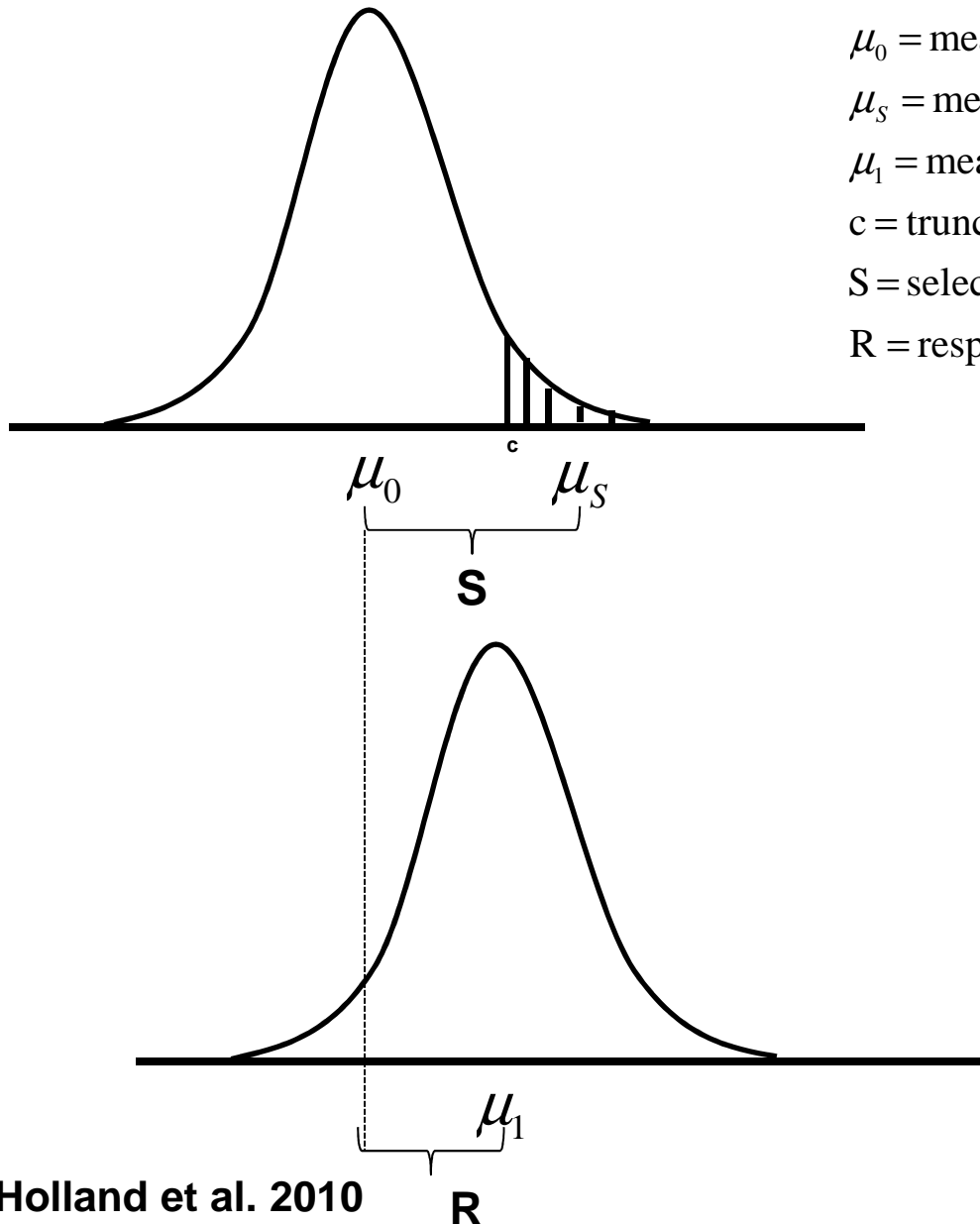


Lecture 6

Heritability and Field Design

Lucia Gutierrez lecture notes
Tucson Winter Institute

Selection Response



μ_0 = mean of the initial Random Mating population
 μ_s = mean of a selected individuals from R.M. population
 μ_1 = mean of a progeny of selected individuals
 c = truncation point
 S = selection differential (ΔX)
 R = response to selection (ΔY)

$$\Delta Y = b(\Delta X)$$

$$R = b_{x,y} S$$

$$b_{x,y} = \frac{R}{S} = h^2$$

Heritability: “Expected proportion of selection differential to be achieved as a gain from selection” (Hanson, 1963).

Heritability

REALIZED HERITABILITY

The relation between the observed response to selection and the selection differential.

$$h^2 = \frac{\hat{R}}{\hat{S}}$$

NARROW SENSE HERITABILITY

The fraction of all trait variation due to variation in breeding values (additive variance).

$$h^2 = \frac{V_A}{V_P}$$

BROAD SENSE HERITABILITY

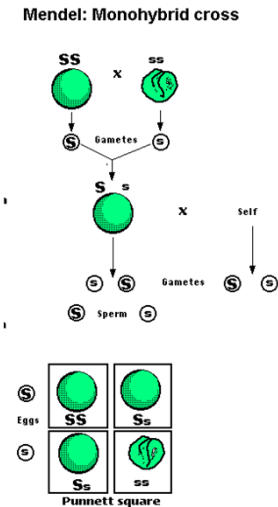
The fraction of all trait variation due to variation in genotypic values (genetic variance).

$$H^2 = \frac{V_G}{V_P}$$

Heritability

FUNCTION OF POPULATION

- They are functions of both genetic and environmental variance, therefore a property of the population.
- Suppose one inbred line with a Mendelian inherited trait. How much is the genetic variance? Will the trait segregate? How much is the h^2 ? Does it mean that the trait is not genetically determined? Since there is no genetic variation within this population, $h^2=H^2=0$. However, it does not mean that the trait is not genetically determined.
- Additionally, it is unreliable to predict future response to selection because while conducting selection the population changes.



Heritability

ENVIRONMENTAL VARIANCE

Since h^2 is also a function of environmental variance, and decreasing environmental variance increases h^2 , controlled conditions would be optimal for identifying superior genotypes (predicting breeding values). However, caution should be exercised because GxE is important for many traits and therefore selecting in a non-targeted environment could be detrimental.



