INDEX

1. CHI SQUARE TEST (χ^2) AND ESTIMATION OF LINKAGE

(i) TEST OF GOODNESS OF FIT

The Chi-Square distribution was first discovered by Helmert in 1876 and later independently by Karl Pearson in 1900. Chi-Square measures the discrepancy between the observed and expected (theoretically) determined frequency. χ^2 value range from 0 to ∞

$$\chi^2 = \sum_i \frac{(O_i - E_i)^2}{E_i} \quad \text{for } i = 1, 2, ..., k \text{ classes}$$

Where, O_i = observed frequency of i^{th} class

E_i = expected frequency of i^{th} class

If $O_i = E_i$ for all the k classes then $\chi^2 = 0$

The calculated χ^2 value is compared with Table value with df = (k-1)

If calculated χ^2 value ≥ Table χ^2 for df = k-1 at 0.05 level of significance then it is said to be significant

Uses of Chi-Square

1. To test **goodness of fit**.
2. To test the **independence of attributes.**
3. To test the **validity of hypothetical ratios.**
4. To test the **homogeneity of several population variances.**
5. To test the **equality of several correlation coefficients.**

Example 1: The observed frequency of four phenotypic classes of a dihybrid is as under. Test whether the frequency distribution follow the segregation ratio of 9:3:3:1 for AB: Ab : aB : ab or not ?

Table 1.1 : χ^2 test for segregation of 9:3:3:1 ratio

Class	Ratio	Observed frequency (O_i)	Expected frequency (E_i)	(O_i - E_i)	$\chi^2 = \sum_i \frac{(O_i - E_i)^2}{E_i}$
AB	9	100	E = (9/16)x160 =90	10.0	1.1111
Ab	3	22	E = (3/16)x160 =30	-8.0	2.1333
aB	3	24	E = (3/16)x160 =30	-6.0	1.2000
ab	1	14	E = (1/16)x160 =10	4.0	1.600
Total	16	160	75.00	$\sum \chi^2$	6.0444

Ho.: The frequency of the four phenotypic classes are distributed as 9:3:3:1

Ha. : The frequency of the four phenotypic classes do not distribute in the ratio of 9:3:3:1

The calculated value of χ^2 is 6.0444 which is compared with the table value of $\chi 2$ at (K-1) = (4-1) degree of freedom at 0.05 level of significance. The table value is 7.81. Thus the calculated χ^2 value is less than table value and we accept the null hypothesis. There is no evidence against the null hypothesis for its rejection and we conclude that the frequency of four phenotypic classes is segregated in the ratio of 9:3:3:1 or we conclude that the fit is good.

(ii) Detection of Linkage

Linkage : The tendency for genes or segments of DNA closely positioned along a chromosome to segregate together at meiosis and therefore be inherited together. Linkage may be defined as " the tendency of two genes of the same chromosome to remain together in the process of inheritance".

Coupling

Coupling refers to the case where dominant alleles are on the same homologous chromosome and both recessive alleles are on the other homologous chromosome. Thus, the parental gametes are **AB** and **ab**. Some authors call this **cis.**

Repulsion

Repulsion refers to the case where each homologous chromosome has one dominant and one recessive allele from the two genes. Thus, the parental gametes are **Ab** and **aB**. Some authors call this **trans.**

Detection of Linkage

Linkage between two dominants genes (eg. A and B) provide significant deviation in the frequencies of different phenotypic classes from those expected on the basis of independent assortment (1:1:1:1 and 9:3:3:1 ratio in the test cross and F_2 generation, respectively) Such a deviation may also result from a departure of the ratio between A_ and aa or B_ and bb phenotypes from the

expected 1:1 or 3:1 ratio therefore, a test for presence of linkage in a test cross or F_2 progeny must proceed as follows.

Step 1: A test for 1:1 or 3:1 ratio between A_ and aa phenotypic classes.

Step 2: A test for 1:1 or 3:1 ratio between B_ and bb phenotypic classes.

Step 3: A test for independent of segregation of A/a from that of B/b1:1 or 3:1 ratio between A_ and aa phenotypic classes.

EXAMPLE 2: The following record of segregation of 2 factors A, a and B, b obtained from inbreeding doubling heterozygote individuals in 3 families.

1. Test whether factor A & B segregates in the ratio of 3:1
2. Test whether there is evidence of linkage or not
3. Test whether the three families are homogeneous with respect to above two point or not ?.

Family	AB	Ab	aB	ab	Total
I	47	8	11	9	75
II	75	14	14	11	114
III	65	13	12	11	101

Family I

(1) Test of segregation of A_a in the ratio of 3:1

The frequency of four phenotypic classes can be classified in a 2 x 2 table as under

	B	b	A-a
A	47 a_{11}	8 a_{12}	55
a	11 a_{13}	9 a_{14}	20
B-b	58	17	75

H_0 = A – a is segregating in the ratio of 3:1

H_a = A – a is not segregating in the ratio of 3:1

Table 1.2 : χ^2 test for segregation of A-a gene as 3:1 ratio

Class	Observed frequency (O_i)	Expected frequency (E_i)	(O_i - E_i)	$\chi^2 = \sum_i \frac{(O_i - E_i)^2}{E_i}$
A	55	E=$\frac{3}{4}$×75=56.25	-1.25	0.0278
a	20	E=$\frac{1}{4}$×75=18.75	1.25	0.0833
Total (n)	75	75.00	$\sum \chi^2$	0.1111

Conclusion : The calculated value of χ^2 = 0.1111 which is compared with table value of χ^2 at (K-1) = 1 degree of freedom at .05 level of significance. The table value is 3.84 and thus calculated χ^2 is non significant so we do not have evidence to reject the null hypothesis. We accept the null hypothesis that A-a segregates in the ratio of 3:1.

This can be examined with a direct method as under

$$\chi^2 = \frac{\{a_{11} + a_{12} - 3(a_{13} + a_{14})\}^2}{3n} \text{ Where } n = a_{11} + a_{12} + a_{13} + a_{14}$$

$$\chi^2 = \frac{\{47 + 8 - 3(11 + 9)\}^2}{3 \text{ x } 75} = \frac{(55 - 60)^2}{225} = \frac{(-5)^2}{225} = 0.1111$$

Test of segregation of B_b in the ratio of 3:1

H_0 = B – b is segregating as 3:1

H_a = B – b is not segregating as 3:1

Table 1.3 : χ^2 test for segregation of B-b gene as 3:1 ratio

Class	Observed frequency (O_i)	Expected frequency (E_i)	(O_i - E_i)	$\chi^2 = \sum_i \frac{(O_i - E_i)^2}{E_i}$
B	58	E=$\frac{3}{4}$×75=56.25	1.75	0.0544
b	17	E=$\frac{1}{4}$×75=18.75	-1.75	0.1633
Total	75	75.00	$\sum \chi^2$	0.2177

Conclusion : The calculated value of χ^2 = 0.2178 which is compared with table value of χ^2 at (K-1) = 1 degree of freedom at .05 level of significance . The table value is 3.84 and thus calculated χ^2 is non significant so we do not have evidence to reject the null hypothesis. We accept the null hypothesis that A-a segregates in the ratio of 3:1.

This can be examined with a direct method as under

Direct method to find out the χ^2 for two classes (B:b) or 3:1 ratio

$$\chi^2 = \frac{\{a_{11} + a_{13} - 3(a_{12} + a_{14})\}^2}{3n} \quad \text{Where } n = a_{11} + a_{12} + a_{13} + a_{14}$$

$$\chi^2 = \frac{\{47 + 11 - 3(8 + 9)\}^2}{3 \text{ x } 75} = \frac{(58 - 51)^2}{225} = \frac{(7)^2}{225} = 0.2177$$

2. Test for the existence of linkage

Table 1.4 : χ^2 test for detecting presence of linkage

Class	Ratio	Observed frequency (O_i)	Expected frequency (E_i)	(O_i - E_i)	$\chi^2 = \sum_i \frac{(O_i - E_i)^2}{E_i}$
AB	9	47	E=$\frac{9}{16}$×75=42.19	-4.81	0.5484
Ab	3	8	E=$\frac{3}{16}$×75=14.06	-6.06	2.6119
aB	3	11	E=$\frac{3}{16}$×75=14.06	-3.06	0.6660
ab	1	9	E=$\frac{1}{16}$×75=4.69	4.31	3.9608
Total	16	75	75.00	$\sum \chi^2$	7.7871

Calculated χ^2 for four phenotypic classes at 3 df is 7.7871

The calculated χ^2 for linkage = Table χ^2 - (χ^2 for A-a) - (χ^2 for B-b)

= 7.7871 – 0.1111 – 0.2177

= 7.4681

Which will have df for χ^2 linkage = (3 df of table χ^2) - (1 df for χ^2 A-a) - (1 df for χ^2 B-b) = 1 df

OR

Calculate χ^2 linkage **by direct method as under**

$$\chi^2 = \frac{\{a_{11} - 3(a_{12}) - 3(a_{13}) + 9a_{14})\}^2}{9n} \quad \text{Where } n = a_{11} + a_{12} + a_{13} + a_{14}$$

$$\chi^2 = \frac{\{47 - (3x8) - (3x11) + (9x9)\}^2}{9 \times 75} = 7.4681$$

Conclusion : The calculated value of χ^2 for linkage = 7.4681 which is compared with table value of χ^2 at 1 degree of freedom at .05 level of significance . The table value is 3.84 and thus calculated χ^2 is significant so we reject the null hypothesis. We conclude that there is existence of linkage between the two gene pair.

Repeat the same procedure for family –II and III .

3. Test of homogeneity of families with regards to segregation and linkage

Table 1.5 : Test the homogeneity of segregation in different families

Family	AB	Ab	aB	ab	Total
I	47	8	11	9	75
II	75	14	14	11	114
III	65	13	12	11	101
Total (overall)	187	35	37	31	290

(i) Carry out test for A-a segregation on pooled basis as 3:1

$$\chi^2 = \frac{\{187 + 35 - 3(37 + 31)\}^2}{3 \times 290} = \frac{(222 - 204)^2}{3x290} = \frac{(18)^2}{3x290} = 0.3724$$

Calculated χ^2 on overall basis or pooled basis = 0.3724 ≤ 3.84 Table χ^2 at 0.05 level of significance at 1 df. Therefore, we conclude that irrespective of families the gene A-a on pooled basis segregates in the ratio of 3:1.

(ii) Test for segregation of B-b as 3:1

$$\chi^2 = \frac{\{187 + 37 - 3(35 + 31)\}^2}{3 \text{ x } 290} = \frac{(224 - 198)^2}{3x290} = \frac{(26)^2}{3x290} = 0.7770$$

Calculated χ^2 on overall basis or pooled basis = 0.7770 ≤ 3.84 Table χ^2 at 0.05 level of significance at 1 df. Therefore we conclude that irrespective of families the gene B-b on pooled basis segregates in the ratio of 3:1.

(iii) Test χ^2 for testing linkage on overall basis

$$\chi^2 = \frac{\{187 - (3x35) - (3x37) + (9x31)\}^2}{9 \text{ x } 290} = 23.9464$$

As the calculated χ^2 = 23.9464 ≥ 3.84 table χ^2 at 0.05 level of probability with 1 df is significant. Thus, there is evidence of linkage irrespective of families.

(iv) Test for Homogeneity of families

Chi square (χ^2) has additive property so all chi square values of different families are added together as under .

Table 1.6 : χ^2 values for segregation of individual gene and for linkage

Family	df	χ^2 (A-a)	χ^2 (B-b)	χ^2 (Linkage)
I	1	0.1111	0.2177	7.4681
II	1	0.5731	0.5731	7.8947
III	1	0.2673	0.0825	8.7140
Total	3	0.9515	0.8733	24.0768

Table 1.7 : ANOVA for test of homogeneity of families

Source	df	χ^2 (A-a)	χ^2 (B-b)	χ^2 (Linkage)	Table χ^2(5%)
Deviation	1	0.3724 ns	0.7770 ns	23.9464*	3.84
Heterogeneity	2	0.5791 ns	0.0964 ns	0.1304ns	5.99
Total	3	0.9515 ns	0.8733 ns	24.0768*	7.81

ns = non-significant; * significant at 5 % level of significance with respective degree of freedom

Note: Heterogeneity χ^2 is calculated by difference between Total χ^2 and deviation χ^2

Conclusion:

1. Single factors (A-a and B-b) are segregating in the expected ratio of 3:1 for each family as well as on overall basis.
2. There is evidence of linkage in each of the family as well as on overall basis.
3. Heterogeneity $\chi 2$ is non significant for A-a, B-b and linkage hence all the families are homogeneous with regards to segregation of genes and linkage.

Estimation of Intensity of Linkage(p) (Recombination fraction)

When the existence of linkage is established one may estimate a measure of its intensity. When linkage exist it is concluded that the two genes do not segregate independently.

The intensity of linkage is measured in inverse sense as the fraction of the total number of chromosomes pairs in which the changes over take place at gamatogenesis. This is known as the recombination fraction. The smaller fraction will indicate more intensity of linkage.

The chromosome theory leads to a measure of the intensity of linkage based on the frequency of breakage and rejoining of the homologous chromosomes between the loci concerned. This is estimated as the proportion of recombination chromosome.

For estimation of linkage different methods are used which are as under

(i) Maximum likelihood Method

(ii) Emerson's Method

(iii) Product Ratio Method

For F_2 data:

(i) Maximum likelihood Method : This method is most efficient method among all the three methods of estimation.

Example :3 Thefollowing indicates the frequency of four phenotypic classes estimate the linkage intensity by different methods.

