Inspiring Greatness

INDEPENDENCE SCHOOL DISTRICT

Math Virtual Learning

AP stats / Intro to Chi squared GOF

May 13th , 2020

Lesson: May 13th, 2020

Objective/Learning Target:
Students will be introduced to the setting for Chi squared goodness of fit tests

Review #1

We open up a bag of M&M's and find the following results:

Cyan blue	orange	green	yellow	red	brown
10	10	8	8	10	5

MARS informs us the M&M's should be in the following proportions:

M&M'S MILK CHOCOLATE: 24% cyan blue, 20% orange, 16% green, 14% bright yellow, 13% red, 13% brown.

Do we have evidence that there are too many red M&M's? Conduct a test.

Review #2

Using the same bag of M&M’s, we see the company puts the same proportion of red and brown in each bag. However, there are far fewer brown. Can we use a 2 proportion z-test to see if the proportions are really different?

Answers

#1 - we can conduct a 1 proportion z t-test. We accept the bag as a random sample. The M&M's should be independent and we are definitely less than 10%. We meet the np rule for normality. P-hat = 19.6%. H_0: p = 0.13, H_a: p>0.13. We use 1-propZtest on TI-84 and get z= 1.40 and p-value=0.08. We do not have evidence that there are too many red at alpha = 0.05. We fail to reject the null.

#2 - The conditions for a 2 prop z-test are not met for this sample. The biggest problem is the lack of independence. This is a single sample not two samples. There are more advanced techniques that can manage this type of data. The other issue is in normality. When we check np we only get 5. This is not as big of an issues. AP statistics teaches np>10, but many textbooks accept np>5.

Testing multiple categories

What if I wanted to know if all 6 colors of M&M's match what the company says? More precisely, what if I want to know if the company is lying about the distribution of colors in a bag of M&M's?

We could conduct a 1-proportion z-test on each of the 6 colors. But we run into two issues… first, that is a lot of work. Second, is we lose confidence in our answer. If I am 95% confident in each test, that is a 95% chance of guessing correct for each test, I am $(0.95\%)^6$ = 73.5% confident in guessing right on all 6 colors. Wow, that is not good! I would have to use an Alpha > 0.99 to be 95% confident in guessing all 6 correct, which loses a lot of power.

Chi Squared Goodness of Fit test

Staticians have come up with several solutions to this issue over the years. In AP stats, we will learn about the Chi-squared family of tests to solve these issues. In more advanced classes, you may learn about different types of regression that are capable of dealing with this type of data.

The first test we learn about in the Chi-squared distribution, is the goodness of fit test (GOF). Today, the goal is to understand the setting in which it works. Tomorrow, you will work through conducting the test. Please watch the following video.

GOF video #1

Extra practice

Reading: pg 678-690

HW: 1, 3, 5, 7, 9, 11, 17