

# **Discriminant Function Analysis (DFA)**

**Dr. N K Chaurasia**  
**Dept. of GPB, MSSoA, CUTM**

- First proposed by **Smith (1936)**- used for plant selection
- He suggested that a better way of exploiting genetic correlation with several traits having high heritability is to construct an index known as **selection index**, which combines information on all the characters associated with the dependent variable like yield
- **Selection index (SI)**: refers to a linear combination of characters associated with yield
- **SI** involve discriminant functions based on the relative economic importance of various characters.
- Since, the desirable genotypes are discriminated from the undesirable ones, based on the combination of various characters, known as **discriminant function analysis**
- Later **Hazel (1943)** developed a simultaneous selection model based on the approach of path analysis

# Features of discriminant function analysis

- Measures the efficiency of various traits combinations in selection
- Provide information on yield components and thus aids indirect selection for genetic improvement of yield
- Analysis is based on assumptions of **linearity and additivity**
- Analysis involves **variances and covariances**
- Selection index leads to simultaneous manipulation of several traits for genetic improvement of economic yield
- Selection indices are commonly three types such as **classical selection index, general selection index and restricted selection index**
- DFA differs from path coefficient analysis

# Difference between DFA and path coefficient analysis

DFA	Path coefficient analysis
Analysis involve variances and covariances	Analysis involve simple correlation
Provide information on weight coefficient, expected genetic advance and relative efficiency	Provide information on direct and indirect effects
Residual effect not estimated	Residual effect estimated
Only one type	Three type: genotypic, phenotype and environmental

# Selection index types

Classical Selection index	General Selection Index	Restricted Selection Index
<p>Developed and was first applied for plant selection by <b>Smith (1936)</b></p> <p>Involves several characters simultaneously in selection index and discriminates b/w desirable and undesirable genotypes on the basis of selection efficiency</p>	<p>First proposed by <b>Hanson and Johnson (1957)</b></p> <p>A modification of the smith's scheme</p> <p>In this model, the weights for various traits are based on the average statistics for several populations</p>	<p>Proposed by <b>Kempthorne and Nordskog (1959)</b></p> <p>Helps in improving a set of characters keeping the value of other characters intact</p>

# Computation of selection index

- Construction of all types of selection indices is based on the estimates involved in the selection index
- Selection indices can be worked out from the **replicated data only**
- Construction of selection index is based on **following assumptions:**
  1. **Statistical assumptions:** random selection of the parents for mating, absence of G x E interactions, and linear or additive nature of component traits to be include in the index
  2. **General assumptions:** diploid segregation, lack of maternal effects, absence of linkage, absence of epistasis, absence of multiple alleles, and equal survival of all the genotypes in a population

# Steps in construction of selection index

1. Calculation of coefficient values or weight coefficients
2. Calculation of expected genetic advance
3. Calculation of relative efficiency

# 1. Coefficient values or weight coefficients:

- Coefficients refers to the relative importance of several traits in the index

$$\text{Selection index (I)} = b_1X_1 + b_2X_2 + b_3X_3 + \dots + b_nX_n$$

Where,  $X_1, X_2, X_3$  are phenotypic values of the traits 1, 2, 3 and n respectively and  $b_1, b_2$  and  $b_3$  are the corresponding weights

- The value of  $b$  are calculated from a series of simultaneous equation involving the appropriate phenotypic and genotypic variances and covariances
- Lets consider if three characters and yield, the simultaneous equations can be set with the help of appropriate variances and covariances of these characters

$$b_1W_{11} + b_2W_{12} + b_3W_{13} = g_1y$$

$$b_1W_{12} + b_2W_{22} + b_3W_{23} = g_2y$$

$$b_1W_{13} + b_2W_{23} + b_3W_{33} = g_3y$$

Where,  $W_{11}, W_{22}$  and  $W_{33}$  are the phenotypic variances of character 1, 2 and 3 , respectively  
 $b_1, b_2$  and  $b_3$  are weight coefficients

$W_{12}, W_{13}, W_{23}$  are phenotypic covariances between traits 1-2, 1-3 and 2-3 respectively

$g_1y, g_2y$  and  $g_3y$  are the genotypic covariances of traits 1, 2, 3 with dependent traits like yield



## 2. Expected genetic advance:

- Genetic advance for discriminant function (GS1) is estimated separately for various selection indices involving single, double, triple and multiple traits using following formula

$$GS1 = K(b_1g_{1y} + b_2g_{2y} + b_n g_{ny})^{1/2}$$

where, K (2.06) is the selection intensity at 5%; b1, b2 and b3 are weight coefficients values for the traits 1, 2 and n, respectively and g<sub>1y</sub>, g<sub>2y</sub> and g<sub>3y</sub> are the corresponding genotypic covariances of traits 1, 2, 3 with dependent traits like yield

Genetic advance by direct method (GS2) is estimated for the dependent trait like yield using the formula,

$$GS2 = VG \sqrt{VP} \times K$$

Where, VG and VP are the genotypic and phenotypic variance, respectively and K is the selection intensity

### 3. Relative efficiency:

- It is the percentage ratio of genetic advance for discriminant function (GS1) to genetic advance for direct selection (GS2)

$$\text{Relative efficiency} = (\text{GS1}/\text{GS2}) \times 100$$

- It measures the effectiveness of various selection indices
- Relative efficiency of direct selection for yield is considered as 100%
- Relative efficiency of selection is worked out separately for various selection indices involving single, double, triple and multiple traits
- Any selection index exhibiting superiority over direct selection index is considered important and the traits involving in such index are considered as major components of yield
- Hence, such combination of traits has to be given weightage in the selection of elite genotypes from the diverse breeding populations

# Merits & demerits of DFA

## Merits

- Provides information on **yield components and thus aids in indirect selection** for genetic improvement for yield
- Can be applied to **both segregating as well as parental populations**

## Demerits

- Selection index construction is **difficult task** because it requires lot of statistical calculation and some times selection indices are **not an effective as yield** alone in spite of making extra efforts for their estimation
- The calculation becomes **still complicated when the index is estimated to over many component characters**
- SI applicable to **individual plant selection only**. However family selection is not greatly improved by the use of an index
- SI have limited applications in practical plant breeding because of **inaccuracies associated with the estimation of variances and covariances**
- Different estimate procedures provide **different values when applies to the same set of population**

# Problems of selection index

- Several causes of non applicability of selection indices in plant breeding programme
  - a. Non fulfillment of the assumptions involved in construction of selection index
  - b. Biased estimates of parameters like the layout and conduct of the experiment and formulation of statistical and genetical models
  - c. Uncertainty of economic weights to be assigned to each trait included in the index
  - d. Lack of an objective method for determining economic weights and the changing concept of economic significance of the traits having been the important problems due to which selection index once constructed does not remain realistic for longer duration

