Exploring χ^2 (Chi Square) Distribution –Test of Association/Independence

The χ^2 test for association/independence is used to determine if two variables are independent or if there may be an association. For this test, the data are in a two-way table and the expected values are calculated for each value within the table.

Let's look at the Titanic data exploration from Chapter 1 and follow the **4 steps for an inference test**. The two-way table for class of passenger and survival status is given below.

		Survival status	
		Survived	Died
Class of Travel	First Class	197	122
	Second Class	94	167
	Third Class	151	476

To use the χ^2 test for association:

Step 1: Each individual is listed in the table and is only represented once within the table.

The **hypotheses** are:

H_0: there is no association between survival status and class of passenger

H_a: there is an association between survival status and class of passenger

Step 2: State the inference procedure: **Chi Square Test of Association/Independence** and verify the conditions have been met.

NOTE: To use inference for the χ^2 Distribution, you MUST verify that all expected counts are at least 1, and 80% or more of the expected counts are > 5. We will need to begin our procedure on the handheld.

If you did this activity in Chapter 1 and saved it, open the document now. If not, enter the data using the steps below.

1. Begin a *New Document* and *Add a Lists & Spreadsheet*.

2. Enter the data from the table above in the next three columns.

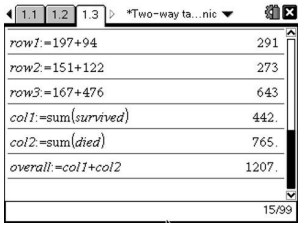
   - Name column A: **class**, and enter the classes of passengers in the first three cells. (Remember to use quotes so the Nspire treats the classes as categorical data.)

- Name column B: **survived** and column C: **died**.

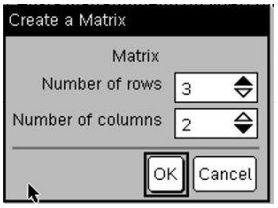
  NOTE: Since we will be using this data to calculate the marginal and conditional distributions, you will need to enter a decimal after every number. This is so the Nspire will display the results as a decimal instead of a fraction.

3. Put the data from the table into a matrix.
   - Insert a *Calculator* page.
   - Define a matrix by typing **observed :=** , then press (menu), *Statistics, List Operations, Convert List to Matrix*.
     - list▸mat will appear.
     - Go to *List Operations* again, and select *augment*.
     - Press (var) **survived** (,) (var) **died** (,) (2).

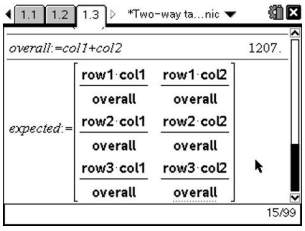
4. Calculate the expected matrix using $\frac{\text{row total * column total}}{\text{overall total}}$ for each cell.



   - On the *Calculator* page, you can define 3 variables for the row totals and 2 variables for column totals. Then for the last variable, **overall**, add the two column totals. *See screen shot*.



   - Define matrix, **expected** :=, and press (▦). Select and enter 3 rows, 2 columns.



   - For each cell, press (ctrl) (÷) to get ⌜▫/▫⌟. Enter the elements of the matrix as shown.

List the elements of the expected matrix.

Check the expected counts and verify that the conditions for inference for Chi Square are satisfied. Use supporting statements.

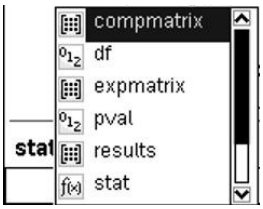
Step 3: Calculate the test statistic and find the p-value.

The χ^2 test uses the following formula: $\sum \frac{(\text{observed} - \text{expected})^2}{\text{expected}}$

1. On the *Calculator* page, press (menu), *Statistics, Stat Tests,* χ^2 2-way.
2. Select **observed** as the observed matrix. Press (enter) and the results are listed.

State the χ^2 statistic. Based on your knowledge of χ^2, what does this number mean?

3. View the component matrix which contains each value of the χ^2 statistic.
   - Type *stat*. and a box will appear. Select *compmatrix*.



State the values in the component matrix. Which elements contributed the most to the χ^2 value? Describe these values in context.

List the p-value calculated.

Step 4: Interpret your results in the context of the problem.

What can you conclude based on the p-value? What does this mean about survival and class of passenger?