

Chapter 12

The Analysis of Categorical Data

12.0 Introduction

The analysis of categorical data involves the use of the Chi square statistic, X^2 . The set up of the data is different than what we have done so far – rather than enter the data in lists, we will enter the data in a different sort of creature, a matrix. You may have studied matrix algebra in other math classes, or otherwise have a familiarity with them. We really will not do any mathematics using matrices, we will only use the matrices for storing original data and expected values from the analysis.

One problem that will come up with this analysis is that depending on the particular TI-83 calculator you have, the **MATRIX** key may be accessed differently. On some TI-83's you will have to access the matrix options by keying in **2nd > Matrix**. Others will only require you to press a **Matrix** key. In our discussion we will use the generic "**Matrix**", and leave it to you to remember that the **2nd** button must be pressed first if that is true for your particular TI-83.

12.1 The Chi Square test

Example 12.7: Risky Behavior

The article “Factors Associated with Sexual Risk-Taking Behaviors Among Adolescents” (*J. Marriage and Family* (1994): 622– 632) examined the relationship between gender and contraceptive use by sexually active teens. Each person in a random sample of sexually active teens was classified according to gender and contraceptive use (with three categories: rarely or never use, use sometimes or most of the time, and always use), resulting in a 3×2 table. Data consistent with percentages given in the article are summarized in Table 12.6.

Table 12.6 *Observed and expected counts for Example 12.7*

Contraceptive Use	Gender		Row Marginal Total
	Female	Male	
Rarely/Never	210	350	560
Sometimes/Most Times	190	320	510
Always	400	530	930
Column Marginal Total	800	1200	2000

The authors were interested in determining whether there is an association between gender and contraceptive use. Using a .05 significance level, we will test

H_0 : Gender and contraceptive use are independent.

H_a : Gender and contraceptive use are not independent.

The data setup

As with our previous inferential procedures, we need to set up the calculations by entering our data. We discussed earlier that the data entry for the X^2 will not be in Lists, but in a matrix. Our first task will be to prepare the matrix for data entry. This may be a bit more complicated than recent data entry, so listen up! The first step is to get to the matrix option screen. Press the following key and we'll look at the resulting screen:

Matrix

NAMES	MATH	EDIT
1: [A]	2x2	
2: [B]	2x2	
3: [C]		
4: [D]		
...		
0: [J]		

The column on the left contains possible names for matrices that the TI-83 has set aside. For our discussion of Chi square we will use only the A and B matrices. You may notice that your screen differs from what you see here. Specifically, you may not see either of those 2x2's, or you might see something else in that part of the screen. The screen as you see it indicates that two matrices, A and B, each with two rows and two columns, are already prepared in the calculator.

This is all well and good, but for our example we need matrices that are 3x2. We will need to reset the dimensions of the A matrix, which will contain our original data. The B matrix will contain the expected values, and will be reset automatically by the calculator. The next step is to prepare a data matrix for the original data and enter the data. Here is the keystroke sequence:

MATRIX > EDIT > [A] > ENTER

Now you should see something like the following screen – yours may have different numbers.

```
MATRIX[A]  2 X 2
[0.0000    0.0000 ]
[0.0000    0.0000 ]
```

Arrow across and down the screen to replace the existing numbers with what we need for our X^2 analysis. When you are finished, the screen – whatever it looks like now – should look like the following:

```
MATRIX[A]  3 X 2
[210.00    350.00 ]
[190.00    320.00 ]
[400.00    530.00 ]
```

After entering the data we are now ready to perform the analysis.

The Chi-square analysis

Now that the data is entered we will test the hypothesis of Independence. The keystroke sequence is:

STAT > TESTS > X^2 -Test...

which will result in a screen to select the correct locations for the data. Choose the following:

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

If you need to change the matrix from what is in this screen, the sequence is: **MATRIX**, followed by pressing **ENTER** when the cursor is at the appropriate matrix.

After your screen looks like the one above, arrow down to `Calculate` and press **ENTER**. You should then see...

```
X2-Test
X2=6.57219
p=.03740
df=2.00000
```

Once again, our calculator results agree with the text except for rounding.

12.2 Afterword

You may have noticed a statistical procedure that is conspicuous in its absence – the Chi square goodness of fit test. There actually is a reason for this. Both the Chi-square tests of independence and homogeneity of proportions are "automatic" procedures, with all calculations determined by the number of rows and columns in the data matrix. The Chi-square Goodness-of-Fit test does not enjoy this luxury, and as a result it cannot be easily programmed.

The Chi-square Goodness-of-Fit test can be performed using Lists, and we refer you to a general discussion of how to do this in POD.