# Application in crop improvement

- Robinson et al., 1951 applied selection index involving three characters like plant height, ears per plant and yield for genetic improvement of maize and found that 30% more efficient than selection based on yield alone
- Brim et al., 1959 applied selection index in soybean for genetic improvement of oil yield. Selection index based on six traits like oil, protein yield, resistance to lodging, fruiting period and seed weight resulted in maximum improvement in oil yield

**Thank you..**

# **Path Analysis**

**Dr. N K Chaurasia**  
**Dept. of GPB, MSSSoA, CUTM**

# Introduction

- The concept of path analysis was originally developed by **Wright** in 1921,
- But technique was first used for plant selection by **Dewey & Lu** in 1959
- Path coefficient analysis is simply a standardized partial regression coefficient which splits the correlation coefficient into the measures of **direct and indirect effects**
- **In other words**, it measures the direct and indirect contribution of various independent characters on a depended character
- For example, in **blackgram grain yield (X5)**, is affected by number of primary (X1), number of secondary branches (X2), pods per plant (X5) and seed per pod (X4)
- Path analysis reveals whether the association of these characters with yield due to their on yield or is a consequence of their indirect effects via other component characters



# Features of path coefficient analysis

- Measures the cause of association between two variables/characters
- Based on all possible simple correlations among various characters
- Provides information about **direct and indirect effects** of independent variables on dependent variable
- Analysis is based on the assumptions of linearity and additivity
- Estimates **residual effects**
- Helps in determining **yield contributing characters** and thus is **useful in indirect selection**

# Types of path analysis

- From **unreplicated** data- only simple path worked out
- From **replicated data**- 3 types

## **Phenotypic path:**

Worked out from phenotypic correlation coefficient

Phenotypic path splits the phenotypic correlation into the measures of **direct and indirect effects**

## **Genotypic path:**

Worked out from genotypic correlation coefficient

It measures the direct and indirect contribution of various independent characters towards a dependent character like yield

## **Environmental path:**

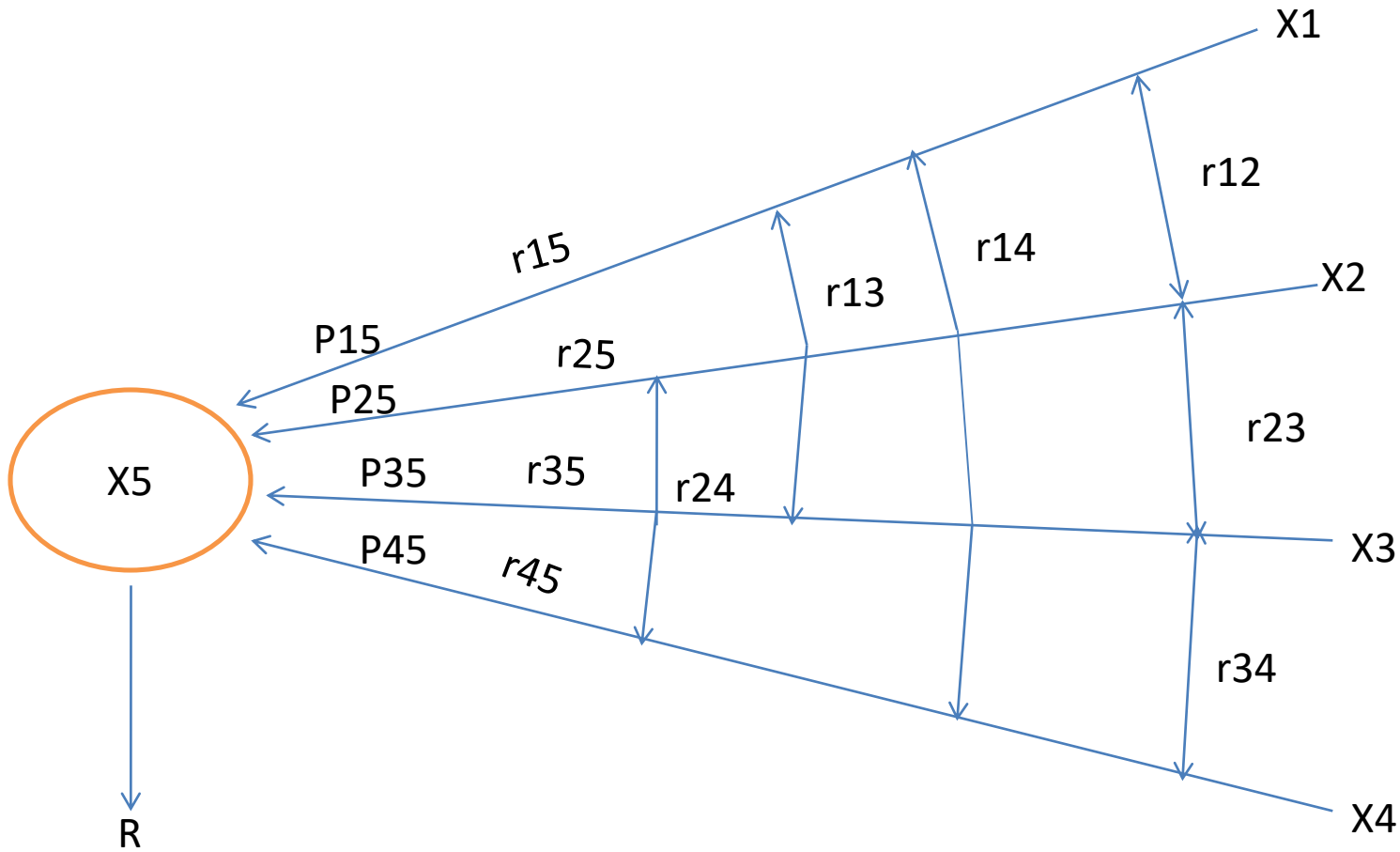
Worked out from the estimates of environmental correlation coefficient

**In plant breeding & genetics, phenotypic & genotypic paths are commonly estimated to determine yield contributing characters which helps in selection of desire genotypes from diverse genetic population**

# Computation of path coefficients

- Selection of genotypes
  - Evaluation of Materials
  - Statistical analysis
- Computation of path coefficients from replicated data consists of following **three important steps**
1. Estimation of variances and covariances for all the characters under studied and their combinations, respectively
  2. Calculation of simple all possible simple correlation coefficient among various traits under study, which is equal to  $\frac{n(n-1)}{2}$  (**n=No. of traits**)
  3. Path analysis following the procedure given by Dewey and Lu (1959)

# Path diagram



Path diagram constructed using the correlation coefficients among yield (X5) and four of its yield contributing traits in blackgram. R is residual effect

# Path diagram

- A linear diagram which is constructed with the help of simple correlation coefficients among various characters under study is referred as **path diagram**
- Constructed before estimation of direct and indirect effects
- Dependent variable like yield is kept at one side and all independent variable on the other side
- ❖ **Path diagram helps in several ways;**
  - Provides a visual picture of cause and effects situation
  - Depicts the association between various characters
  - Helps in understanding the direct and indirect contribution of independent variable towards dependent variables (like yield)
  -

