# Chi-Square Tests



## Math Objectives

- Students will recognize that chi-squared tests are for counts of categorical data.
- Students will identify the appropriate chi-squared test to use for a given situation: $\chi^2$ Goodness of Fit Test, $\chi^2$ Test of Independence, or $\chi^2$ Test of Homogeneity.
- Students will learn how to calculate the degrees of freedom for each type of chi-squared test.
- Students will interpret the results of a chi-square test.
- Students will reason abstractly and quantitatively.

## Vocabulary

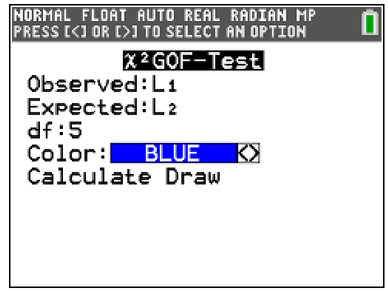
- Alpha
- categorical data
- chi-squared ($\chi^2$) distribution
- degrees of freedom
- expected counts
- goodness-of-fit
- observed counts
- p-value
- test of homogeneity
- test of independence

## About the Lesson

- This lesson is aligning with the curriculum of IB Mathematics Applications and Interpretations SL/HL
- This falls under the IB Mathematics Core Content Topic 4 Statistics and Probability:

  **4.11a** Formulation of null and alternative hypotheses, $H_0$ and $H_a$, using appropriate significance levels, and finding and concluding with *p*-values.

  **4.11b** Finding expected and observed frequencies, using the $\chi^2$ test for independence (contigency tables, degrees of freedom, and critical values), and using the $\chi^2$ goodness of fit test.
- This lesson involves investigating chi-squared tests and distributions.
- Students will compare different scenarios and determine which chi-square test is appropriate
- Students will write the appropriate null and alternative hypotheses for the given scenario.
- Students will determine the degrees of freedom for the chi-square test.




**Tech Tips:**

- This activity includes screen captures taken from the TI-84 Plus CE. It is also appropriate for use with the rest of the TI-84 Plus family. Slight variations to these directions given within may be required if using other calculator models.
- Access free tutorials at http://education.ti.com/calculators/pd/US/Online-Learning/Tutorials

**Lesson Files:**

*Student Activity*

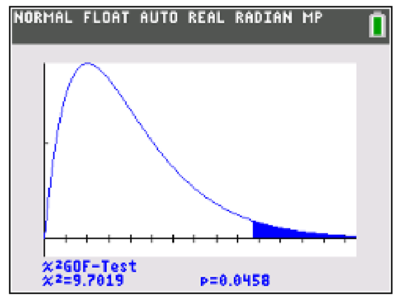
- Chi-Squared Tests_ Student-84.pdf
- Chi-Squared Tests_ Student-84.doc

- Students will look at chi-square test results and make the correct decision to reject or fail to reject the null hypothesis and write their conclusions in context.

## Activity Materials

- Compatible TI Technologies:
  TI-84 Plus*, TI-84 Plus Silver Edition*, TI-84 Plus C Silver Edition, TI-84 Plus CE

In this activity, you will look at a problem situation that involves categorical data and will determine which is the appropriate chi-square test to use: the chi-squared goodness of fit or the chi-squared two-way test.




Three different chi-squared tests will be discussed in this activity:

- $\chi^2$ **Goodness-of-Fit (1)**: Compares sample counts (sometimes given as proportions) to expected counts based on a given population distribution.
- $\chi^2$ **Two-way tables (2 & 3)**: There are two chi-squared tests using two-way tables—independence and homogeneity. The two tests differ in their hypotheses and conclusions but are mechanically identical. Determining which to use depends on how the data were collected.
  - **Test of Independence**: Compares two categorical variables in a *single population* to determine whether there is a significant association between the two variables.
  - **Test of homogeneity**: Compares categorical variables from *two or more different populations* to determine whether proportions are the same across different populations.

1. a. Suppose that in a typical week the number of absences from a large high school was 805. State about how many you would expect per day. Explain your reasoning.

**Sample Answer:** Some students might suggest that about the same number might be absent each day, so 805 divided by 5; others might think more students would be absent on Monday and Friday than on other days of the week.

b. The school wants to see whether student absences are the same on different days of a randomly selected week of school. State the type of hypothesis test that should be used. Explain your answer.

**Sample Answer:** This is a chi-square goodness of fit test because the data are categorical counts of absences on each day of the school week. We want to see if the sample absences fit a population pattern- is the number of students absent about the same each day?

c. Write the null and alternative hypotheses for this test.

