INTRODUCTION TO
DESIGN OF EXPERIMENT [DoE]

DoE: Outline

- Preliminary
- Introduction
- Common designs in plant breeding
- Estimate of error
- Checks
- Treatment
- Sites
- Blocking
- RCBD
- Lattice & Alpha design
- Augmented design
- P-rep design
- Sparse testing design

DoE: Preliminary

- Statistics can be described as the science
  - collecting,
  - analyzing, and
  - drawing conclusion from data.
- Data is usually collected via:
  - Sampling Survey
  - Observational studies, or
  - Experiments.

DoE: Introduction

- The term **experimental design** refers to **a set of experiments** or runs that are **prepared in advance of their execution**.
- There are many types of experimental designs – CRD, RCBD, IBD, etc.
- In experimental design it is important to identify the **sources of variation**.
- The specific runs chosen in an experimental design will be determined by the design's objective.

DoE: Introduction

- Design objective can be classified into two main categorize:
  - Study the source of variability
  - Establish cause and effect relationships
- Purpose of experimental design
  - Determine cause of variation in measured responses
  - Identifying the variables that cause a maximum or minimal response
  - Compare response between controlled variables
  - Minimize the effect of experimental error
  - Model future response values

Introduction: Definition of Some DoE Terms

- **Experiment (Run):** A procedure in which the researcher modifies ***at least*** one of the variables under investigation and then examines the results.
- **Experimental Unit (EU):** Item that is being studied and on which something is modified
- **Sub-Sample/Sub-Unit/Observational Unit:** It is formed when the experimental unit is split after the action has been performed on it. NB: *Sub-samples or sub-units of the same experimental unit are usually correlated and should be averaged before analysis.*

Introduction: Definition of Some DoE Terms

- **Treatment Factor (Independent Variable):** Variable under investigation/control that is kept at near or perfect value, or level during experiment.
- **Background/Lurking Variable:** Variable the experimenter is unaware of or unable to control that could affect the outcome of the experiment.
- **Response (Dependent Variable):** The feature of the experimental unit that is measured after each experiment or runs. NB: *The size of the responses is influenced by the settings of the independent variables or factors, as well as any hidden variables.*

Introduction: Definition of Some DoE Terms cont'd

- **Replicate:** Two or more experiments using different experimental units, but with the same factor or independent variable settings. NB: *Due to changes in lurking variables and inherent variations in experimental units, the measured dependent variable may differ among replicate runs.*
- **Randomization:** Is the process of randomly assigning treatments to experimental units. As a result of the random process, every treatment factors has the same probability.
- **Blocking:** Is the procedure for gathering similar experimental units into a relatively homogeneous group.

Introduction: Definition of Some DoE Terms cont'd

- **Biased Factor:** An experimenter adjusts an independent variable at the same time as changes in background or lurking variables occurs. NB: *It becomes difficult to determine the cause of the changes in the response variable.*
- **Experimental Error:** The variation between the observed response for a specific experiment and the long run average of all tests conducted with the same independent variables or factors. NB: *Because of background or lurking variables, experimental errors are not equal to zero.*

Introduction: Definition of Some DoE Terms cont'd

- **Effect:** The change in the response variable that is caused by change in the treatment factor or independent variable. NB: *It's termed **calculated effect** when it's computed after an experiment from observed data, while it's called **effect size** when it's determined before an experiment.*