*Continued*

- Path diagram also help in setting of simultaneous equations as given below, which are used for the **estimation of direct effects**

$$r_{15} = P_{15} + r_{12}P_{25} + r_{13}P_{35} + r_{14}P_{45}$$

$$r_{25} = r_{12}P_{15} + P_{25} + r_{23}P_{35} + r_{24}P_{45}$$

$$r_{35} = r_{13}P_{15} + r_{23}P_{25} + P_{35} + r_{34}P_{45}$$

$$r_{45} = r_{14}P_{15} + r_{24}P_{25} + r_{34}P_{35} + P_{45}$$

- Here,  $r_{12}$ ,  $r_{13}$ ,  $r_{14}$ , etc are estimates of simple correlation coefficients between variables  $X_1$  and  $X_2$ ,  $X_1$  and  $X_3$ ,  $X_1$  and  $X_4$ , etc., respectively and  $P_{15}$ ,  $P_{25}$ ,  $P_{35}$ ,  $P_{45}$  are the estimates of direct effects of variables,  $X_1$ ,  $X_2$ ,  $X_3$  and  $X_4$ , respectively on the dependent variable  $X_5$  (yield)

**Direct effects:** every component character will have a direct effect on yield

## **Indirect effects:**

- Effects of an independent character on the dependent character via other independent characters are known as **indirect effects**
- Indirect effects are calculated by putting the values of correlation coefficients and those of direct effects as follows:
- **Indirect effect of primary branches (X1) via,**
  - Secondary branches (X2) =  $r_{12}.P_{25}$
  - Pods per plant (X3) =  $r_{13}.P_{35}$
  - Seeds per pod (X4) =  $r_{14}.P_{45}$

Similarly, the indirect effects of secondary branches (X2) will be as follows via;

- Primary branches (X1) =  $r_{21}.P_{15}$
- Pods per plant (X3) =  $r_{23}.P_{35}$
- Seeds per pod (X4) =  $r_{24}.P_{45}$

Similarly, the indirect effects of other component characters such as pods per plant (X3) and seeds per pod (X4) can be estimated

## Residual effects:

- ✓ Residual effects measures the role of other possible independent variables which were not included in the study on the dependent variable
- ✓ Residual effect is calculated with the help of direct effects and simple correlation coefficients

$$\text{Residual effect (R)} = P^2R_5 + P15.r15 + P25.r25 + P35.r35 + P45.r45$$

Where,  $P^2R_5$  is the square of residual effect



# Interpretation of results

- If the correlation between yield and a trait is due to direct effect of a trait, reveals true relationship between them and **direct selection for this characters will be rewarding**
- If the correlation is mainly due to indirect effects of the character through another component trait, **indirect selection** through such trait will be live in **yield improvement**
- If direct effect is positive and high, but the correlation is negative, in such situation **direct selection for such trait should be practiced to reduce the undesirable indirect effect**
- If the value of **residual effect is moderately or high**, indicting that there are some other attributes for yield other than character under study

# Application in crop improvement

- **In case of oats**, indirect selection through seed width resulted in a 9% increase in grain yield
- **In case of maize**, indirect mass selection for number of ears per plant resulted in an increased yield

## Merits

- Provide information regarding cause and effect situation in understanding the cause of association between two variables
- Permits the examination of direct effect and indirect effects
- Provide basis for selection of superior genotypes from diverse breeding populations

# Demerits

Analysis is designed to ideal with variable having additive effects. Its application to variables having non-additive effects may leads to wrong result (Kempthorne, 1957)

Computation is somewhat difficult and inclusion of many variable leads to more complication in calculation

# Comparison between correlation and path analysis

Particulars	Path analysis	Correlation
Analysis is based on	Path analysis	Degree and covariances
Provide information about	Direct and indirect effect	Degree and direction of association
Types	Genotypic, phenotypic and environmental	Simple, partial and multiple
Based on assumption	Linearity and additivity	Linearity and additivity
Residual effects	Calculated	Not estimated
Determines	Yield components	Yield components

# **Diallel Analysis**

**Dr. N K Chaurasia**  
**Dept. of GPB, MSSSoA, CUTM**

# Introduction

- Diallel cross refers to mating of selected parents in all possible combinations and evaluation of a set of diallel crosses is known as diallel analysis
- Diallel mating design is used to evaluate several inbred lines in terms of combining ability variances and effects
- The procedure of diallel cross analysis for the study of polygenic traits in self pollinated species was developed by **Jinks (1954) and Hayman (1954)**
- The total number of crosses =  $n(n-1)/2$  (direct cross/ one way cross)  
=  $n(n-1)$  [when reciprocals cross included]

Where,  $n$  = number of parental line

# Types of diallel cross

- The diallel mating design involves three materials, namely, **direct crosses or one way crosses, reciprocals and parents**
- Depending upon number of crosses to be made, diallel cross is of two types (Simmonds, 1979 and Fehr, 1987) given below:

**1. Full diallel**

**2. Half diallel**

# Types of diallel cross

## 1. Full Diallel

- All possible combinations (**both direction i.e direct & reciprocals**) are made among the selected parental lines
- Here, each parent is used as male and female for each mating
- The total number of single crosses =  $n(n-1)$  [where,  $n$ = number of parents]
- Used when,
  - a. reciprocal differences are significant and**
  - b. parents do not have male sterility or self incompatibility**
- Permits estimation of maternal effects



# Types of diallel cross

## ➤ Two methods of evaluation of full diallel cross:

### a. Full diallel with parents (F1's, reciprocal & parents)

- Includes both **two crosses and their parent**
- The **total number of entries/genotypes** to be evaluated =  **$n^2$**  [ where, **n=number of parent**]
- Used when self incompatibility and male sterility absent
- Inclusion of parents in analysis **permits** the estimation of heterosis

### b. Full diallel without parents (F1's and Reciprocal crosses)

- Includes all possible single crosses made among 'n' parents in both the direction (direct & reciprocal)
- Total number of single crosses =  **$n(n-1)$**  [where, **n= number of parent**]
- Used when self incompatibility and male sterility present
- Not permits the **estimation of heterosis**

# Types of diallel cross

## 2. Half diallel

All possible crosses among the selected parental lines are made in one way direction i.e. **direct crosses**

**Here, parent is used either as male or female in the mating system**

**The number of single crosses =  $n(n-1)$**

**Used when reciprocal differences are not significant**

**Used when parents have male sterility and self incompatibility**

# Types of diallel cross

➤ **Half diallel crosses are evaluated in two ways:**

## a. Half diallel with parent (F1's & Parents)

Includes one way crosses and parental lines

The total number of entries to be evaluated =  $n(n+1)/2$

Used when parent don't have self incompatibility and male sterility

Inclusion of parents permits the estimation of heterosis

## b. Half diallel without parent (only F1's)

All possible single crosses are made in only one way direction

The total number of single crosses are made for evaluation =  $n(n-1)/2$

Used when self incompatibility and male sterility present

Not permits the estimates of heterosis

# General assumptions

- Normal diploid segregation
- Lack of maternal effects
- Absence of multiple alleles
- Homozygosity of parents
- Absence of linkage
- Absence of epistasis
- Random mating