Table 1.8 : Estimation of linkage intensity from F_2 data

Family	AB	Ab	aB	ab	Total
F_2	86	8	14	20	128
	a_1	a_2	a_3	a_4	

First, test the evidence of linkage

χ^2 for testing linkage by direct method (9:3:3:1)

$$\chi^2 = \frac{\{a_{11} - 3(a_{12}) - 3(a_{13}) + 9a_{14})\}^2}{9n} \text{ Where } n = a_{11} + a_{12} + a_{13} + a_{14}$$

$$\chi^2 = \frac{\mathbf{\{86 - (3x8) - (3x14) + (9x20)\}^2}}{\mathbf{9 \times 128}} = \mathbf{34.7222}$$

As the calculated χ^2 = 34.7222 ≥ 3.84 table χ^2 at 0.05 level of significance at df=1 , it is said to be significant and existence of linkage is established. When linkage exist then we can find linkage intensity.

Procedure for Estimation of linkage intensity (p) by Maximum likelihood method.

The following equation is used in this method

$n\theta^2 - \theta (a_1 - 2(a_2 + a_3) - a_4) - 2a_4 = 0$

Where a_1, a_2, a_3, a_4 = frequency of the respective classes and $n = a_1 + a_2 + a_3 + a_4$

For the solution of the above equation we have to find out roots of θ as under

$$\theta = \frac{-b \pm \sqrt{b^2 - 4ac}}{2a}$$

Positive root of θ is used for further calculation

Fraction for coupling phase : value of $p = 1 - \sqrt{\theta}$

Fraction for repulsion phase : value of $p = \sqrt{\theta}$

θ is our estimate so we need to estimate its std. error. First we need to calculate its variance

$$\text{Var}(\theta)=\frac{2\theta\,(1-\theta)(2+\theta)}{n(1+2\theta)}$$

From the above equation we get variance and standard error of P

Var (p) = $\frac{\text{var}(\theta)}{4\theta}$ and SE (p) = $\sqrt{\text{var}(p)}$

with the help of SE (p) we provide the 95 % Confidence limits for the estimate of p for both the phases as under:

$$p - 1.96 \times SE(p) < P < p + 1.96 \times Se(p)$$

Example : from the data given in table 1.8 the linkage intensity can be estimated as under

$$128\theta^2-\theta(86-2(8+14)-20)-2\times 20=0$$

$$128\theta^2-\theta(86-44)-40=0$$

$$128\theta^2-22\theta-40=0$$

Where a_1, a_2, a_3, a_4 = frequency of the respective class $n = a_{11} + a_{12} + a_{13} + a_{14}$

Now find out the roots of θ as under

$$\theta=\frac{-(-22)\pm\sqrt{(-22)^2-4\times 128\times 40}}{2\times 128}$$

Root of θ = 0.6515 and –0.4796

Only the positive root of θ is used for further calculation so we consider 0.6515

Fraction for coupling phase : value of p = 1 - $\sqrt{\theta}$ = 1 - $\sqrt{0.6515}$ = 0.1928

Fraction for repulsion phase : value of p = $\sqrt{\theta}$ = $\sqrt{0.6515}$ = 0.8072

θ is our estimate so to estimate its std. error we need to calculate its variance

$$\text{Var}(\theta)=\frac{2\times 0.6515(1-0.6515)(2+0.6515)}{128(1+2\times 0.6515)}=0.004085$$

From the above equation we get variance and standard error of P

Var (p) = $\frac{Var(\theta)}{4\theta}$ = $\frac{0.004085}{4x0.6515}$ = 0.0015

SE (p) = $\sqrt{\text{var}(p)}$ = $\sqrt{0.0015}$ = 0.0387

95 % Confidence limits for the estimate p for coupling phase

p – 1.96 x SE (p) < P < p +1.96 x SE (p)

0.1928 – 1.96 x 0.0387 < P < 0.1928 + 1.96 x 0.0387

0.1169 < P < 0.2687

95 % Confidence limits for the estimate p for repulsion phase

p – 1.96 x SE (p) < P < p +1.96 x SE (p)

0.8072 – 1.96 x 0.0387 < P < 0.8072 + 1.96 x 0.0387

0.7313 < P < 0.8831

(ii) Emerson's Method for estimation of linkage intensity

$\theta = \frac{E - M}{n}$ Where $E = a_1 + a_2$; $M = a_3 + a_4$; $n = a_1 + a_2 + a_3 + a_4$

From data given in table 1.8 estimate the linkage intensity as under

Now E = 86+20 = 106

M = 8 + 14 = 32 and n= 128

$$\theta = \frac{106 - 32}{126} = 0.6562$$

Fraction for coupling phase : value of p = 1 - $\sqrt{0.6562}$ = 0.1900

Fraction for repulsion phase : value of p = $\sqrt{0.6562}$ = 0.8100

Variance of θ

$$\text{Var}(\theta) = \frac{1 - (0.6562)^2}{128} = 0.004488$$

Now we get variance and standard error of p as

Var (p) = $\frac{Var(\theta)}{4\theta}$ = $\frac{0.004488}{4X0.6562}$ = 0.0017

SE (p) = $\sqrt{\text{var}(p)}$ = $\sqrt{0.0017}$ = 0.0412

95 % Confidence limits for the estimate p for coupling phase

$p - 1.96 \times SE(p) < P < p + 1.96 \times SE(p)$

$0.1900 - 1.96 \times 0.0412 < P < 0.1900 + 1.96 \times 0.0412$

$0.1092 < P < 0.2708$

95 % Confidence limits for the estimate p for repulsion phase

$p - 1.96 \times SE(p) < P < p + 1.96 \times SE(p)$

$0.8100 - 1.96 \times 0.0412 < P < 0.8100 + 1.96 \times 0.0412$

$0.7292 < P < 0.8908$

(iii)) Product Ratio Method for estimation of linkage intensity

In this method the value of θ is estimated as under

$$\theta = \frac{1 + Q - \sqrt{1 + 3Q}}{Q - 1} \qquad \text{Where } Q = \frac{a_1 \times a_4}{a_2 \times a_3}$$

From the data given in table 1.8 linkage intensity can be estimated as under

Now calculate the value of $Q = \frac{a_1 \times a_4}{a_2 \times a_3} = \frac{86 \times 20}{8 \times 14} = 15.3771$

$$\theta = \frac{1 + 15.3771 - \sqrt{1 + 3 \times 15.3771}}{15.3771 - 1} = 0.6614$$

Fraction for coupling phase : value of $p = 1 - \sqrt{0.6614} = 0.1868$

Fraction for repulsion phase : value of $p = \sqrt{0.6614} = 0.8132$

Note : To find out the variance of θ, p and Standard error of p, the method is same as Maximum Likelihood method

$$\text{Var}(\theta) = \frac{2 \times 0.6614(1 - 0.6515)(2 + 0.6614)}{128(1 + 2 \times 0.6614)} = 0.0040$$

From the above equation we get variance and standard error of p

$$\text{Var (p)} = \frac{\text{var}(\theta)}{4\theta}$$

$$= \frac{0.004}{4x0.6614}$$

$= 0.00151$ and

SE (p) = $\sqrt{\text{var}(p)}$

= $\sqrt{0.00151}$

= 0.03887

95 % Confidence limits for the estimate p for coupling phase

p – 1.96 x SE (p) < P < p +1.96 x SE (p)

0.1868 – 1.96 x 0.03887 < P < 0.1828 + 1.96 x 0.03887

0.1162 < P < 0.2630

95 % Confidence limits for the estimate p for repulsion phase

p – 1.96 x SE (p) < P < p +1.96 x SE (p)

0.8132 – 1.96 x 0.03887 < P < 0.8132 + 1.96 x 0.03887

0.7370 < P < 0.8894

Note: higher the value of SE wider will be the range of confidence limits

Linkage from back cross data:

Example 4 : With a view to study linkage between genes for the throat colour and filament colour, a set of experiments was carried out at Central Tobacco Research Institute, Rajhamundry, AP. The distributions for four phenotypic classes are as under.

Family	AB	Ab	aB	ab	Total
F_1x A 5(P) (Back Cross data)	25 b_1	5 b_2	7 b_3	23 b_4	60 n

1. Test the existence for linkage
2. If there is linkage estimate linkage intensity (p) by
   (i) Maximum likelihood Method
   (ii) Emerson's Method and
   (iii) Product Ratio Method

1. χ^2 for testing the segregation of gene A-a in the ratio of 1:1

Table 1.9 : χ^2 for testing the segregation of factor A-a as 1:1

Class	Observed frequency (O_i)	Expected frequency (E_i)	(O_i - E_i)	$\chi^2 = \sum_i \frac{(O_i - E_i)^2}{E_i}$
A	30 ($b_1 + b_2$)	E=$\frac{1}{2}$×60=30	0.00	0.00
a	30 ($b_3 + b_4$)	E=$\frac{1}{2}$×60=30	0.00	0.00
Total	60	60.00	$\sum \chi^2$	0.00

Direct method to find out the χ^2 for two classes (A-a) or 1:1 ratio

$$\chi^2 = \frac{\{b_1 + b_2 - (b_3 + b_4)\}^2}{n} \text{ Where } n = b_1 + b_2 + b_3 + b_4$$

$$\chi^2 = \frac{\{25 + 5 - (7 + 23)\}^2}{60} = \frac{(30 - 30)^2}{60} = 0.00$$

The calculated value of χ^2 = 0.00 ≤ 3.84 table χ^2 at 0.05 level of significance for df = 1. Thus it is said to be non significant, there is no evidence against the hypothesis that A-a is segregating as 1:1 i.e. a-a segregates in the ratio of 1;1.

2. χ^2 for testing the segregation of gene B-b as 1:1

H_0 = B – b is segregating as 3:1

H_a = B – b is not segregating as 3:1

Table 1.10: χ^2 for testing the segregation of factor B-b as 1:1

Class	Observed frequency (O_i)	Expected frequency (E_i)	(O_i - E_i)	$\chi^2 = \sum_i \frac{(O_i - E_i)^2}{E_i}$
B	32 ($b_{11} + b_{13}$)	E = (1/2) x 60 =30	2	0.1333
b	28 ($b_{12} + b_{14}$)	E = (1/2) x 60 =30	-2	0.1333
	60	60.00	$\sum \chi^2$	0.2667

Direct method to find out the χ^2 for two classes (B-b) in 1:1 ratio

$$\chi^2 = \frac{\{b_1 + b_3 - (b_2 + b_4)\}^2}{n} \text{ Where } n = b_1 + b_2 + {}_3 + b_4$$

$$\chi^2 = \frac{\{25 + 7 - (5 + 23)\}^2}{60} = \frac{(4)^2}{60} = 0.2667$$

Here calculated $\chi^2 = 0.2667 \leq 3.84$ table X^2 at 0.05 level of significance for df = 1 thus, it is said to be non significant, there is no evidence against the hypothesis that B-b is segregating as 1:1.i.e. we do not reject the null hypothesis.

2. χ^2 for testing linkage by direct method

H_0 = There is no existence of linkage

H_a = There is existence of linkage

$$\chi^2 = \frac{(b_1 - b_2 - b_3 + b_4)^2}{n} \text{ Where } n = b_1 + b_2 + b_3 + {}_4$$

$$\chi^2 = \frac{(25 - 5 - 7 + 23)^2}{60} = \frac{(36)^2}{60} = 21.6000$$

As the calculated value of $\chi^2 = 21.6000 \geq 3.84$ table χ^2 at 0.05 level of significance for df=1, it is said to be significant and it is concluded that there is existence of linkage. i.e. The null hypothesis is rejected. When linkage existence is established we can estimate linkage intensity for back cross data as under

Procedure for estimation of linkage intensity from Back cross data

Linkage intensity (p)for coupling stage = $(b_2+b_3)/n$

$= (5 + 7) / 60 = 0.2$

Linkage intensity (p)for repulsion = $(b_1+b_4)/n$

$= (25 + 23) / 60 = 0.8$

$$\text{Var (p)} = \frac{p(1-p)}{n} = \frac{0.2(1-0.2)}{60} = 0.0027$$

Variance for both the phase is same as numerator p(1-p) is same

Se (p) = $\sqrt{Var(p)} = \sqrt{0.0027} = 0.0516$

95 % Confidence limits for linkage intensity (p) for coupling phase

$p - 1.96 \times SE(p) < P < p + 1.96 \times SE(p)$

$0.2 - 1.96 \times 0.0516 < P < 0.2 + 1.96 \times 0.0516$

$0.0989 < P < 0.3011$

95 % Confidence limits for linkage intensity (p) for coupling phase

$p - 1.96 \times SE(p) < P < p + 1.96 \times SE(p)$

$0.8 - 1.96 \times 0.0516 < P < 0.8 + 1.96 \times 0.0516$

$0.6989 < P < 0.9011$

2. GENE, GENOTYPIC FREQUENCY & HARDY WEINBERG LAW

Let us assume that there are two alleles A and a at one locus then there will be three genotypes in the population i.e. AA , Aa and aa . It is assumed that these genotypes are easily distinguishable then one can count the frequency of genotype. If these frequencies of the genotypes are expressed in proportion their sum in the population will be equal to unity. They can also be expressed in terms of percentages but for this course we will use gene and genotypic frequency in terms of proportion.

Example

Genotype	Frequency of genotypes	Genotypic frequency In terms of proportion	Symbols used
AA	203	(203/700) = 0.29	p^2 OR D
Aa	406	(406/700) = 0.58	2pq OR H
aa	91	(91/700) = 0.13	q^2 OR R
Total	700	1.00	

p^2 or D is the frequency of AA genotype in the population

2pq or H is the frequency of Aa genotype in the population

q^2 or R is the frequency of aa genotype in the population

Gene frequency :

It is defined as the proportion of a given allele in the population. The gene frequency of a dominant allele (A) is symbolized as p while for recessive allele (a) as q. The sum of gene frequency (p+q) =1.0 at a given locus for a gene pair with two alleles. It can be calculated as under

Gene frequency for Allele A (p) = (D + ½ H) or = p^2 + pq and

Gene frequency for Allele a (q) = (R +½ H) or = q^2 + pq and

Accordingly freq. for allele A (p) = {0.29 + ½ (0.58)} = 0.58 and

freq. for allele a (q) = {0.13 + ½ (0.58)} = 0.42

Knowing the freq. value either for p or q one can estimate the other as q as q = (1- p) or p= (1- q) because (p +q) = 1.0. The sum of gene frequency at a given locus is unity in case of a gene pair having diploid behaviour (two alleles only).

