Inference for Two-way Tables

ID: 12729

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 ***2-way 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 enter their responses directly into the TI-Nspire handheld or write on the accompanying handout.*
- ***To download the student and solution TI-Nspire documents (.tns files) and student worksheet, go to education.ti.com/exchange and enter "12729" in the keyword search box.***

Associated Materials

- *TwoWayTables_Student.doc*
- *TwoWayTables.tns*
- *TwoWayTables_Soln.tns*

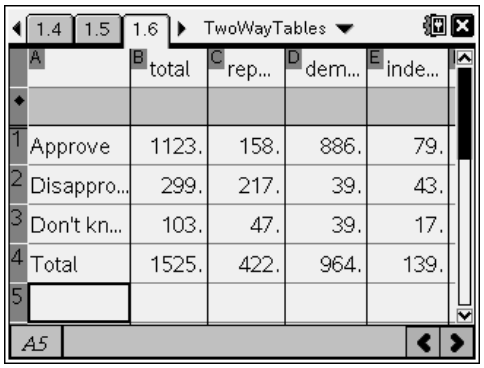
Suggested Related Activities

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

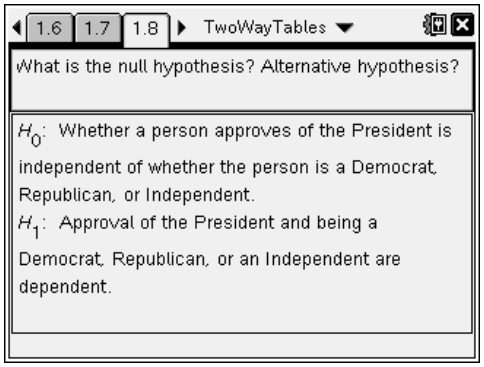
Students begin by thinking about relationships between categorical variables. Questions are posed based on a survey by Quinnipiac University. Discuss with students why they think the variables are independent or dependent.



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.

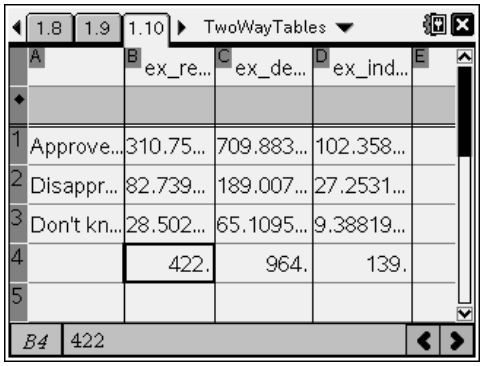
For example, from the table students might conclude that the approval rating is dependent on party affiliation but that does not mean that the affiliation category has some direct causative effect on approval rating.

Students are asked to write the null and alternative hypotheses on page 1.8.



Now students will use page 1.10 to calculate the expected frequencies. This should be calculated using the following formula, also given on page 1.9.

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

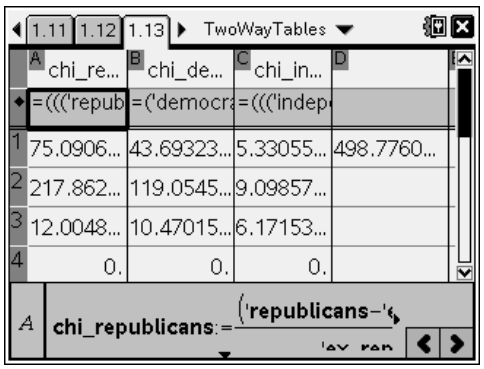


Explain to students for cell B1, it is the same as calculating the overall percentage of people who approve of the President's handling of the economy and then calculating the number of Republicans you would expect if the same percentage of Republicans had responded with approval.

Have students verify that each column totals to the number of respondents to the survey.

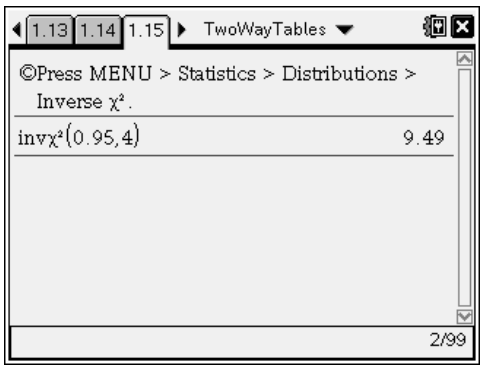
The next step is to find a test statistic to allow us to compare the observed counts with the expected counts.

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



Students will calculate this value on page 1.13 in two steps: (1) using the names of the columns in previous spreadsheets to calculate $(O - E)^2/E$ and then (2) finding the sum of the cells in cell D1.

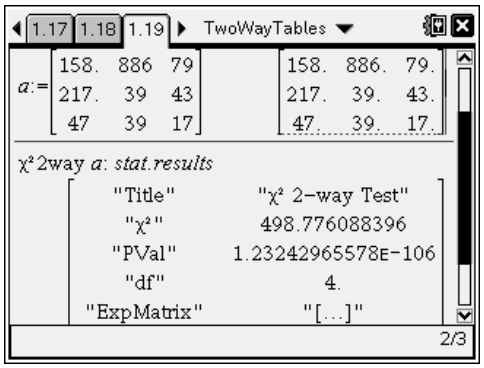
Now students are to calculate the critical value using the **Inverseχ^2** command. Since they are testing with a significance level of 0.05 and the critical region is located in the right tail only, students will use 0.95 as the area.



Having students sketch a graph of a χ^2 distribution and plotting 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.

To confirm their decision, students can use the **χ^2 2-way Test** command on page 1.19. But first they need to store the data from the spreadsheet on page 1.6, totals omitted, in a matrix.

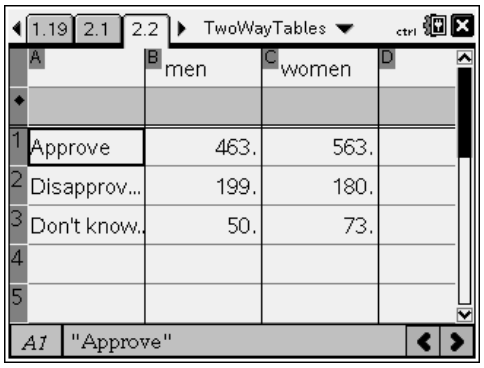


Notice that the command returns an “expMatrix”. This matrix will give the expected frequencies. The matrix can be found through the **var** menu ((var)). It can only be found after the test has been completed.

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.

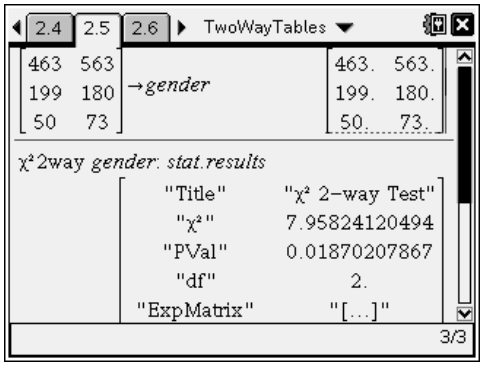
Homework problem

Students work with a similar claim as before except they are using male vs. female data. The data is given on page 2.2.



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

Determine which method you would like students to use—the spreadsheet and formulas *or* the χ^2 **2-way 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.