# Plant of crossing for a diallel design

Parent	1	2	3	4	5	6	7	8	9	10
1	*	X	X	X	X	X	X	X	X	X
2	+	*	X	X	X	X	X	X	X	X
3	+	+	*	X	X	X	X	X	X	X
4	+	+	+	*	X	X	X	X	X	X
5	+	+	+	+	*	X	X	X	X	X
6	+	+	+	+	+	*	X	X	X	X
7	+	+	+	+	+	+	*	X	X	X
8	+	+	+	+	+	+	+	*	X	X
9	+	+	+	+	+	+	+	+	*	X
10	+	+	+	+	+	+	+	+	+	*

Here, X: Direct crosses; \*: Selfing and + : Reciprocal crosses

# Planting arrangements for diallel

## 1. Unpaired Parents

- A common planting arrangement for any type of mating programme is the use of unpaired parents
- Here, each parent is planted once regardless of the number of parents with which it will be crossed

### Adv:

- ❖ Requires the least space and parent seed
- ❖ The decision on which parents to be crossed can be made after planting
- ❖ Parents of each mating are known
- ❖ Can be used with any number of parents

### Disadv:

- More time is required per mating during hybridization
- Greater chance of wrong mating, particularly by inexperienced persons

## 2. Paired Parents

- Planting adjacent to each other the pair of parents to be mated
- When the paired parent arrangement is used, breeders employ various techniques to facilitate movement through field during hybridization and to prevent persons from getting in wrong place and making incorrect an incorrect mating

### Adv:

- ❖ Less time required per mating during hybridization
- ❖ Less chances of wrong mating
- ❖ The parents of each mating are known
- ❖ Can be used with any number of parent

### Disadv:

- Required more space and parental seeds
- Decision on parents to be mated must be made before planting

# Approaches of Diallel Analysis

➤ Two approaches of diallel analysis:

1. Hayman's graphical approach
2. Griffing's numerical approach

# 1. Hayman's graphical approach

- This approach is based on the estimation of components of variation
- Initially developed by Jinks and Hayman in 1953 and later on elaborated independently by Jinks (1954,56) and Hayman (1954, 57)

## Main features of Hayman's approach are:

- A graphical approach involves **Vr-Wr graph**
- Analysis is based on the estimation of **components of variance**
- Provides information about six parameters like, **D, H<sub>1</sub>, H<sub>2</sub>, E, F and h<sup>2</sup>**

## Where,

**D:** additive genetic variance

**H<sub>1</sub>:** dominance variance

**H<sub>2</sub>:** **H<sub>1</sub>[1-(μ-v)]**, where, **μ & v** are the **proportion of positive & negative genes**, respectively in the parents

**E:** expected environmental component of variance

**F:** mean of Fr over the array, where, Fr is the covariance of additive and dominant effects in a single array

**h<sup>2</sup>:** dominance effects, as algebraic sum over all the loci in heterozygous phase in all the crosses

- All these six components help in estimation of various genetic ratios



# Estimation of genetic ratios using data from $F_1$ generation of diallel cross

1. Average degree of dominance =  $(H_1/D)^{1/2}$

**If,**

The value of this ratio is = **zero, No dominance**

**> zero but less than 1, partial dominance**

**= 1, complete dominance**

**> 1, over dominance**

2. Ratio of dominant and recessive genes in parents is

$$= [(4DH_1)^{1/2} + F] / [(4DH_1)^{1/2} - F]$$

**If,**

The ratio is = **1, dominant & recessive genes in the parents are in equal proportion**

**< 1, indicates an excess of recessive genes**

**> 1, indicates an excess of dominant genes**

*Continued*

3. The number of gene which control the character and exhibit dominance is  
 $= h^2/H_2$

4. The proportion of genes with positive and negative effects in the parents is  
 $= H_2/4H_1$

If, the **positive and negative alleles are symmetrically distributed** then ratio will be equal to **0.25**

## Estimation of genetic ratios in $F_2$ generation of diallel cross

- Jinks (1956) gave the procedure for **estimating genetic components from  $F_2$  data**
- The coefficient of  $H_1$  &  $H_2$  are  $\frac{1}{4}$  in  $F_2$  , while they are **1** in  $F_1$
- The  $h$  &  $F$  have a coefficient of  $\frac{1}{2}$  because of the one generation of inbreeding to get  $F_2$

*Continued*

➤ Accordingly, the genetic ratios in  $F_2$  are worked out as follows:

1. Degree of dominance =  $[1/4(H_1/D)]^{1/2}$

1. The proportion of dominant & recessive genes in the parents is  
=  $[1/4(4H_1) + (1/2)F] / [1/4(4H_1) - (1/2)F]$

3. The number of gene which control the character and exhibit dominance is  
=  $h^2/H_2$

4. The proportion of genes with positive and negative effects in the parents is  
=  $H_2/4H_1$

➤ The graphical analysis also involves parents

➤ Graphical approach does not help in identification of superior cross combinations

# Vr-Wr Graph

In the Hayman's approach of diallel analysis, a graph is drawn with the help of variance arrays ( $V_r$ ) and covariances between parents and their offspring ( $W_r$ ), referred as Vr-Wr graph

## Construction of Vr-Wr Graph

For each character a separate graph is constructed

The  $V_r$  value is taken on the X-axis and  $W_r$  values on the Y-axis

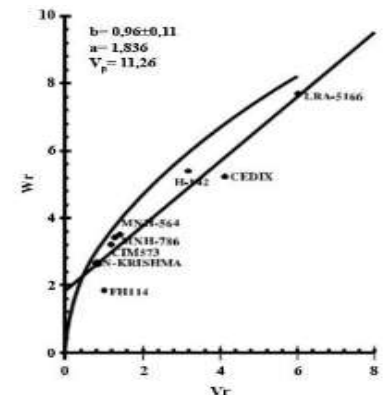
- The  $W_r$  values are estimated for all the array with the help of following formula,

$$W_{ri} = (V_{ri} \times VOLO)1/2$$

Where,

$V_{ri}$  is the variance of  $r^{\text{th}}$  array and VOLO is variance of parents

- $W_r$  values used to draw the limiting parabola



- For drawing regression line, the expected values of  $W_r$  are required which can be estimated for all the arrays separately with the help of following formula,

$$W_r = W_r - bV_r + bV_r$$

Where,

$W_r$  = array means of variances,  $V_r$  = array means of covariances, and  $b$  = regression coefficient

- These values are plotted against  $V_r$  values. Then a straight line passing through these points is drawn, which is known as regression line
- The regression line of a characters may be occupied either of the four positions on the  **$V_r$ - $W_r$  graph** like,
  1. It may pass through origin
  2. May pass below origin cutting  $V_r$ -axis
  3. May pass above the origin cutting  $W_r$ -axis but within the parabola
  4. May touch the parabola limit cutting  $W_r$ -axis