Parent-Offspring Regression

$$z_{oi} = \mu + b_{o|p} (z_{pi} - \mu) + e_i$$

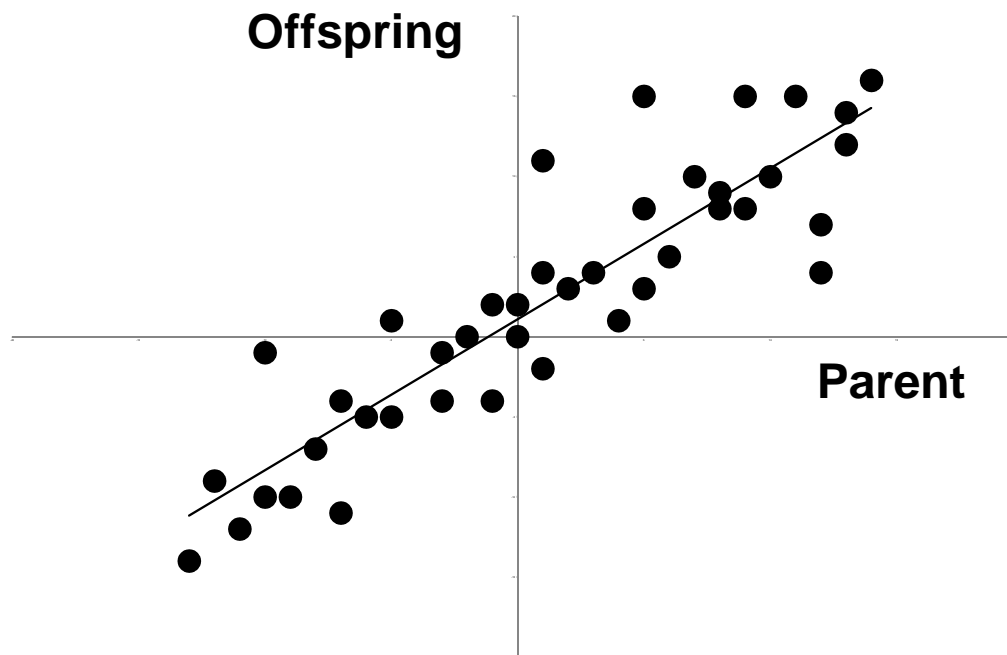
z_{oi} = phenotype of the i - th offspring

μ = population mean

$b_{o|p}$ = regression parent - offspring slope

z_{pi} = phenotype of parent of i - th offspring

e_i = deviations



Parent-Offspring Regression

Regression one parent – offspring (one offspring or the mean of multiple offspring).

$$E(\hat{b}_{o|p}) = \frac{\sigma(z_o, z_p)}{\sigma^2(z_p)} \cong \frac{\frac{1}{2}\sigma_A^2 + \frac{1}{4}\sigma_{AA}^2 + \sigma(E_o, E_p)}{\sigma_z^2} \cong \frac{1}{2}h^2, \quad h^2 \cong 2b_{o|p}$$

Regression one parent on offspring – no environment correlation among parent and offspring.

$$E(\hat{b}_{o|p}) = \frac{\sigma(z_o, z_p)}{\sigma^2(z_p)} \cong \frac{\frac{1}{2}\sigma_A^2 + \frac{1}{4}\sigma_{AA}^2}{\sigma_z^2} \cong \frac{1}{2}h^2, \quad h^2 \cong 2b_{o|p}$$

Regression mid parent on offspring – no environment correlation among parent and offspring.

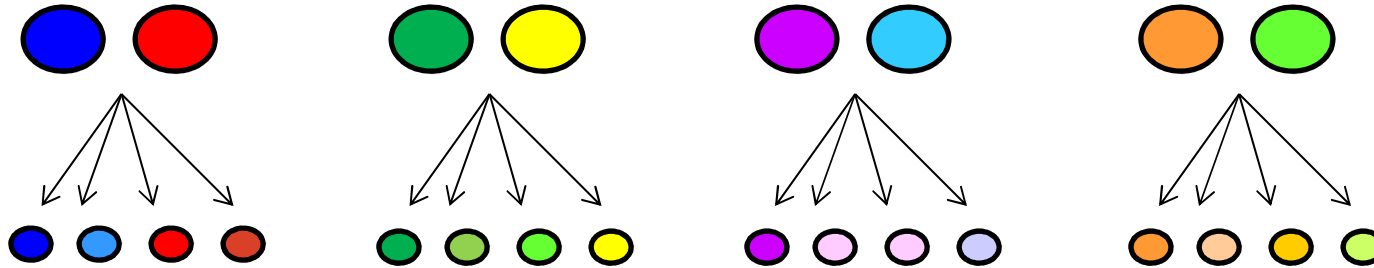
$$E(\hat{b}_{o|\bar{p}}) = \frac{\sigma(z_o, \bar{z}_p)}{\sigma^2(\bar{z}_p)} = \frac{\sigma(z_o, \frac{1}{2}z_{P1} + \frac{1}{2}z_{P2})}{\sigma^2(\frac{1}{2}z_{P1} + \frac{1}{2}z_{P2})} = \frac{\frac{1}{2}\sigma_A^2 + \frac{1}{4}\sigma_{AA}^2}{\frac{1}{2}\sigma_z^2} = h^2, \quad h^2 \cong b_{o|\bar{p}}$$

Regression parent -offspring inbreeding – no environment correlation.

$$E(\hat{b}_{S0:1|S0}) = \frac{\sigma(z_{S0}, z_{S0:1})}{\sigma^2(z_{S0})} = \frac{\sigma_A^2 + \frac{1}{2}\sigma_D^2 + \frac{1}{2}\sigma_{D1}^2 + \sigma_{AA}^2}{\sigma_z^2} \cong h^2, \quad h^2 \cong b_{S0:1|S0}$$

Mating Designs

FULL-SIB DESIGN: N full-sib families with n offspring each.



$$z_{ij} = \mu + \underline{f_i} + \underline{w_{ij}}$$

z_{ij} = phenotype of the j -th offspring of the i -th family

μ = population mean

f_i = effect of the i -th family

w_{ij} = residual error (segregation, dominance, environmental contribution)
within - family variance

| SoV | df | SS | MS | EMS |
|-----------------|----------|---|-------------------|----------------------------------|
| Among-families | $N-1$ | $SS_f = n \sum_i (\bar{z}_{i.} - \bar{z}_{..})^2$ | $SS_f / df_{(f)}$ | $\sigma_{w(FS)}^2 + n\sigma_f^2$ |
| Within-families | $n(N-1)$ | $SS_w = \sum_{i,j} (z_{ij} - \bar{z}_{i.})^2$ | $SS_w / df_{(w)}$ | $\sigma_{w(FS)}^2$ |

Mating Designs

FULL-SIB DESIGN: N full-sib families with n offspring each.

$$z_{ij} = \mu + \underline{f_i} + \underline{w_{ij}}$$

$$\begin{aligned} \text{Cov}(FS) &= \sigma(z_{ij}, z_{ij}) \\ &= \sigma(\mu + f_i + w_{ij}, \mu + f_i + w_{ij}) \\ &= \sigma(f_i, f_i) + \sigma(f_i, w_{ij}) + \sigma(w_{ij}, f_i) + \sigma(w_{ij}, w_{ij}) \\ &= \sigma_f^2 \end{aligned}$$

$$\sigma_f^2 = \frac{1}{2} \sigma_A^2 + \frac{1}{4} \sigma_D^2 + \sigma_{Ec}^2$$

$$\sigma_P^2 = \sigma_f^2 + \sigma_{w(FS)}^2$$

$$\begin{aligned} \sigma_{w(FS)}^2 &= \sigma_P^2 - \left(\frac{1}{2} \sigma_A^2 + \frac{1}{4} \sigma_D^2 + \sigma_{Ec}^2 \right) \\ &= \sigma_A^2 + \sigma_D^2 + \sigma_E^2 - \left(\frac{1}{2} \sigma_A^2 + \frac{1}{4} \sigma_D^2 + \sigma_{Ec}^2 \right) \\ &= \frac{1}{2} \sigma_A^2 + \frac{3}{4} \sigma_D^2 + \sigma_E^2 + \sigma_{Ec}^2 \end{aligned}$$

Mating Designs

FULL-SIB DESIGN: N full-sib families with n offspring each.

