

Chi-square testing of homogeneity

March 3, 2022

Is the distribution of gummy bear color the same for Haribo gummy bears and Meijer gummy bears? Below is some data that was conducted. Does this data provide convincing evidence of a difference?

	Haribo	Meijer	Total
Red	20	31	51
Green	11	11	22
Yellow	8	17	25
Orange	11	26	37
White	13	19	32
Total	63	104	167

1. How many samples do we have? What population is each respective sample from? Explain.

Solution: 2 samples are drawn from 2 populations, the two candy manufacturers.

2. How many variables are we examining?

Solution: We are examining one variable: colour.

3. Write down the hypothesis for a significance test below:

Solution:

H_0 : There is no difference in the colour distribution for Haribo and Meijer gummy bears

H_a : There is a difference in the color distribution for Haribo and Meijer gummy bears.

4. We will use a chi-square test to see if there is a difference in the two populations. Complete the table for expected values under the null hypothesis below:

	Haribo	Meijer	Total
Red	$(51/167) \times (63)$		51
Green			22
Yellow			25
Orange			37
White			32
Total	63	104	167

5. Use your work from the previous page to complete a 4-step significance test.

Solution: State: (Hypothesis from previous page). $\alpha = 0.05$.

Plan: Carrying out a chi-square test for homogeneity

Conditions: Random ✓, 10% (independence) ✓, Large counts (expected counts greater than or equal to 5) ✓.

Do: Test Statistic: $\chi^2 = \sum \frac{(O-E)^2}{E}$

Write down your work here $\chi^2 = 2.93$. The degrees of freedom for the chi-square distribution is $(r-1)(c-1) = (5-1)(2-1) = 4$. Draw the picture as well!

P-Value: $P(\chi_4^2 > 2.93) = \text{chisqcdf}(2.93, 100000, 4) = 0.57$

Conclude: Because the p-value is greater than 0.05 (α) we fail to reject the null hypothesis and do not have convincing evidence that there is a difference in the colour distribution for Haribp and Meijer gummy bears.

6. Explain how this test is different from a chi-square test for goodness of fit?

Solution: We have 2 samples from 2 populations. χ^2 GOF testing has 1 sample from 1 population.