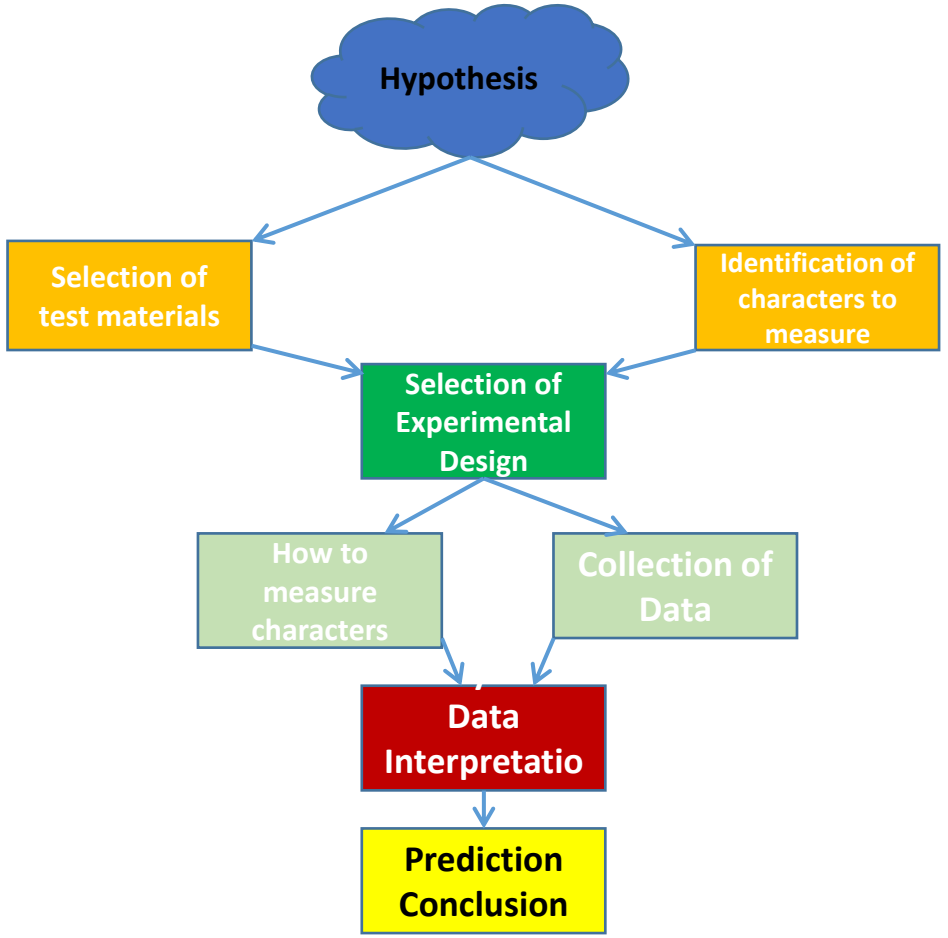
  - **Fixed Effect:** We have a fixed effect if the treatments are well defined and easily replicable and are expected to yield the same impact on average in each replicate. The model is called a ***fixed effect model***.
  - **Random Effect:** We have a random effect if the treatments cannot be considered to come from a predefined or known set, they are supposed to be a random sampling from a larger population of possible treatments. The model is called a ***random effects model***.

DoE: Some Considerations

- The choice of experimental design to study the source of variability **depends on the number of sources under study.**
- The appropriate experimental design to study cause and effect relationship depends on:
  - Type & number of treatment factors
  - Degree of homogeneity of experimental units
  - Ease of randomization
  - Ability to block experimental units into homogenous groups
- The final experimental design chosen will determine
  - How data will be collected
  - Model to be fit to analyze the data
  - Data interpretation
  - Conclusions drawn from experiment

Introduction: Steps in planning DoE

- Define Objectives - Hypothesis
- Identify Experimental Units
- Define Measurable Response – Dependent Variable
- List Factor/Independent & Lurking Variables
- Run Pilot Tests
- Make a Flow Diagram of Experimental Procedure
- Choose Experimental Design
- Determine the number of Replicates Required
- Randomize the Experimental Condition to Experimental Units
- Describe a Method for Data Analysis
- Timetable and Budget for Resources Needed to Complete the Experiments



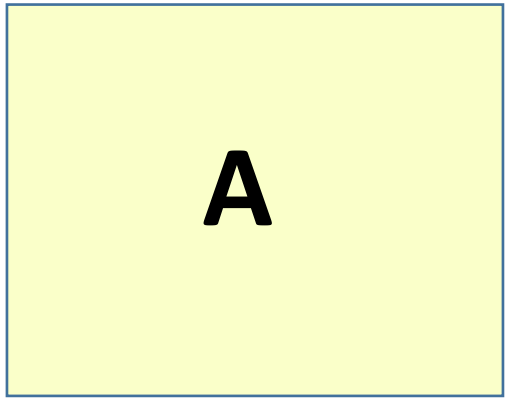
DoE: Common designs in Plant Breeding

Mostly used designs in Plant Breeding:

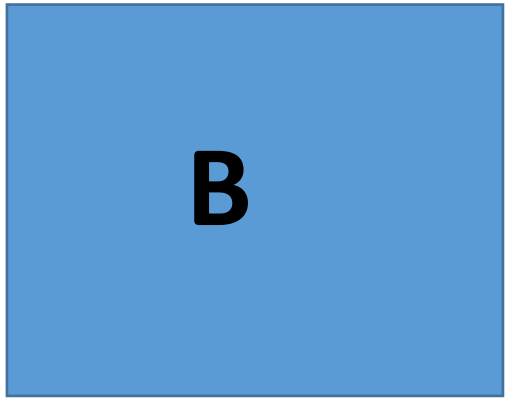
- **incomplete block designs** (including alpha-designs), **row-column designs**
    - can be difficult to use, especially in EGVTs (constraints on space, population size and resources availability)
- **augmented designs** with unreplicated lines and replicated checks to enable the estimation of block effects and error variances
- **p-rep designs** - check plots from a grid-plot design replaced by plots of replicated lines leading to **p**artially **rep**licated designs -- improved the accuracy of selection

DoE: Estimate of error

- Consider the following simple experiment:



Maize variety A
Yield: 6 t ha $^{-1}$

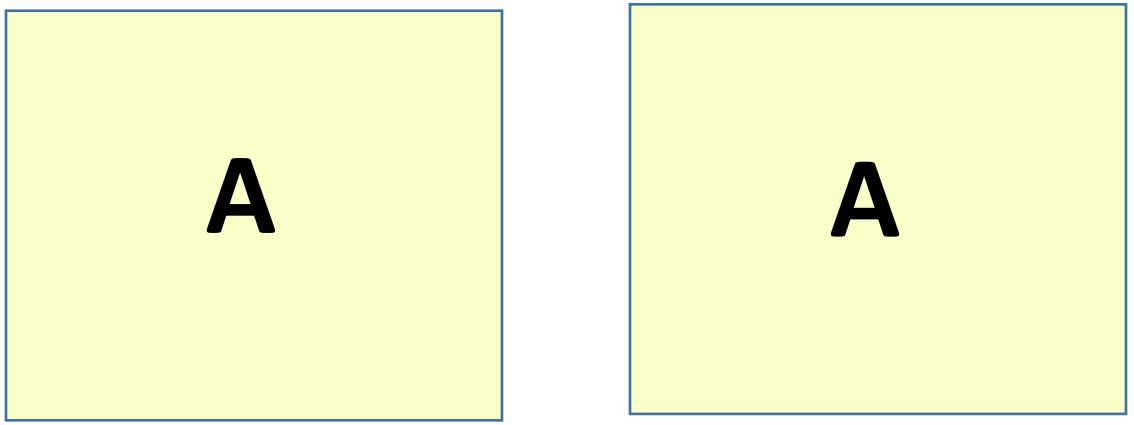


Maize variety B
Yield: 2.5 t ha $^{-1}$

- Conclusion? Variety A better than variety B?
- If Yes, what are the assumptions?
- Are these assumptions valid?

DoE: Estimate of error

- The assumptions are “Any difference between the yields of the two plots is caused by the genotypes and nothing else”
- This is not certainly true!
- Would we have same yield if we had the same variety on both plots?

