

(A) Analysis of Seed Samples by Chi-square Method

Application of statistical methods in biological experiments is called *Biometry*. A very important use of biometry in Genetics is for testing whether an observed ratio may be taken as fit for an expected ratio. For examples, the expected F_2 phenotypic monohybrid ratio is 3 : 1. But the actual observed ratio as obtained by Mendel himself for the Tall X Dwarf monohybrid cross in pea plant was 2.84 : 1 and for the coloured and white seed coat cross was 3.15 : 1. Some deviations between observed and expected ratios is always expected while experimenting with living organisms due to accidents and chance. The problem is — how much deviation can be allowed due to these reasons, or, in other words, whether 2.84 : 1 and 3.15 : 1 can be taken as equivalent to 3 : 1? This can be found out by a process called *Testing Goodness of Fit* and there are two methods of finding it out : (a) *The Standard Error Method* and (b) *The Chi-square (χ^2) Method*.

(B) Probability

In order to understand the application of the Chi-square method for testing goodness of fit of a sample to a particular ratio, one must have a clear conception about the *Law of Probability*. The Law of Probability enables us to forecast the frequency of occurrence of a particular event when 2 or more alternative results are possible. As for example, if a coin is tossed the result is either 'head' or 'tail', both the events having an equal chance of occurrence. Thus, if the coin is tossed 10,000 times, there will be near about 5,000 heads and 5,000 tails, i.e. the two events 'head' and 'tail' will occur in 1 : 1 ratio. Accuracy in Biometry depends upon the size of the sample, which should be sufficiently large. Thus, if the coin is tossed 5, 10 or 50 times the result may not be exactly in 1 : 1 ratio.

If one takes two coins and tosses them simultaneously for a large number of times, the result will be — two Heads : one Head and one Tail : Two Tails in the ratio 1 : 2 : 1. Probability in these cases is given by the coefficients of expansion of the binomial formula $(a + b)^n$, where 'n' represents the number of objects. So, if four coins are tossed simultaneously, the probabilities can be calculated as — $(a + b)^n$, i.e.

$(a + b)^4$, i.e., $a^4 + 4a^3b + 6a^2b^2 + 4ab^3 + b^4$. Therefore, the proportion of 4 Heads : 3 Heads and 1 Tail : 2 Heads and 2 Tails : 1 Head and 3 Tails : 4 Tails will be — 1 : 4 : 6 : 4 : 1.

The Chi-square (χ^2) Method

This method, named after the Greek letter χ (Chi), is more easily — and, therefore, more frequently — applicable to a number of biological problems. The formula used in this method is :

$$\chi^2 = \sum \frac{d^2}{e} \text{ or } \sum \frac{(O - e)^2}{e}$$
 which is a summation of the square of deviation in each class (d^2), i.e., $(O - e)^2$ where O = the observed number and e = the expected number in each class and is divided by the expected number (e) for each class.

For example, let us consider Mendel's Tall X Dwarf cross with pea plant. The observed numbers of plants in F_2 were 787 tall and 277 dwarf. For an exact 3 : 1 ratio, the expected number should have been 798 tall and 266 dwarf. Now, considering the two classes tall and dwarf

$$\begin{aligned} \therefore \chi^2 &= \frac{(787 - 789)^2}{798} + \frac{(277 - 266)^2}{266} \\ &= \frac{121}{798} + \frac{121}{266} \\ &= 0.61 \end{aligned}$$

To find out whether this χ^2 value fits a 3 : 1 ratio, the attached Chi-square table is to be consulted (Table in Appendix).

Degree of freedom (d.f.) means the number of alternative classes which will be one less than the number of actual class (i.e., $n - 1$). So, for a 3 : 1 ratio the *d.f.* is 1. The right side column of the table shows the probability of getting a high value of χ^2 because of chance alone if the sample is from a population conforming to this particular ratio. Consulting this table it is found that the χ^2 value of 0.61 has *P* between 0.30 and 0.50, i.e., the probability of getting such a χ^2 value lies between 30% and 50%. So the result obtained is a *very good fit* for 3 : 1 ratio. A '*P*' value above 0.05 (about 5% level, i.e., 5 times in 100 trials) is considered a good fit. Higher the value of *P*, i.e. smaller the Chi-square value, better is the fit.

It should always be remembered that in statistical analysis the sample taken must be reasonably large. Otherwise the conclusion may become erroneous. The ratio may be found to be good fit for 2 or 3 different genetic factors. In class-work, if such a problem arises, all the genetic factors should be discussed. If the *P* value lies between 5% and 70% or even 80% it should be considered a *good fit*. If the *P* value falls below 5% level then it is a *bad fit*. And if the *P* value lies above the 80% limit then it indicates artificiality or some other defect in the experiment.

