

4. Chi-square statistics and Probability

For determining chi-square or goodness of fit, the size of the population must be considered. Suppose in one cross of tall and dwarf plants, out of the 100 plants of F_2 , 70 are tall and 30 dwarf instead of 75 and 25 as expected from a 3: 1 ratio. Obviously, there is a deviation of 5 from the normal. In a second cross where 1000 plants of F_2 progeny showed the same numerical deviation of 5, there were 745 tall plants and 255 dwarf ones. In the third cross, the 1000 plants of F_2 generation appeared in the same proportion as in the first cross (70: 30 or 7: 3) so that 700 plants were tall and 300 dwarf. The chi-square formula given below will show whether the differences in the observed results of the three crosses are significantly different from the 3: 1 ratio or not (See Table)

Thus, chi square is defined as the sum of the quotients obtained from division of squares of deviations of observed frequencies from the expected frequencies of various classes with the respective expected frequencies. This may have expressed as follows:

$$\chi^2 = \sum \frac{(o-e)^2}{e}$$

Where, Σ = summation, O = observed frequency of a class, and E = Expected frequency of the class.

No.of plants in F_2 generation	Cross no. 1 100		Cross no. 2 1000		Cross no. 3 1000	
	Tall	dwarf	Tall	dwarf	Tall	dwarf
Observed	70	30	745	255	700	300
Expected (e)	75	25	750	250	750	250
Deviation (d)	5	5	5	5	50	50
d^2	25	25	25	25	2500	2500
d^2/e	0.33	1	0.033	1	3.3	10
$\chi^2 = \Sigma (d^2/e)$	1+0.33 = 1.33		0.033 + 1 = 1.033		3.3+10 = 13.3	

The smaller the chi-square value the more likely it is that deviation has occurred due to chance. Now to find out whether the differences between expected and observed results are due to chance alone or not, we must be familiar with two more concepts namely

degree of freedom and level of significance.

Degrees of Freedom: The number of degrees of freedom is calculated as the number of classes whose value is required to describe the outcome from all classes. The concept of degrees of freedom is important in experiments and genetic ratios because one must consider the total number of observed individuals in the experiment as a fixed or given quantity. This fixed quantity is composed of one or more classes some of which are variable. In the experiment between tall and dwarf pea plants there are only two classes, tall and dwarf. As soon as the number of one class is set, the other can be determined. Thus, when two classes are scored, there is one degree of freedom. In an experiment where three classes are scored, there are two degrees of freedom, and so on. The rule states that for the kind of genetic experiments described, the degrees of freedom are equal to one less than the number of classes.

Level of Significance: In the experiment described the actual ratio departs from that which is expected. We must now determine how significant this discrepancy is so that we can decide to accept or reject the results. Small discrepancies are not significant; large discrepancies are significant and lead to rejection of a result or hypothesis. Therefore, values are assigned to these two kinds of discrepancies—the large discrepancies are the largest 5% and small discrepancies are remaining 95%. On this basis if the discrepancy lies in the large class it is significant and the result may be discarded. The 5% frequency value that enables us to reject the result is called the *5% level of significance*. The level of significance can be changed. If 5% is too high we can decide on a low level of significance say 1%. In this case it is not so easy to reject a result. Contrarily, if we decide on a high level of significance say 10%, it is easier to reject a result. Usually the accepted level of significance is between the two extremes, that is 5%.

Chi-square: After determining the degrees of freedom in an experiment and deciding on the level of significance, the actual size of the discrepancy between expected and

observed is found by chi-square. Statisticians have prepared tables that relate the number of degrees of freedom with the probability that particular groups of chi-square values will be found (Table).

We can now examine the results of the experiment described in Table 1. The chi-square values of the first two crosses are 1.33 and 0.133. Both are acceptable discrepancies because these values are smaller than the chi-square value for one degree of freedom given as 3.84 in Table 2. The results of the first two crosses therefore may be considered to be consistent with Mendel's hypothesis, the difference between expected and observed being due to chance.

TABLE 2. Values of chi-square corresponding to 5% and 1% levels of significance and up to 10 degrees of freedom

Degrees of freedom	Level of significance	
	5%	1%
1	3.84	6.64
2	5.99	9.21
3	7.82	11.34
4	9.49	13.28
5	11.07	15.09
6	12.59	16.81
7	14.07	18.48
8	15.51	20.09
9	16.92	21.67
10	18.31	23.21

Probability:

In the experiments described above the *F2* generation segregates in a particular ratio. The numbers of individuals in each ratio result from *chance* segregation of genes during gamete formation, and their chance combinations to form zygotes. Since these are chance events, accurate predictions about the results cannot be made. This is especially true in cases where the progeny is limited to a small number such as in experiments in animal breeding and even more so in human pedigree studies. It is hardly possible to predict the appearance of a certain phenotype or genotype. But what we can say is that there is a certain probability of occurrence of a given genetic event.