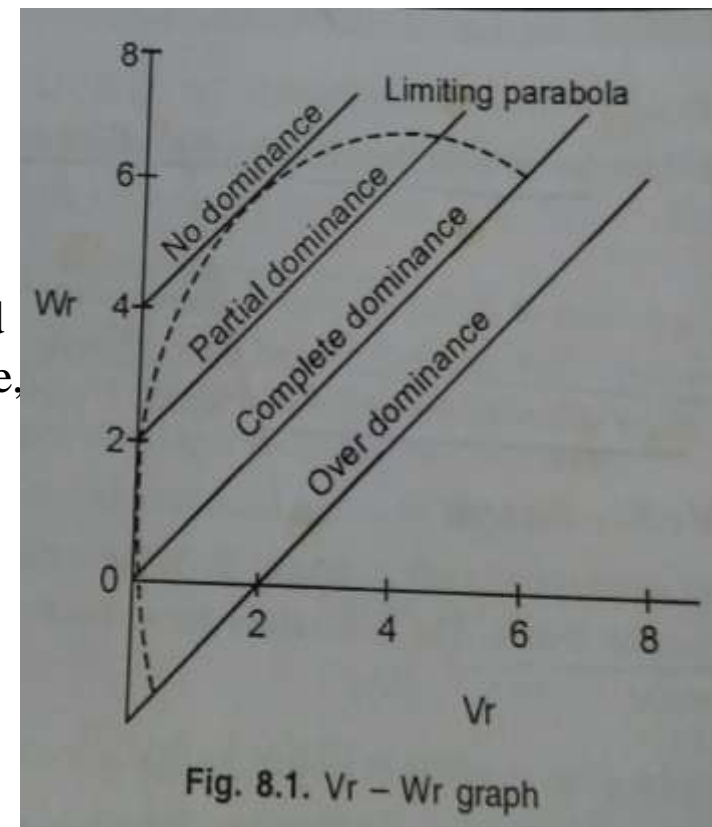


Fig. 8.1.  $V_r - W_r$  graph

# Inferences:

- Depending upon the position of regression line, the following inferences can be drawn from Vr-Wr graph
1. The position of regression line on Vr-Wr graph provides information about the **average degree of dominance**
    - ❖ When the regression line passes through the origin- **indicates complete dominance ( $D=H1$ )**
    - ❖ When regression line passes above the origin cutting the Wr-axis- **indicates partial dominance ( $D>H1$ )**
    - ❖ When regression line passes above the origin cutting Wr-axis and touching the limiting parabola- **indicates absence of dominance**
    - ❖ When regression line passes below the origin, cutting the Vr-axis- **indicates the presence of over dominance ( $D<H1$ )**

*Continued*

2. The position of parental points along with the regression line indicates the dominance order of the parent
- The parents with more dominant genes are located near to the origin, while those with more recessive gene fall far from the origin
  - The parent with equal frequencies of dominant & recessive genes occupy an intermediate position

# 2. Griffing's Numerical Approach

- Griffing, 1956, proposed numerical method of diallel analysis which is based on the **estimates of combining ability variance and effects**
- Griffing (1956) given **four methods** of estimation of combining ability from a diallel cross:
  1. **Method 1** :  $[(P + R + F1) = n^2] = 100$  (if,  $n = 10$ )
  2. **Method 2** :  $[(P + F1) = n(n+1)/2] = 55$  (if,  $n = 10$ )
  3. **Method 3** :  $[(R + F1) = n(n-1)] = 90$  (if,  $n = 10$ )
  4. **Method 4** :  $[F1] = n(n-1)/2 = 45$  (if,  $n = 10$ )
- Method 2 utilizes direct crosses and parents in analysis is most commonly used



# Degree of Freedom and Partition of Variation In Four Methods of Diallel Analysis

Sources of variation	Method 1 (F1 + P + R)	Method 2 (F1 + P)	Method 3 (F1 + R)	Method 4 (F1)
Replication	r-1	r-1	r-1	r-1
Treatments	t-1	t-1	t-1	t-1
GCA	p-1	p-1	p-1	p-1
SCA	c	c	c-p	c-p
Reciprocal	c	-	c	-
Error	(r-1)(t-1)	(r-1)(t-1)	(r-1)(t-1)	(r-1)(t-1)

Where,

r, t, p & c are number of replications, treatment, parents & all possible cross combinations, respectively

# Main features of Griffing's approach of diallel analysis

- A numerical approach which is based on the estimates of combining ability variance and effects
- Provide information about D and H components through gca and sca variance and effects
- Analysis is possible even when parents are not present
- Helps in selection of best parental lines and cross combinations
- Calculation of genetic ratios not possible

➤ The variation among the crosses in diallel can be divided into -

**a. Variation among half-sib families**

**b. Variation among full-sib families**

**a. Variation among half-sib families:**

➤ There is **one half-sib family** for each parent in a diallel

➤ The **performance of a half-sib family** is determined by **averaging the performance of all crosses with one parent in common**

➤ The variation among the half-sib families in a diallel is an **estimate of GCA**

**b. Variation among full-sib families**

➤ A full-sib family is the **mating of two parent**;

➤ Therefore, the **number of full-sib families in a diallel = number of single crosses** that are evaluated

➤ The performance of full-sib families is used to **estimate of SCA**

- The components of variance associated with the **covariances of half-sib families (Cov HS)** and **full-sib families (Cov FS)** depends on the inbreeding (F) of genotypes used as parents in diallel

**Case 1:** When the parents are inbred ( $F = 1$ ), then genetic components are,

$$\text{Cov HS} = 1/2V_A + 1/4 V_{AA} + \text{higher order additive epistasis}$$

$$\text{Cov FS} = V_A + V_D + V_{AA} + \text{higher order of additive \& dominance epistasis}$$

**Let's assume, there is no epistasis,**

$$\text{Then, } V_A = 2 \times \text{Cov HS}$$

$$V_A = 2 \times (1/2V_A)$$

**An estimate of  $V_D$  is obtained as,**

$$\text{Cov FS} - 2 \times \text{Cov HS}$$

$$(V_A + V_D) - 2 \times (1/2V_A)$$

**Case 2:** When the parents are not inbred ( $F = 0$ ), then genetic components are,

$$\text{Cov HS} = 1/4V_A + 1/16 V_{AA} + \text{higher order of epistasis}$$

$$\text{Cov FS} = 1/2V_A + 1/4V_D + 1/4V_{AA} + \text{additive \& dominance epistasis}$$