For class-work, usually artificial seed samples composed of different pulses and beans are supplied. These samples are often so small that they appear to be good fits for more than one genetical ratio. Some genetical ratios are discussed :

1. Seed Sample No. 1

The given sample (pea seeds) shows the following classes of phenotypes :

No. of Class	Character	Observed no. of seeds	Expected no. of seeds
1	Cotyledon yellow	87	90
2	Cotyledon green	33	30

The observed number of seeds appear to be very close to 3 : 1 ratio.

The observed no. of seeds = $87 + 33 = 120$. This no. is divisible by 4, then the expected numbers be 90 and 30 which will give the genetic ratio 3 : 1.

$$\begin{aligned}\therefore \chi^2 &= \sum \frac{(O - e)^2}{e} \\ &= \frac{(87 - 90)^2}{90} + \frac{(33 - 30)^2}{30} \\ &= \frac{(-3)^2}{90} + \frac{(3)^2}{30} \\ &= \frac{9}{90} + \frac{9}{30} \\ &= 0.10 + 0.30 \\ &= 0.40\end{aligned}$$

The no. of classes in this experiment is 2. So, degree of freedom (*d.f.*) is $(n - 1) = 1$. For *d.f.* 1, Chi-square value of 0.40 has P (probability) lying between 0.50 and 0.70. Hence, the probability of getting such a Chi-square value lies between 50% and 70%. So the given seed sample is a *very good fit* for 3 : 1 ratio, which is the F_2 phenotypic ratio for a monohybrid cross, i.e. a cross between two homozygous parents involving one pair of alleles.

2. Seed Sample No. 2

The given sample shows the following classes of phenotypes :

No. of Class	Character	Observed no. of seeds	Expected no. of seeds
1	Seeds dark brown in colour	67	63
2	Seeds light yellow in colour	45	49

The observed number of seeds appear to be very close to 9 : 7 ratio. The observed number of seeds = $67 + 45 = 112$. This number is divisible by 16 and the quotient is 7. For 9 : 7 ratio, the expected numbers will be $7 \times 9 = 63$, $7 \times 7 = 49$.

$$\begin{aligned}\therefore \chi^2 &= \sum \frac{(O - e)^2}{e} \\ &= \frac{(67 - 63)^2}{63} + \frac{(45 - 49)^2}{49} \\ &= \frac{(4)^2}{63} + \frac{(-4)^2}{49} \\ &= \frac{16}{63} + \frac{16}{49} = 0.253 + 0.326 \\ &= 0.579 \text{ or } 0.58\end{aligned}$$

The number of classes in this experiment is 2. So *d.f.* is $n - 1$. For, *d.f.* 1, a Chi-square volume of 0.58 has P (Probability) lying between 0.30 and 0.50. Hence, the probability of getting such a Chi-square value

lies between 30% and 50%. So the given seed sample is a good fit for 9 : 7 ratio, which is the F_2 phenotypic ratio for *complementary* factor. When two pairs of alleles have no individual effect, but together they complement each other to produce a phenotypic effect, it is called a complementary factor.

3. Seed Sample No. 3

The given sample shows the following classes of phenotypes :

No. of Class	Character	Observed no. of seeds	Expected no. of seeds
1	Seeds dark brown in colour	92	96
2	Seeds yellow in colour	27	24
3	Seeds light greenish in colour	9	8

The observed number of seeds appear to be very close to 12 : 3 : 1 ratio. The observed number of seeds is $92 + 27 + 9 = 128$. This number is divisible by 16 and the quotient is 8. For 12 : 3 : 1 ratio, the expected number will be $8 \times 12 = 96$, $8 \times 3 = 24$ and $8 \times 1 = 8$.

$$\begin{aligned} \therefore \chi^2 &= \sum \frac{(O-e)^2}{e} \\ &= \frac{(92-96)^2}{96} + \frac{(27-24)^2}{24} + \frac{(9-8)^2}{8} \\ &= \frac{(-4)^2}{96} + \frac{3^2}{24} + \frac{1^2}{8} \\ &= \frac{16}{96} + \frac{9}{24} + \frac{1}{8} = 0.166 + 0.375 + 0.125 = 0.666 \end{aligned}$$

The number of classes in this experiment is 3. So, d.f. is $n - 1 = 3 - 1 = 2$. For d.f. 2, a chi-square value of 0.666 has P (Probability) lying between 0.70 and 0.80.