In a general way, we can say that the probability or chance that an event will occur can be defined as the proportion of times in which that event occurs in a very large number of trials. If there are x trials and the event occurs on the average y times during the x trials, then the probability is expressed as y/x . The probability would be between zero and one. The closer the probability is to zero, the less chance is there for the event to occur. When probability is one, the event is certain to occur.

Rules of probability:

The multiplication rule: While considering probabilities it is important to note that the inheritance of certain genes such as for height of pea plants (tall or dwarf), flower colour (red or white), seed texture (round or wrinkled) are independent events. If we consider each gene separately, the probability of any one F_2 plant being tall is $3/4$, and of its being dwarf $1/4$.

Similarly, there is a probability of $3/4$ that an F_2 plant will bear red flowers and $1/4$ that the flower will be white. Now what is the probability that an F_2 plant would be both tall and coloured? Assuming that the inheritance of each gene is an independent event, the probability that a plant be coloured and tall is equal to the product of their individual probabilities, *i.e.*, $3/4 \times 3/4 = 9/16$. Thus there are 9 chances out of 16 that an F_2 plant be coloured and tall. Stated in a general way, when the probability of an event is independent of that of another event, and the occurrence of one does not influence the occurrence of the other, probability that both events will occur together is the product of their individual probabilities.

The Addition Rule: The probability that *one* of several mutually exclusive events will occur is the *sum* of their individual probabilities. This law is applicable when different types of events cannot occur together. If one occurs the other is *excluded*. When a coin is tossed there are two alternatives, either heads or tails will appear. If the probability for heads is $1/2$, for tails $1/2$, then the probability that either heads or tails will appear is $1/2 + 1/2 = 1$.

DOMINANCE RELATIONSHIPS

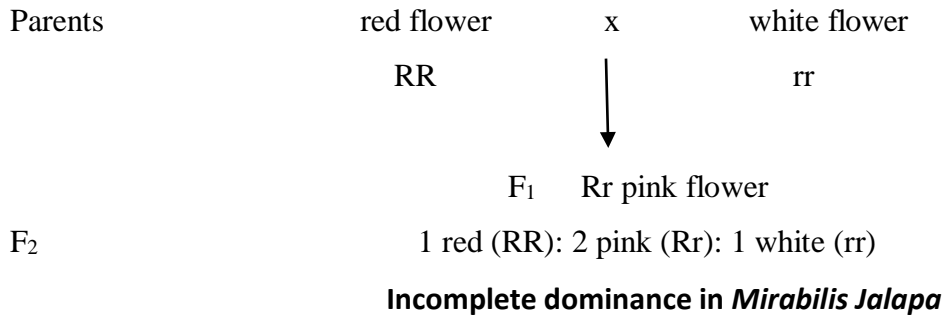
i) Complete dominance:

In case of complete dominance, the phenotype produced by heterozygotes is identical with that produced by homozygotes for the concerned dominant allele. The dominant allele in such a situation is known as completely or fully dominant. An example of

complete dominance occurs for seed shape in peas. Round seed shape is produced by the dominant allele W, while wrinkled shape is determined by its recessive allele w. Seeds having the genotype Ww are round and indistinguishable from those having the genotype WW. As a result, characters showing complete dominance yield the typical 3:1 monohybrid ratio in F₂.

ii) Incomplete dominance:

Mendel always observed complete dominance of one allele over the other for all the seven characters which he studied in garden pea. Later on, cases of incomplete dominance were reported. For example, in four ‘O’ clock plant (*Mirabilis Jalapa*) there are two types of flowers, viz., red and white. A cross between red and white flowered plants produced plants with intermediate flower colour, i.e., pink colour in F₁ and a modified ratio of 1 red: 2 pink: 1 white was observed in F₂.



iii) Codominance:

In case of co dominance both alleles express their phenotypes in heterozygote. The example is AB blood group in human. The people who have blood type AB are heterozygous exhibiting phenotypes for both the I^A and I^B alleles. In other words, heterozygotes for co- dominant alleles are phenotypically similar to both parental types. The main difference between co -dominance and incomplete dominance lies in

the way in which genes act. In case of co- dominance, both alleles are active, while in case of incomplete dominance only one allele (dominant) is active.

iv) Overdominance:

In case of some genes, the intensity of character expression is greater in the heterozygotes than in the two concerned homozygotes. This situation is known as overdominance. True overdominance is known in the case of some genes. Many cases of presumed overdominance may actually be cases of repulsion phase linkage and/or epistasis. The white eye (*w*) gene of *Drosophila* exhibits overdominance for some of the pigments, e.g., sepiapteridine and himmelblaus. Allele *w* produces white eye in homozygous state (*ww*), while its completely dominant allele *W* gives rise to the normal dull red eye colour. Eye pigments sepiapteridine and himmelblaus are present in low concentrations in *ww* homozygotes, while *WW* homozygotes have relatively higher concentrations of these pigments. However, flies heterozygous for this gene (*Ww*) have an appreciably higher concentration of these two pigments than the two homozygotes (*WW* and *ww*). Genes showing overdominance would also give rise to 1:2:1 ratio in F_2 .