This can be examined in other way as

Table 2.1 : Gene and genotypic frequency for single locus

Genotype	Frequency of genotypes	Total genes	No of genes A	a
AA	203 D	203x2 = 406	406	------
Aa	406 H	(406x2) = 812	406	406
Aa	91 R	(91x2) = 182	------	182
Total	700 N	1400	812	588

Freq. of gene A i.e. p = Total no. of A genes at a locus in the population

= 812/1400

= 0.58 **OR**

p = (2D + H) /2N

= { (2x203) + 406 }/2x70

= 812 / 1400 =0.58

Freq. of gene a i.e. q = Total no. of a genes at a locus in the population 588/1400

= 0.42 **OR**

q = (2R + H) /2N

= { (2x91) + 406 }/2x70

= 588 / 1400

=0.42

generally gene and genotypic frequencies are expressed in terms of proportions

Gametic and genotypic array :

If offspring is produced by random mating in parental generation such that all parents contributes equally in pool of gametes then genotypic array of offspring generation will be square of gametic array. The gametic array is (pA + qa) where p and q are gene frequency of A and a allele in parental generation. The genotypic array of the parental generation will be $(pa + qa)^2$ .i.e. square of

gametic array. The genotypes of this will be p^2AA, 2pq Aa , q^2aa in next generation

If p=0.7 and q=0.3 then in next generation the genotypic frequency will be

$(0.7A + 0.3 a)^2$ = 0.49 AA, 0.42 Aa and 0.09aa

and sum of the three genotypic frequency = 0.49 + 0.42 + 0.09 = 1.00 i.e.

$(p+q)^2 = p^2 + 2pq + q^2$ and $(p+q) = 1.0$

therefore $p^2 + 2pq + q^2 = 1.0$

Multiple alleles :

More than two alternative form of a gene are known as multiple allele i.e. for a single locus if there exist three forms (A1, A2, A3) of a gene and we are interested in determining genotypic frequency of say A1A1 than

Gametic array = (pA1 +q (A2+A3) and where p= gene freq. of A1 and q for (a2+A3)

genotypic array = $(pA1 +q (A2+A3)^2$ will have

$= p^2$ A1A1 + 2pq (A1A2 + A1A3) + q^2 (A2A2+A2A3) = 1.00

where we can have p^2 being the frequency of A1A1 genotypes only

Sex linked gene :

If number of males and females are equal in the population then male (XY)contributes one X chromosome and female (XX)contributes two X chromosomes in common gene pool. Thus proportion of X chromosome for male is 1/3 and for female it is 2/3 in the population. If A-a alleles are there then gene freq. for A in female will be pf = D+1/2 H and in entire population its freq. will be p= 1/3{2D+H +W}

Table 2.2 Genotypic frequency for sex linked genes

Female			Male	
A_1A_1 p^2 (D)	A_1A_2 2pq(H)	A_2A_2 q^2(R)	A_1 W	A_2 X
0.4	0.4	0.2	0.3	0.7

Gene freq. of A in entire population (p) = 1/3{ (2x 0.4) + 0.4 + 0.3} = 0.5 and for a allele (q) = 1-p = 0.5 or q =1/3{ (2R + H +X) = 1/3{(2 x 0.2) + 0.4 + 0.7} = 0.5 and for a allele

Hardy Weinberg Law of equilibrium of genes

This law was given by a British mathematician G H Hardy and a German physician Dr W Weinberg. It refers to gene frequency in a random mating population.

" In a large random mating population in absence of mutation, migration, selection and random genetic drift, the gene and genotypic frequencies do not change from one generation to next generation". This law is known as HW Law of equilibrium of genes.

If this situation is observed than the population is said to be in HW equilibrium .In Random mating population - Each member of population has equal chance to produce offspring and any female gamete be equally likely be fertilized by any male gametes.

Table 2.3 : Random mating for one gene pair :

Female gametes	Male gametes	
	A (p)	a(q)
A (p)	AA (p^2)	Aa (pq)
a (q)	Aa (pq)	aa (q^2)

p and q are gene freq. of allele A and a and by random union of gametes Aa, 2 Aa and aa genotypes are produced with their corresponding freq. p^2, 2pq and q^2 and if from this if we calculate gene freq. for gene A = p^2 +pq = p(p+q) where (p+q)=1 therefore gene freq. of A = p like wise it will be q for gene a . In this way gene frequency do not change from parental generation to next generation produced by random mating.

⇒ The frequency of a gene in given generation depends upon the freq. of that gene if previous generation for a large random mating population in absence of mutation, migration and selection.

⇒ The genotype freq. of a generation produced by random mating among parents depend upon gene freq. of the previous generation.

⇒ H W equilibrium frequency are attained after one generation of random mating

Assumptions of HW law :

1. Population is large and random mating
2. No selection operates at any stage and all genotypes contributes equally to the pool of gametes, all zygote are equally viable
3. No immigration and emigration in the population i.e. population is closed
4. No mutation takes place or there is no differential mutation take place i.e. rate of the mutation is equal for all genes
5. Meiosis is normal

Table 2.4 :Random mating between possible genotypes in population

		Frequency in offspring generation		
Mating type	Frequency	AA	Aa	Aa
AA x AA	D^2	D^2	-	-
AA x Aa	2DH	DH	DH	-
AA x aa	2DR	-	2DR	-
Aa x Aa	H^2	¼ H^2	½ H^2	¼ H^2
Aa x aa	2HR	-	HR	HR
aa x aa	R^2	-	-	R^2
Total	$(D+H+R)^2$	$(D + ½ H)^2$	2(D + ½ H) (R+½ H)	$(R+½ H)^2$
		P^2	2pq	q^2

After allowing random mating in all possible ways among the three genotypes of single locus , the genotypic frequency in offspring will be = $(D + ½ H^2)^2$ or = p^2 for AA genotype where (D + ½ H) is gene freq. of allele A i.e. p . If the population is not in equilibrium then population is allowed to random mating then in next generation the H.W. equilibrium is reached.

Example : For a single locus following genotypic frequency were observed . What will be the gene and genotypic freq. at HW equilibrium?

Genotype	Geno. Freq.
AA	0.18 D
Aa	0.04 H
aa	0.78 R

Gene A freq. (p) = D + ½ H = 0.18 + 0.02 = 0.20 and

Gene a freq. (q) = R + ½ H = 0.78 + 0.02 = 0.80

When random mating is allowed

gemmating array = (pA + qa)

= (0.2A + 0.8a)

genotypic array =$(pA + qa)^2$

= $(0.2A + 0.8a)^2$

= p^2 AA + 2 pq Aa + q^2 aa

= (0.04) AA + 2 x 0.2 x 0.8 Aa + 0.64 aa

= 0.04 AA + 0.32 Aa + 0.64 aa→ will be the frequency of genotypes after one generation of random mating.

From these genotype if we calculate gene freq. for gene A (p) it will be = 0.04+0.16 = 0.2 and q=0.8 these freq. are same as parental generation but the genotypic frequency of this generation are the same as parental generation. Therefore, the population is not in HW equilibrium. If this generation is again allowed for random mating then the next generation will have same gene and genotypic frequency . Therefore at HW equilibrium gene frequency will be p=0.2, q =0.8 and genotypic frequency will be p^2= 0.16, 2pq = 0.32 and q^2 = 0.64.

Equilibrium at two locus : (A-a) and (B-b)

For two locus system when the population is in equilibrium we observe (ps)(qr) = (pr)(qs) where p and q are gene freq. of A and a and r and s are gene freq. of B and b. If (ps)(qr) ≠ (pr)(qs) then the population is not in equilibrium and we estimate a measure of departure from equilibrium as d

d = {(ps)(qr)} - {(pr)(qs)}

Assume that gene frequency of four genes A, a, B and b in population have frequency of 0.8,0.2,0.8 and 0.2 respectively. If the population comprise of four genotypes AABB, AABb, aaBB and aabb with their frequency as 0.70, 0.10, 0.10 and 0.10, respectively. Four kinds of gametes will be produced as given in Table 2.5.

Table 2.5 : Estimation of departure (d) form equilibrium in two locus system

Genotypes	Gametes	Initial Frequency	Frequency of gametes
AABB	AB	0.7	0.70
AAbb	Ab	0.1	0.10
aaBB	aB	0.1	0.10
aabb	ab	0.1	0.10

d = { (ps) (qr) } – {(pr) (qs) } = (0.1) (0.1) - (0.7) x (0.1) = 0.01 - 0.07 = -0.06

If value of d is positive repulsion gametes are more frequent than coupling gametes and if value of d is negative coupling gametes are more frequent than repulsion gametes. If sign of d is positive half the difference is added equally to the frequency to two kind of coupling gametes. If d is negative half the difference is added to frequency in two types of repulsion gametes and corresponding amount subtracted from two kinds of coupling gametes, this process will continue till the value of d become zero. Generally four to five generation of random matting (in absence of linkage) are required to approach the value of d to zero value. i.e. at equilibrium position.

3. Models and expected mean squares

Model : Model is defined as a system of postulates, data and inferences presented as mathematical description of an entity. Model gives complete explanation of make up of an observation under consideration. The simplest linear additive model is

$X_i = \mu + \varepsilon_i$ where X_i is an individual observation, μ = general mean and ε_i = error associated with X_i observation. ε_i are assumed to be normally independently distributed with mean =0 and variance σ^2 ($\varepsilon_i \sim$ NID 0, σ^2_e). Statistics deals with sampling technique. Sample studies are being conducted and the conclusion are drawn for the population from which the samples have been drawn. Hence, there is error tem in statistical model whereas in mathematical model there is no error term.

There are three types of statistical model

(1) Fixed effect Model (2) Random effect Mode (3) Mixed effect Model

(1) Fixed effect Model :

In this model the nature of treatments is considered to be fixed. When the same treatments are applied while repeating the experiment with same set of treatments we apply the same treatments without change in the treatment effects i.e. If we consider a set of genotypes of inbred lines in a RBD design we consider the genotypes to have fixed effect because when we repeat the same experiment we apply the treatment with same set of genotypes whose genotypic constituents are the same over repetition. When a model consist of fixed effect treatments it is known as fixed effect model or Type I model. In this model we are interested to examine treatments effects. i.e. mean of treatment comparison. Majority of field experiments are of fixed effect nature. e.g. Nitrogen levels, p levels,

(2) Random effect model :

In this model the nature of treatment effect is random i.e. when we repeat the same experiments then because treatments were randomly selected we don't have the same set of treatment. This situation is observed in random survey where selection of treatments is based on random sampling from the population of treatments. When a model consists all random sources for treatment effect, it is known as random effect model or Type II model. Generally we are interested to compare the population of treatments in this type of model.

(3) Mixed effect model :

When a model consist of fixed effect as well as random effects of treatments then it is called mixed model or Type – III model. e.g. When we carry out pooled or combined analysis of a same set of varieties tested over different years then varieties have fixed effects and years will have random effects, therefore it is an example of mixed effect model.

Rules for writing Expected Mean squares of different sources in ANOVA:

(1) The bottom most row is the source of experimental error in ANOVA and its expected mean square is symbolized as σ^2_e

(2) Appropriate subscript corresponding to the source of variation should be utilized with σ^2 e.g. for genotype σ^2_g for year σ^2_y etc.

(3) In writing out expected mean square of a given source of variation following conditions are to be observed.

(i) It should contain σ^2_e for each source of variation

(ii) It should contain variance component representing the source for which we are writing exp. mean square i.e. for variety σ^2_v must be there

(iii) Higher order components of variance will appear if following conditions are satisfied.

Ignoring the subscript of the source for which we are writing expected mean square, if the remaining subscript corresponds to random effect or effects then it will appear in that source. If the remaining subscript contain one or more letter of fixed effect then it does not appear in that source.

(4) For any component the coefficient can be determined by examining the subscript of that component. The complementary part of the subscript or the components appears as the coefficients of the source.

Example 1 : Fixed effect model :

write expected mean square for an experiment having following ANOVA

Replications = 4, Treatments = 10 (treatments effects are fixed)

Table 3.1 : Exp. mean square in randomized block design

Source of variation	df	SS	MS	Exp. Mean squares
Replication	(r-1)=3	240	80	$\sigma^2_e + t\,\sigma^2_R$
Treatment	(t-1)=9	450	50	$\sigma^2_e + r\,\sigma^2_T$
Error	(r-1)(t-1) =27	135	5	σ^2_e
Total	(rt-1)=39	835		

Test Ho: $\sigma^2_T = 0$ against Ha : $\sigma^2_T \neq 0$

The F test will be used in Anova and it must satisfy that when null hypothesis is true the ratio should be unity accordingly

$$F = \frac{\text{Treatment ms}}{\text{Error ms}} = \frac{\sigma^2_e + r\,\sigma^2_T}{\sigma^2_e}$$

Therefore, when null hypothesis is true $\sigma^2_T = 0$

$$F = \frac{\sigma^2_e + (r\,\sigma^2_T = 0)}{\sigma^2_e} = \frac{\sigma^2_e}{\sigma^2_e} = 1.0$$ Thus F test is valid for this testing.

