

Inference for Two-way Tables

ID: 12728

 Time required
 20 minutes

Activity Overview

In this activity, students will use the chi-square test to analyze whether two categorical variables are independent or dependent. First they will calculate the expected frequencies, test statistic and critical values. Then students will use the χ^2 -Test command to calculate the P-value.

Topic: Inferential statistics

- *Two-way tables*
- *Chi-square test for independence*

Teacher Preparation and Notes

- *Students should have knowledge of a chi-square distribution, p-values, and hypotheses testing in general.*
- *If there is extra time available, have students calculate the conditional distributions as a review and compare their findings from the activity.*
- *Students will write their responses on the accompanying handout.*
- ***To download the student worksheet and optional program, go to education.ti.com/exchange and enter "12728" in the keyword search box.***

Associated Materials

- *TwoWayTables_Student.doc*
- *INVERSX2.8xp (optional program)*

Suggested Related Activities

To download any activity listed, go to education.ti.com/exchange and enter the number in the keyword search box.

- *Testing Goodness-of-Fit and Two-Way Table Chi-Square (TI-84 Plus family) — 4590*
- *TI Using the Chi-Squared Significance Test (TI-84 Plus family) — 3502*
- *Chi-Square Test for Independence and Homogeneity (TI-84 Plus family with TI-Navigator) — 1986*

Problem 1 – Party Affiliation

At the start of the activity ask students the following questions to get them thinking about the relationships between categorical variables. Discuss with students why they think the variables are independent or dependent.

- Would you expect there to be a relationship between a person’s political affiliation and their approval rating of the President’s handling of the economy?
- Would you expect that there would be a relationship between a person’s job status (employed or unemployed) and their approval rating of the President’s handling of the economy?
- List two categories that would not necessarily have a relationship with the approval rating of the President’s handling of the economy.

The results of the survey by Quinnipiac University are given in the first table on the worksheet. In this problem students will conduct a test of independence to determine whether a contingency table’s row variable is independent of its column variable. Explain to students that the test cannot be used to establish a direct cause-and-effect link between two variables.

For example, from the table students might conclude that the approval rating is dependent on party affiliation but that does not mean that the affiliation category has some direct causative effect on approval rating.

Students are asked to write the null and alternative hypotheses on the worksheet.

Before beginning the activity, have students set up the List Editor. From the Home screen, press **[STAT]**, select **SetUpEditor**, and press **[ENTER]**.

Then press **[STAT]** **[ENTER]**, arrow up to highlight L1, and then arrow right until an untitled list appears. Then enter L7, move right again, enter L8, and move right again, enter L9.

L7	L8	L9	9
-----	-----	████████	
L9(1)=			

Students need to enter the data from the table (excluding the totals) into lists L1, L2, and L3 by pressing **[STAT]** and choosing **Edit(**.

L1	L2	L3	3
158	886	79	
217	39	43	
47	39	17	
-----	-----	████████	
L3(4) =			

Now students will calculate the expected frequencies in lists **L4**, **L5**, and **L6**. This should be done using the following formula:

$$\frac{(\text{row total})(\text{column total})}{(\text{grand total})}$$

Students can use the totals found in the table on the worksheet. For example the value for **L4(1)** is **(1123*422)/1525**.

Explain to students the value in **L4(1)** is the same as calculating the overall percentage of people who approve of the President's handling of the economy and then calculating the number of Republicans you would expect if the same percentage of Republicans had responded with approval.

If desired students can verify that each column totals to the number of respondents to the survey using the **sum** command (**2nd** [LIST]) on the Home screen.

The next step for students is to find a test statistic to allow them to compare the observed counts with the expected counts.

$$\text{Test statistic} = \sum \frac{(\text{observed} - \text{expected})^2}{\text{expected}}$$

Students will calculate this value in two steps: (1) using the values in the previous lists to calculate $(O - E)^2/E$ in lists **L7**, **L8**, and **L9**, and then (2) finding the sum of all the values.