$$z_{ij} = \mu + \underline{f}_i + \underline{w}_{ij}$$

| SoV | df | SS | MS | EMS |
|-----------------|----------|---|-----------------|----------------------------------|
| Among-families | $N-1$ | $SS_f = n \sum_i (\bar{z}_{i.} - \bar{z}_{..})^2$ | $SS_f/df_{(f)}$ | $\sigma_{w(FS)}^2 + n\sigma_f^2$ |
| Within-families | $n(N-1)$ | $SS_w = \sum_{i,j} (z_{ij} - \bar{z}_{i.})^2$ | $SS_w/df_{(w)}$ | $\sigma_{w(FS)}^2$ |

$$\text{Var}(f) = \frac{MS_f - MS_w}{n}$$

$$\text{Var}(w) = MS_w$$

$$\text{Var}(z) = \text{Var}(f) + \text{Var}(w)$$

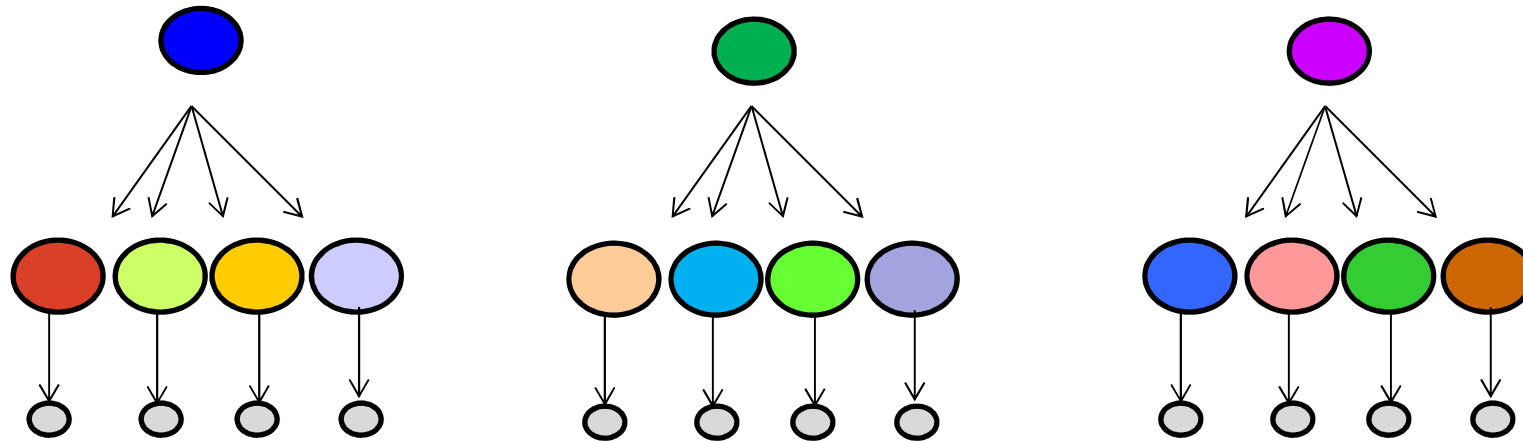
$$t_{FS} = \frac{\text{Var}(f)}{\text{Var}(z)} = \frac{\frac{1}{2}\sigma_A^2 + \frac{1}{4}\sigma_D^2 + \sigma_{Ec}^2}{\sigma_z^2} = \frac{1}{2}h^2 + \frac{\frac{1}{4}\sigma_D^2 + \sigma_{Ec}^2}{\sigma_z^2}$$

$$h^2 \cong 2t_{FS}$$

$$SE(h^2) \cong 2(1-t_{FS})[1 + (n-1)t_{FS}] \sqrt{\frac{2}{[Nn(n-1)]}}$$

Mating Designs

HALF-SIB DESIGN: N half-sib families with n offspring each.



$$z_{ij} = \mu + \underline{f_i} + \underline{w_{ij}}$$

z_{ij} = phenotype of the j -th offspring of the i -th family

μ = population mean

f_i = effect of the i -th family

w_{ij} = residual error (segregation, dominance, environmental contribution)
within - family variance

| SoV | df | SS | MS | EMS |
|-----------------|----------|---|-----------------|----------------------------------|
| Among-families | $N-1$ | $SS_f = n \sum_i (\bar{z}_{i.} - \bar{z}_{..})^2$ | $SS_f/df_{(f)}$ | $\sigma_{w(FS)}^2 + n\sigma_f^2$ |
| Within-families | $n(N-1)$ | $SS_w = \sum_{i,j} (z_{ij} - \bar{z}_{i.})^2$ | $SS_w/df_{(w)}$ | $\sigma_{w(FS)}^2$ |

Mating Designs

HALF-SIB DESIGN: N half-sib families with n offspring each.

$$z_{ij} = \mu + \underline{f_i} + \underline{w_{ij}}$$

z_{ij} = phenotype of the j -th offspring of the i -th family

μ = population mean

f_i = effect of the i -th family

w_{ij} = residual error (segregation, dominance, environmental contribution)
within - family variance

| SoV | df | SS | MS | EMS |
|-----------------|----------|---|-------------|----------------------------------|
| Among-families | $N-1$ | $SS_f = n \sum_i (\bar{z}_{i.} - \bar{z}_{..})^2$ | SS_f/df_f | $\sigma_{w(FS)}^2 + n\sigma_f^2$ |
| Within-families | $n(N-1)$ | $SS_w = \sum_{i,j} (z_{ij} - \bar{z}_{i.})^2$ | SS_w/df_w | $\sigma_{w(FS)}^2$ |

$$\text{Var}(f) = \frac{MS_f - MS_w}{n}$$

$$\text{Var}(w) = MS_w$$

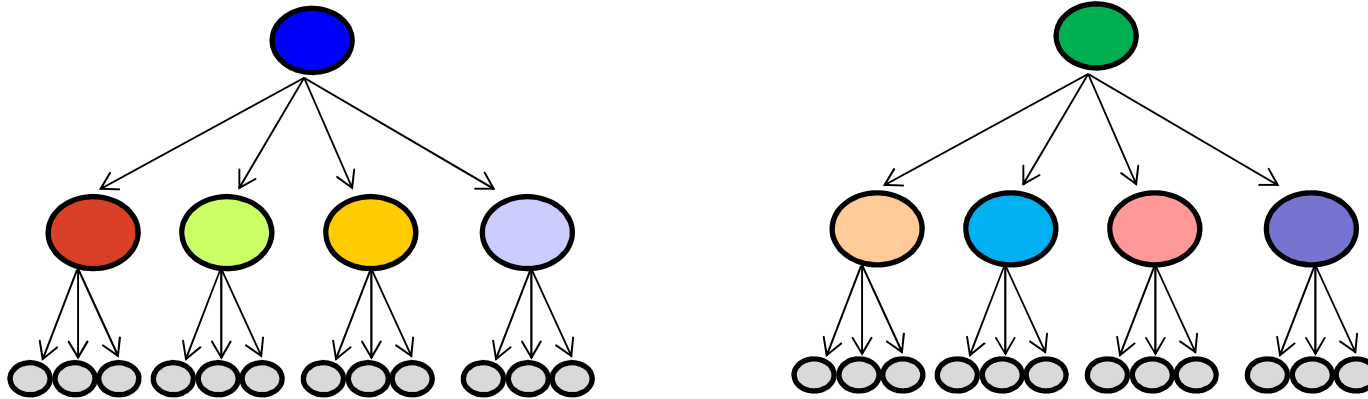
$$\text{Var}(z) = \text{Var}(f) + \text{Var}(w)$$

$$t_{HS} = \frac{\text{Var}(f)}{\text{Var}(z)} = \frac{1/4 \sigma_A^2}{\sigma_z^2} = \frac{1}{4} h^2$$

$$h^2 \cong 4t_{HS}$$

Mating Designs

NORTH CAROLINA DESIGN I: Each male (N sire) is mated to several unrelated females (M dams) to produce n offspring per dam.



