What is Chi Square Analysis? Why is it Important?

The chi-square (x^2) test, also known as the "goodness-of-fit" test, was developed to compare the experimental observation with the expected outcome. Chi-square analysis is typically used to determine whether frequencies of the data observed in a study are due to chance or if there is a relationship between the variables.

The possible outcomes from the chi-square test are:

- A small chi-square value which means that the observed data fits the expected data hence, there is likely to be a relationship between the variables
- A large chi-square value which means that the observed data does not fit the expected data hence, there is not likely to be a relationship between the variables

Below is the equation used for chi-square analysis:

$$x^2 = \sum \frac{(O-E)^2}{E}$$

x^2 = chi-square value
O = observed values
E = expected values
$\sum$ = the 'sum of'

The interpretation of the chi-square test results is based on a probability (P) value. The P value represents whether the test results are significant or not. The P value depends on two factors:

1) x^2 value
2) Degree of freedom (df)

The degree of freedom represents the number of independent values that are free to vary. The calculation for the degree of freedom is:

Number of outcome classes (n) - 1

The x^2 value and degree of freedom are used to determine the P value, which determines whether the null, or chance, hypothesis can be rejected.

The null hypothesis states that there is no relationship between the two variables being studied. If the P value is lower than 0.05, that means there is less than a 5% probability that if the results were random, the observed data would be different. The null hypothesis would be rejected in this case, because the results are likely due to chance. If the P value is greater than 0.05, the

null hypothesis cannot be rejected and the results are not random. Note that the null hypothesis cannot be accepted, only rejected or failed to be rejected.

This information is obtained using a chi-square table. In a chi-square table, the X^2 values are in the body of the table, the df values are on the far left column, and the P values are on the top row. Below is an example of the first five rows of a chi-square table:

Degree of freedom

Probability level (alpha)

df	0.5	0.10	0.05	0.02	0.01	0.001	P value
1	0.455	2.706	3.841	5.412	6.635	10.827	
2	1.386	4.605	5.991	7.824	9.210	13.815	
3	2.366	6.251	7.815	9.837	11.345	16.268	
4	3.357	7.779	9.488	11.668	13.277	18.465	
5	4.351	9.236	11.070	13.388	15.086	20.517	

Chi-square values (X²)

It is important to know that a high P value does not mean that the hypothesis of the experiment is correct. The P value is only a method used to make reasonable decisions based on the probability that the results are significant.

Below is an example of a chi-square analysis done on the influence of gender on voting behaviour based on the results of the 2001 Senate election in Vermont State.

2001 Vermont State Senate Election Results

		Vote		
		Yes	No	Total
Women		35 (28.83)	9 (15.17)	44
Men		60 (66.17)	41 (34.83)	101
Total		95	50	145

Chi-square Calculation for the Relationship Between Gender and Voting Behaviour

$$x^2 = \sum\sum \frac{(O_{ij} - E_{ij})^2}{E_{ij}} = \frac{(35 - 28.83)^2}{28.83} + \cdots + \frac{(41 - 34.83)^2}{34.83} = 5.50$$

The values inside brackets represent the expected values and the values outside of the brackets represent the actual values. Note that in the equation above, the i and j index represented the rows and columns of the table.

Listed below is a summary of the steps for finding the P value for the exemplary study:

1. Using the observed and expected data, the value of X^2 was calculated to be 5.50
2. The degree of freedom is 1 because there are two independent variables that can vary
3. Using a chi-square table and the chi-square (X^2) and df values, the probability was found to be 0.019
4. The P value was interpreted to develop a conclusion, in this case the P value is less than 0.05

Therefore, the null hypothesis was rejected and it was concluded that gender did not have an impact on one's voting behaviour. In other words, the two variables were independent of each other.

Sources:
https://www.researchgate.net/figure/Excerpts-from-the-Chi-square-distribution-table_tbl1_277935900

https://www.uvm.edu/~statdhtx/StatPages/R/Chi-Square-Folder/Chi%20square%20test%20analysis%20of%20contingency%20tables_David_Howell%20.pdf