Hence, the probability of getting such a chi-square value lies between 70% and 80%. So the given seed sample is a good fit for 12 : 3 : 1 ratio, which is F_2 phenotypic ratio for *Dominant Epistasis*. When the dominant allele belonging to one allelic pair is also dominant over both the alleles of another allelic pair it is called a case of dominant epistasis.

Example 7:- Test the goodness of fit in the following sample from an experiment with garden pea plants.

Round, yellow	31
Round, green	26
wrinkled, yellow	27
wrinkled, green	26

No. of classes	Sample character	Observed no.
1	Round, yellow	31
2	Round, green	26
3	wrinkled, yellow	27
4	wrinkled, green	26

Solution:-

Total count = 110

Step 1: Determination of expected segregation ratio

Observed value	9 : 3 : 3 : 1		1 : 1 : 1 : 1	
	Expected	d	Expected	d
Round, yellow 31	$\frac{9}{16} \times 110$ = 61.875	30.875	$\frac{1}{4} \times 110$ = 27.5	3.5
Round, green 26	$\frac{3}{16} \times 110$ = 20.625	5.375	$\frac{1}{4} \times 110$ = 27.5	1.5
wrinkled, yellow 27	$\frac{3}{16} \times 110$ = 20.625	6.375	$\frac{1}{4} \times 110$ = 27.5	0.5
wrinkled, green 26	$\frac{1}{16} \times 110$ = 6.875	19.125	$\frac{1}{4} \times 110$ = 27.5	1.5
Total deviation		61.75		7.0

According to the table, we find that the deviation is minimum in case of 1:1:1:1 ratio. So, it is assumed that the observed sample should fit with this ratio.

Step 2: Determination of Chi-square value

No. of classes	Sample characters	Observed value (O)	Expected value (E)	Deviation (O-E)	(O-E) ²	$\frac{(O-E)^2}{E}$
1	Round, yellow	31	27.5	3.5	12.25	$\frac{12.25}{27.5} = 0.445$
2	Round, green	26	27.5	-1.5	2.25	$\frac{2.25}{27.5} = 0.082$
3	Wrinkled, yellow	27	27.5	-0.5	0.25	$\frac{0.25}{27.5} = 0.009$
4	Wrinkled, green	26	27.5	-1.5	2.25	$\frac{2.25}{27.5} = 0.082$

$$\chi^2 = \sum \frac{(O-E)^2}{E} = 0.445 + 0.082 + 0.009 + 0.082 = 0.618$$

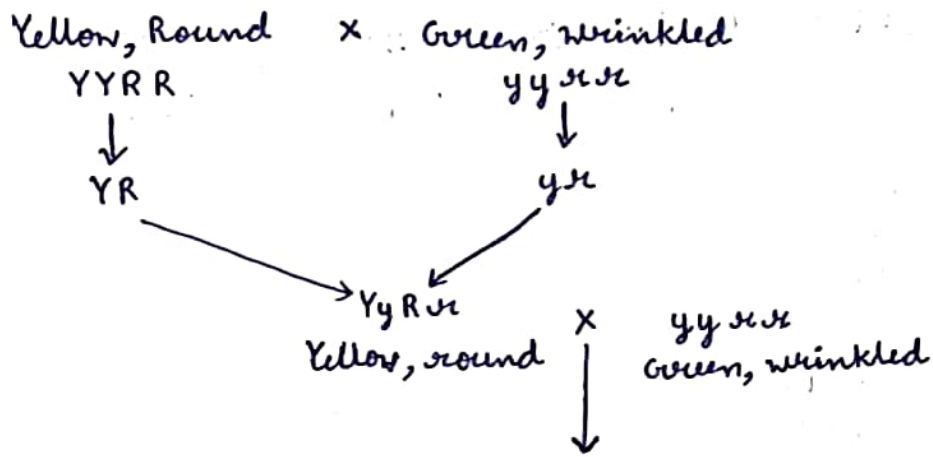
Step 3: Conclusion

Since there are 4 classes, degrees of freedom = 4 - 1 = 3. The calculated χ^2 value is 0.618, which is much less than the tabulated χ^2 value for 3 degrees of freedom at 0.05 probability level.

The calculated value lies between 70-90% probability level. Therefore the deviation from expected value is insignificant and the observed ratio has a good fit with the expected ratio. So, the genes involved follow Mendel's law of segregation and independent assortment.