**Let's assume, there is no epistasis,**

$$\text{Then, } V_A = 4 \times \text{Cov HS}$$

$$V_A = 4 \times (1/4V_A)$$

**An estimate of  $V_D$  is obtained as,**

$$V_D = (\text{Cov FS} - 2 \times \text{Cov HS})$$

$$= 4[(1/2V_A + 1/4V_D) - 2 \times (1/4V_A)]$$

# Models of Combining Ability Analysis

## 1. Fixed Effect Model (Model-I)

- A method of diallel analysis which is used with diallel set of crosses made among fixed sample of inbred lines
- Here, experimental material includes a set of fixed genotypes such as set of inbreds or varieties
- Inferences are drawn about individual line/variety

## 2. Random effect Model (Model-II)

- A method of diallel analysis which is used with diallel set of crosses made among random samples of individuals from random mating population
- Here, random sample of a random mating population used
- Inferences can not be drawn about individual line but about the population as a whole

# Comparison between Hayman's and Griffing's Approaches

Hayman's Approach	Griffing's Approach
➤ A graphical approach	➤ A numerical approach
➤ Components of variances can be estimated	➤ Combining ability effects and variances both can be estimated
➤ D, H <sub>1</sub> , H <sub>2</sub> , E, F & h <sup>2</sup> components estimated	➤ D and H components estimated
➤ Estimation of genetic ratios possible	➤ Estimation of genetic ratios not possible
➤ Inclusion of parents in analysis is essential	➤ Inclusion of parents in analysis is not essential
➤ Identification of superior cross combinations not possible	➤ Identification of superior cross combinations possible

# Merits:

- Used for evaluation of several single crosses in terms of genetic component of variance
- Provide information regarding GCA & SCA variance and effects and additive (A) and dominance (H) components of genetic variance
- Each parent has equal opportunity to mate and recombine with every other parents
- Combining ability analysis involves first order statistics and estimation of D and H components involves second order statistics, hence result obtained with more precision
- Heterosis, heritability and genetic advance can be estimated
- Inbreeding depression can be estimated when F<sub>2</sub> population are available
- Helps in selection of suitable parents used for hybridization
- Helps in selection of suitable breeding programme for genetic improvement of various quantitative traits



# Demerits:

- Test only limited number of parents at a time because of large number of crosses generated if you increased the number of parents

**Eg: If,  $n = 20$ ,**

Then, total number of F1's generated =  $n(n-1)/2 = 20(20-1)/2 = 190$   
in case of diallel which is nearly double than partial diallel crosses  
[where number of F1 generated =  $ns/2 = (20 \times 11)/2 = 110$ ]

All assumptions of a diallel cross are seldom fulfilled. The assumption of the absence of linkage and epistasis are generally not satisfied

Analysis of diallel crosses by hand is very difficult

**Thank You...**

# **Line x Tester Analysis**

**Dr. N K Chaurasia**  
**Dept. of GPB, MSSSoA, CUTM**

# Introduction

- The concept of L x T analysis was developed by Kempthorne in 1957
- It is a modified form of top cross scheme
  - In case of top cross,  
only one tester is used, while,
  - In case of L x T cross,  
several testers are used
- The total number of crosses =  $m.f$

Where,  $f$  = number of female line &  $m$  = number of male line (tester)

# Features of Line x Tester Analysis

- In L x T analysis, the **total number of crosses = mf** ( where, m and f are number of male and female lines, respectively)
- If, **m = 5 and f = 10**,  
then, total number crosses obtained = **50**
- Evaluate large number of germplasm lines in terms of **gca & sca variance and effects and D and H components**
- Helps in identification of **good general combiners and specific cross combinations**
- Helps in selection of **breeding procedure** for genetic improvement of various quantitative traits
- **Analysis is simple** as compared with diallel and partial diallel analysis
- Results obtained have **high level of precision** because it involve both first and second order statistics

# Steps involved in L x T analysis

1. Selection of parents
2. Generation of crosses
3. Evaluation of materials
4. Biometrical analysis

# Planting of parents for crossing

## 1. Unpaired Parents

- Here, each female and male parent is planted only once in the crossing block
- Generally, the number of male parents kept much higher (5-6 times) than female parents
- Requires the least space and parent seed
- More time is required per mating during hybridization
- Greater chance of wrong mating, particularly by inexperienced persons

## 2. Paired Parents

- The parents to be mated are planted adjacent to each other or in pairs
- Required more space and parental seeds
- Less time required per mating during hybridization
- Less chances of wrong mating

If, T or m = 5 and L or f = 10

then total number of crosses =  $LT = mf = [5 \times 10] = 50$

Where, T or m = number of male parent and L or f = number of female parent

### Plan of crossing for L x T design

Female	Male parent				
	T1	T2	T3	T4	T5
L1	X	X	X	X	X
L2	X	X	X	X	X
L3	X	X	X	X	X
L4	X	X	X	X	X
L5	X	X	X	X	X
L6	X	X	X	X	X
L7	X	X	X	X	X
L8	X	X	X	X	X
L9	X	X	X	X	X
L10	X	X	X	X	X

TOTAL NO. OF CROSSES =  $m.f = L.T = 50$



# Biometrical Analysis

- Biometrical analysis is carried out as per the procedure suggested by Kempthorne, 1957

**ANOVA Table for combining ability analysis in Line x Tester design**

Sources of variation	d.f.	Mean Squares	Expected Mean Squares
Replication	$r-1$		
Male	$m-1$	M1	$VE + rV_{fm} + rfV_m$
Female	$f-1$	M2	$VE + rV_{fm} + rmV_f$
Male x Female	$(m-1)(f-1)$	M3	$VE + rV_{fm}$
Error	$(r-1)(mf-1)$	M4	VE
Total	$rmf-1$		

Where,

$r, m, \& f$  = number of replications, male parents and female parent, respectively  
 $V_m, V_f \& V_{fm}$  = variance due to males, females and male x female, respectively

# Components of genetic variance

- L x T design divides the variation of single crosses into three components like,
  - Variance due to male parents
  - Variance due to female parents and
  - Variance due to male x female interaction
- There are full-sib and half-sib families
- When **male parent is common to all crosses**, the covariance among half-sib families designated as **Cov HSm**
- When **female parent is common to all crosses**, the covariance among half-sib families designated as **Cov HSf**
- The components of genetic variance associated with the two covariances are the same

*Continued*

- ❖ When the parents are **inbred lines (F=1)**, then,  

$$\text{Cov HS}_m \text{ and Cov HS}_f = \frac{1}{2}V_A + \frac{1}{4}V_{AA} + \text{other form of additive epistasis}$$
- ❖ The separated covariance for half-sib families of male and female parents provide separate estimates of  $V_A$
- ❖ The estimate of  $V_D = \text{Cov FS} - (\text{Cov HS}_m + \text{Cov HS}_f)$
- ❖ When the parents are **non-inbred (F=0)**, then,  

$$\text{Cov HS}_m \text{ \& Cov HS}_f = \frac{1}{4}V_A + \frac{1}{16}V_{AA} + \text{other form of additive epistasis}$$

## Genetic information:

- L x T analysis provides information on GCA and SCA variance and effects

GCA variance = Cov HS and

SCA variance = Cov FS – 2Cov HS

Additive genetic variance (VA) = general combining ability (GCA) variance

Dominance genetic variance (VD) = specific combining ability (SCA) variance

# Characteristic of tester?

## 1. Broad Genetic Base:

Broad genetic based tester includes heterogeneous cultivars, populations or crosses which determine GCA in early year of hybrid development programme

## 2. Wider adaptability:

Tester having wider adaptability can be used for evaluation of inbred lines for combining ability under changing environmental conditions

## 3. Low yield potential:

Best inbred line has masking effects due to desirable dominant alleles, hence it should not be used as tester

Thus, in order to get real expression of a lines, low performing tester should be used

## 4. Low performance for other traits

A tester should be poor in the trait for which lines are to be evaluated.

