The Chi-Square Test

Objectives

- To calculate and use the Chi-Square Test to determine whether a given set of data approximates a theoretically expected ratio.
- To interpret a calculated χ^2 value, given the appropriate number of degrees of freedom and a table of χ^2 values.

Introduction

The analysis of the inheritance patterns is often dependent on a particular ratio, for example 3:1, being observed in the F_2 generation. A statistical test known as the **Chi-Square Test** (χ^2) is frequently used to determine whether experimentally obtained data constitutes a good fit (or a satisfactory approximation) to the expected data. That is, the χ^2 test enables one to determine if it is reasonable to attribute deviations from a perfect fit to chance. If the deviations are small, they can be more reasonably attributed to chance than if they are large. The question the χ^2 test answers is "How small must the deviations be to be attributed to chance alone?". The formula for the χ^2 test is as follows:

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

where O = the number of observed individuals of a particular phenotype and E = the number of expected individuals of that phenotype. Σ = the summation of all possible values of $(O - E)^2/E$ for the various phenotypic categories.

For example, suppose that you suspect that the allele for wild eye color which gives red eyes (+) is dominant to the allele for apricot eyes (a). A cross of heterozygous F_1 parents yielded 102 red eyed flies and 44 apricot eyed flies in the F_2 generation. Is this a good fit to the expected 3:1 ratio? To answer this question, χ^2 was calculated. The calculations are summarized in Table 1:

Table 1.

Phenotype	Genotype	O	E	(O-E)	$(O-E)^2$	$(O-E)^2/E$
Red	++ or +a	102	109.5	-7.5	56.25	0.5137
Apricot	aa	44	36.5	7.5	56.25	1.5411
	Totals	146	146.0			χ^2 = 2.0548

What does the χ^2 value of 2.0548 mean? If the observed numbers (O) were exactly equal to the theoretically expected numbers (E), the fit would be perfect and χ^2 would be zero. Thus, a small value for χ^2 indicates that the observed and expected numbers are in close agreement.

A large value indicates marked deviation from the expected values. Because chance deviations from the theoretical values are expected to occur, the question to be answered is "Are the observed deviations within the limits expected by chance?" Generally, statisticians have agreed on the arbitrary limits of odds of 1 chance in 20 (probability = .05) for the drawing the line between acceptance and rejection of the data as a satisfactory fit to the expected ratio.

A table of χ^2 values is given in Table 2. These values are the standards to which the calculated χ^2 values are compared. In order to use the table, the **degrees of freedom** (df) must be calculated. The degrees of freedom is one less than the number of categories used in the χ^2 calculation. In our example, the number of categories (phenotypes) was two. Therefore, the degrees of freedom is: 2 - 1 or 1. The value of χ^2 at df = 1 and .05 probability is 3.841. This is the maximum value of χ^2 that we can accept and still attribute that the deviations observed are due to chance alone.

The comparison of the calculated χ^2 value from the data and the table χ^2 value can be summarized as follows:

Accept the fit if: calculated χ^2 value is $\leq$ table χ^2 value

Reject the fit if: calculated χ^2 value is $>$ table χ^2 value

In our example, our calculated χ^2 of 2.0548 is less than the table χ^2 of 3.841. We would then accept the hypothesis that the observed data are a good fit to the expected data. The F_2 generation does show a 3:1 ratio of red eyes to apricot eyes.

Table 2.

Degrees of Freedom (df)	P = .05
1	3.841
2	5.991
3	7.815
4	9.488
5	11.070
6	12.592