**Sample Answer:** $H_o$: the proportions of absences are the same for each day of the week ($p_{monday} = p_{tuesday} = p_{wednesday} = p_{thursday} = p_{friday} = 0.2$) $H_a$: at least one proportion is different

2. In the table below, Column 2 represents the absences (observed) throughout a typical week. Please fill in Column 3 (expected absences) and the totals.

| Day of week | Observed absences | Expected absences |
|---|---|---|
| Monday | 173 | **161** |
| Tuesday | 157 | **161** |
| Wednesday | 138 | **161** |
| Thursday | 149 | **161** |
| Friday | 188 | **161** |
| Total: | 805 | 805 |

a. Explain how the observed number of absences compares to your conjecture in question 1a.

**Sample Answer:** Most absences do seem to be on Friday, and then Monday. The fewest absences were on Wednesday.

b. Describe how the expected number of absences are calculated, and state what they represent.

**Sample Answers:** The expected counts were found by taking the total of the observed absences (805) and dividing that by the typical number of days in a school week (5). The resulting answer (161) represents the number of absences that would be expected each day if the absences were the same for every day.

c. State the conditions for this test. State if the conditions are met.

**Sample Answer:** The conditions for a $\chi^2$ Goodness-of-Fit Test are that the sampling was random, less than 20% of the expected values are less than 5, and all of the expected values are greater than 1. Yes the conditions are met.

d. The chi-square statistic is dependent on the degrees of freedom. The number of degrees of freedom for a $\chi^2$ Goodness of Fit test is found using the number of categories minus one. State the degrees of freedom that should be used in this situation.

**Sample Answer:** There are five categories – each day of the school week. So, 5 – 1 = 4 degrees of freedom.

3. The chi-square test statistic and the associated *p*-value can be found by inputting the observed absences in $L_1$ (**Stat > 1: Edit**), the expected absences in $L_2$, and then pressing **Stat > Tests > D: $\chi^2$ GOF-Test**. Make sure you have filled in the degrees of freedom appropriately. Place your cursor on **Draw** and press enter. (**Note**: Make sure your plots are off.)

a. Describe the graph.

**Sample Answer:** The curve is skewed to the right with the area above the $\chi^2$ value of about 9.7 shaded.

b. Describe why the chi-squared is always a positive value.

**Sample Answer:** The chi-squared statistic is found by summing the $\frac{(observed-expected)^2}{expected}$ values. Because the differences (observed – expected) are squared, the answer will always by positive.

c. State the area of the shaded region. Explain your answer in the context of the problem.

**Sample Answer:** The shaded region has an area of about 0.046. If represents the p-value – the probability of getting a chi-squared test statistic this extreme or greater if the absences are the same each day.

d. Make a decision to reject or fail to reject your null hypothesis using an alpha value of 0.01. Write your conclusion in context.

**Sample Answer:** Fail to reject the null hypothesis because the p-value of 0.0457 is not less than the alpha value of 0.01. Based on the average weekly data, the evidence at the 0.01 level is not sufficient to suggest that student absences are different on different days of the week.

**Teacher Tip:** Point out to the students that at the .05 level, the null hypothesis would be rejected.

4. An advertiser for television shows suspected males and females had different television viewing preferences. The company commissioned a survey of 100 males and 120 females asking their preferences among crime, reality and comedy formats.
   a. Describe why the advertiser would care about such a difference.

**Sample Answer:** If a difference between viewing habits of shows and gender existed, the advertiser could tailor advertisements towards that specific audience.

   b. State the type of hypothesis test the advertiser should use to analyze the results. Explain your answer.

**Sample Answer:** This example compares categorical variables from male/female populations to determine whether proportions of TV format preferences are the same across the two genders, so it uses a Chi-squared Test of Homogeneity.

   c. Write the null and alternative hypotheses for this test.

**Sample Answer:** $H_0$: The three television formats have the same proportion of male and female viewers ($p_{males} = p_{females}$ for each format); $H_a$: the proportions are not the same.

5. a. The table below shows the survey results. Fill in the totals of the first table and the expected values of the second table.

| Program Format | Males from survey | Females from survey | Totals |
|---|---|---|---|
| Crime | 29 | 49 | **78** |
| Reality | 31 | 45 | **76** |
| Comedy | 40 | 26 | **66** |
| Totals | **100** | **120** | **220** |

| Program Format | Males expected | Females expected | Totals |
|---|---|---|---|
| Crime | **35.4545** | **42.5455** | **78** |
| Reality | **34.5455** | **41.4545** | **76** |
| Comedy | **30** | **36** | **66** |
| Totals | **100** | **120** | **220** |

b. Describe how you think the expected count for Males—Crime was calculated. Explain why this makes sense.

