Chi-Square Tests
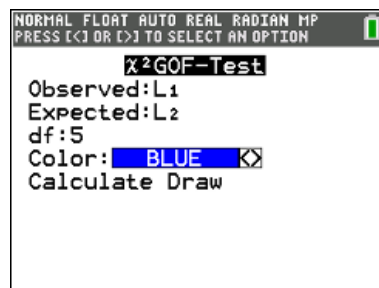**Student Activity**

Name _____

Class _____

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.

```
NORMAL FLOAT AUTO REAL RADIAN MP
PRESS [<] OR [>] TO SELECT AN OPTION
            χ²GOF-Test
Observed:L₁
Expected:L₂
df:5
Color:    BLUE    <>
Calculate Draw
```

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

- χ^2 **Goodness-of-Fit (1)**: Compares sample counts (sometimes given as proportions) to expected counts based on a given population distribution.
- χ^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.

   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.

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

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	
Tuesday	157	
Wednesday	138	
Thursday	149	
Friday	188	
Total:		

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

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

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

d. The chi-square statistic is dependent on the degrees of freedom. The number of degrees of freedom for a χ^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.

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: χ^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.

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

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

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.

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.

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

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

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	
Reality	31	45	
Comedy	40	26	
Totals			

Program Format	Males expected	Females expected	Totals
Crime			
Reality			
Comedy			
Totals			

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

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

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

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.

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

6. a. Using your handheld, conduct the χ^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: χ^2-Test**, input the appropriate matrices and press calculate.)

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

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

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.

b. A restaurant samples customers to determine if there is a relationship between customer 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.

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.

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.

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

a. The χ^2 curve is right-tailed.

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

c. The number of degrees of freedom is $n - 1$ for χ^2 tests, where *n* is the sample size.