z_{ijk} = phenotype of the k - th offspring from the family of the i - th sire and j - th dam

μ = population mean

s_i = effect of the i - th sire

d_{ij} = effect of the j - th dam mated to the i - th sire

w_{ijk} = residual error (within - family variance deviations)

$$z_{ijk} = \mu + \underline{s_i} + \underline{d_{j(i)}} + \underline{w_{ijk}}$$

Mating Designs

NORTH CAROLINA DESIGN I: Each male (N sire) is mated to several unrelated females (M dams) to produce n offspring per dam.

$$z_{ijk} = \mu + \underline{s}_i + \underline{d}_{j(i)} + \underline{w}_{ijk}$$

| SoV | df | SS | MS | EMS |
|------------|--------|--|-----------------|---|
| Sires | N-1 | $SS_s = Mn \sum_{i,j} (\bar{z}_{i.} - \bar{z}_{..})^2$ | $MS_s/df_{(s)}$ | $\sigma_w^2 + n\sigma_d^2 + Mn\sigma_s^2$ |
| Dams(Sire) | N(M-1) | $SS_d = \sum_{i,j} (z_{ij} - \bar{z}_{i.})^2$ | $MS_d/df_{(d)}$ | $\sigma_w^2 + n\sigma_d^2$ |
| Sibs(dams) | T-NM | $SS_w = \sum_{i,j,k} (z_{ijk} - \bar{z}_{ij.})^2$ | $MS_w/df_{(w)}$ | σ_w^2 |

$$\text{Var}(s) = \frac{MS_s - MS_d}{Mn}$$

$$\text{Var}(d) = \frac{MS_d - MS_w}{n}$$

$$\text{Var}(w) = MS_w$$

$$\text{Var}(z) = \text{Var}(s) + \text{Var}(d) + \text{Var}(w)$$

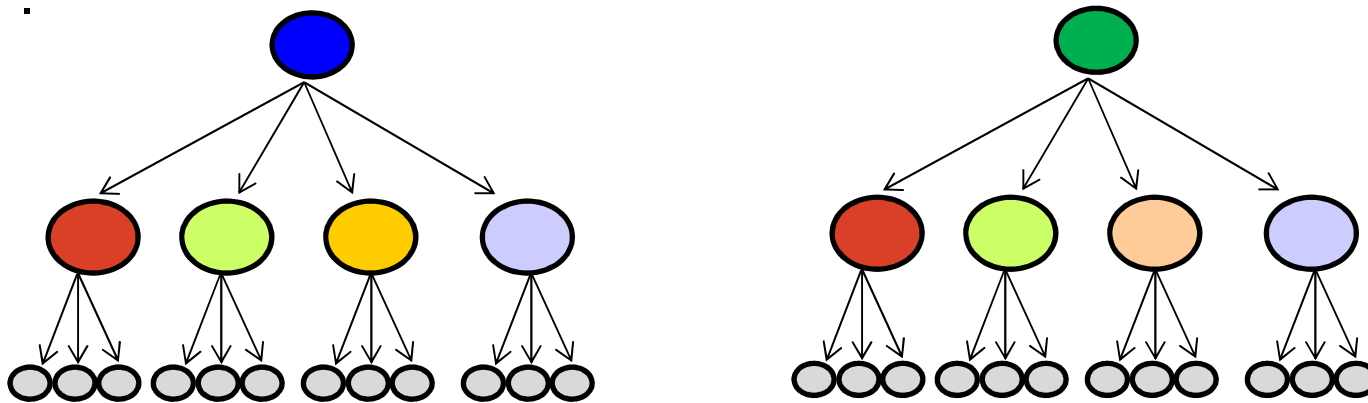
$$t_{PHS} = \frac{\text{Var}(s)}{\text{Var}(z)} = \frac{1/4 \sigma_A^2}{\sigma_z^2} = \frac{1}{4} h^2$$

$$t_{FS} = \frac{\text{Var}(s) + \text{Var}(d)}{\text{Var}(z)} = \frac{1/2 \sigma_A^2 + 1/4 \sigma_D^2 + \sigma_{Ec}^2}{\sigma_z^2} = \frac{1}{2} h^2 + \frac{1/4 \sigma_D^2 + \sigma_{Ec}^2}{\sigma_z^2}$$

$$h^2 \cong 4t_{PHS}$$

Mating Designs

NORTH CAROLINA DESIGN II: A group of sires (N_S sires) are mated to an independent group of dams (N_D dams) to produce n offspring



$$z_{ijk} = \mu + \underline{s_i} + \underline{d_j} + \underline{I_{ij}} + \underline{w_{ijk}}$$

z_{ijk} = phenotype of the k - th offspring from the family of the i - th sire and j - th dam

μ = population mean

s_i = effect of the i - th sire

d_j = effect of the j - th dam

I_{ij} = effect of the interaction between the i - th sire and the j - th dam

w_{ijk} = residual error (within - family variance deviations)

Mating Designs

NORTH CAROLINA DESIGN II: A group of sires (N_s sires) are mated to an independent group of dams (N_d dams) to produce n offspring

$$z_{ijk} = \mu + \underline{s}_i + \underline{d}_j + \underline{I}_{ij} + \underline{w}_{ijk}$$

| SoV | df | SS | EMS |
|-------------|----------------------|---|---|
| Sires | $N_s - 1$ | $SS_s = nN_d \sum_i (\bar{z}_{i..} - \bar{z}_{...})^2$ | $\sigma_w^2 + n\sigma_I^2 + nN_d\sigma_s^2$ |
| Dams | $N_d - 1$ | $SS_d = nN_s \sum_j (\bar{z}_{.j.} - \bar{z}_{...})^2$ | $\sigma_w^2 + n\sigma_I^2 + nN_s\sigma_d^2$ |
| Interaction | $(N_s - 1)(N_d - 1)$ | $SS_I = \sum_{i,j} (\bar{z}_{ij.} - \bar{z}_{i..} - \bar{z}_{.j.} - \bar{z}_{...})^2$ | $\sigma_w^2 + n\sigma_I^2$ |
| Sibs | $N_s N_d (n - 1)$ | $SS_w = \sum_{i,j,k} (z_{ijk} - \bar{z}_{ij.})^2$ | σ_w^2 |