The formula for **L7** is $(L1-L4)^2/L4$.

When using the sum command, students can enter the list names by pressing **2nd** [LIST] and selecting the appropriate name from the list.

Now students are to determine the critical value; to do this they may use a Chi-square distribution chart or the **INVERSX2** program. Since they are testing with a significance level of 0.05 and the critical region is located in the right tail only, students will use 0.95 as the area. They should find the critical value to be 9.49.

Having students sketch a graph of a χ^2 distribution and plot the test statistic and critical value may help them visualize the result.

Discuss with students whether they should reject or fail to reject the null hypothesis. Since the test statistic is greater than the critical value, they should reject the null hypothesis.

L4	L5	L6	6
310.76	709.88	102.36	
82.74	189.01	27.253	
28.502	65.11	9.3974	
-----	-----	-----	
L6(3) = (103*139)/...			

L7	L8	L9	7
75.091	43.693	5.3306	
217.86	119.05	9.0986	
12.005	10.47	6.1715	
-----	-----	-----	
L7 = (L1-L4)^2/L4			

```
sum(L7+L8+L9)
498.7760884
```

```
PRGM INVERSX2
Cumul. Area? 0.95
Degr. Free. ?4
9.487729037
9.487729037
```

To confirm their decision, students can use the χ^2 **Test** command. First they need to store the data from the first table on the worksheet (or lists L1, L2, and L3), totals omitted, in a matrix.

Press $\boxed{2\text{nd}}$ $\boxed{[MATRIX]}$ and choose **EDIT**. Students should enter the size of the matrix (in this case 3 x 3) and use the arrow keys to navigate through the matrix, entering the values as shown.

```
MATRIX[A] 3 x3
[ 158   886   79 ]
[ 217   39   43 ]
[ 47    39   87 ]
3, 3=17
```

Once the data is entered into the matrix students can use the χ^2 -**Test** command by pressing $\boxed{[STAT]}$ and choosing χ^2 -**Test** from the **TESTS** menu.

```
 $\chi^2$ -Test
Observed: [A]
Expected: [B]
Calculate Draw
```

Students will be prompted for an observed matrix; they should enter **A** (or wherever they entered the observed data). For the expected matrix they may enter **B** (or another *blank* matrix where they would like the values stored). They should then highlight **Calculate** and press $\boxed{[ENTER]}$.

```
MATRIX[B] 3 x3
[ 310.76  709.88  102.36 ]
[ 82.74   189.01  27.253 ]
[ 28.502  65.11   9.3882 ]
```

Note: The expected matrix will show the expected frequencies. It can only be found after the test has been completed.

```
 $\chi^2$ -Test
 $\chi^2=498.7760884$ 
P=0
df=4
```

Discuss the meaning of the *P*-value that is returned. In general for a test of independence, if there is a small χ^2 value and a large *P*-value, you fail to reject independence; and if there is a large χ^2 value and a small *P*-value, you reject independence. The screenshot at the right shows a very small *P*-value which confirms the decision to reject the null hypothesis.

Problem 2 (Homework) – Gender

Students work with a similar claim as before except they are using male vs. female data. The data is included on the worksheet.

Students should first answer the question, based on the data alone: *Do you think the two variables are independent or dependent?*

They will use a 0.01 significance level to test the claim that the President’s approval rating is independent of a person’s gender.

Determine which method you would like students to use—the lists and formulas or the χ^2 Test command.

Students will then test the claim of independence. They should justify the answer by comparing the test statistic to the critical value or using the *P*-value. Students should also write a conclusion in the context of the claim.

Since the *P*-value is large, they should fail to reject the null hypothesis. It appears that the approval rating and whether a person is male or female is independent.

L1	L2	L3	Z
463	563	-----	
199	180		
50	73		

L2(4) =

MATRIX[A] 3 x 2

[463	563]
[199	180]
[50	73]

3, 2=73

χ^2 -Test

$\chi^2=7.958241205$

$P=.0187020787$

$df=2$