Chi-Square Statistic

A multinomial population is a population with a single characteristic of interest but more than two possible results.

For example we can view an election as

Binomial

$p =$ probability of voting liberal

$q =$ probability of not voting liberal

Multinomial

$p_1 =$ probability of voting NDP

$p_2 =$ probability of voting liberal

$p_3 =$ probability of voting conservative

$p_4 =$ probability of voting for another party

Suppose we wanted to know, out of 1000 voters, how many would vote for each of these parties. We might have some expectations based on the previous election. Let's say we expect, out of 1000 voters we expect e_1, e_2, e_3 and e_4 people to vote for the NDP, Liberal, Conservative and other parties respectively.

After the election we take a random sample of 1000 voters and ask them who they voted for. It turns out that f_1, f_2, f_3 and f_4 people voted for the NDP, Liberal, Conservative and other parties respectively.

How do we measure how far off we were with our expectations? One way is to use the **chi-square statistic**

$$\chi^2 = \frac{(f_1 - e_1)^2}{e_1} + \frac{(f_2 - e_2)^2}{e_2} + \frac{(f_3 - e_3)^2}{e_3} + \ldots + \frac{(f_n - e_n)^2}{e_n}$$

$$= \sum \frac{(f_i - e_i)^2}{e_i} \qquad \text{sometimes written as} \qquad \sum \frac{O_i - E_i}{E_i}$$

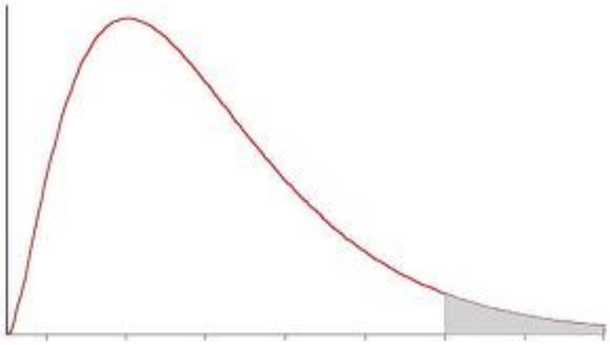
(in our case n=4)

Notice that the numerator is a positive number measuring how far off the observed frequency f_i is from our expected frequency e_i.

The denominator puts the size of the numerator into perspective. For example, if $f_i - e_i$ = 10 from $f_i = 110$ and $e_i = 100$ then this is quite different then if $f_i = 15$ and $e_i = 5$.

The chi-square distributions are a family of distributions each one depending on the degrees of freedom, like the t-distributions.

Chi-Square distribution



How does this distribution work?

Let's suppose for a second that the *actual* population proportions for k categories are $p_1, p_2, p_3, \ldots, p_k$. This means that, for a sample size n, we should expect frequencies $e_i = n \cdot p_i$. It is unlikely that for a sample of size n, that we get frequencies f_i that are far away from e_i. If we do, it will result in a large χ^2 value. The chi-square distribution tells us how likely we are to get a sample that results in a χ^2 value at least as large as the one we got.

(notice that the chi-square values start at 0 since they are always non-negative and only 0 if all of our observed frequencies are the same as our expected frequencies)

Note: Our calculated value for χ^2 will have sampling distribution that can be approximated by the chi-square probability distribution as long as all expected frequencies are greater than or equal to 5.

Remember, the distribution depends on the degrees of freedom $df = k - 1$ where k is the number of categories (or cells, as they are called in the text).

Inferences Concerning Multinomial Experiments

Suppose we want to test a die (at $\alpha = 0.05$) and decide whether to reject or fail to reject the claim "this die is fair." The die is rolled 60 times with the following observed frequencies.

Number	1	2	3	4	5	6
Observed frequency	7	12	10	12	8	11

This was an example of a **multinomial experiment**. A multinomial experiment has the following characteristics:

1)

2)

3)

4)

For multinomial experiments we will always use a one-tailed critical region (on the right side of the chi-square distribution.

Example: Suppose a poll in a particular riding predicted the following election results:

Liberal 35%

Conservative 25%

NDP 15%

BQ 20%

Other 5%

Is this claim supported by the following random sample of 1000 voters? (Test at 5% significance).

Liberal 323

Conservative 271

NDP 137

BQ 206

Other 63

Example: A manager believes that production levels are the same in each of the 6 factories in her district. Does the following random sample support her claim? Test at $\alpha = 0.01$.

Factory	1	2	3	4	5	6
Production	38	34	22	20	36	21