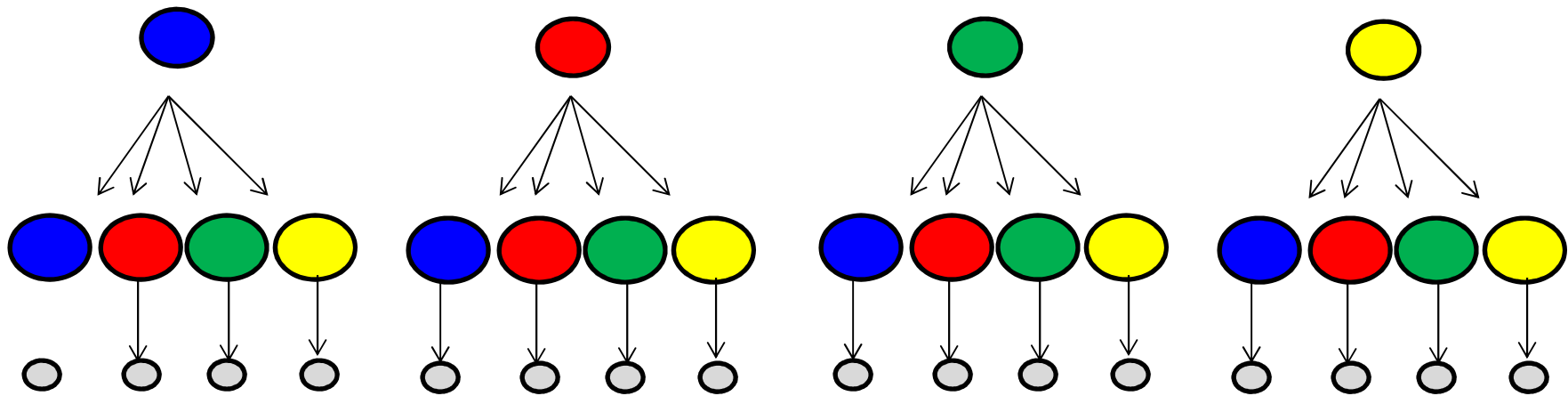
$$t_{PHS} = \frac{1/4 \sigma_A^2}{\sigma_z^2}$$

$$t_{MHS} = \frac{1/4 \sigma_A^2 + \sigma_{Gm}^2 + \sigma_{Ec}^2}{\sigma_z^2}$$

$$t_I = \frac{1/4 \sigma_D^2}{\sigma_z^2}$$

Mating Designs

DIALLELS: A group of individuals (N) are mated to the same set of individuals (N) to produce n offspring



Full Diallele (all selfed and reciprocal crosses are made)

Incomplete Diallele – no selfed crosses

Incomplete Diallele – no selfed, no reciprocal crosses

z_{ijk} = phenotype of the k - th offspring from the i - th and j - th parents and j - th dam

μ = population mean

g_i = general combining ability of parent i - th

g_j = general combining ability of parent j - th

s_{ij} = specific combining ability of parents i - th and j - th

w_{ijk} = residual error (within - family variance deviations)

Mating Designs

DIALLELS: A group of individuals (N) are mated to the same set of individuals (N) to produce n offspring. Analysis for incomplete diallele without selfed or reciprocal crosses.

$$z_{ijk} = \mu + \underline{g}_i + \underline{g}_j + \underline{s}_{ij} + \underline{w}_{ijk}$$

| SoV | df | SS | EMS |
|------|-------------------|--|---|
| GCA | N-1 | $SS_{GCA} = \frac{n(N-1)^2}{N-2} \sum_i (\bar{z}_{i..} - \bar{z}_{...})^2$ | $\sigma_w^2 + n\sigma_{SGA}^2 + n(N-2)\sigma_{GCA}^2$ |
| SCA | N(N-3)/2 | $SS_{SCA} = n \sum_{i<j} (\bar{z}_{ij.} - \bar{z}_{...})^2 - SS_{GCA}$ | $\sigma_w^2 + n\sigma_{SCA}^2$ |
| Sibs | (n-1)[N(N-1)/2-1] | $SS_w = \sum_{i<j,k} (z_{ijk} - \bar{z}_{ij.})^2$ | σ_w^2 |

$$t_{GCA} = \frac{1/4 \sigma_A^2}{\sigma_z^2}$$

$$t_{SCA} = \frac{1/4 \sigma_D^2}{\sigma_z^2}$$

Genotypic Means

GENOTYPIC MEANS:

$$z_{ijkl} = G_i + E_j + GE_{ij} + p_{ijkl} + \varepsilon_{ijkl}$$

The environment includes non-genetic factors that affect the phenotype, and usually has a large influence on quantitative traits.

- **Micro-environment.** Environment of a single plant. Need to be controlled with experimental design.
- **Macro-environment.** Environment associated to a location and time. GxE is the norm and not the exception in plants. Therefore defining the target environments is a crucial part in plant breeding, both for variance component estimation and identifying superior genotypes.

Micro-environmental variation

HERITABILITY: For related individuals, heritability can be calculated as previously shown. The previous calculation assumes individual plants are measured, and heritability on an individual plant basis is estimated. However, because quantitative traits measured on individual plants have large non-genetic effects, heritability on a mean basis is higher.

Individual plant basis, $n=1$, $r=1$, $e=1$

$$\sigma_P^2 = \sigma_F^2 + \sigma_{FE}^2 + \sigma_e^2 + \sigma_w^2$$

$$\hat{h}_F^2 = \frac{\left(\frac{4}{1+F_P}\right)\hat{\sigma}_F^2}{\hat{\sigma}_p^2} = \frac{\left(\frac{4}{1+F_P}\right)\hat{\sigma}_F^2}{\hat{\sigma}_F^2 + \hat{\sigma}_{FE}^2 + \hat{\sigma}_e^2 + \hat{\sigma}_w^2}$$

Plot basis, $n=n$, $r=1$, $e=1$

$$\sigma_P^2 = \sigma_F^2 + \sigma_{FE}^2 + \sigma_e^2 + \frac{\sigma_w^2}{n}$$

$$h_F^2 = \frac{\hat{\sigma}_F^2}{\hat{\sigma}_p^2} = \frac{\hat{\sigma}_F^2}{\hat{\sigma}_F^2 + \hat{\sigma}_{FE}^2 + \hat{\sigma}_e^2 + \frac{\hat{\sigma}_w^2}{n}}$$

Plot basis, multiple env, $n=n$, $r=r$, $e=e$

$$\sigma_P^2 = \sigma_F^2 + \frac{\sigma_{FE}^2}{e} + \frac{\sigma_e^2}{er} + \frac{\sigma_w^2}{ern}$$

$$h_F^2 = \frac{\hat{\sigma}_F^2}{\hat{\sigma}_p^2} = \frac{\hat{\sigma}_F^2}{\hat{\sigma}_F^2 + \frac{\hat{\sigma}_{FE}^2}{e} + \frac{\hat{\sigma}_e^2}{er} + \frac{\hat{\sigma}_w^2}{ern}}$$

Experimental Design and Analysis

So we need good experimental designs to test genotypes!

FISHER'S PRINCIPLES. Statistical theory for efficient estimation (i.e. unbiased and of minimum variance) of treatment mean differences are based on 3 principles:

- **Randomization**, random assignment of treatments to experimental units provides valid estimation of experimental error, unbiased comparisons of treatments, and justifies statistical inference.
- **Replication** allows estimation of experimental error variance, and decreases mean variances (i.e. variance of a mean = variance of the observation divided by the number of replications).
- **Local control** is the grouping of homogenous experimental units. Well chosen blocks will decrease error variance.

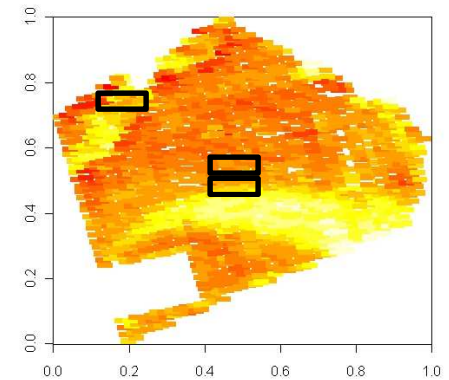
Experimental Design and Analysis

CLASSIFICATION OF EXPERIMENTAL DESIGNS:

1. Experimental Units.
 1. Homogenous – Complete Randomized Design (CRD)
 2. Heterogenous in one way – Randomized Complete Block Design (RCBD)
 3. Heterogenous in more than one way – Latin squares or latinized designs.