**Sample Answer:** 100 x (78/220) = 35.4545 – to get the answer on the handheld. If the null hypothesis is true, the expected count for males would have to give the same proportion for males as the total proportion for crime, which is 78/220 and there are 100 males so it would be 78/220 times the 100.

c. Explain what is meant by the expected count for the cell Males—Crime.

**Sample Answer:** The predicted number of men watching crime would be 35.4545 or about 35 if the proportions of men/women watching crime were the same.

d. State the conditions for this hypothesis test and state if the conditions are met. Explain your answer.

**Sample Answer:** The conditions are that the sampling was random, each expected cell count should be greater than 1 and no more than 20% of them should be less than 5. Yes, all conditions are met.

e. State if it appears from the results of the survey that there is a difference in the viewing preferences of men and women. Explain your reasoning.

**Sample Answer:** It kind of does look like there is a difference, especially for crime and comedy shows. Fewer males like crime than expected, and more females liked it. More males liked comedy than expected, and fewer women liked it.

f. The degrees of freedom for a $\chi^2$ two-way table is found using (# rows – 1)(# columns –1). What is the number of degrees of freedom for this test?

**Sample Answer:** (3 – 1)(2 – 1) = 2 degrees of freedom.

6. a. Using your handheld, conduct the $\chi^2$ test and interpret the results. (Input your observed values in matrix A with dimensions 3 x 2 and your expected values in matrix B also with dimensions 3 x 2. Press **Stat > Tests > C: $\chi^2$-Test**, input the appropriate matrices and press calculate.)

**Sample Answer:** The $\chi^2$ test statistic of 8.93249 has a p-value of 0.01149, which means that a $\chi^2$ as large as or larger than 8.9 would occur by chance in about 1.1% of the samples.

b. Make a decision to reject or fail to reject your null hypothesis using an alpha value of 0.05. Explain your reasoning.

**Sample Answer:** Using an alpha value of 0.05, you would reject the null hypothesis because the decision is based on having a chi-square occur at least as great as 8.9 in. less than 0.05 of the sample outcomes. The p-value for the survey results of 0.01 is less than 0.05, so having a chi-square of 8.9 is unlikely.

c. Write your conclusion in the context of the problem.

**Sample Answer:** There is strong evidence at the 0.05 level that males and females had different television viewing preferences.

## Wrap-Up/Assessment

The following questions can be used as part of the lesson as a self-check for students or can be used as an assessment to determine how well students understand the concepts.

1. Choose the appropriate chi-squared test for each situation and explain your choice:

a. A school wants to compare how its students did on the AP Statistics exam this year compared to the national scores.

**Answer:** $\chi^2$ Goodness of Fit – the school's results are being compared to the total population (national scores)

b. A restaurant samples customers to determine if there is a relationship between customer age and satisfaction with the restaurant's service.

**Answer:** $\chi^2$ Test of Independence – one group of customers is surveyed and asked both their age and satisfaction with the restaurant's service.

c. A consumer safety organization wants to see if there is a difference in seat belt use in Los Angeles, California; Miami, Florida; and Dallas, Texas.

**Answer:** $\chi^2$ Test of Homogeneity – three samples are chosen, one from each city, and the sample proportions are compared.

d. A survey asked men and women how confident they were, on a scale from 1 to 5, that they could change a flat tire.

**Answer:** $\chi^2$ Test of Homogeneity because two samples are chosen, one of men and one of women, and the sample proportions in each category of the confidence scale are compared.

e. The proportion of each color of M&M's in a bag are compared to the color distribution that the manufacturer claims to make.

**Answer:** $\chi^2$ Goodness of Fit – the color distribution of M&M's in a bag are compared to the manufacturer's claims to make.

2. Decide whether the following statements are always, sometimes or never true. Explain your reasoning in each case.

a. The $\chi^2$ curve is right-tailed.

**Sample Answer:** Always true – the (observed – expected) values are squared to get the $\chi^2$ test statistic so it is always positive.

b. A *p*-value is the probability of making a correct decision.

**Sample Answer:** Never true – a p-value is the probability of getting a test statistic as extreme as the one we got if the null hypothesis is true.

c. The number of degrees of freedom is $n$ – 1 for $\chi^2$ tests, where $n$ is the sample size.

**Sample Answer:** Never true – the number of degrees of freedom is (# categories – 1) for the $\chi^2$ Goodness-of-Fit test and (# rows – 1)(# columns – 1) for the $\chi^2$ two-way tables.

***Note: This activity has been developed independently by Texas Instruments and aligned with the IB Mathematics curriculum, but is not endorsed by IB™. IB is a registered trademark owned by the International Baccalaureate Organization.*