet, go to education.ti.com/exchange and enter "10039" in the keyword search box.***

Associated Materials

- *CandyPieces_Student.doc*

In any situation, the outcomes (like the proportions of ingredients in a sample of candy) will vary due to chance. If you toss a coin 1000 times, you would expect to get somewhere around 500 heads.

The chi-square test for goodness-of-fit can be used to determine if there is a statistical difference between the proportions you would expect to get in a sample and the proportions you actually get. This activity contains an informal introduction to the mechanics of this test.

Calculating chi-square is one of many procedures statisticians use for testing hypotheses. The starting hypothesis (called a null hypothesis) in this case assumes there is no difference in the distribution of colors. The data are inspected to see if they support this assumption and chi-square is the statistical tool used to make a decision. The alternate hypothesis is that the proportions of colors are not all equal.

A chi-square test requires that…

- the samples are chosen randomly
- the samples are independent
- the sample size is large enough for the expected values to be at least 5 (i.e., if you toss a 6-sided die 60 times, the expected numbers for each outcome would be 10)

For this example we will assume that all these requirements are met.

Problem 1 – Introduction to chi-square

Suppose a certain popular brand of candy pieces comes in five colors. A student counted the number of pieces of each color in a bag and found the results shown in the table below.

Color	Yellow	Red	Blue	Orange	Green
Amount	11	19	25	17	13

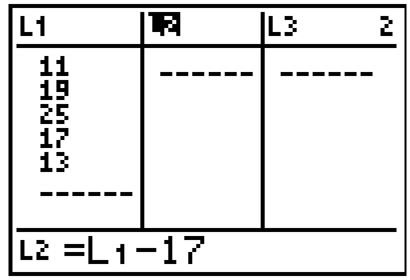
Based on this data, is it likely that this bag of candy came from a manufacturing process that was designed to produce equal proportions of each color? One way to answer this question is to perform a statistical procedure called a chi-square test for goodness-of-fit, checking to see how well the sample distribution fits the theoretical distribution.

1. How many pieces of candy were in the bag?

2. If the proportions of each color were the same, how many pieces of each color you would expect to find?

These computed values are called the **expected values**. The actual counts of each color of candy in the table are called **observed values**.

Open the List Editor by pressing STAT. Enter the numbers of each color from the table into list **L1**.



3. In the formula bar of list **L2**, compute the (Observed – Expected) value for each candy color.

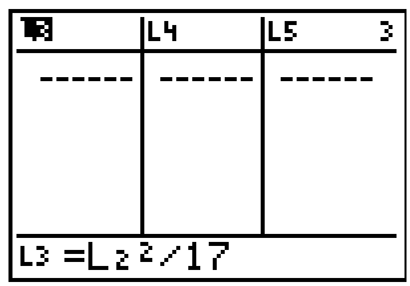
4. Do you think that there is a considerable difference between the observed values and the expected values?

The chi-square test involves quantifying the extent of the difference between the observed and expected values for each of the colors.

5. Does it make sense to find the sum of the differences (Observed – Expected) to describe the total difference? Why or why not?

6. Use the formula bar of list **L3** to compute $\frac{(\text{Observed} - \text{Expected})^2}{\text{Expected}}$ for each color.

Then in **L4(1)** use the **sum(** command from the Catalog to find the sum of these values.



This sum is called the chi-square, represented by χ^2.

The question now is whether the chi-square found is large or small. To find out, you can compare your observed chi-square to the theoretical distribution of the probabilities that a chi-square would be greater than or equal to any given value. The cumulative chi-square density can be obtained from a command on your calculator.

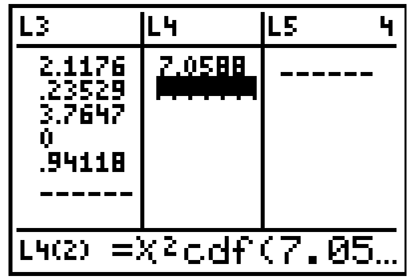
To determine if the chi-square found is too large or small, three quantities must be known: the lower bound, the upper bound, and the degrees of freedom.

The **Lower Bound** is the chi-square value calculated in Question 6.

The **Upper Bound** is positive infinity, so that you can find the probability that a chi-square would fall between your value and infinity by chance.

The **Degrees of Freedom** is one less than the number of categories; in this case, the number of degrees is 4. Since there are 5 colors, the number of candies possible for the fifth color is determined by the number of candies for each of the other four colors. That is, there are only four independent variables; the last color depends on how many of the other four there are.

7. In **L4(2)**, compute the cumulative chi-square density by pressing [2nd] [DISTR], and selecting χ^2 **cdf(**. Type the three quantities described above, in that order, in parentheses.



The value found in Question 7 is the region under the graph of the chi-square density function from the chi-square value to the right (positive infinity). This area is called the *p*-value. It is the probability that chi-square would be as large, or larger, than the observed value simply due to chance.

8. Write the *p*-value for this bag of candy as a percent.

If the *p*-value is more than 5%, a statistician would conclude that there is not enough evidence to reject the hypothesis, or in this case, that the bag of candy did not come from a process that produced equal numbers of the colors. If the *p*-value is less than 5%, a statistician would conclude that there is sufficient evidence to reject the hypothesis.

9. Based on your *p*-value in Question 8, is there sufficient evidence to reject the hypothesis that the bag of candy came from a process that produced equal numbers of colors? Explain.

Problem 2 – A second bag

A student opened a bag of a second brand of candy, counted the number of pieces of each color, and found the results shown in the table below.

Color	Brown	Yellow	Red	Blue	Orange	Green
Amount	15	14	16	35	29	24

10. Use the command **ClearAllLists** from the Catalog to clear lists **L1**, **L2**, **L3**, and **L4**. Then use these lists to compute the chi-square value.

11. Find the *p*-value.

12. Determine if it is likely that this bag of candy came from a manufacturing process that was designed to produce equal numbers of each color. Explain.

The student looked at the company's website and found that they claim to produce the colors in the following proportions.

Brown: 13%, Yellow: 14%, Red: 13%, Blue 24%, Orange 20%, Green: 16%

13. Compute the chi-square value using these proportions of colors. (Hint: First, find the new expected values.)

14. Find the *p*-value.

15. Determine if it is likely that this bag of candy came from a manufacturing process that was designed to produce the colors in the proportions given on the website. Explain.

Solutions

1. 85

2. Divide 85 by 5: 17 of each color.

3. –6, 2, 8, 0, –4

4. Answers may vary.

5. No, the sum of the differences is 0.

6. Yellow: 2.1176, Red: 0.2353, Blue: 3.7647, Orange: 0, Green: 0.9412;
Sum is about 7.0588

7. about 0.1328

8. 13.28%

9. No, not enough evidence because 13.28% > 5%.

10. 16.729

11. about 0.00504 = 0.504%

12. Since the *p*-value is less than 0.05 there is enough evidence to reject the hypothesis that the colors were manufactured in equal numbers.

13. 2.407

14. about 0.7904 = 79.04%

15. Since the *p*-value is greater than 0.05 there is not enough evidence to suggest that the bag of candy did not come from a process that produced the stated proportion of colors.