When we carry out an experiment with g genotypes in r replications over y years with same set of g genotypes then the statistical model for the pooled analysis will be

$Y_{ijk} = m + r_i + t_j + y_k + (Y \times t)_{ijk} + e_{ijk}$ and its ANOVA will be as under

(Assume R = 4 , g = 8, Y = 3 and if all effects are assumed to be random then it will be a random model and expected mean squares will be as under)

Example 2: Random effect model

Table 3.2 : Expected mean squares in randomized block design (pooled over years)

Source of variation	d.f.	SS	MS	EMS (Random Model)
Replication (year)	(r-1) y = 9	SSR = 45	MSR = 5	$\sigma_e^2 + \text{ry}\,\sigma^2_R$
Genotypes	(g-1) = 7	SSG = 280	MSG = 40	$\sigma_e^2 + \text{r}\,\sigma^2_{YG} + ry\sigma^2_G$
Year	(y-1) = 2	SSY = 350	MSY = 175	$\sigma_e^2 + r\sigma^2_{YG} + rg\sigma^2_Y$
(y x g)	(g-1) (y-1) = 14	SSYG = 98	MSYG = 7	$\sigma_e^2 + r\sigma^2_{YG}$
Error	y(r-1)(g-1) = 63	SSE = 63	MSE = 1	σ_e^2
Total	(rgy-1) = 95			

SSR = sum of replication SS of individual year
SSE = sum of error SS of individual year

(1) The appropriate test for Ho: $\sigma^2_{YG} = 0$ will be
F =(MSYG / MSE)

(2) The appropriate test for Ho: $\sigma^2_Y = 0$ will be
F = (MSY / MSYG)

(3) The appropriate test for Ho: $\sigma^2_G = 0$ will be

F = (MSG / MSYG)

In Type – II model, mean square of a source of variation will be divided by its higher order interaction for valid F Test. Here mean square of genotype is a component of σ^2_e , σ^2_{YG} and σ^2_G while it is a component of σ^2_e and σ^2_G in case of fixed effect model.

Mixed effect model :

Generally genotypes are always fixed in plant breeding trials and they are tested over different years or seasons and hence pooled analysis consist genotype as fixed effect, year or season as random effect and thus will be an example of mixed effect model or type III model.

Table 3.3 : Expected mean squares for mixed effect model in RBD

Source of variation	d.f.	SS	MS	EMS (Mixed Model)
Replication (year)	(r-1) y = 9	SSR = 45	MSR = 5	$\sigma_e^2 + \text{gy}\ \sigma_R^2$
Genotypes	(g-1) = 7	SSG = 280	MSG = 40	$\sigma_e^2 + \text{r}\ \sigma_{YG}^2 + ry\sigma_G^2$
Year	(y-1) = 2	SSY = 350	MSY = 175	$\sigma_e^2 + rg\sigma_Y^2$
(y x g)	(g-1) (y-1) = 14	SSYG = 98	MSYG = 7	$\sigma_e^2 + r\sigma_{YG}^2$
Error	y(r-1)(g-1) = 63	SSE = 63	MSE = 1	σ_e^2
Total	(rgy-1) = 95			

(1) The appropriate test for Ho: $\sigma^2_{YG} = 0$ will be

F = (MSYG / MSE)

(2) The appropriate test for Ho: $\sigma^2_Y = 0$ will be

F = (MSY / MSE)

(3) The appropriate test for Ho: $\sigma^2_G = 0$ will be

F = (MSG / MSYG)

If one wish to estimate the pure component of genotypic variance then it can be estimated as under

r = 4 , g = 8 and y =3 and Ms value of genotype = 40

MSG = $\sigma^2_e + r\sigma^2_{YG} + ry\sigma^2_G$ where $\sigma^2_e + r\sigma^2_{YG} = 7$

$40 = 7 + (4x3)\ \sigma^2_G$

$\{(40 – 7) / 12\} = \sigma^2_G$

Thus we find the pure component of genetic variance from which we can calculate genotypic coefficient of variation and phenotypic coefficient of variation.

Genotypic Coefficient of Variation (GCV) (%) = $\frac{\sqrt{\sigma_G^2}}{\overline{Y}} \times 100$

where σ^2_G = genotypic variance

$\overline{Y}$ = general mean

GCV is a measure of variability existing at genes level.

Phenotypic Coefficient of Variation (PCV) (%) = $\frac{\sqrt{\sigma_P^2}}{\overline{Y}} \times 100$

where σ^2_P = Phenotypic variance

$\overline{Y}$ = general mean

PCV is a measure of variability existing at phenotypic level i.e. sum total of genetic and environmental influence. In general the PCV values are higher then the GCV values because $\sigma^2_p = \sigma^2_G + \sigma^2_e$. The genotypic variance can further be partitioned in to Additive variance (σ^2_A), Dominance variance (σ^2_D) and Interaction variance or epistatic variance (σ^2_I). i.e.

$$\sigma_G^2 = \sigma_A^2 + \sigma_D^2 + \sigma_I^2$$

4. Population mean

There are two types of characters – Quantitative and qualitative characters..

Quantitative characters :

They are governed by polygene and show continuous variation and follow continuous distribution. Each gene has small and cumulative effects which can not be separated out foe each gene.

Qualitative characters:

They are govern by single or few genes and the show discrete variation and follow discrete distribution. Their effects are distinct and easily classified in to groups.

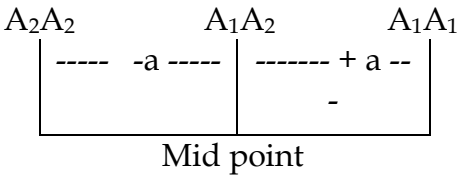
We quantify the phenotypic value of quantitative characters in different metrical units and we can estimate their average values for the whole population which is known as population mean for that character. In genetics the population mean is estimated on the basis of the effects of genes governing that particular character. Phenotype observations are measured from the population which are the sum total of genetic effects and environmental deviation .

Phenotypic value (P) = Genotypic value (G) + environmental deviations (E)
Thus,

P = G + E and $\overline{p} = \overline{G + E}$

The environmental agencies includes the non heritable variation and their deviations are positive or negative and the sum of deviations will therefore be zero. $\overline{E} = \mathbf{0}$ and thus the entire population mean is dependent on genotypic value. i.e. mean phenotypic value and mean genotypic values will be equal. So if gene frequency gets changed then only value of p is changed. It is difficult to estimate separate effect of genes governing the quantitative character. Let us take an example of single gene pair (A_1, A_2)governing the quantitative character. Let us take

(A_1A_2 - Mid Point) = +a
(A_2A_2 - mid point) = -a
(A1A2-midpoint) = d



allele A_1 tends to increase phenotypic value

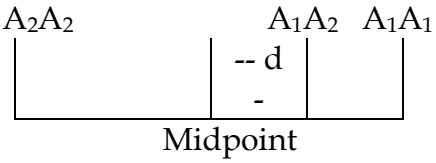
allele A_2 tends to decrease phenotypic value

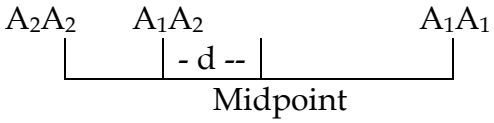
d measures the departure from mid point value i.e. departure of gene effect from additivity at a locus

If a hybrid is produced by cross of A_1A_1 x A_2A_2 and if its value is exactly at the mid point of the parental values then we say that there is additive effect and

$$\text{Pheno. value } A_1A_2 = \frac{A1A1 + A2A2}{2}$$

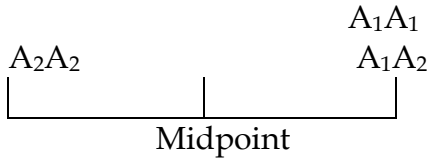
Partial Dominance





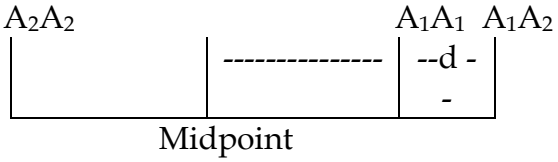
If the value of A_1A_2 is not at the mid point and between mid point to A_1A_1 or mid point to A_2A_2 then there is partial dominance.

Complete Dominance



If the value of A_1A_2 is $=A_1A_1$ value then it is complete dominance

Over dominance



If value of A_1A_2 is higher than A_1A_1 then it is over dominance.

Let us consider an example of single locus having three genotypes with their genotypic frequency as p^2 , 2pq and q^2 for A_1A_1 A_1A_2 and A_2A_2 respectively and it is governing a character. Then population mean can be estimated as under

Table 4.1 Population mean for a single locus

Genotype	Frequency	Genotypic Value (X)	fx
A_1A_1	P^2	+ a	P^2a
A_1A_2	2pq	d	2pqd
A_2A_2	q^2	-a	$-q^2a$

As statistically mean for grouped data can be estimated as

$\overline{X} = \Sigma f_i X_i / \Sigma f_i$ in similar one can find out population mean as under

$\Sigma f_i X_i = P^2a + 2pqd - q^2a$
$= P^2a - q^2a + 2pqd$
$= a(p^2-q^2) + 2pqd$
$= a(p+q)(p-q) + 2pqd$
$= a(p-q) + 2pqd$

Σf_i the sum of frequency = $p^2+2pq+q^2 = 1.0$ therefore, $\Sigma f_i x_i$ = population mean which can be expressed in terms of gene frequency as under

Population mean = = a (p-q) + 2pqd where a(p-q) is the contribution from homozygous parents (additive effect) and 2pqd is the contribution from heterozygous (dominance effect) genotypes in single locus system.

If there is no dominance (d=0)

Then $\overline{X}$ = a(p-q) + 2pqd where d=0 no dominance
= a (p-q)
= a (p-1+p)
= a (2p-1) or q (1-2q)

Example

Genotype	No. of Plant	X geno value	Proportion/	Geno. freq.	f_i	$f_i x_i$
A_1A_1	252	16	252/700	0.36	p^2	5.76
A_1A_2	336	12	336/700	0.48	2pq	5.76
A_2A_2	112	4	112/700	0.16	q^2	0.64
Total	700			1.00		12.16

gene frequency A_1 (P) = D + ½ H = 0.36 + 0.24 = 0.6
A_2 (q) = R + ½ H = 0.16 + 24 = 0.4
(p+q) = 1.0

Mid parent value = $\frac{16+4}{2}$ = 10.0

a = A_1A_1 – Mid parent value = 16-10 = 6
d = A1A2- Mid parent value = 12-10 = 2
-a = A2A2- Mid parent value = 4-10 = -6

Population mean = a(p-q) + 2pqd
= 6 (0.6 -0.4) + 2 (0.6) (0.4) (2)
= 6 (0.2) + 4 (0.24)
= 1.2 + 0.96 = 2.16

As the population mean is estimated as the deviation of mid parent value the original population mean will be

Population mean = Mid parent value + population mean
= 10 + 2.16 = 12.16

If there is no dominance, Population mean = a (p-q)
= 6 (0.6 - 0.4) = 1.2

Population mean = Mid parent value + population mean
= 10 + 1.2 = 11.20

If in same example if gene A_1 frequency increases to 0.7 then what will be the population mean. ? Let us examine that when p=0.7 then q=0.3 therefore

Population mean = a (p-q) + 2 pqd
= 6 (0.7 -0.3) + 2 (0.7) (0.3) (2)
= 6 (0.4) + 4 (0.21)
= 2.4 + 0.84
= 3.24

Population mean = 10 + 3.24
= 13.24

If gene frequency A_1 is reduced to 0.5 what will be the population mean ?

Population mean = 6 (0.5-0.5) + 2 (0.5) (0.5) (2)
= 0 + 4 (0.25)
= 1.0

Population mean = 10 + 1.0
= 11

These results are summarized as under

Frequency		Mean
p = 0.6	q = 0.4	12.16
p = 0.7	q = 0.3	13.24
p = 0.5	q=0.5	11.00

Result given in above table regarding mean and gene frequency indicates that as gene A1 frequency increases the population mean is also increases thus gene A1 has increasing effect and selection in favour of gene A1 will have higher genetic advance while favouring gene A2 the genetic advance will be in negative direction. i.e. gene A2 has decreasing effect .

For the characters governed by poly genes (Quantitative characters), Population mean will be as under

Population mean = $\sum ai(Pi - qi) + \mathbf{2}\sum Piqidi$

i = number of gene responsible for a character under study or number of locus responsible to manufacture a quantitative characters

When a characters is governed by 5 gene → polygenes we can not separate out the effect due to each gene separately.

Example :

A population having genotype frequency A_1A_1 = 0.02 A_1A_2 = 0.36 and A_2A_2 = 0.62

1 Find out the gene frequency at equilibrium

2 If gene frequencies increase as P = 0.3, what will be happen the population mean

5. Components of genetic variance :

The phenotypic variance $\sigma^2_P = \sigma^2_G + \sigma^2_E$ where σ^2_G is genotypic variance. In 1918 Fisher conceived that genotypic variances (σ^2_G)can be partitioned in three components (1) additive portion (σ^2_A)arising due to differenced between homozygotes (2) dominance part ($\sigma^2_{D)}$ arising due to intra-allelic interaction and (3) an epistatic component (σ^2_I)arising from inter-allelic interactions.

Thus, $\sigma^2_G = \sigma^2_A + \sigma^2_D + \sigma^2_I$ i.e. additive, dominance and epistatic variances respectively. And $\sigma^2_P = \sigma^2_A + \sigma^2_D + \sigma^2_I + \sigma^2_E$.

How to find out the value of these variances . Let us examine it with respect to a single locus system .

(1) Breeding value of Genotypes :

Parents passes their genes to the offspring. Average effects of genes of parents determine the mean genotypic value of offspring. Value of an individual judged by the mean value of its progeny is known as **breeding value** of that individual. It can be defined in absolute unit but it is more convenient to define it as deviation from population mean. The sum of square of breeding value and genotypic frequency is known as additive variance.

