In this section we will review a multiple sample tests, which is called a Chi square test.

>>

Chi square tests are multiple sample tests used when dealing with count or categorical data. For example, supposing that I have sent out a direct mail offer and I want to check is there a difference in the response rate by income category. So if I look at the number of people that have responded to the offer vs the people that haven't responded to the offer based on say low income level, medium income level, and high income level.

Now the number of people is count data and therefore what I will do in that case is use a multiple sample test for difference in response rates across three different income categories and that is the Chi square test. The dependant variable or the outcome variable is a frequency count. There are many applications of Chi square test, two very common ones are test of association and goodness of fit test. But these are essentially Chi square test dealing with count the data.

>>

To understand Chi square test, let's look at a very simple, example. We are going to use a retail example. Let's say that you are working for a large retail chain with many stores. The retail chain wants to understand how it can optimize or improve the utilization of health space in its store.

Supposing we take any one particular category of products that are sold at a store. Let's say carbonated beverages like Coke or Pepsi. How does a retailer decide how much and what brand to display on the shelf. Remember a retailer has fixed shelf space. They have to figure out ok if I have ten brands, how many units of each brand should I put up on the shelf. That is not an arbitrary decision.

Sometimes retailers may want to put products that sell more on the shelf. That way they make sure that people find the product what they are looking from. Sometime they may do it on the basis of the margin. The product may not be the most popular product in that category. But it generates the highest margin. Sometimes a manufacturer may pay a retailer to have additional display for their particular brand on the shelf.

So clearly the retailer has to really think through what the brands to display on the limited shelf space that they have. In other words they want to optimize the shelf space display. If you working on this project for a particular retailer, where your aim is to figure out optimal shelf space utilization across multiple categories. The way you want to do it is let's say you take one category and in a particular store you want to change the display of shelf space utilization and see what impact it has on sales and sort of use at test store to figure out the optimal strategy.

Let's start with one category carbonated beverages. In your population of stores across all

the stores that you have let's say that the transaction share of this particular carbonated beverages category. Let's say there is a brand A that has a 52% share of transactions, Brand B has the 35% share of transactions and Brand C which is may be all other brands included as a 13% share of transactions. Remember this is population behaviour.

You need to pick a test store in which you are going to try all these different things. Let's say you pick one particular store as a test store. Before you go ahead and change things at the store, you want to make sure that this test store is representative of your population of stores. How would you do that? You could check if the store representative or not.

One way to check is take a random sample of transactions and see what behaviour do you see in the sample and it is representative up to population. Let's say that I randomly picked 300 transactions for carbonated beverages at this particular store XX. In that 3000 transactions, I see the brand A has a 177 transactions, brand B has 78 transactions, and brand C is 45 transactions. Out of 300, that is 59% for brand A, 26% percent for brand B, and 15% for brand C. you can see that these shares of transactions are different from the population share. But how different are they? Are they significantly different or is they simply because of random chance variation. The first thing I want to do is establish all the differences in the share of transactions explainable by random chance variation or should you conclude that this

store is not representative and therefore you should pick some other store for your test. So we want to do a hypothesis test. Because this is number of transactions count to data across three different samples. Brand A is one sample, brand B is another sample, and brand C is the third sample. We will use a Chi square test.

>>

How does a Chi square test work? Essentially we want to say if this sample was representative of the population I would have expected to see a certain number of transactions for A, B, and C. what are those transactions? If this store was exactly like the population, then remember brand A had a 52% share of transactions in the population. So out of 300 transactions, I would have seen 52% of 300, which is 156 transactions for brand A if the store was representative.

Notice in this table here that there is a row for observed values and expected values. Observed is what we seen in the sample, expected is what we should have seen in the sample if the sample was like the population. So for brand A, I should have expected to see 52% share out of 300, 156 transactions. For brand B I should have expected to see 35% share. Therefore 35% of 300, 105 transactions, and for brand C, I should have expected to see 13% share of transactions, which is 39.

These are my observed values what I am really see in the sample. Below that are my expected values

what I should have seen if the sample was representative of the population. Our job is to figure out are these differences statistically significant or not and we can do that by calculating a Chi square test statistic. Mathematically we can prove that

$$\sum \frac{(f_o - f_e)^2}{f_e}$$

Follows what is called a Chi square distribution.

>>

What is Chi square distribution? It is a continuous distribution. It is an asymmetric distribution that depends on sample size. Chi square distribution are generated as the square of standard normal distributions and the larger the sample size, the more the Chi square tends to normal. The image here shows you Chi square distributions at different sample sizes and you can see as sample size increases, Chi square tends to normal.

>>

Mathematically we can prove that

$$\chi^2 = \sum \frac{(f_o - f_e)^2}{f_e}$$

Will follow a Chi square distribution with k-1 degrees of freedom and k is the number of samples or cells we have? So in our particular example, we can actually calculate this value.
(177-156)^2/156, for brand A,
(78-105)^2/105, for brand B, and

(45-39)^2/39, for brand C. that gives us a Chi square statistic of 10.69.

Our degrees of freedom is number of cells – 1, which is 2.

>>

We simply need to look up a Chi square table. At the 5% level of significance, the critical value for two degrees of freedom is 5.99. Remember our calculated test statistic is 10.69. Because the calculated test statistic in terms of distance is further away from the mean than the critical value of 5.99, we will reject the null hypothesis. The sample is not representative of the population of stores.

>>

Of course we can also directly do this using a tool like excel. In excel we want to use the CHITEST function and the CHITEST function asks for observed values and expected values. So this is my data and I want to figure out that there is a difference. I simply say that CHITEST and notice that excel is asking for an actual range and expected range. Actual range is the observed values and expected range is what we should have expected if the null hypothesis was true. So now if I enter this, I will now get a p value of 0.00476.

Therefore because this p value is lot lower than my level of significance α 5%, I will reject the null hypothesis that this sample is representative of

the population stores. In other words this stores behaviour is not reflective of or not representative of the population stores. There is something in the store that makes it different in behaviour from all the other stores.

>>

Therefore in this particular example, I should pick some other store is run my test. Because this store is not really representative of the population stores. So that is how a Chi square test works. Remember it relies on observed values and expected values. Let's look now at a slightly more complex example.

>>