CHI-SQUARE GOODNESS-OF-FIT, PART TWO

0:00 / 4:39 1.0x

SPEAKER: MICHAEL J. MAHOMETA, Ph.D.

The chi-square Goodness of Fit test can be

used to test an observed pattern of categorical responses

against a RANDOM distribution.

And we've seen this.

But the test can also handle a specific NULL distribution

- not just a random one.

Remember, the test is all about comparing counts - what's observed,

and what's expected.

It turns out, we the investigators can decide

exactly what the EXPECTED distribution should be.

Help

Now notice how that question is a little bid different from the first one -

this time taking into account a possible specific distribution.

Here's our Table of Counts again - 381 independent observations

of respiratory events - a cough or a sneeze.

The Centers for Disease and Prevention have recommendations

for what to do when a respiratory event occurs.

They recommend either using a tissue, or using your elbow

if no tissue is available.

Using your hand is not recommended, as hand to surface transmission

is then a real possibility.

Now, knowing these recommendations, and the fact

that the CDC creates posters and flyers for colleges and universities to use,

it seems reasonable that students may be following these guidelines

- so let's go ahead and test a different NULL distribution possibility.

Help

Instead of a random distribution of expected percentages

being 25% in each cell - as we have 4 possible categorical responses -

we can set those percentages to be something more in line

with the CDC recommendations.

Here's the Percentage of expectation that we

might use - say because the recommendations are for the two

behaviors, those will share 70% of the overall distribution, with 35% in each.

Then the last two cells: "hand" and "no cover"

get to split 30% - so 15% in each.

The total for the percentage of the expected distribution is 100%.

Now, we tackle this just like before.

We run through the assumptions, and the steps of the Hypothesis test.

The assumptions still hold, but there are slight changes to the Hypothesis.

First the Null and Alternative.

The Null states that the probability of the first response is 35%,

Help

the second is 35%, the third is 15% and the fourth is 15%.

The Alternative is that the probabilities specified in the Null

are not true.

Now the second step is the alpha; that's fine.

It's the same as before at 0.05.

The third step of the hypothesis test is still

the calculation of the chi-square statistic.

But now, we use the NEW specified percentages

to calculate the expected counts.

The formulation and the summation is the exactly the same.

When we're all done, we get a chi-square statistic of 541.28.

Again, at 3 degrees of freedom, the critical value corresponding

to our alpha level is 7.81.

Step four, the conclusion.

Since our chi-square statistic is greater than our critical chi-square,

we Reject the Null Hypothesis - the observed distribution

of the categorical variable of behavior does not follow the specified

distribution of behavior.

Again we can get the same result using R and the chisq.test() function,

but this time telling the function what the expected probabilities should be.

Again, our p-value is lower than our alpha,

so we Reject our Null Hypothesis.

The take away from all this?

One: A hypothesis test for a single categorical variable

uses the chi-square Goodness-of-Fit test.

Two: That test can use either random probabilities for the expected

distribution, or specified probabilities that we can define

- the mechanics and the underlying test for both are still the same.

And three: I wouldn't shake hands with a college student during flu season

if I were you

Help

1. Jurors are selected from the list of registered voters, so the ages for jurors should have the same distribution as the ages of voters. A law professor obtains voter registration records and finds that 20% of registered voters are 18-29, 45% are 30-49, and 35% are age 50 or older. The professor then monitors jury composition over a month-long period and finds the following distribution of jurors:

18-29 years old	30-49 years old	50 years old and over

12	36	32

Help

(1 point possible)

1a. We want to match the distribution of a categorical variable to a hypothesized distribution model. Which Chi Square test should we run?

- Goodness-of-fit ✔
- Test of Independence

Hide Answer

(1 point possible)

1b. The null hypothesis is H0: The ages of jurors are distributed the same way as the ages of voters. What is the alternative hypothesis?

- There are more older jurors than younger jurors.
- There is no relationship between the age of jurors and the age of voters.
- The age of jurors are not distributed the same way as the ages of voters. ✔

Hide Answer

1c. One law student divides the total number of jurors by the number of categories and comes up with the following expected values for each category:

18-29 years old	**30-49 years old**	**50 years old and over**
26.67	26.67	26.67

Help

What did he do wrong?

(1 point possible)

- He found expected values using the method for a test of independence instead of a goodness of fit test.
- He came up with expected values that are not whole numbers; it's not possible to have 26.67 people.
- He did his division wrong.
- He assumed the jurors should be evenly distributed across the categories, but this is not the hypothesized model. ✔

Hide Answer

(3 points possible)

1d. Find the expected values for each category, assuming the age distributions of jurors and voters are the same.

18-29 years old: *(Report as a whole number.)*

Help

Answer: 16

30-49 years old: *(Report as a whole number.)*

Answer: 36

50 years or older: *(Report as a whole number.)*

Answer: 28

Hide Answer

(1 point possible)

1e. Find the chi-square statistic using the following formula:

$$\chi^2 = \sum \frac{(obs - exp)^2}{exp}$$

- 1.289
- 1.571 ✔
- 3.028
- 0.023

Help

Hide Answer

(1 point possible)

1f. What is the degrees of freedom for this hypothesis test?

Answer: 2

Hide Answer

(1 point possible)

1g. What is the critical Chi-square value for this hypothesis test? (Rounded to 2 decimal places)

Help

- 2.33
- 5.99 ✓
- 6.43
- -4.35

Hide Answer

(1 point possible)

1h. Based on your above answers, you should _________ the null hypothesis.

fail to reject

Hide Answer

(1 point possible)

1i. What is the appropriate interpretation of this hypothesis test?

- We have no evidence to suggest that the age distributions of jurors are any different of those of registered voters. ✓
- Jurors are older than registered voters; these distributions are not the same.
- There is no relationship between the age of jurors and the age of voters.
- Age of jurors and age of voters are related.

Hide Answer

Help