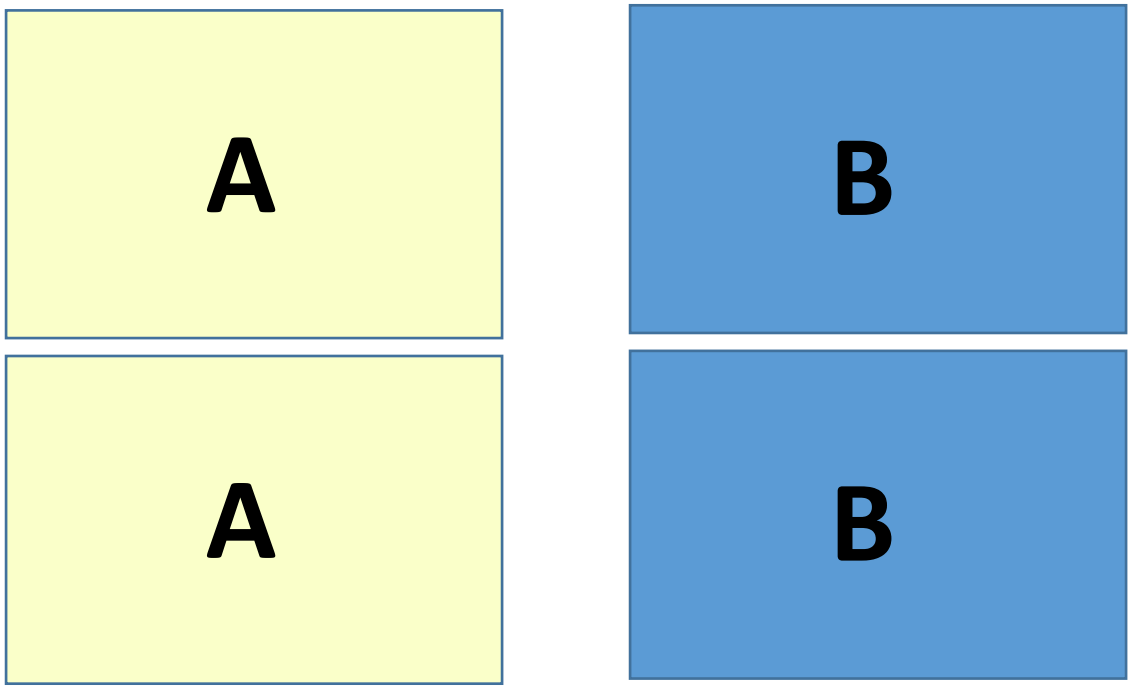


Maize variety A Maize variety A

DoE: Estimate of error

- Other factors, such as soil fertility, and insects, birds, diseases damage can also affect yields
- One important question: how can we know that the difference on the yield is due to:
  - ***variety*** or
  - ***other sources of variation*** we are not testing?

DoE: Estimate of error



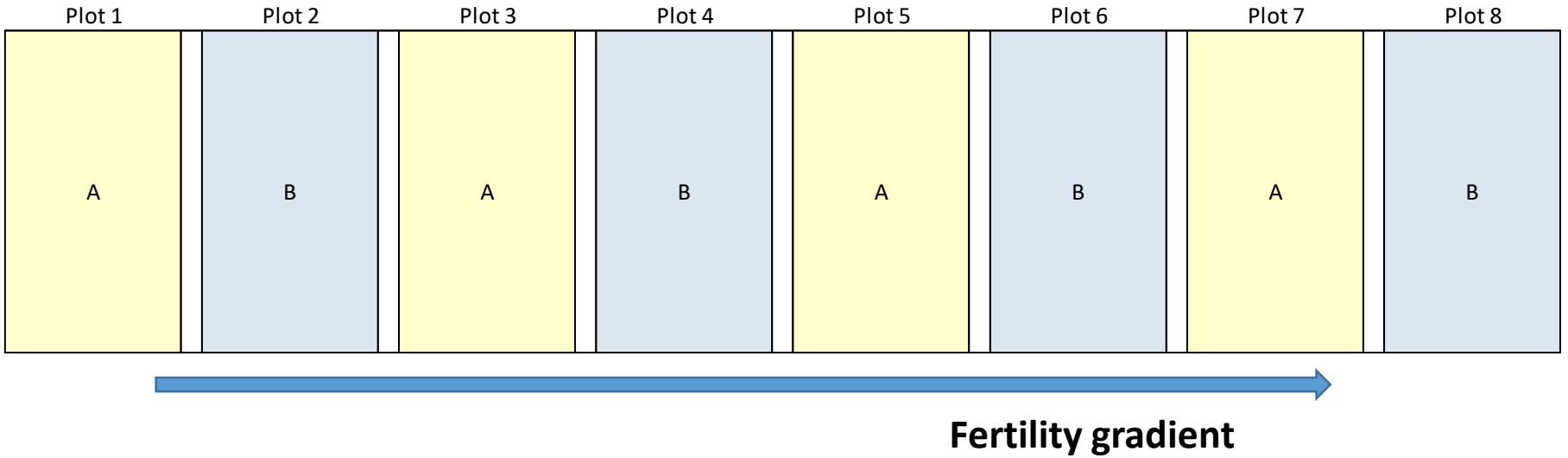
- Intuitively, ***Variety A is different to Variety B*** if difference between A and B is ***larger*** than difference with ***same varieties***

DoE: Estimate of error

- Then we need to know
  - not only the yield difference between plots planted to different varieties
  - but also the yield difference between plots planted to the same variety
- The difference among experimental plots planted to the same variety is called **experimental error**
- Then, to obtain a measure of experimental error, what is needed?
  - **replication**

DoE: Estimate of error

- Knowing that replication is needed, suppose we plant varieties A in plots 1, 3, 5 and 7 and variety B in plots 2, 4, 6 and 8 and we have a fertility gradient (*Gomez and Gomez, 1984*). What would happen?