2. Large number of treatments.
 1. Incomplete Block Designs (IBD or Alpha)
 2. Unreplicated experiments (or Federer)

3. Modeling (post-blocking, spatial analysis)



CRD

| | | | |
|---|---|---|---|
| 1 | 2 | 1 | 4 |
| 3 | 3 | 4 | 2 |

$$y_{ij} = \mu + \alpha_i + \varepsilon_{ij}$$

RCBD

| | | | |
|---|---|---|---|
| 1 | 2 | 4 | 3 |
| 3 | 4 | 1 | 2 |

$$y_{ij} = \mu + \alpha_i + \beta_j + \varepsilon_{ij}$$

IBD

| | | | |
|---|---|---|---|
| 1 | 2 | 4 | 3 |
| 3 | 4 | 1 | 2 |

$$y_{ijk} = \mu + \alpha_i + \beta_j + \gamma_{k(j)} + \varepsilon_{ijk}$$

Complete Randomized Design (CRD)

RICE EXAMPLE. YIELD COMPARISON OF 4 RICE CULTIVARS

Treatments: 4 cultivars

Replications: 4 homogenous experimental plots per treatment

Experimental design: CRD

Dependent variable : $Y = \text{Grain yield (Kg ha}^{-1}\text{)}$

TREATMENT ASSIGNMENT: Each treatment is assigned at random to the experimental units.

- Experimental unit: one plot.

| | | | | | | | |
|----|---|----|---|----|---|----|---|
| 1 | 3 | 2 | 2 | 3 | 2 | 4 | 1 |
| 5 | 2 | 6 | 1 | 7 | 4 | 8 | 3 |
| 9 | 1 | 10 | 4 | 11 | 3 | 12 | 4 |
| 13 | 1 | 14 | 4 | 15 | 2 | 16 | 3 |

EXPERIMENTAL PRINCIPLES

1. Randomization
2. Replication
3. Local control



Complete Randomized Design (CRD)

$$Y_{ij} = \mu + \alpha_i + \varepsilon_{ij}$$

Y_{ij} = response of the i -th treatment on the j -th rep

μ = population mean

α_i = effect of the i -th treatment

ε_{ij} = experimental error (residual)

ASSUMPTIONS:

1. TO THE MODEL:

- Is correct (in relation to the experimental units)
- Is additive

2. EXPERIMENTAL ERRORS:

- Are random variables
- $\varepsilon_{ij} \sim N$
- $E(\varepsilon_{ij}) = 0$ for every i, j
- $V(\varepsilon_{ij}) = \sigma^2$ for every i, j
- Are independent

3. “BY DEFINITION” $\alpha_i = \mu_i - \mu$



Complete Randomized Design (CRD)

| T1 | T2 | T3 | T4 |
|----|----|----|----|
| 47 | 50 | 57 | 54 |
| 52 | 54 | 53 | 65 |
| 62 | 67 | 69 | 74 |
| 51 | 57 | 57 | 59 |
| 53 | 57 | 59 | 63 |

58

| SoV | df | SS | MS | F | p |
|-------|--------|---|----------------------|---------------|---|
| Geno | (t-1) | $SS_T = r \sum_i (\bar{y}_i - \bar{y}_{..})^2$ | $SS_T / gl(T)$ | MS_T / MS_E | |
| Error | t(r-1) | $SS_E = \sum_{i=1}^t \sum_{j=1}^r (y_{ij} - \bar{y}_i)^2$ | $SS_E / gl(E)$ | | |
| Total | tr-1 | $SS_{TOT} = \sum_{ij} (y_{ij} - \bar{y}_{..})^2$ | $SS_{TOT} / gl(TOT)$ | | |

| SoV | df | SS | MS | F | p |
|-------|----|-----|------|------|-------|
| Geno | 3 | 208 | 69.3 | 1.29 | 0.323 |
| Error | 3 | 646 | 59.8 | | |
| Total | 15 | 854 | | | |



Randomized Complete Block Design (RCBD)

WHEAT EXAMPLE. PLANT HEIGHT OF 5 ADVANCED LINES OF WHEAT IN 5 BLOCKS

Treatment: 5 advanced wheat lines

Block: 5 different blocks

Experimental design: RCBD

Dependent variable: $Y = \text{cm}$

TREATMENT ASSIGNMENT: Within a block, each treatment is assigned randomly to the experimental units within a block. Randomization in each block is independent.

- Experimental unit: one plot.

| B 1 | B 2 | B 3 | B 4 | B 5 |
|-----|-----|-----|-----|-----|
| 1 | 2 | 5 | 2 | 5 |
| 2 | 3 | 1 | 4 | 2 |
| 4 | 1 | 2 | 1 | 4 |
| 3 | 5 | 4 | 3 | 1 |
| 5 | 4 | 3 | 5 | 3 |

EXPERIMENTAL PRINCIPLES

1. Randomization
2. Replication
3. Local control



Randomized Complete Block Design (RCBD)

$$Y_{ij} = \mu + \alpha_i + \beta_j + \varepsilon_{ij}$$

Y_{ij} = response of the i -th treatment on the j -th block

μ = population mean

α_i = effect of the i -th treatment

β_j = effect of the j -th block

ε_{ij} = experimental error (residual)

ASSUMPTIONS:

1. TO THE MODEL:

- Is correct (in relation to the experimental units)
- Is additive
- There is **NO** treatment by block interaction

2. EXPERIMENTAL ERRORS:

- Are random variables
- $\varepsilon_{ij} \sim N$
- $E(\varepsilon_{ij}) = 0$ for every i, j
- $V(\varepsilon_{ij}) = \sigma^2$ for every i, j
- Are independent

3. "BY DEFINITION" $\alpha_i = \mu_i - \mu$



Randomized Complete Block Design (RCBD)

| | T1 | T2 | T3 | T4 | T5 | |
|----|-----|-----|-----|-----|-----|------|
| B1 | 8 | 10 | 8 | 9 | 10 | 9 |
| B2 | 7 | 9 | 8 | 8 | 9 | 8.2 |
| B3 | 6 | 8 | 9 | 8 | 8 | 7.8 |
| B4 | 6 | 7 | 9 | 8 | 7 | 7.4 |
| B5 | 7 | 9 | 10 | 7 | 9 | 8.4 |
| | 6.8 | 8.6 | 8.8 | 8.0 | 8.6 | 8.16 |

| SoV | df | SS | MS | F | p |
|-------|----|-------|-------|------|--------|
| Block | 4 | 7.36 | 1.84 | 2.77 | 0.0637 |
| Geno | 4 | 13.36 | 3.34 | 5.02 | 0.0081 |
| Error | 16 | 10.64 | 0.665 | | |
| Total | 24 | 31.36 | | | |

| SoV | df | SS | MS | F | p |
|-------|------------|--|-------------------|---------------|---|
| Block | (r-1) | $SS_B = t \sum_j (\bar{y}_{.j} - \bar{y}_{..})^2$ | $SS_B / gl_{(B)}$ | | |
| Geno | (t-1) | $SS_T = r \sum_i (\bar{y}_{i.} - \bar{y}_{..})^2$ | $SS_T / gl_{(T)}$ | MS_T / MS_E | |
| Error | (t-1)(r-1) | $SS_E = \sum_{i=1}^t \sum_{j=1}^r (y_{ij} - \bar{y}_{i.} - \bar{y}_{.j} + \bar{y}_{..})^2$ | $SS_E / gl_{(E)}$ | | |
| Total | tr-1 | $SS_{TOT} = \sum_{ij} (y_{ij} - \bar{y}_{..})^2$ | | | |



Number of treatments

HOW TO DEAL WITH HIGH NUMBER OF TREATMENTS?