Step 4: Comment

As the experimental result shows good fit with 1:1:1:1 ratio, so it is assumed that the result is obtained from a dihybrid test cross. The dominant characters are yellow and round. Parental genotypes are assumed as:



$\begin{matrix} \searrow \\ \swarrow \end{matrix}$	YR	Yr	yR	yr	$YyRr \rightarrow$ Yellow, round
$\begin{matrix} \rightarrow \\ \leftarrow \end{matrix}$	YR	Yr	yR	yr	$Yyrr \rightarrow$ Yellow, wrinkled.
	YR	Yr	yR	yr	$yyRr \rightarrow$ green, round
	YR	Yr	yR	yr	$yyrr \rightarrow$ green, wrinkled

Example 8:- A geneticist obtained 209 plants after selfing a hybrid, among which with red flower = 52 plants, with pink flower = 128 plants and with white flower = 29 plants. Show the goodness of fit with 1:2:1 ratio.

Character	No. of plants
Red	52
Pink	128
White	29
Total no. of plants	209

Solution :-

Step 1: Determination of expected frequency of different plant.

$$\text{Red flower} = 209 \times \frac{1}{4} = 52.25$$

$$\text{Pink flower} = 209 \times \frac{2}{4} = 104.5$$

$$\text{White flower} = 209 \times \frac{1}{4} = 52.25$$

Step 2: Calculation of Chi-square value

Character	Observed value (O)	Expected value (E)	Deviation (O-E)	(O-E) ²	$\frac{(O-E)^2}{E}$	$\chi^2 = \sum \frac{(O-E)^2}{E}$
Red	52	52.25	-0.25	0.0625	0.00119	15.626
Pink	128	104.5	23.5	552.25	5.28	
White	29	52.25	-23.25	540.56	10.345	

Step 3: Conclusion

The calculated value of $\chi^2 = 15.626$. Since there are 3 classes, so the degree of freedom = $(3-1) = 2$. The tabulated χ^2 value for 2 degrees of freedom at 0.05 probability level is 5.99. As the observed value much higher than the the tabulated value, so the result does not in agreement with 1:2:1 ratio, i.e., the observation does not show goodness of fit with the expected ratio.

In conclusion, the principle of incomplete dominance is not applicable here. So, the genes involved here do not follow Mendel's laws.

Table 11.5 Chi-Square Probabilities

df	Probabilities									
	0.95	0.90	0.70	0.50	0.30	0.20	0.10	0.05	0.01	0.001
1	0.004	0.016	0.15	0.46	1.07	1.64	2.71	3.84	6.64	10.83
2	0.10	0.21	0.71	1.39	2.41	3.22	4.61	5.99	9.21	13.82
3	0.35	0.58	1.42	2.37	3.67	4.64	6.25	7.82	11.35	16.27
4	0.71	1.06	2.20	3.36	4.88	5.99	7.78	9.49	13.28	18.47
5	1.15	1.61	3.00	4.35	6.06	7.29	9.24	11.07	15.09	20.52
6	1.64	2.20	3.83	5.35	7.23	8.56	10.65	12.59	16.81	22.46
7	2.17	2.83	4.67	6.35	8.38	9.80	12.02	14.07	18.48	24.32
8	2.73	3.49	5.53	7.34	9.52	11.03	13.36	15.51	20.09	26.13
9	3.33	4.17	6.39	8.34	10.66	12.24	14.68	16.92	21.67	27.88
10	3.94	4.87	7.27	9.34	11.78	13.44	15.99	18.31	23.21	29.59
11	4.58	5.58	8.15	10.34	12.90	14.63	17.28	19.68	24.73	31.26
12	5.23	6.30	9.03	11.34	14.01	15.81	18.55	21.03	26.22	32.91
13	5.89	7.04	9.93	12.34	15.12	16.99	19.81	22.36	27.69	34.53
14	6.57	7.79	10.82	13.34	16.22	18.15	21.06	23.69	29.14	36.12
15	7.26	8.55	11.72	14.34	17.32	19.31	22.31	25.00	30.58	37.70
20	10.85	12.44	16.27	19.34	22.78	25.04	28.41	31.41	37.57	45.32
25	14.61	16.47	20.87	24.34	28.17	30.68	34.38	37.65	44.31	52.62
30	18.49	20.60	25.51	29.34	33.53	36.25	40.26	43.77	50.89	59.70
50	34.76	37.69	44.31	49.34	54.72	58.16	63.17	67.51	76.15	86.66

← | →
 Fail to reject | Reject
 at 0.05 level