Table 5.1 Breeding value of genotypes in single locus system

Genotype	Frequency(f)	Breeding Value(BV)	f. (BV)	$(BV)^2$ x f
A_1A_1	P^2	$2q\alpha$	$2p^2q\alpha$	$P^2(4q^2\alpha^2)$
A_1A_2	2pq	$(q-p)\ \alpha$	$2pq(q-p)\ \alpha$	$(q-p)^2\alpha^2 2pq$
A_2A_2	q^2	$-2p\alpha$	$-2q^2p\alpha$	$(4p^2\alpha^2)q^2$
			$\sum(f\ x\ BV) = \mathbf{0}$	$\sum f\ x\ (BV)^2 = \sigma^2{}_A$

Σ(f x BV) = $2p^2q\alpha + 2pq\alpha\ (q-p) - 2pq^2\alpha$

= $2p^2q\alpha + 2pq^2\alpha - 2p^2q\alpha - 2pq^2\alpha$

= $2p^2q\alpha - 2p^2q\alpha + 2pq^2\alpha - 2pq^2\alpha$

= 0 (the sum of product of Frequency and breeding value)

If gametes produced by genotypes bearing gene A1 unite randomly with gametes in the population then there will be two types of genotypes produced i.e. A_1A_1, A_1A_2 with p and q as their frequencies.

Table 5.2 : Mean of genotypes in offspring produced by gene A1

Offspring	freq.(f)	geno. value (x)	fx
A_1A_1	P	a	Pa
A_1A_2	q	d	qd
Total	1.0		(pa+qd)

The Mean of offspring will be = (pa + qd) and

Entire population mean = a (p – q) + 2 pqd

α = Av. effect of gene substitution i.e. which can be estimated as under

α = Off spring mean – population mean

$$\alpha = (pa + qd) - [a(p - q) + 2pad)$$

= ~~pa~~ + qd – ~~ap~~ + aq – 2pqd

= q (a + d – 2 pd)

= q (a + d (1-2e)

= q (a + d (1-p-p))

$$\alpha_1 = q(a + d(q - p))$$

α_1 = effect of A_1 allele and in similar way for A2 it will be α_2

$$\alpha_2 = -p(a + d(q - p))$$

α_2 = effect of A_2 allele.

Thus average effect of gene substitution α can be estimated as

$$\alpha = \alpha_1 - \alpha_2$$

= q (a+d(q-p)) – (-p(a+d (q-p))

= a + d (q-p)

We assume the value of p = 0.6 and q = 0.4 a = 6 and d =2 therefore

α = 6 + 2 (0.4 – 0.6)

= 6 – 0.4 = 5.6

breeding value of different genotypes will be

2qα = 2 x 0.4 x 5.6 = 112.4 = 4.48

(q- p)α = (0.4 – 0.6) x 5.6 = -1.12

-2pα = -2x (0.6) x 5.6 = -6.72 substituting these values

Table 5.3 : Estimation of additive variance

Geno. Freq.(f)	Breeding Value (BV)	f(BV)	$(BV)^2$ x f
$P^2 = 0.36$	2qα = 4.48	0.36 x 4.48 =1.6128	$(4.48)^2$ x 0.36 = 7.2253
2pq= 0.48	(q-p)α = - 1.12	0.48 x 1.12 = -0.5376	$(-1.12)^2$ x 0.48 = 0.6021
q^2= 0.16	-2qα = - 6.72	0.16 x -6.72 = -1.0752	(-6.72) x 0.16 = 7.2233
		=0.0000	σ^2A = 15.0527 Single locus

$\Sigma f (BV)^2$ = 7.2253 + 0.6021 + 7.2233 = 15.0527 = σ^2A

(2) Dominance deviation :

Non additivity of two alleles at a locus is the cause of dominance deviation. The difference between genotypic value of a genotype and its breeding value is called dominance deviation (DD) provided a single locus is under consideration

Table 5.4 : Dominance deviation for genotypes in single locus system

Genotype	Frequency(f)	Dominance deviation (DD)	f. (DD)	$(DD)^2$ x f
A_1A_1	P^2	$-2q^2d$	$2p^2d^2d$	$P^2(4q^2\alpha^2)$
A_1A_2	2pq	-2pqd	$4p^2q^2d$	$(q-p)^2\alpha^2 2pq$
A_2A_2	q^2	$-2p^2d$	$-2p^2q^2d$	$(4p^2\alpha^2)q^2$
			$\sum(f\ x\ DD) = 0$	$\sum f\ x\ (DD)^2 = \sigma^2{}_D$

- If d = 0 then genotypic value of a genotype is breeding value of that genotype.
- There is no correlation between (BV & DD)

We assume the value of p = 0.6 and q = 0.4 a = 6 and d =2 (same as in breeding value calculation)

Table 5.5 : Estimation of dominance deviation for genotypes in single locus system

Geno type	Freq Uency (f)	Domin ance deviation (DD)	f. (DD)	$(DD)^2$ x f
A_1A_1	0.36	-0.64	0.36 x -0.64 = -0.2304	$(-0.64)^2$ x 0.36 = 0.1475
A_1A_2	0.48	0.96	0.48x0.96 = 0.4608	$(0.96)^2$ x 48 = 0.4424
A_2A_2	0.16	-1.44	0.16x-1.44 = -0.2304	$(-1.44)^2$ x 0.16 = 0.3318
			$\sum(f\ x\ DD) = 0$	$\sum f\ x\ (DD)^2 = \sigma^2{}_D$ = 0.9217

$-2q^2d = -2(0.4)^2 \times 2$
$= -2\ (0.16) \times 2$
$= -0.64$

$2pqd = 2 \times 06 \times 0.4 \times 2$
$= 0.96$

$-2p^2d = -2 \times 0.36 \times 2$
$= -1.44$

We assume the value of p = 0.6 and q = 0.4 a = 6 and d =2 (same as in breeding value and dominance deviation calculation). The dominance variance (σ^2_D) will be = 0.9217

(3) Interaction Deviation :

If we consider two locus A and B the genotypic value at locus A will be G_A and for B it will be G_B and if there is absence of non allelic interaction then the aggregate genotypic value will be $G = G_A + G_B$. If there exist non allelic interaction between two locus then non additivity between two locus will take place which is a component of genotypic value and can be written as $G = G_A + G_B + I_{AB}$. Non additivity means intra allelic interaction within a locus and interallelic interaction(epistasis).

(4) Genotypic variance

Table 5.6 : Estimation of genotypic variance in single locus system

Genotype	Frequency(f)	Genotypic value (X)	$(X_i \overline{X})$	$f(X_i \overline{X})^2$
A_1A_1	0.36	16	16 – 12.16 =3.84	5.3084
A_1A_2	0.48	12	12-12.16 = -0.16	0.0123
A_2A_2	0.16	4	4-12.16 = -8.16	10.6537
Total	1.00			15.9774

$\overline{X}$ = (Σ fx / Σf)

= [(0.36x16)+(0.48x12)+(0.16x4)}/1.0

= 5.76 + 5.76 + 0.64

= 12.16

$\sigma^2_G = \Sigma f(X_i \overline{X})^2$ because genotypic freq. are expressed as proportion and thus

σ^2_G = 15.9774 is the value of genotypic variance

Let us examine

$\sigma^2_G = \sigma^2_A + \sigma^2_D + \sigma^2_I$

15.9744 = 15.0527 + 0.9217 + 0 = 15.9744 we can see that there is no interaction because of single locus the interaction variance is taken as zero and thus we can say that

$\sigma^2G = \sigma^2A + \sigma^2D + \sigma^2I$

Example :

A population have gene frequency A_1A_1 = 0.02 A_1A_2 = 0.36 and A_2A_2 = 0.62 and if genotypic values for A_1A_1 = 28, A_1A_2 = 22 and A_1A_2 = 12 then find the population mean, additive variance, dominance deviation and genotypic variances.

Components of genotypic variance in two locus system :

Using the concept of factorial design Cockerham (1954) worked out three components of genotypic variances. He explained the theory by considering two locus having A/a and B/b gene pair. In this case there will be nine different genotypes as under

Table 5.7 Different genotypes and their genotypic values in two locus system

Locus A	**Locus B**		
	$BB(b_0)$	$Bb(b_1)$	$Bb(b_2)$
$AA(a0)$	AABB (3)	AABb (2)	AAbb (1)
$Aa(a_1)$	AaBB (3)	AaBb (2)	Aabb (1)
$Aa(a_2)$	aaBB (1)	aaBb (1)	Aabb (1)

( ) indicates the genotypic value

two locus were considered as two factor and each locus will have three genotypes in the population i.e. AA, Aa or aa and BB, Bb and bb. Each gene pair has two types of effects i.e. additive (linear) and dominance (quadratic) and

interaction between genes will give rise to epistatic effect. The theory is explained in following table.

Table 5. 8 : Estimation of genetic variance in two locus system

Genotypes	Freq.	Gene Value	f_1x_1	$f_1x_1^2$
AABB	1/16	3	3/16	9/16
AABb	2/16	2	4/16	8/16
AAbb	1/16	1	1/16	1/16
AaBB	2/16	3	6/16	18/16
AaBb	4/16	2	8/16	16/16
Aabb	2/16	1	2/16	2/16
aaBB	1/16	1	1/16	1/16
aaBb	2/16	1	2/16	2/16
Aabb	1/16	1	1/16	1/16
			Σf_1x_1 28/16	$f_1x_1^2$ 58/16

$S^2 = \Sigma f_1x_1^2 - (\Sigma f_1x_1)^2 = 58/16 - (28/16)^2 = 9/16$ as X_i are the genotypic value the values of $S^2 = \sigma^2_G = 9/16 = \sigma^2_A + \sigma^2_D + \sigma^2_I$

Table 5.9 :ANOVA as two way factorial design

Factors	Source	Symbol	d.f.	σ^2 component
A	Linear (additive gene effect)	A_L	1	σ^2A_L
	Quadratic (dominance deviation)	A_Q	1	σ^2A_Q
	Total for A-a locus		2	$\sigma^2L_1 = \sigma^2A_L + \sigma^2A_Q$
B	Linear (additive gene effect)	B_2	1	σ^2B_L
	Quadratic (dominance deviation)	B_Q	1	σ^2B_Q
	Total for B-b locus		2	$\sigma^2L_2 = \sigma^2B_L + \sigma^2B_Q$
A x B	Lin x Lin (add. x add)	A_LB_L	1	$\sigma^2 A_LB_L$
	Lin x Qua.(add x domi.)	A_LB_Q	1	$\sigma^2 A_LB_Q$
	Qua x Lin.(domi.x add)	A_QB_L	1	$\sigma^2 A_QB_L$
	Qua x Qua.(domi x domi.)	A_QB_Q	1	$\sigma^2 A_QB_Q$
	Total epistasis		4	$\sigma^2I = \sigma^2A_LB_L + \sigma^2A_LB_Q + \sigma^2A_QB_L + \sigma^2A_QB_Q$
	Overall total		8	$\sigma^2G = \sigma^2L_1 + \sigma^2L_2 + \sigma^2I$

As there are 9 genotypes the total degree of freedom will (9-1) = 8 .The components of additive, dominance and epistasis variances can be estimated using orthogonal polynomial coefficients as under

Table 5. 10 : Orthogonal polynomial coefficient (ξ_I) for different genotypes in two locus system

Geno.	Gen. freq. (f)	Orthogonal polynomial coefficients (ξ_I)							
		A_L	B_L	A_Q	B_Q	A_LB_L	A_LB_Q	A_QB_L	A_QB_Q
AABB(a_0b_0)	1/16	1	1	1	1	1	1	1	1
AABb(a_0b_1)	2/16	1	0	1	-1	0	-1	0	-1
AAbb(a_0b_2)	1/16	1	-1	1	1	-1	1	-1	1
AaBB(a_1b_0)	2/16	0	1	-1	1	0	0	-1	-1
AaBb(a_1b_1)	4/16	0	0	-1	-1	0	0	0	1
Aabb (a_1b_2)	2/16	0	-1	-1	1	0	0	1	-1
aaBB(a_2b_0)	1/16	-1	1	1	1	-1	-1	1	1
aaBb(a_2b_1)	2/16	-1	0	1	-1	0	1	0	-1
aabb(a_2b_2)	1/16	-1	-1	1	1	1	-1	-1	1
$\Sigma f_i\xi^2_i$ (Divisor)		½	½	1	1	¼	½	½	1

Contrast/orthogonal polynomial for single locus

	AA or BB (Homo.)	AAU or Bb (Hetero.)	aa or bb (Homo.)
Linear effect of a gene (Additive)	1	0	-1
Quadratic effect of a gene (dominance)	1	-1	1

The divisor (D)for each component can be estimated as

D = Σ f_i ξ^2_i where f_i is the frequency of genotype and ξ_i is the orthogonal polynomial contrast of different genotypes

For A_L it will be =

$$(1/16)(1^2)+(2/16)(1^2)+(1/16)(1^2)+(2/16)(0)+(4/16)(0)+(2/16)(0)+(1/16)(1)^2+(2/16)(-1)^2+(1/16)(-1)^2$$

=1/2

In similar way we can find out for all eight sources of variation. These values are given in table as $\Sigma f_i\xi^2_i$

We can find out effect of each source as under

Effect of A_L = 1/16 (a_0-a_2) (b_0+2b_1+ b_2)

= 1/16 (a_0b_0+2a_0b_1+a_0b_2 – a_2b_0 - 2a_2b_1- a_2b_2) – substitute the genotypic value from table 5.7 for the genotypes

= 1/16 (3 + 2(2) + 1 -1-2(1) -1)