1. **STRATIFICATION:** Group genotypes with similar characteristics (maturity, color, family), compare within groups. NO BETWEEN GROUP COMPARISONS.
2. **PRODUCE HOMOGENOUS EXPERIMENTAL UNITS:** Make every effort to homogenize experimental area (look for soil similarity, field conditions to reduce variation, choose seeds of similar vigor).
3. **USE REPEATED CHECKS:** You may use checks in a systematic way to control or model soil heterogeneity.
4. **EXPERIMENTAL DESIGN WITH SPATIAL CONSIDERATIONS.** Use experimental designs that include a large number of treatments while controlling variability (i.e. alpha designs, unrep, etc.).

Incomplete Block Designs (IBD)

IBD CLASIFICACION.

1. Based on their balance.

A. Balanced: each treatment is compared with another one the same number of times in an incomplete block (usually once). A large number of reps are needed.

| BI 1 | BI 2 | BI 3 | BI 4 | BI 5 | BI 6 | BI 7 | BI 8 | BI 9 | BI 10 | BI 11 | BI 12 |
|------|------|------|------|------|------|------|------|------|-------|-------|-------|
| 1 | 4 | 7 | 1 | 2 | 3 | 1 | 2 | 3 | 1 | 2 | 3 |
| 2 | 5 | 8 | 4 | 5 | 6 | 5 | 6 | 4 | 6 | 4 | 5 |
| 3 | 6 | 9 | 7 | 8 | 9 | 9 | 7 | 8 | 8 | 9 | 7 |

B. Partially unbalanced: Not all pair of treatments are compared in an incomplete block the same number of times.

Precision for mean comparisons is different depending on the pair. Statistical analysis more complex.

| BI 1 | BI 2 | BI 3 | BI 4 | BI 5 | BI 6 |
|------|------|------|------|------|------|
| 1 | 4 | 7 | 1 | 2 | 3 |
| 2 | 5 | 8 | 4 | 5 | 6 |
| 3 | 6 | 9 | 7 | 8 | 9 |



Incomplete Block Designs (IBD)

2. Based on their structure.

A. Resolubles: incomplete blocks could be grouped in complete replications.

| BI 1 | BI 2 | BI 3 | BI 4 | BI 5 | BI 6 | BI 7 | BI 8 | BI 9 | BI 10 | BI 11 | BI 12 |
|------|------|------|------|------|------|------|------|------|-------|-------|-------|
| 1 | 4 | 7 | 1 | 2 | 3 | 1 | 2 | 3 | 1 | 2 | 3 |
| 2 | 5 | 8 | 4 | 5 | 6 | 5 | 6 | 4 | 6 | 4 | 5 |
| 3 | 6 | 9 | 7 | 8 | 9 | 9 | 7 | 8 | 8 | 9 | 7 |

B. Non resolubles: incomplete blocks cannot be grouped in complete replications.

| BI 1 | BI 2 | BI 4 | BI 5 | BI 7 | BI 8 | BI 10 | BI 11 |
|------|------|------|------|------|------|-------|-------|
| 1 | 4 | 1 | 2 | 1 | 2 | 1 | 2 |
| 2 | 5 | 4 | 5 | 5 | 6 | 6 | 4 |
| 3 | 6 | 7 | 8 | 9 | 7 | 8 | 9 |



Incomplete Block Designs (IBD)

SUNFLOWER EXAMPLE. OIL YIELD COMPARISON OF 20 ADVANCED INBRED LINES.

Treatment: 20 sunflower IL

Experimental design: IBD

Resoluble with $r=3$ and $s=5$

Dependent variable : $Y = L \text{ ha}^{-1}$

Incomplete block

| | 1 | 2 | 3 | 4 | 5 |
|----|----|----|----|----|----|
| | 1 | 2 | 3 | 4 | 5 |
| R1 | 6 | 7 | 8 | 9 | 10 |
| | 11 | 12 | 13 | 14 | 15 |
| | 16 | 17 | 18 | 19 | 20 |
| | 1 | 2 | 3 | 4 | 5 |
| R2 | 7 | 8 | 9 | 10 | 6 |
| | 13 | 14 | 15 | 11 | 12 |
| | 19 | 20 | 16 | 17 | 18 |
| | 1 | 2 | 3 | 4 | 5 |
| R3 | 8 | 9 | 10 | 6 | 7 |
| | 15 | 11 | 12 | 13 | 14 |
| | 17 | 18 | 19 | 20 | 16 |

TREATMENT ASSIGNMENT:

Each treatment is assigned randomly to the experimental units in the first rep. In the following reps, restrictions in the randomization are conducted such that each pair of treatment is compared the same number of times within an incomplete block.



Incomplete Block Designs (IBD)

$$Y_{ijk} = \mu + \alpha_i + \beta_j + \gamma_{k(j)} + \varepsilon_{ijk}$$

Y_{ijk} = response of the i -th treatment on the j -th rep and the k -th incomplete block

μ = population mean

α_i = effect of the i -th treatment

β_j = effect of the j -th rep

$\gamma_{k(j)}$ = effect of the k -th incomplete block within the j -th rep

ε_{ijk} = experimental error (residual)



Field Designs and Heritability

FACTORIAL DESIGNS (HS, FS, RIL, DH, Clones)

$$y_{ijk} = \mu + E_j + \beta_{k(j)} + P_i + PE_{ijk} + \varepsilon_{ijk}$$

| SoV | df | EMS |
|--------------|-------------|---|
| Environment | (e-1) | |
| Block(Env) | e(r-1) | |
| Progeny | (n-1) | $MS_{\text{Progeny}} = V_{\varepsilon} + rV_{\text{PE}} + reV_{\text{Progeny}}$ |
| ProgenyXEnv | (n-1)(e-1) | $MS_{\text{PE}} = V_{\varepsilon} + rV_{\text{PE}}$ |
| Pooled error | (n-1)(r-1)e | $MS_{\text{Error}} = V_{\varepsilon}$ |

$$h_F^2 = \frac{\hat{\sigma}_F^2}{\hat{\sigma}_p^2} = \frac{\hat{\sigma}_F^2}{\hat{\sigma}_F^2 + \frac{\hat{\sigma}_{FE}^2}{e} + \frac{\hat{\sigma}_e^2}{er} + \frac{\hat{\sigma}_w^2}{ern}}$$

