# Chi-square Analysis

The Chi-square is a statistical test that makes a comparison between the data collected in an experiment versus the data you expected to find. In the case of genetics (and coin tosses) the expected result can be calculated using the Laws of Probability (and possibly the help of a Punnett square).

Variability is always present in the real world. The Chi-square test is a way to evaluate this variability to get an idea if the difference between real and expected results is due to normal random chance, or if there is some other factor involved.

Genetics uses the Chi-square to evaluate data from experimental crosses to determine if the assumed genetic explanation is supported by the data. The Chi-square test helps you to decide is the difference between your observed results and your expected results is probably die to random chance alone, or if there is some other factor influencing the results.

Chi-square can also be applied to situations where organism choice or selection is involved. Hypothesizing that equal outcomes are expected and then comparing that to the results will tell you if, for example, natural selection is occurring or organism preference or choice is occurring.

The Chi-square test will not, in fact, prove or disprove if random chance is the only thing causing observed differences, but it will give an estimate of the likelihood that chance alone is at work.

**Determining the Chi-square Value**

Chi-square is calculated based on the formula below:

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

*O = the frequencies observed*

*E = the frequencies expected*

$\sum$ *= the 'sum of'*

http://geography-site.co.uk/pages/skills/fieldwork/stats/chi.html

**Determining the Null Hypothesis**

A null hypothesis states that no statistical significance exists between the observed and expected results. It attempts to show there is no variation between two sets of numbers. It is presumed true until statistical evidence nullifies it for an alternative hypothesis.

In genetics, the null hypothesis includes a statement about the expected genetic ratio and the type of inheritance presumed to be acting on the results. In other areas of science, the null hypothesis is used to show no difference between two sets of data or equal distribution of organisms.

## Interpreting the Chi-square Value

When looking at the Chi-square distribution table, you must first know how to determine the degrees of freedom. This will then be used with the Chi-square value to determine the probability value (p-value).

## Degrees of Freedom

The rows in the Chi-square distribution table refer to the degrees of freedom. The degrees of freedom are calculated as one less than the number of possible results in your experiment. In a sense degrees of freedom is measuring how many classes of results can "freely" vary their numbers. In other words, how many group's data do you have to know (with the total given) in order to know the numbers for all the groups.

For example, if you have the total number observed and there were 3 groups, then you have to know the data for 2 groups to be able to calculate all 3 group's numbers. So, the degrees of freedom would be 2.

## Probability = p

The columns in the Chi-square distribution table with the decimals from 0.99 to 0.01 refer to probability levels of the Chi-square. To find the p-value, find the row of degrees of freedom. Then, find the number in the chart that is the closest to the calculated chi-square value. Follow up the column to the p-value.

The larger the chi-square value, the larger the difference between expected and observed and therefore the smaller the p-value. If a p-value of 0.05 is determined it means the variance between our observed results and our expected results would occur from random chance alone only about 5% of the time. Therefore, we would conclude that chance factors alone are not likely to be the cause of this variance. Some other factor is causing the difference.

Biologists generally accept p=0.05 as the cutoff for accepting and rejecting a null hypothesis. If the difference between your observed data and your expected data would occur due to chance alone fewer than 1 time in 20 (p=0.05) then the acceptability of your null hypothesis may be questioned. Biologists consider a p-value of 0.05 or less to be a "statistically significant" difference.

## Coin Data: Chi Square Analysis

If 2 coins are flipped 100 times and there results tabulated, then a chi-square test can be done to see if the expected values are close to the observed values.

| | A | B | C | D | E |
|---|---|---|---|---|---|
| | Obs | Exp | Obs - Exp | (Obs - Exp)$^2$ | $\frac{(Obs - Exp)^2}{Exp}$ |
| H/H | 22 | | | | |
| H/T | 48 | | | | |
| T/T | 30 | | | | |
| | | | | $\chi^2$ Total | |
| | | | | Degrees of Freedom | |
| | | | | p-value | |

Null Hypothesis:

Is p-value accepted or rejected? Why?

## Chi-square Distribution Table

| Degrees of freedom | Probability (p) value | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | 0.99 | 0.95 | 0.80 | 0.70 | 0.50 | 0.30 | 0.20 | 0.10 | 0.05 | 0.01 |
| 1 | 0.001 | 0.004 | 0.06 | 0.15 | 0.46 | 1.07 | 1.64 | 2.71 | 3.84 | 6.64 |
| 2 | 0.20 | 0.10 | 0.45 | 0.71 | 1.30 | 2.41 | 3.22 | 4.60 | 5.99 | 9.21 |
| 3 | 0.12 | 0.35 | 1.00 | 1.42 | 2.37 | 3.67 | 4.64 | 6.25 | 7.82 | 11.34 |
| 4 | 0.30 | 0.71 | 1.65 | 2.20 | 3.36 | 4.88 | 5.99 | 7.78 | 9.49 | 13.28 |
| 5 | 0.55 | 1.14 | 2.34 | 3.00 | 4.35 | 6.06 | 7.29 | 9.24 | 11.07 | 15.09 |
| 6 | 0.87 | 1.64 | 3.07 | 3.38 | 5.35 | 7.23 | 8.56 | 10.65 | 12.59 | 16.81 |
| 7 | 1.24 | 2.17 | 3.84 | 4.67 | 6.35 | 8.38 | 9.80 | 12.02 | 14.07 | 18.48 |