= 1/16 (8-4) = 4/16 = ¼

Then $\sigma^2A_L = \dfrac{(\mathbf{A_L\ effect})^2}{\mathbf{Divisor\ of\ A_L}} = \dfrac{(1/4)^2}{1/2} = \dfrac{1/16}{1/2} = 1/8$

Effect of A_Q = $\frac{1}{16}$ $(a_0 - 2a_1 + a_2)(b_0 + 2b_1 + b_2)$

= $\frac{1}{16}$ $(a_0b_0 + 2a_0b_1 + a_0b_2 - 2a_1b_0 - 2a_1b_0 - 4a_1b_1 - 2a_1b_2)$

= $\frac{1}{16}$ $(3 + 2(2) + 1 - 2(3) - 4(2) - 2(1)$

= $\frac{1}{16}$ $(8 - 16 + 4) = -\frac{4}{16} = -\frac{1}{4}$

$$= \sigma^2 A_Q = \frac{(A_Q \ effect)^2}{\textbf{Divisior of } \mathbf{A_Q}} = (\tfrac{1}{4})^2 / 1 = \tfrac{1}{16}$$

So for the single locus having gene A-a the total variation will be

$$\sigma^2 L_1 = \sigma^2 A_L + \sigma^2 A_Q = 1/8 + 1/16 = \frac{2+3}{16} = \frac{3}{16}$$

where $\sigma^2 A_L$ indicates the contribution of additive variance and $\sigma^2 A_Q$ indicate contribution of dominance variance due to gene pair A-a . In same way we can estimate for gene pair B-b which will be

Effect of B_L = $\frac{1}{16}$ $(b_0 - b_2)(a_0 + 2a_1 + a_2)$

= $1/16(a_0b_0 + 2a_1b_0 + a_2b_0 - a_0b_2 - 2a_1b_2 - a_2b_2)$

= $1/16(3 + 2(3) + 1 - 1 - 2(1) - 1)$

= $1/16\ (6)$

= $3/8$

$$\text{Then } \sigma^2 B_L = \frac{(\mathbf{B_L \ effect})^2}{\textbf{Divisor of } \mathbf{B_L}} = \frac{(3/8)^2}{1/2} = \frac{9/64}{1/2} = \tfrac{9}{32}$$

Effect of B_Q = $\frac{1}{16}$ $(b_0 - 2b_1 + b_2)(a_0 + 2a_1 + a_2)$

= $1/16\ (a_0b_0 + 2a_1b_0 + a_2b_0 - 2a_0b_1 - 4a_1b_1 - 2a_2b_1 + a_0b_2 + 2a_1b_2 + a_2b_2)$

= $1/16(3 + 2(3) + 1 - 2(2) - 4(2) - 2(1) + 1 + 2(1) + 1)$

= $1/16\ (0)$

= 0

$$\text{Then } \sigma^2 B_Q = \frac{(\mathbf{B_Q \ effect})^2}{\textbf{Divisor of } \mathbf{B_Q}} = \frac{(0)^2}{1} = \frac{0}{1} = 0$$

we can find that

$\sigma^2 L_2 = \sigma^2 B_L + \sigma^2 B_Q$

$\sigma^2 L_2 = 9/32 + 0 = 9/32$

Now in two locus system there will be interaction between the gene pair. Their variance can be estimated as

Effect of $A_LB_L = \frac{1}{16}(a_0-a_2)(b_0 - b_2)$

$= 1/16\,(a_0b_0 - a_0b_2 - a_2b_0 + a_2b_2)$

$= 1/16(3-1-1+1)$

$= 2/16 = 1/8$

$$\sigma^2A_LB_L = \frac{\textbf{(Effect of } \mathbf{A_LB_L})^2}{\textbf{Divisor of } \mathbf{A_LB_L}} = \frac{1/8^2}{1/4} = \frac{1}{64}x\frac{4}{1} = \frac{1}{16}$$

Effect of $A_LB_Q = \frac{1}{16}(a_0 - a_2)(b_0 - 2b_1 + b_2)$

$$= \frac{1}{16}(a_0b_0 - 2a_0b_1 + a_0b_2 - a_2b_0 - 2a_2b_1 - a_2b_2)$$

$$= \frac{1}{16}(3-2\,(2) + 1 -1 +2\,(1) -1) = = \frac{1}{16}(0) = 0$$

$$\sigma^2A_LB_Q = \frac{\textbf{(Effect of } \mathbf{A_LB_Q})^2}{\textbf{Divisor of } \mathbf{A_LB_Q}} = \frac{(0)^2}{1/4} = \frac{0}{1/4} = 0$$

$\therefore\ \sigma^2A_LB_Q = 0$

Effect of $A_QB_L = \frac{1}{16}(a_0-2a_1 + a_2)(b_0 - b_2)$

$$= \frac{1}{16}(a_0b_0 - a_0b_2 - 2a_1b_0 + 2a_1b_2 + a_2b_0 - a_2b_2)$$

$$= \frac{1}{16}(3-2\,(3) + 1 -1 +2\,(1) -1)$$

$$= \frac{1}{16}(-2) = -\frac{1}{8}$$

$$\therefore\ \sigma^2A_QB_L = \frac{\textbf{(Effect of } \mathbf{A_QB_L})^2}{\textbf{Divisor of } \mathbf{A_QB_L}} = \frac{-(1/8)^2}{1/2} = \frac{1}{64}x\frac{2}{1} = \frac{1}{32}$$

Effect of $A_QB_Q = \frac{1}{16}(a_0-2a_1 + a_2)(b_0 - 2b_1 + b_2)$

$$= \frac{1}{16}(3-2\,(2) + 1 -2\,(3) + 4(2) - 2(1) + 1-2\,(2) +1)$$

$$= \frac{1}{16}(3-4+3-6 + 8 -2 +1 =2+1)$$

$= 0$

$\therefore \sigma^2 A_Q B_Q = 0$

So, $\sigma^2_I = \sigma^2 A_L B_L + \sigma^2 A_L B_Q + \sigma^2 A_Q B_L + \sigma^2 A_Q B_Q$

$$= \frac{1}{16} + 0 + \frac{1}{32} + 0 = \frac{3}{32}$$

The genotypic variance value estimated in table 5.7 = 9/16 and we can examine that for a two locus system the sum of additive , dominance and epistatic variance i.e. $\sigma^2_G = \sigma^2 L_1 + \sigma^2 L_2 + \sigma^2_I$ or

$$\sigma^2_G = \sigma^2_A + \sigma^2_D + \sigma^2_I$$

Where $\sigma^2_A = \sigma^2 A_L + \sigma^2 B_L$

$= 1/8 + 9/32$

$= 13/32$

Where $\sigma^2_D = \sigma^2 A_Q + \sigma^2 B_Q$

$= 1/16 + 0$

$= 1/16$

Where $\sigma^2_I = \sigma^2 A_L B_L + \sigma^2 A_L B_Q + \sigma^2 A_Q B_L + \sigma^2 A_Q B_Q$

$= 1/16 + 0 + 1/32 + 0$

$= 3/32$

Now total gene effect $= \sigma^2_A + \sigma^2_D + \sigma^2_I$

$$\sigma^2_G = \frac{13}{32} + \frac{1}{16} + \frac{3}{32} = \frac{18}{32} = \frac{9}{16}$$

6. Components of co-variances and estimation of correlation

For a single character we calculate variance which is a measure of variation. Like wise co-variation can be estimated between two characters which is known as co-variance . Let us take an example from an randomized block design from which we have recorded two characters i.e. X and Y . One can analyze the data of X and Y character separately with ANOVA technique. The procedure to estimate the covariance from these two character is as under :

Procedure for estimation of co-variances in RBD

1) Correction Factor (CF) = $\frac{x..y..}{r.t}$ where X.. and Y.. are grand total

Where i = 1,2,.....r replications and j= 1, 2,.....g genotypes

2) Total Sum of cross product (SCP) with (rt-1) df = $\sum_{i.j}^{r.g}\{(X_{ij}Y_{ij}) - CF$

3) Replication sum of cross product (RSCP) with (r-1) df = $\frac{\sum_{i=1}^{r}((x_{i.})(y_{i.}))}{g} - CF$

4) Genotype sum of cross product (GSCP) with (g-1) df = $\frac{\sum_{j=i}^{g}(x_{.j} - y_{.j})}{r} - CF$

5) Error SCP (r-1) (g-1) = Total SCP – Repl. SCP – Geno. SCP

Note : The sum of cross products may be positive or negative

Table 6.1 : ANCOVA

Source	d.f.	SCP	MSCP	Expected Mean Sum of Cross Product
Replications	2	RSCP	RSCP/df =RMSCP	σe_{xy} + gσr_{xy}
Genotypes	7	GSCP	GSCP/df = GMSCP	σe_{xy} + rσg_{xy}
Error	14	ESCP	ESCP/df = EMSCP	σe_{xy}

MSCP of genotype = σe_{xy} + rσg_{xy}

we can estimate genotypic covariance between X and Y character as

GMSCP = σe_{xy} + rσg_{xy} substituting value of error SCP

{(GMSCP – ESCP) / r } = σg_{xy}

σg_{xy} is genotypic covariance and phenotypic covariance (σp_{xy})will be = σg_{xy} + σe_{xy}

Example : From a varietal trial following information is available

Table 6.2 : ANCOVA table

Source	d.f.	SS		SCP	MS		MSCP
		X	Y		X	Y	
Replications	2	214.8	6.6	74.9	107.4	3.3	37.5
Genotypes	7	3946.0	130.4	387.1	563.7	18.6	55.3
Error	14	1177.9	34.9	134.4	84.1	2.5	9.6

Genotypic variance ($\sigma^2 g$)
For X character ($\sigma^2 g_x$) = (MSG-MSE)/r = (563.7-84.1)/3 = 159.9
For Y character ($\sigma^2 g_y$) = (MSG-MSE)/r = (18.6-2.5)/3 = 5.4
Phenotypic variance ($\sigma^2 p$)
For X character ($\sigma^2 p_x$) = $\sigma^2 g_x + \sigma^2 e_x$ = 159.9 + 84.1 = 244.0
For Y character ($\sigma^2 p_y$) = $\sigma^2 g_y + \sigma^2 e_y$ = 5.4 + 2.5 = 7.9
Genotypic co-variance ($\sigma g_x g_y$) = (55.3 – 9.6)/3 = 15.2
Phenotypic co-variance ($\sigma p_x p_y$) = 15.2 + 9.6 = 24.8
These covariance are used to estimate correlations and co-heritability as under

Correlation coefficients
1) Simple correlation(r) : The degree of association between two characters is measured as simple correlation coefficient. Its limit are -1 to +1.

$$r_{xy} \text{ or } r_{12} = \frac{\sum xy}{\sqrt{\sum x^2 \sum y^2}} or \frac{\text{cov}.xy}{\sqrt{\text{var}\, x \,\text{var}\, y}}$$

$$\sum xy = \sum xy - \frac{\sum x \sum y}{n}$$

$$\sum x^2 = \sum x^2 - \frac{(\sum x)^2}{n} \qquad \sum y^2 = \sum y^2 - \frac{(\sum y)^2}{n}$$

In genetics three different types of correlations are worked out. They are

(1) Genotypic correlation coefficient ($rg_i g_j$) :

This may result from pleiotropic effect of gene or linkage of genes governing inheritance of two or more characters. Genotypic correlation indicates the degree of association between two characters under study at genetic level. Positive correlation indicates that genes which tends to increase the value of X

character also increases the value of Y character and negative correlation means genes which tends to increase the value of X character decreases the value of Y character in the population.

It can be estimated as under

$$r_{g_i g_j} = \frac{\sigma g_i g_j}{\sqrt{(\sigma^2 g_i)(\sigma^2 g_j)}}$$

$\sigma g_i g_j$ = genotypic co-variance between i^{th} and j^{th} character

and $\sigma^2 g_i$ and $\sigma^2 g_j$ are the genetic variance of i^{th} and j^{th} character

The value of genetic co-variance may be positive or negative and accordingly genotypic correlation will have positive or negative sign.

(2) Phenotypic correlation coefficient (rp_ip_j):

Phenotypic correlation indicates the degree of association between two characters under study at phenotypic level which includes both genotypic and environmental influence of the characters .

$$r_{p_i p_j} = \frac{\sigma p_i p_j}{\sqrt{(\sigma^2 p_i)(\sigma^2 p_j)}}$$

$\sigma p_i p_j$ = phenotypic co-variance between i^{th} and j^{th} character

and $\sigma^2 p_i$ and $\sigma^2 p_j$ are the phenotypic variance of i^{th} and j^{th} character

The value of phenotypic co-variance may be positive or negative and accordingly phenotypic correlation will have positive or negative sign.

(3) Environmental correlation. (re_ie_j):

It indicates degree of association between the environments under which the two characters are developed

$$r_{e_i e_j} = \frac{\sigma e_i e_j}{\sqrt{(\sigma^2 e_i)(\sigma^2 e_j)}}$$

$\sigma e_i e_j$ = environmental co-variance between i^{th} and j^{th} character

and $\sigma^2 e_i$ and $\sigma^2 e_j$ are the environmental variance of i^{th} and j^{th} character

The value of environmental co-variance may be positive or negative and accordingly environmental correlation will have positive or negative sign.