$$\text{HS} : V_{\text{Progeny}} = \frac{1+F}{4} V_A$$

$$\text{FS} : V_{\text{Progeny}} = \frac{1+F}{2} V_A + \frac{(1+F)^2}{4} V_D$$

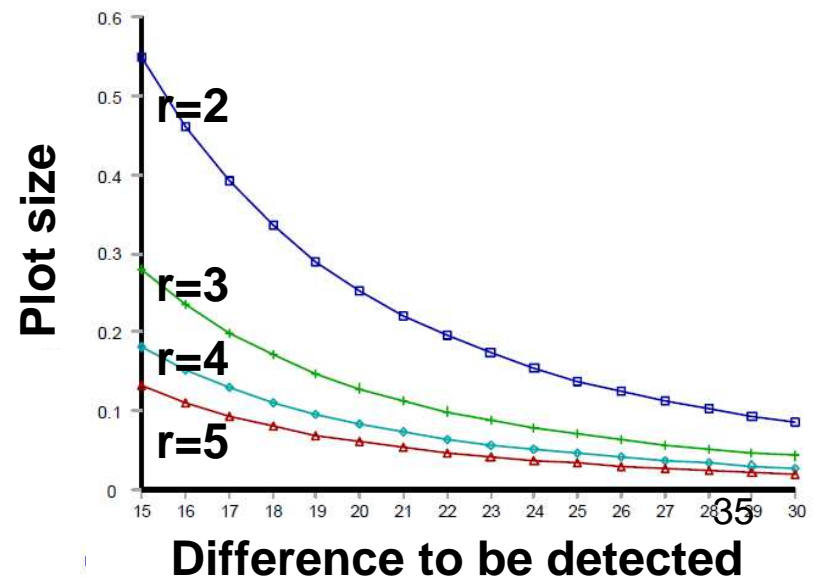
$$\text{RIL} : V_{\text{Progeny}} = 2V_A$$

Other Considerations (Design)

1. PLOT SIZE:

- Big enough to have plants growing normally (i.e. one-plant does not work in crops but works fine for trees).
- Big enough to represent trait variation (i.e. you might need larger plots for yield than for maturity).
- Not too big as to have considerable within-plot variation (i.e. increasing plot size increases within-plot variation).
- Balance between cost of increasing plot size and increasing number of plots.

EXAMPLE. Mohammad et al. (2001) compared plot size and number of replications to detect differences in wheat. Bigger plots require less reps.



Other Considerations (Design)

2. PLOT SHAPE:

- Balance between plot border and plants within plot (i.e. rectangular plots have more border than squared ones).
- Take into account gradients (i.e. if unidirectional gradient like fertility long and narrow plots might be better).

EXAMPLE. Haapanen
(1992) compared plot size
and shape for height in pines

| Plot type | n | Var _(plot) | Var _(within-plot) | Var _(p) | Rel. eff., % | Rel. size, % |
|-----------|----|-----------------------|------------------------------|--------------------|--------------|--------------|
| 1×1 | 1 | 29.32 | 0.00 | 29.32 | 100 | 100 |
| 3×1 | 3 | 1.39 | 28.02 | 11.90 | 81 | 122 |
| 7×1 | 7 | 1.71 | 27.86 | 6.25 | 67 | 150 |
| 8nc | 8 | 0.83 | 28.54 | 4.92 | 75 | 135 |
| 14×1 | 14 | 1.43 | 28.14 | 3.72 | 56 | 180 |
| 7×2 | 14 | 2.18 | 27.77 | 4.44 | 47 | 214 |
| 16nc | 16 | 0.35 | 29.03 | 2.42 | 76 | 134 |
| 7×3 | 21 | 1.90 | 27.99 | 3.42 | 41 | 248 |
| 14×2 | 28 | 1.30 | 28.41 | 2.46 | 43 | 240 |
| 6×5 | 30 | 1.36 | 27.90 | 2.40 | 40 | 251 |
| 14×3 | 42 | 1.07 | 29.65 | 1.88 | 37 | 277 |
| 7×6 | 42 | 0.86 | 28.76 | 1.64 | 43 | 243 |
| 7×7 | 49 | 1.26 | 28.12 | 1.91 | 31 | 331 |
| 49nc | 49 | 0.45 | 28.94 | 1.12 | 53 | 195 |

Other Considerations (Design)

3. ROW vs. HILL PLOTS:

Table 2. Coefficients of variation and ranges (as a percentage of experiment mean) for agronomic traits measured in multiple environment hill and row plot evaluations of spring malting barley

| Statistic | Plant height | Harvest index | Grain yield | %plump seeds | No. of seeds/spike | No. of spikes/area | 100-seed weight |
|----------------------|--------------|---------------|-------------|--------------|--------------------|--------------------|-----------------|
| Row plots | | | | | | | |
| CV | 7.6 | 14.7 | 10.8 | 8.1 | 8.7 | 17.0 | 6.6 |
| Range (% of mean) | 79-121 | 83-113 | 68-113 | 63-111 | 62-203 | 53-113 | 81-115 |
| Hill plots | | | | | | | |
| CV | 20.0 | 14.3 | 21.0 | 9.7 | 11.7 | 24.5 | 10.0 |
| Range (% of mean) | 81-136 | 78-118 | 64-114 | 79-114 | 57-189 | 65-131 | 81-116 |

EXAMPLE. Tragoonrung et al., 1990 compared row and hill-plots for different traits in barley. Hill-plots are ok for most traits but not for yield.

Other Considerations (Design)

- 4. BORDER ROW.** It is a good idea to include a row of plants as a border row to avoid having plots in the margins of the experiment performing differently due to lack of competition.
- 5. OPERATORS NOISE.** If measurements on the experiment will be performed by different people it is a good idea to consider operators noise by having different people measuring for example in different blocks.
- 6. EARLY TESTING.** Small amount of seeds are available in early generations. Therefore replicating is a challenge. Taking this into consideration when deciding which traits will be measured early is important.
- 7. OTHER DESIGNS AND SPATIAL CORRELATIONS.** More efficient designs exist for testing large numbers of genotypes in fields that might not be completely heterogeneous. Additionally, spatial corrections might improve estimations of genotypic means.

Other Considerations (Estimation)

1. **MIXED MODELS.** Careful interpretation of Expected Means Squares should be exercised especially when using Mixed Models. EMS is different if factors (i.e. environments, genotypes, etc.) are defined as random or fixed effects:

a. E, G, GxE random: $EMS_{(G)} = \sigma_e^2 + n\sigma_I^2 + nN\sigma_G^2$

within env:
$$h^2 = \frac{\sigma_G^2}{2F(\sigma_e^2 + \sigma_G^2 + \sigma_I^2 + \sigma_E^2)}$$

a. G, GxE random, E fixed: $EMS_{(G)} = \sigma_e^2 + nN\sigma_G^2$

across env.:
$$h^2 = \frac{\sigma_G^2 + \sigma_I^2}{2F(\sigma_e^2 + \sigma_G^2 + \sigma_I^2)}$$

Other Considerations (Estimation)

2. **NON-BALANCED DESIGNS.** When designs are not balanced due to their design, missing plots or missing data, MS calculations are not straight-forward. There are four ways to estimate MS.

MIXED MODELS - INCOMPLETE BLOCKS – UNBALANCED

With unbalanced data, mixed models, or correlation among genotypes, variance component estimation of heritability is not related to the gain from selection. Using the concept of “effective error variance” (Cochran and Cox, 1957), Cullis et al (2006) proposed to use a generalized method for heritability:

$$h^2 = 1 - \frac{\bar{v}_{BLUP}}{2\sigma_G^2}$$

\bar{v}_{BLUP} : mean variance of a difference of two BLUP