Chi-Square Tests

William H. Knapp III

March 25, 2016

1 When to Use the Chi-Square

If your design compares two or more groups of people and the dependent variable is discrete (i.e. only takes on a few values), the Chi-Square test is probably what you need. For example, imagine that you have an experiment with an experimental group and a control group. Further imagine that the dependent variable is whether or not they agree with some statement. Here you have discrete independent and dependent variables. There are four possible combinations (i.e. control and agree, control and disagree, experimental and agree, and experimental and disagree), but each person only fits into one of those groups (e.g. control and disagree).

2 Displaying Data

When using the Chi-square test, it's typical to create a table that contains the number of observations that fit into each category. One variable's levels occur over different columns and the other variable's levels occur over over different rows. I'll show you how to create a quick table using R (in the RStudio environment, of course) that you can then take the data from and enter into a more nicely formatted table using a word processor (e.g. MS Word) or spreadsheet (e.g. MS Excel).

Before we begin, however, it's important to do a couple of things. First we need to set the current working directory. Remember to do this open up the example or homework file you're working with, then click "Session," then "Set Working Directory," and finally "To Source File Location." Next we need to load up the data file we're going to use.

```
> dat<-read.csv(example4.csv)
```

As usual, I recommend checking out the structure of your data before you begin trying to analyze anything. We can see that the

```
> str(dat)
```

```
'data.frame':        60 obs. of  2 variables:
 $ major : Factor w/ 2 levels "engineering",..: 1 2 1 2 1 2 1 2 1 2 ...
 $ gender: Factor w/ 2 levels "female","male": 1 1 1 1 1 1 1 1 2 1 ...
```

Looking at the structure we can see that we have two variables that are both discrete factor variables with two levels each. We can also see that we have 60 observations for each of these variables.

Now all we need to do is create a table for the data that shows how many people (i.e. frequncy data) fit into each category. So let's create a table using the table() function.

```
> mytable<-table(dat$gender,dat$major)
```

Since we assigned the table to a variable, it didn't show up after the last command. To see the table, all we need to do is ask R to display it by entering mytable.

```
> mytable
```

```
         engineering psychology
  female          11         18
  male            19         12
```

So the big question at this point is whether or not gender and major are independent of one another or related to each other. From looking at the table, there appears to be some relationship. Specifically, there are more male engineering majors and more female psychology majors. But we need to use statistics to back up or refute our intuition. To do this we'll use the chisq.test() function.

```
> chisq.test(dat$gender,dat$major)
```

```
        Pearson's Chi-squared test with Yates' continuity correction

data:  dat$gender and dat$major
X-squared = 2.4027, df = 1, p-value = 0.1211
```

The output from the chi-square test tells us everything we need to know to create our statistical statement. Statistical statements take the following form:

```
TestSymbol(DegreesOfFreedom) = TestValue, p = p-value.
```

The p-value is probably the most important part of any data analysis as it tells you the probability of observing data as or more extreme as what you observed when the null hypothesis is true (e.g. they're unrelated or no different). By convention, we'll use the traditional alpha level of .05. When the p-value is less than or equal to .05, we reject the null hypothesis in favor of the alternative. In other words, you conclude that the data are related or that there is an effect of one variable on the other.

According to the test I performed earlier, gender and major are unrelated $\chi^2(1) = 2.4$, p = .12.