Example : Substituting the values of co-variances and variances estimated from table 6.2 we can estimate different correlations as under

(1) genotypic correlation coefficient (rg_ig_j)

$$rg_ig_j = \frac{\sigma g_i g_j}{\sqrt{(\sigma^2 g_i)(\sigma^2 g_j)}} = \frac{15.2}{\sqrt{(159.9)(5.4)}} = \mathbf{0.52}$$

(2) Phenotypic correlation coefficient (rp_ip_j **)**

$$rp_ip_j = \frac{\sigma p_i p_j}{\sqrt{(\sigma^2 p_i)(\sigma^2 p_j)}} = \frac{24.8}{\sqrt{(244.7)(7.9)}} = \mathbf{0.56}$$

(3) Environmental correlation (re_ie_j)

$$re_ie_j = \frac{\sigma e_i e_j}{\sqrt{(\sigma^2 e_i)(\sigma^2 e_j)}} = \frac{9.6}{\sqrt{(84.1)(2.5)}} = \mathbf{0.66}$$

Test of significance of correlation coefficient:

The null hypothesis is

Ho: $\rho = 0$

Ha: $\rho \neq 0$

In statistical table, Table values at 5% and 1% level of significance are available which can be used to test the significance of the correlation coefficients. The degree of freedom for genotypic correlation will be (g-2) and that for phenotypic correlation it will be {(rg)-2}. OR it can be tested with student t test. For above example genotypic correlation can be tested as

$$t = \frac{r\sqrt{(n-2)}}{\sqrt{(1-r^2)}} = \frac{0.52\sqrt{(8-2)}}{\sqrt{(1-(0.52)^2}} = 1.75$$

If calculated value of t is higher than table value of t at 0.05 level of significance at corresponding degree of freedom the r value is called significant otherwise it is non significant. In this example the calculated value of t is 1.75 and table value at (8-2)=6 degree of freedom is 2.447. Therefore we accept null hypothesis .i.e. absence of genetic correlation among two characters under study.

Estimation of Co-heritability :

The combined inheritance of two character from one generation to the next generation is called co-heritability. It is a ratio of genotypic co-variances to phenotypic co-variances. It can be estimated as under :

$$\text{Co-heritability} = \frac{\sigma g_i g_j}{\sigma p_i p_j} = \frac{15.2}{24.8} = \mathbf{0.61}$$

7. Path Analysis

Crop yield is influenced by different biometrical characters. Grafius expressed his opinion that there may not be genes for yield *per se* there could be genes which governs inheritance of component characters. The degree of influence of one variable on the other can expressed in quantitative term. Path coefficient analysis was developed by Sewall Wright in (1921). Dewey and Lu 91959) employed this method for first time in plants to disentangle the direct and indirect influence of components of seed yield. It's a multivariate analysis technique which deals with a closed system of variables which are linearly related. Path analysis provides the direct effect of a character on the dependent character as well as its indirect effects through other variables in the system. Path analysis is analogous to the analysis of variance and may be called analysis of correlation coefficient (Li, 1956). Path coefficients are standardized partial regression coefficients which has no units. Therefore the direct effect of different variables can be ranked on the basis of their magnitude. This information may help the plant breeders in choosing the different characters in selection so as to achieve higher genetic gain.

Let us assume that there are for characters y being the yield of forage maize, x1 = plant height , x2 = leaf length ,x3 = leaf width and x4 = number of leaves per plant. we can write four simultaneous equations as under

$r_{yx1} = P_{1y} + P_{2y}\, r\, x_1x_2 + P_{3y}\, rx_1x_3 + P_{4y}\, rx_1x_3$

$r_{yx2} = P_{1y}\, r\, x_2x_1 + P_{2y} + P_{3y}\, rx_2x_3 + P_{4y}\, rx_1x_3$

$r_{yx3} = P_{1y}\, r\, x_3x_1 + P_{2y}\, r\, x_3x_2 + P_{3y} + P_{4y}\, rx_1x_3$

$r_{yx4} = P_{1y}\, r\, x_4x_1 + P_{2y}\, r\, x_4x_2 + P_{3y}\, rx_4x_3 + P_{4y}$

This information can be arranged in matrix form as under

$$\begin{bmatrix} 1 & r\,x_1x_2 & r\,x_1x_3 & r\,x_1x_4 \\ r\,x_2x_1 & 1 & r\,x_2x_3 & r\,x_2x_4 \\ r\,x_3x_1 & r\,x_3x_2 & 1 & r\,x_3x_4 \\ r\,x_4x_1 & r\,x_4x_2 & r\,x_4x_3 & 1 \end{bmatrix} \begin{bmatrix} P_1y \\ P_2y \\ P_3y \\ P_4y \end{bmatrix} = \begin{bmatrix} rx_1y \\ rx_2y \\ rx_3y \\ rx_4y \end{bmatrix}$$

A B C

We have information of matrix A which indicates the correlation coefficients between xi variables i.e. component characters and c matrix which indicates the correlation coefficient between component characters and yield. Thus, we need to find out the values of P_{iy} i.e. path coefficient values which can be solved by

$B = A^{-1} C$ where A^{-1} is the inverse of matrix A

We find out Path coefficient values i.e. P_{1y}, P_{2y}, P_{3y} and P_{4y} which are infect the direct effect of variable X_1 , X_2, X_3 and X_4 respectively.

Indirect effect of X1

via x2 = $P_2y\ rx_1x_2$ via x3 = $P_3y\ rx_1x_3$ via x4 = $P_4y\ rx_1x_4$

Indirect effect of X2

via x1 = $P_1y\ rx_1x_2$ via x3 = $P_3y\ rx_2x_3$ via x4 = $P_4y\ rx_2x_4$

Indirect effect of X3

via x1 = $P_1y\ rx_1x_3$ via x2 = $P_2y\ rx_2x_3$ via x4 = $P_4y\ rx_3x_4$

Indirect effect of X4

via x1 = $P_1y\ rx_1x_4$ via x2 = $P_2y\ rx_2x_4$ via x3 = $P_3y\ rx_3x_4$

Let us consider an example with four component character and Y being the green fodder yield of maize , X_1 = plant height, X_2 = leaf length, X_3 = leaf width, X_4 = no. of leaves per plant.

Table 7.1 :

X_1	X_2	X_3	X_4			
1	0.076	0.136	-0.382	P1Y	=	0.699
0.076	1	0.629	0.844	P2Y		0.994
0.136	0.629	1	-0.503	P3Y		0.628
-0.382	0.844	-0.503	1	P4Y		0.594
A				B		C

The inverse of matrix can be obtained as under with the computer programme

Table 7.2 : A inverse matrix

X_1	X_2	X_3	X_4
0.961334	0.38318	-0.46817	-0.19166
0.38318	-0.68185	0.99046	1.220058
-0.46817	0.99046	-0.0934	-1.06177
-0.19166	1.220058	-1.06177	-0.63701

Now we can find out the value of matrix B by multiplying a inverse with c matrix

we can estimate value of P_1Y as under which are the direct effect of the xi variables

P_1Y= (0.961334 x 0.699) + (0.38318 x 0.994) +(-0.46817 x 0.628) +(-0.19166 x 0.594)
=0.644996

P_2Y= (0.38318 x 0.699) + (-0.68185 x 0.994) +(0.99046 x 0.628) +(1.220058 x 0.594)
=0.936807

P_3Y= (-0.46817 x 0.699) + (0.99046 x 0.994) +(-0.0934 x 0.628) +(-1.06177 x 0.594)
=-0.03208

P_4Y= (+(-0.19166 x 0.699) + (1.220058 x 0.994) +(-1.06177 x 0.628) +(-0.63701 x 0.594)
=0.033589

The indirect effect can be estimated as under

Indirect effect of X_1 via X_2 = (P_2Y x rx1x2) = (0.936807 x 0.076) =0.071197

Indirect effect of X_1 via X_3 = (P_3Y x rx1x3) = (-0.03208 x 0.136) =-0.00436

Indirect effect of X_1 via X_4 = P_4Y x rx1x4) = (0.033589 x -0.382) = -0.012831

Total indirect effect of x1 = 0.054006

Indirect effect of X_2 via X_1 = (P_1Y x rx1x2) = (0.644996 x 0.076) =0.0490196

Indirect effect of X_2 via X_3 = (P_3Y x rx2x3) = (-0.03208 x 0.629) = -0.0201783

Indirect effect of X_2 via X_4 = (P_4Y x rx2x4) = (0.033589 x 0.844) = 0.0283491

Total indirect effect of x2 = 0.0571904

Indirect effect of X_3 via X_1 = (P_1Y x rx1x3) = (0.644996 x 0.136) = 0.0877194

Indirect effect of X_3 via X_2 = (P_2Y x rx2x3) = (0.936807 x 0.629) = 0.5892516

Indirect effect of X_3 via X_4 = (P_4Y x rx3x4) = (0.033589 x -0.503) = -0.168952

Total indirect effect of x3 = 0.6600758

Indirect effect of X_4 via X_1 = (P_1Y x rx1x4) = (0.644996 x -0.382) = -0.2463884

Indirect effect of X_4 via X_2 = (P_2Y x rx2x4) = (0.936807 x 0.844) = 0.7906651

Indirect effect of X_4 via X_3 = (P_3Y x rx3x4) = ((-0.03208 x -0.503) = 0.0161362

Total indirect effect of x4 = 0.5604129

It can be verified that the direct effect of X1 + Total indirect effect of X1

= 0.644996 +0.054006

= 0.699002

which is equal to the correlation coefficient value of X1 with Y and thus we can say that Path analysis is partitioning of correlation in to direct and its indirect effect through the other variables in the system. The above results are presented in table as under

Table 7.3 : Direct and indirect effect of different variables in forage maize

	X_1 Plant height	X_2 Leaf length	X_3 Leaf width	X_4 No. of leaves	Corr. with forage yield of maize(Y)
X_1	**0.6450**	0.0712	-0.0044	-0.0128	0.699**
X_2	0.0490	**0.9368**	-0.0202	0.0283	0.994**
X_3	0.0877	0.5893	**-0.0321**	-0.1690	0.628**
X_4	-0.2464	0.7907	0.0161	**0.0336**	0.594**

Bold figures indicates the direct effect and rest are the indirect effects

The direct effects indicates that leaf length has the highest direct influence on forage yield followed by plant height. As these two variables have positive and highly significant association with forage yield selection of these two variables for direct selection will result in higher genetic gain. The indirect effect via leaf length in leaf width and no. of leaves are also higher thus selection of leaf length will improve these two characters also.

As the path coefficients are partial regression coefficient. regression technique is used we can estimate the total variation accounted by the variables in the analysis as R^2 and the unaccounted variation as (1- R^2) which is known as residual.

Residual variation $= 1 - (P_{1Y})^2 - (P_{2Y})^2 - (P_{2Y})^2 - (P_{4Y})^2 - (2P_{1Y}P_{2Y}rX_1X_2) - (2P_{1Y}P_{3Y}rX_1X_3) - (2P_{1Y}P_{4Y}rX_1X_4) - (2P_{2Y}P_{3Y}rX_2X_3) - (2P_{2Y}P_{4Y}rX_2X_4) - (2P_{3Y}P_{4Y}rX_3X_4)$

8. Selection

Non random differential reproduction mating is called selection. It can acts effectively on heritable differences and it can not create variability. The fundamental effect of selection is to change in gene frequency. The phenotypically observable responses to selection are change in genotypic frequency then the previous generation and change in population mean, appearance of new genotypes and change in variability of the character in population. Effect of selection in quantitative characters are measured with the help of mean, variances and co-variances. Change in gene frequency may be due to artificial selection, through natural differences of fertility and viability.

Fitness : Proportionate contribution of offspring to the next generation is called fitness of that individual or also called as adaptive value or selective value .

Coefficient of selection(s) : The force acting to reduce adaptive value of a genotype is called coefficient of selection. The fitness of unfavoured genotype which is reduced by s will be (1-s).

(1) Selection at gametic stage :

	Genotypes of gametes		Total
	A1	A2	
Initial frequency	p0	q_0	1
Fitness	1	(1-s)	-
Frequency after selection	p0	$q_0(1-s)$	$=p0 + q_0(1-s)$ $= p0 + q_0-sq_0$ $=1 - sq_0$

If gene A2 is lethal i.e. s=1 then the gene will be eliminated in next generation. If $s \neq 1$ then after one generation of selection freq. (q1) of gene A2 will be

$q1 = \{q_0(1-s)\}/ (1 - sq_0)$ so one can calculate change in gene frequency as (δq)

$$\begin{aligned}\delta q &= q1 - q0 \\ &= [\{q_0(1-s)\}/ (1 - sq_0)] - q0 \\ &= q0 - sq0 - q0 (1 - sq_0) / (1 - sq_0) \\ &= q0 - sq0 - q0 + sq^20 / (1 - sq0) \\ &= -sq0 + sq^20 / (1 - sq0)\end{aligned}$$

$= -sq0 (1 – q0) / (1 –sq0)$

Example : If gene A1 freq =0.4 and for a2 = 0.6 and coefficient of selection (s=0.2) then find gene freq. of A2 after one generation.

$\delta q = -sq0 (1 – q0) / (1 –sq0)$

$= - (0.2x0.6) (1-0.6) / (1-(0.2x)0.6)$

$= -0.048 / 0.88$

$= -0.0545$

The value is negative so in the next generation a2 freq. will be reduced by 0.0545.

Thus $(q0 - \delta q)$ will be gene freq. after one generation of selection = 0.5455

(2) Zygotic selection :

(i) (Consider s=1 for recessive genotype)

	Genotypes of gametes			Total
	A1A1	A1A2	A2A2	
Initial frequency	p^2_0	$2p_0q_0$	q^2_0	1
Fitness	1	1	(1-s) =0	
Frequency after selection	p^2_0	$2p_0q_0$	$q^2_0(1-s)$ = 0	$=p^2_0 + 2p_0q_0$ $= p_0 (p_0 + 2q_0)$ $=p_0 (1+ q_0)$

gene freq. of a2 gene after one generation of selection (q1)

$q1 = p_0q_0 / p_0 (1+ q_0) = q_0 / (1+ q_0)$ and

$\delta q = q1 – q_0$

$= \{q_0 / (1+ q_0)\} – q_0$

$= - q^2_0 / (1 + q_0)$

(ii) in the same way we can estimate the change when $s \neq 1$ for A2A2 genotype.