Eg: a synthetic susceptible to lodging should be used to test the capacity of lines to withstand lodging

# L x T Analysis Vs. Diallel Analysis

## L x T analysis

- Total no. of crosses =  $mf$
- Only one method of combining ability analysis
- Evaluate more parents at a time (up to 50)
- Some parent used as male whereas some used as female
- No graphical approach
- Involves direct crosses in analysis
- Each parent does not have equal chance to mate with every other parent

## Diallel analysis

- Total No. of crosses =  $n(n-1)/2$
- Four methods of combining ability analysis
- Evaluated up to 10-12 parents
- All parent can be used as male as well as female in crossing
- Graphical approach applicable
- Involves both direct as well as reciprocal cross in analysis
- Each parent has equal chance to mate every other parent

# Merits:

- Helps in selection of desirable parents
- Helps in selection of appropriate breeding procedures by measuring the GCA and SCA effects
- Good technique for evaluation of large number of germplasm at a time in terms of their combining ability variance and effects (Singh, 1978)
- This technique can be used when self-incompatibility and male sterility present
- Heterosis, heritability and genetic advanced can be estimated.
- Inbreeding depression can be calculated if F2 population available

## Demerits:

- Does not provide the estimates of epistatic variance
- Each parent does not have equal opportunity to mate and recombines with every other parent
- With inclusion of more number of testers, the number of hybrids become too large for evaluation



**Thank You...**

# **Partial Diallel Analysis**

**Dr. N K Chaurasia**  
**Dept. of GPB, MSSSoA, CUTM**

# Introduction

- The concept of partial diallel mating design was developed by Kempthorne in 1957 and was further elaborated by Kempthorne and Curnow (1961)
- Partial diallel analysis is **modified form of a diallel cross** in which only a part of all possible crosses made among 'n' parents is utilized for evaluation and biometrical analysis
- The total number of crosses =  $ns/2$

Where, n= number of parental line & s = sampled crosses which should be greater than or equal to  $n/2$

# Features of partial diallel

- Utilize **only a part of all possible crosses** from a diallel
- Here, each parent is crossed to some of the other parents, but not all
- Here, total number of crosses generated =  $ns/2$  (where,  $n$  &  $s$  are number of parents and sample crosses, respectively)
- Provides information about gca and sca variances and gca effects and D & H components but **not provide information regarding sca effects**
- Result obtained through partial diallel have **lesser precision** than diallel analysis
- Provide opportunity for **evaluation of more number of inbred lines at a time** than complete diallel cross

# General assumptions

- Normal diploid segregation
- Lack of maternal effects
- Absence of multiple alleles
- Homozygosity of parents
- Absence of linkage
- Absence of epistasis
- Random mating

# **Steps involved in partial diallel analysis**

- 1. Selection of parents**
- 2. Generation of sample crosses**
- 3. Evaluation of materials**
- 4. Biometrical analysis**

# Sampling procedure for partial diallel

- In partial diallel, only a part of all possible crosses made in a diallel fashion is utilized for evaluation and analysis
- The total number of sampled crosses is equal to  $ns/2$  (where,  $n$  = number of parents and  $s$  = number of sampled crosses per parent or per array)
- The 's' should be a whole number
- The  $s$  should be either greater than or equal to  $n/2$
- Both 'n' and 's' can neither be odd nor even. If 'n' is odd, then 's' should be even and vice-versa
- For the purpose of sampling, first a constant  $K$  is worked out as follows;

$$K = (n + 1 - s)/2$$

Where,

$n$  &  $s$  = number of parents and sampled crosses, respectively

*Continued*

If,  $n = 10$  and  $s = 5$

$$\text{then, } K = (n + 1 - s)/2 = [(10 + 1 - 5)/2] = 3$$

It means that, in each array five crosses are to be made ( $s = 5$ ) and sampling is to be being after 3 arrays, i.e., from the 4<sup>th</sup> array as given below

### Procedure of sampling and plan of crossing in a partial diallel design

Parents	P1	P2	P3	P4	P5	P6	P7	P8	P9	P10	
P1				X	X	X	X	X			
P2					X	X	X	X	X		
P3						X	X	X	X	X	
P4							X	X	X	X	
P5								X	X	X	
P6									X	X	
P7										X	
P8											
P9											
P10			TOTAL NO. OF CROSSES = 25								



# Biometrical Analysis

- Biometrical analysis is carried out as per the procedure suggested by Kempthorne, 1957

## ANOVA for combining ability in a partial diallel analysis

Sources of variation	d.f.	Mean Squares	Expected Mean Squares
Replication	$r-1$		
GCA	$n-1$	M1	$VE + rVS + [rs(n-2)(n-1)]VG$
SCA	$(ns/2) - 1$	M2	$VE + rVS$
Error	$(r-1)[(ns/2)-1]$	M3	VE
Total	$[(rns/2)-1]$		

Where,

$r, n, \& s$  = number of replications, parents and sample crosses, respectively

VG = Additive variance & additive epistasis

VS = dominance variance and genetic variance

VE = error variance

*Continued*

- The variation among the crosses in partial diallel cross can be divided into two fractions,
  1. variation among half-sib families
  2. variation among full-sib families
- There is one half-sib family for each parent in a partial diallel
- The average of all crosses with one parent in common provides the performance of a half-sib family
- The variance among the half-sib families in a partial diallel is an estimates of GCA variance and the full-sib families gives an estimates of SCA variance
- Number of full-sib families in a partial diallel =  $ns/2$  (total number of crosses)

# Differences between Half-sib & Full-sib families

Half-sib families	Full-sib families
▪ One parent common	▪ Both parent common
▪ Provides estimates of additive genetic variance	▪ Provides estimates of both additive and dominance variances
▪ $\text{Cov HS} = 1/2V_A$	▪ $\text{Cov FS} = V_A + V_D$
▪ $V_D = \text{Cov FS} - 2\text{Cov HS}$	▪ $V_A = 2\text{Cov HS}$

