Using the Chi-Square Test for Statistical Analysis of Experimental Data

Example 1

Statistics can be used to determine if differences among groups are significant, or simply the result of predictable error. The statistical test most frequently used to determine whether data obtained experimentally provide a good fit or approximation to the expected or theoretical data is the chi-square test. This test can be used to determine if deviations from the expected values are due to chance alone, or to some other circumstance. For example, consider corn seedlings resulting from an F_1 cross between parents that are heterozygous for color.

A Punnett square of the F_1 cross **Gg X Gg** would predict that the expected proportion of green:albino seedlings would be 3:1. Use this information to fill in the Expected (e) column and the (o−e) column in Table 7.3.

Table 7.3

Phenotype	Genotype	# Observed (o)	# Expected (e)	(o−e)
Green	GG or Gg	72		
Albino	gg	12		
	Total:	84		

There is a small difference between the observed and expected results, but are these data close enough that the difference can be explained by random chance or variation in the sample?

To determine if the observed data fall within acceptable limits, a chi-square analysis is performed to test the validity of a **null hypothesis** (that there is no statistically significant difference between the observed and expected data). If the chi-square analysis indicates that the data vary too much from the expected 3:1, an **alternative hypothesis** is accepted.

The formula for chi-square is:

$$\chi_2 = \Sigma \frac{(o-e)^2}{e}$$

where **o** = **observed** number of individuals

e = **expected** number of individuals

Σ = the **sum of the values** (in this case, the differences, squared, divided by the number expected)

1. This statistical test will examine the null hypothesis, which predicts that the data from the experimental cross above will be expected to fit the 3:1 ratio.
2. Use the data from Table 7.3 to complete Table 7.4.

Table 7.4

Phenotype	# Observed (o)	# Expected (e)	(o−e)	$(o-e)^2$	$\frac{(o-e)^2}{e}$
Green	72				
Albino	12				
				$\chi^2 = \Sigma \frac{(o-e)^2}{e} =$	

3. Your calculations should give you a value for $\chi^2 = 5.14$. This value is then compared to Table 7.5.

Table 7.5: Critical Values of the Chi-Square Distribution

	Degrees of Freedom (df)				
Probability (p)	1	2	3	4	5
0.05	3.84	5.99	7.82	9.49	11.1
0.01	6.64	9.21	11.3	13.2	15.1
0.001	10.8	13.8	16.3	18.5	20.5

How to Use the Critical Values Table

1. Determine the **degrees of freedom (df)** for your experiment. It is the number of phenotypic classes minus 1. Since there are two possible genotypes, for this experiment df = 1 (2 samples –1). If the experiment had gathered data for a dihybrid cross, there would be four possible phenotypes and therefore 3 degrees of freedom.

2. Find the p value. Under the 1 df column, find the critical value in the probability (p) = 0.05 row: it is 3.84. What does this mean? **If the calculated chi-square value is greater than or equal to the critical value from the table, then the null hypothesis is rejected.** Since for our example $\chi^2 = 5.14$ and $5.14 > 3.84$, we reject our null hypothesis that there is no statistically significant difference between the observed and expected data. In other words, chance alone cannot explain the deviations we observed and there is, therefore, reason to doubt our original hypothesis (or to question our data collection accuracy.) The minimum probability for rejecting a null hypothesis in the sciences is generally 0.05, so this is the row to use in the chi-square table.

3. These results are said to be significant at a probability of $p = 0.05$. This means that only 5% of the time would you expect to see similar data if the null hypothesis was correct; thus, you are 95% sure that the data do not fit a 3:1 ratio.

4. Since these data do not fit the expected 3:1 ratio, you must consider reasons for this variation. Additional experimentation would be necessary. Perhaps the sample size is too small, or errors were made in data collection. In this example, perhaps the albino seedlings are underrepresented because they died before the counting was performed.

Example 2

In a study of incomplete dominance in tobacco seedlings, the counts in Table 7.6 were made from a cross between two heterozygous (Gg) plants:

Table 7.6

Phenotype	Genotype	# Observed (0)
Green	GG	22
Yellow Green	Gg	50
Albino	gg	12
	Total:	84

A Punnett square for this cross indicates that the expected counts should be in a 1 green:2 yellow green:1 albino ratio (Table 7.7). The expected values for a total count of 84 organisms are therefore:

1 green	= $\frac{1}{4} \times 84 = 21$
2 yellow green	= $\frac{1}{2} \times 84 = 42$
1 yellow	= $\frac{1}{4} \times 84 = 21$
	84

Table 7.7

Phenotype	# Observed (o)	# Expected (e)	(o−e)	$(o-e)^2$	$\frac{(o-e)^2}{e}$
Green	22	21	1	1	0.05
Yellow Green	50	42	8	64	1.52
Albino	12	21	9	81	3.86
				$\chi^2 = \Sigma\frac{(o-e)^2}{e} =$	5.43

Go to the chi-square table, this time for two degrees of freedom (there are three phenotypes: 3 – 1 = 2 df). If the χ^2 value were greater than or equal to the critical value of 5.99 we would reject our hypothesis. Since 5.43 is less than the critical value at $p = .05$, we accept the null hypothesis (this second data set does fit the expected 1:2:1 ratio).

Practice Problem

An investigator observes that when pure-breeding, long-wing *Drosophila* are mated with pure-breeding, short-wing flies, the F_1 offspring have an intermediate wing length.

When several intermediate-wing-length flies are allowed to interbreed the following results are obtained:

Observed

230 long wings
510 intermediate-length wings
260 short wings

a. What is the genotype of the F_1 intermediate-wing-length flies? ______________

b. Write a hypothesis describing the mode of inheritance of wing length in *Drosophila* (this is your null hypothesis).

____

____

____

c. Complete Table 7.8.

Table 7.8

Phenotype	# Observed (o)	# Expected (e)	(o − e)	$(o - e)^2$	$\frac{(o-e)^2}{e}$
				$\chi^2 = \Sigma\frac{(o-e)^2}{e} =$	

(i) Calculate the chi-square value for these data.

1. How many degrees of freedom (df) are there? ___________

2. χ^2 (chi-square) = ___________

3. Referring to the critical values chart, what is the probability value for these data?

____

(ii) According to the critical value of χ^2, can you accept or reject the null hypothesis? Explain why.

____

____

____

____

____