$q1 = p_0q_0 / p_0 (1+ q_0) = q_0/ (1+ q_0)$ and

$\delta q = q1 – q_0$

$= -sq^2_0 (1 - q_0) / (1 - sq^2_0)$

(iii) Fitness of A1A1 and A2A2 are reduced by s1 and s2 , respectively and heterozygote has fitness = 1

$q1 = \{p_0q_0 + q^2_0(1-s_2)\} / \{1 –s_1p^2_0 -s_2q^2_0 \}$

$\delta q = q1 – q_0$

$= \{p_0q_0 (s_1 p_0 -s_2q_0) \} / \{1 –s_1 p^2_0 -s_2q^2_0 \}$

There are three situations when change in gene frequency (δq) can be zero (1) frequency of A1 allele = 0 or (2) frequency of A2 allele = 0 or (3) s1p = s2q

Number of generations :

Number of round of selection required to attain a particular level of gene frequency :

As per example 2(i)

$q_1 = q_0 / (1+ q_0)$ q1 is the gene frequency of A2 after one round of selection in the same way after second round it will be

$q_2 = q_1 / (1 + q_1)$ and after n round it will be

substituting the value as $q_1 = q_0 / (1+ q_0)$ in equation for q2 it will be

$q_2 = q_1 / (1 + 2q_1)$

$q_2 = q_0 / (1+ q_0) / (1 + (q_0 / (1+ q_0))$

$q_2 = q_0 / (1+ 2q_0)$

$qn = q_0 / (1 +nq_0)$

$qn (1 +nq_0) = q_0$

$qn + nq_0qn = q_0$

$n = (q_0 - qn) / q_0qn$

thus number of round (n) required to attain a level of gene frequency can be estimated.

If $q_0 = 0.8$ how many round of selection will be required to reach frequency to 0.2?

$n = (q_0 - qn) / q_0qn$

$= (0.8 -0.2) / (0.8*0.2)$

$= 0.6 / 0.16$

$= 3.75$ i.e. after four round of selection gene frequency of A2 will reach to a level of 0.02

Response to selection :

By selection the gene and genotypic frequency are changed whose effects are measured as change in population mean. Response may be defined as

the difference between phenotypic value of offspring of selected parents and whole of the parental generation before selection.

Selection differential mean phenotypic value of selected parents expressed as deviation from population mean i.e. Mean of all individuals in the parental generation before selection that is known as selection differential.

$$S = \overline{Xs} - \overline{X}0 \text{ or } \mu 1 - \mu 0$$

$$S = \frac{Z}{P} \sigma P$$

When Z = height of coordinate of normal dust curve
P = proportionate area

Genetic gain $R = i \cdot h^2 \sigma P$ or
$= i h \cdot \sigma g$

i = selection intensity (Weber 1967 has given the tabulated value for proportionate area selected. For 5 % area the value of i = 2.06 and for 10 % area = value 1.76

Broad sense heritability can be estimated as under

$$h^2 = \frac{\sigma^2 g}{\sigma^2 p} \text{ or } h^2 = \frac{\sigma^2 g}{\sigma^2 p} x100 \text{ (percentage)}$$

$$\text{So } h = \sqrt{\frac{\sigma^2 g}{\sigma^2 p}}$$

Different characters are measured in different metrical units therefore genetic gain can not be compared directly so we need to work out genetic advance as percentage of mean (GA% of mean) .

GA as % of mean = $\frac{R}{Y} x100$

Correlated response :

"Two characters x & y are correlated than a change is mean of x characters through selection will cause associated change in mean value of y. This chance in y got through indirect selection of x is known as correlated response CRy"

CRy = $i_x h_x h_y . r_g . \sigma_{py}$

Y = yield
X = No. leaves/plant

i_x = intensity of selection
(table value for 5 % selection intensity = 2.06)

Where $h_x = \sqrt{\frac{\sigma^2_{gx}}{\sigma^2_{px}}}$ $h_y = \sqrt{\frac{\sigma^2_{gy}}{\sigma^2 y}}$

r_g = genotypic correlation between x & y character

σ_{py} = Phenotypic standard deviation of y)

Genetic Advance (δG) :
Genetic advance for character 1 can be estimated as

$$G1 = \frac{z}{v} x \frac{\sum biGi1}{\sqrt{aGb}}$$

where z /v is the selection intensity (i)
Expected genetic gain can be predicted as under

$$(\delta G) = \{ (z/v) \Sigma\Sigma a_i b_j G_{ij} \} / (\Sigma\Sigma b_i b_j P_{ij})^{1/2}$$

9. Selection index

Selection index : Basically there are three selection methods which could be employed for improving several characters simultaneously. They are as under

(1) Tandem selection : In this different characters are improved one by one and so on . At a time a single character is improved the response of selection will be of a single character only (assuming no genetic correlation).

(2) Independent culling level : In this method the selected individuals have to surpass a certain minimum value for each of the characters to be improved. A plant breeder may decide the limit of 50 gram yield per plant, 90 to 100 days to maturity and 180 to 190 cm plant height. Thus in selection those plants whose yield is 50 or more than 50 gram, maturing in 90 to 100 days with plant height 180 to 190 cm will only be selected.

(3) Selection index : The aim of plant breeding programme is simultaneous improvement of several characters. Most rapid improvement in the economic value of a plant is expected from selection applied simultaneously to all characters provided appropriate weights are assigned to each character according to their relative economic importance, heritability and correlations. If the component characters are combined together into a score or an index in such a way that when selection is applied to the index, as is index is the character to be improved, most rapid improvement of economic value is expected. Such an index was first proposed by Smith (1937) based on Discriminant function of Fisher (1936). Hazel (1943) also developed a method of index selection based on the path coefficient .

The characters of economic importance are mostly quantitative in nature. There is no sound basis to assign weight to the characters. Smith advocated use of Discriminant function of observable characters which will best indicate the genetic worth of a plant.

Let us assume that economic value of a plant is determined by three variables *viz.* x1, x2 and x3 and the merit of a plant is H .Then

$H = a_1G_1 + a_2G_2 + a_3G_3$ if the character is governed by say n characters then $H = a_1G_1 + a_2G_2 + a_3G_3 + \ldots\ldots\ldots + a_nG_n$ where $G_1, G_2, \ldots\ldots G_n$ are the genotypic value of the character and $a_1, a_2 \ldots a_n$ are the weight assigned to the respective characters. The merit (H) can not be directly evaluated because we measure the phenotypic values of the characters. The phenotype (I) is linearly expressed as

$I = b_1x_1 + b_2x_2 + b_3x_3$ and for n characters it will be

$I = b_1x_1 + b_2x_2 + b_3x_3 + \ldots\ldots\ldots\ldots\ldots\ldots + b_3x_n$

Values of b are calculated in such a way that regression of I on H is maximized so that selection phenotype on the basis of I as a discriminant function will ensure maximum gain. Let us take an example of three characters

$$b_1\sigma^2_1 + b_2\sigma_{12} + b_3\sigma_{13} = A1$$

$$b_1\sigma_{21} + b_2\sigma^2_2 + b_3\sigma_{23} = A2$$

$$b_1\sigma_{31} + b_2\sigma_{32} + b_3\sigma^2_3 = A3$$

σ^2_1, σ^2_2 and σ^2_3 are phenotypic variances of characters x1, x2 and x3 and σ^2_{12} , σ^2_{13}, σ^2_{23} are the phenotypic co-variances where $\sigma^2_{12} = \sigma^2_{21}$, $\sigma^2_{13} = \sigma^2_{31}$, $\sigma^2_{23} = \sigma^2_{32}$

$$A1 = a_1G^2_1 + a_2G_{12} + a_3G_{13}$$

$$A2 = a_1G_{21} + a_2G^2_2 + a_3G_{23}$$

$$A3 = a_1G_{31} + a_2G_{32} + a_3G^2_3$$

$$\begin{bmatrix} \sigma^2_1 & \sigma_{12} & \sigma_{13} \\ \sigma_{21} & \sigma^2_2 & \sigma_{23} \\ \sigma_{31} & \sigma_{32} & \sigma^2_3 \end{bmatrix} \begin{bmatrix} b1 \\ b2 \\ b3 \end{bmatrix} \begin{bmatrix} G^2_1 & G_{12} & G_{13} \\ G_{21} & G^2_2 & G_{23} \\ G_{31} & G_{32} & G^3_3 \end{bmatrix} \begin{bmatrix} a1 \\ a2 \\ a3 \end{bmatrix}$$

$$[\ X\ ] \qquad [\ b\ ] \qquad [\ G\ ] \qquad [\ a\ ]$$

To obtain the value of b we need to find out inverse of X matrix. X is phenotypic variance co-variance matrix , G is genotypic variance co-variance matrix, b is coefficient matrix to be estimated and a is weight matrix. Here we have taken equal weight for all characters i.e. 1.

Thus $b = G \times a \times X^{-1}$ this matrix solution will give us the value of b for each variable. With the help of this we can calculate selection index by substituting value of the character and b values in following equation

$I = b_1x_1 + b_2x_2 + b_3x_3 + \ldots\ldots\ldots\ldots\ldots\ldots + b_3x_n$

Selection index for individual character, combination of two characters, three characters and so on can be calculated and the combination which provides the higher genetic gain is selected.

Example : Let there are four characters X1, X2, x3 and X4 . The genotypic and phenotypic variance covariance matrix are as under

Genotypic variance covariance matrix (G)

$$\begin{bmatrix} 18.45 & -4.82 & -1.3 & 26.49 \\ -4.82 & 1032.793 & 9.34 & -1.82 \\ -1.3 & 9.34 & 1.77 & -2.58 \\ 26.49 & -1.82 & -2.58 & 51.68 \end{bmatrix}$$

Phenotypic variance covariance matrix (X)

19.13	-5.45	-1.269	26.426
-5.45	1057.253	9.01	-2.166
-1.269	9.01	2.716	-2.408
26.426	-2.166	-2.408	53.478

Inverse of Phenotypic variance covariance matrix (X^{-1})

0.165212	0.000661	0.002751	-0.08149
0.000661	0.000977	-0.00332	-0.00044
0.002751	-0.00332	0.394923	0.016289
-0.08149	-0.00044	0.016289	0.059682

Estimated B values

For X1	1.1068
For X2	0.9808
For X3	0.7279
For X4	0.9051

I = 1.1068 x1 + 0.9808 x2 + 0.7279 x3 +0.9051 x4 will be the equation for getting the index score. The plants having higher score values are to be selected for further selection programme so as to have higher genetic improvement .

10 Discriminant function

The discriminant function was given by R A Fisher in 1936 with a purpose to discriminate the individuals belonging to two different populations showing some degree of overlapping. A function say z is defined as

$Z = b1x1 + b2x2 + b3x3 + \ldots\ldots\ldots\ldots + bnxn$

here x1, x2 ,x3 are the measured characters and b1, b2, b3…..bn are weighing coefficients. The bi values are estimated such that the ration of variance between populations to within population would be maximized. The maximization leads to following simultaneous equations.

Let us take an example with three characters x1, x2 and x3, the equation will be as under

$b'_1 + b'_2 r_{12} + b'_3 r_{13} = d_1'$

$b'_1 r_{21} + b'_2 + b'_3 r_{23} = d_2'$

$b'_1 r_{31} + b'_2 r_{32} + b'_3 = d_3'$

where b′ are the discriminant coefficient to be estimated, $r_{12} = r_{21}$, $r_{13} = r_{31}$, $r_{23} = r_{32}$ being the correlation coefficients

The above equations can be written in matrix form as under

$$\begin{bmatrix} 1 & r12 & r13 \\ r21 & 1 & r23 \\ r31 & r32 & 1 \end{bmatrix} \begin{bmatrix} b'1 \\ b'2 \\ b'3 \end{bmatrix} = \begin{bmatrix} d'1 \\ d'2 \\ d'3 \end{bmatrix}$$

(R) (b) (d)

where $d'_1 = d_1 / ((\sum x_1)^2)^{1/2}$, $d'_2 = d_2 / ((\sum x_2)^2)^{1/2}$, $d'_3 = d_3 / ((\sum x_3)^{1/2}$

$d'_1 = x^1_1 - x^2_1$, $d'_2 = x^1_2 - x^2_2$ and $d'_3 = x^1_3 - x^2_3$

where $b'_1 = b_1 / ((\sum x_1)^2)^{1/2}$, $b'_2 = b_2 / ((\sum x_2)^2)^{1/2}$, $b'_3 = b_3 / ((\sum x_3)^2)^{1/2}$

x^1_1 and x^2_1 indicates mean of character x_1 in population 1 and 2 (superscript for population and subscript for character)

R is the correlation matrix , b is the matrix of discriminate coefficient to be estimated and d is the deviation matrix. Solution of b matrix can be obtained as under

$b = (R^{-1})(d)$ where (R^{-1}) is the inverse of R matrix

After obtaining the values of discriminant coefficient we can write the discriminate function as $Z = b_1x_1 + b_2x_2 + b_3x_3$ replacing the values of b we can have the function in which replacing the values of each individual variable i.e. x_1, x_2 and x_3 for both the population we obtain Z score values for all individuals To test the significance whether the function is capable for discrimination or not ? we perform Anova as under

ANOVA :

Source	df	Sum of squares
Between populations	α +m -2	$\{(n^1n^2)/(n1+n2)\}(Z)^2$
Within populations	n - α - m+1	Z

Where m = number of the characters

α = number of populations

n = number of observations $(n_1 + n_2)$

n_1 = number of observations in population 1

n_2 = number of observations in population 2