# Differences between diallel and partial diallel crosses

Diallel cross	Partial diallel cross
<ul style="list-style-type: none"><li>▪ Total no. of crosses = <math>n(n-1)/2</math></li></ul>	<ul style="list-style-type: none"><li>▪ Total No. of crosses = <math>ns/2</math></li></ul>
<ul style="list-style-type: none"><li>▪ Two types, i.e. full and half diallel</li></ul>	<ul style="list-style-type: none"><li>▪ Only one type</li></ul>
<ul style="list-style-type: none"><li>▪ Effective up to 10-12 parents</li></ul>	<ul style="list-style-type: none"><li>▪ Effective up to 20 parents</li></ul>
<ul style="list-style-type: none"><li>▪ Don't required sampling</li></ul>	<ul style="list-style-type: none"><li>▪ Required sampling</li></ul>
<ul style="list-style-type: none"><li>▪ Have four method of analysis</li></ul>	<ul style="list-style-type: none"><li>▪ Have only one method of analysis</li></ul>
<ul style="list-style-type: none"><li>▪ Equal chance of parents to mate with each other</li></ul>	<ul style="list-style-type: none"><li>▪ Mating chances of parents are restricted</li></ul>
<ul style="list-style-type: none"><li>▪ Estimates gca, sca variance and their effects</li></ul>	<ul style="list-style-type: none"><li>▪ Cannot estimates sca effects</li></ul>
<ul style="list-style-type: none"><li>▪ Graphical approach applicable</li></ul>	<ul style="list-style-type: none"><li>▪ Graphical approach not applicable</li></ul>

# Merits:

- Provide information regarding **gca and sca variance and gca effects** and also component of genetic variance
- **More parent can be evaluated** at a time than diallel analysis
- Heritability, genetic advance and heterosis can also be estimated
- Can be used with open pollinated species having **self-incompatibility** and **male sterility**

# Demerits:

- Each parent does not have the opportunity to mate and recombine with each other
- Estimates have less precision than those obtained from diallel analysis

**Thank You...**

# Eberhart and Russell Model (1966)

**Eberhart and Russell defined** a stable variety is one which having regression coefficient unity ( $b=1$ ) and deviation from regression line ( $s^2d = 0$ )

Stability analysis by partitioning the **genotype by environment interaction** of each genotype into-

1. slope of regression line and
2. deviations from the regression line

- This model consists of three parameters like, Mean performance of the genotype over environment, regression coefficient ( $b$ ) and deviation from regression ( $s^2d$ )
- Analysis of stability parameters is **simple** as compared to other model of stability analysis
- Degree of freedom for environment is **one**

- Less expensive model than Freeman and Perkins model because Eberhart and Russell Model required **less experimental area**
- Does not provide independent estimation for mean performance and environmental index (EI)
- Relative adaptability is measured by 3 parameters: **a. mean, b. response(regression coefficient) and c. stability performance (deviation from regression)**
- Here, firstly total variance is divided into **two components viz; genotypes and environment plus G x E interaction (E + G x E)**
- Further (E + G x E) is subdivided into **three components i.e. Environment linear, G x E (linear) and pooled deviation**
- Further the S.S due to pooled deviations are **subdivided into S.S due to individual genotype**



# ANOVA for stability (**Eberhart and Russell Model** )

Source of variation	df	SS	MS
Genotype	$g-1$		
Environment (E) + (G x E interaction)	$g(e-1)$		
Environment (linear)	$1$		
G x E (linear)	$g-1$		
Pooled deviation	$g(e-2)$		
Genotype-1	$e-2$		
Genotype-2	$e-2$		
Genotype-3	$e-2$		
Pooled error	$e(g-1)(r-1)$		

Where, g: genotype; e: environment; r: replication

# Environmental Index

- Deviation of the mean of all the varieties at a given environment from the overall mean
- $EI = \text{Mean}$
- It indicates which environment is most favourable for yield and its component traits
- The EI directly reflects the environment by positive (indicate favourable environment) and negative values(unfavourable environment) for expression of characters

# Perkins and Jinks Model(1968)

- This model also includes three parameters of stability analysis
- Less expensive model of stability analysis than Freeman and Perkins because it requires less area for experimentation
- More difficulty in analysis than Eberhart and Russell model but simpler than Freeman and Perkins model
- Does not provide independent estimation for mean performance and environmental index (EI)
- First the total genetic variance is divided into three components i.e. genotype, environment and G x E interaction
- Then, G x E interaction is subdivided into heterogeneity due to regression and sum of squares due to remainder and
- Then, S.S due to remainder divided into S.S due to individual genotype
- The degree of freedom for environment is  $e-2$

# ANOVA for stability(**Perkins and Jinks Model**)

Source of variation	df	SS	MS
Genotype	$g-1$		
Environment (joint regression)	$e-1$		
G x E	$(g-1)(e-1)$		
Heterogeneity among regression	$(g-1)$		
Reminder	$(g-1)(e-2)$		
Error	$ge(r-1)$		

Where, g: genotype; e: environment; r: replication

# Freeman and Perkins Model (1971)

- A better method for portioning of sum of squares in stability analysis
- This model also includes three parameters like Eberhart and Russell model
- This model provides independent estimation of mean performance and environmental index
- The degree of freedom for environment is  $e-2$
- Analysis is more difficult
- Analysis is more expensive than Perkins and Jinks model

- Here, firstly the total variance is divided into three components i.e. **genotype, environments, and G x G interaction**
- Further, the environmental S.S is subdivided into **combined regression and residual-1**
- The interaction variance subdivided into **heterogeneity of regression and residual-2**

# ANOVA for stability (**Freeman and Perkins Model** )

Sources of variation	df	SS	MS
Genotype	$(g-1)$		
Environment	$(e-1)$		
Combined regression	$1$		
Residual-1	$(e-2)$		
G x E interaction	$(g-1)(e-1)$		
Heterogeneity of regression	$(g-1)$		
Residual-2	$(g-1)(e-2)$		
Error (between replication )	$ge(r-1)$		

Where, g: genotype; e: environment; r: replication

# Comparison among stability model

Eberhart & Russell	Perkins & Jinks	Freeman & Perkins
3-parameter model	parameter model	parameter model
Two fraction of variation: genotype and G x E	Three fraction of variation: genotype, Env. and G x E	Three fraction of variation: genotype env. and G x E
Independent estimation of mean performance and environmental index is <b>not possible</b>	Independent estimation of mean performance and environmental index is <b>not possible</b>	Independent estimation of mean performance and environmental index is <b>possible</b>
Degree of freedom for env. is <b>1</b>	Degree of freedom for env. is <b>e-2</b>	Degree of freedom for env. is <b>e-2</b>
Calculation is simple	Calculation is difficult	Calculation is more difficult
Involved less expenditure	Involved less expenditure	Involved more expenditure



# **Additive Main Effects and Multiplicative interaction (AMMI) Model**

# Advantage of stability analysis

- Helps in understanding the adaptability of crop varieties over a wide range of environment
- Used of adaptable genotype for general cultivation over wide range of environment
- Use of stable genotype in hybridization programme
- Predicting the response of various genotype over changing environments
- Selection of stable genotypes for yield and its contributing traits