- One variety would be in more favorable conditions
- Varieties must be **randomly** arranged into **blocks**

DoE: Estimate of error

We have some techniques to control the experimental error:

- **Blocking**: arrangement of experimental units into groups (blocks); in this case, variation among blocks can be measured and removed from the experimental error
- **Proper plot technique**: all other factors such as soil nutrients, pest incidence, are maintained uniformly
- **Data analysis**: sometimes, only blocking may not help, proper choice of data analysis can be used (covariance analysis can reduce the variability among experimental units by adjusting their values)

DoE: Key concept

Let's review the definitions of a few key concepts in experimental design that are connected to breeding.

- The **population of inference** is the set of ALL entities (genotypes) to which the Breeder intends to have the results of the experiment be applicable
- The **experimental unit** is the smallest entity to which the treatment is applied
- A **factor** is a procedure or condition whose effect is to be measured.

DoE: Some illustrations

- We consider three illustrative examples:
    1. A trial that evaluates 80 genotypes
    2. A trial that evaluates 40 genotypes under 2 different water regimes
    3. A trial that evaluates 20 genotypes at 2 nitrogen levels for 2 planting dates
- For each of the examples, what is treatment, factor, level?

DoE: Some illustrations

- These three trials all have 80 **treatments**
- In the first example (*a trial that evaluates 80 genotypes*), there is only one factor, genotype, with 80 levels
- In this simple case, the different genotypes are
  - the treatments
  - or the **levels** of a treatment factor

DoE: Some illustrations

- In the second experiment (*a trial that evaluates 40 genotypes under 2 different water regimes*), there are 2 factors:
  - genotype with 40 **levels**
  - water with 2 levels
- Each treatment consists of a combination of a particular genotype and a particular water level
- There are 80 different combinations or treatments. This is sometimes known as a 40 by 2 **factorial treatment** structure

DoE: Some illustrations

- Similarly, the third trial (*a trial that evaluates 20 genotypes at 2 nitrogen levels for 2 planting dates*) has 3 factors and the 80 treatments are arranged in a 20 by 2 by 2 factorial treatment structure
- Often in Plant Breeding, the goal is to evaluate different genotypes

DoE: How many genotype per trial?

- The statement of the objectives (in the “Protocol”) should include an indication of the number of treatments — in Plant Breeding, the genotypes — to be included in the trial
- There is no prescribed limit for this
- As in the early stages, some trials may have many hundreds of genotypes, each in small, sometimes single-line, plots

DoE: How many genotype per trial?

- There is sometimes a choice within a given site between putting all genotypes in a single large trial or having many smaller ones ...
- Here the guideline is to put them in the same trial only if they all ***NEED to be compared with each other***
- Otherwise, they can be distributed among a set of smaller experiments

DoE: How many genotype per trial?

- For example, if genotypes are known to be in different maturity groups and recommendations are required for each of a range of season lengths
- There is no requirement to compare short with long season genotypes
- They can therefore be included in separate trials

DoE: How many genotype per trial?

- In contrast, disease resistance could be present in genotypes of any season length, and it will then be more appropriate to evaluate all the genotypes together
- Even in this latter case, if it is expected recommendations at each of different season lengths, then a set of smaller trials should be used

DoE: Checks

- In addition to the genotypes being evaluated there will often be one or more standard lines that are considered as “**controls**” or “**checks**”
- Their presence and the way they are incorporated in the trial are linked to the objectives

DoE: Checks

- For example, a trial on resistance to a given disease might include three controls: one **resistant**, one **tolerant** and one **susceptible**
- If, however, there is interest in highlighting genotypes that are highly resistant, then the only control might be a well-known resistant variety
- The controls are sometimes replicated more times than the other genotypes

DoE: Checks (gg assessment)

Checks	**#**	**Features**
Fixed gg checks	2	dynamically stable: should be present in all the trials regardless of location and time
Commercial varieties	4	commercial varieties or ‘competing’ products currently being grown in the wide target market and to be beaten according to the replacement strategy in the product profile, and that show a GxE pattern similar to the cohorts being tested; to be renewed at the rate of one per year
Local checks	1	locally grown due to their specific preference (e.g., resistance, performance, etc.); to be renewed as desired

DoE: Treatment (practical consideration)

- A common question is: is it possible/correct to analyze a trial with many genotypes? --- Yes, but a problem!
- This is no statistical problem
- There is sometimes a risk that the large volumes of data collected may be a problem for field staff, resulting in lower quality data than would be the case with smaller trials

DoE: Treatment (practical consideration)

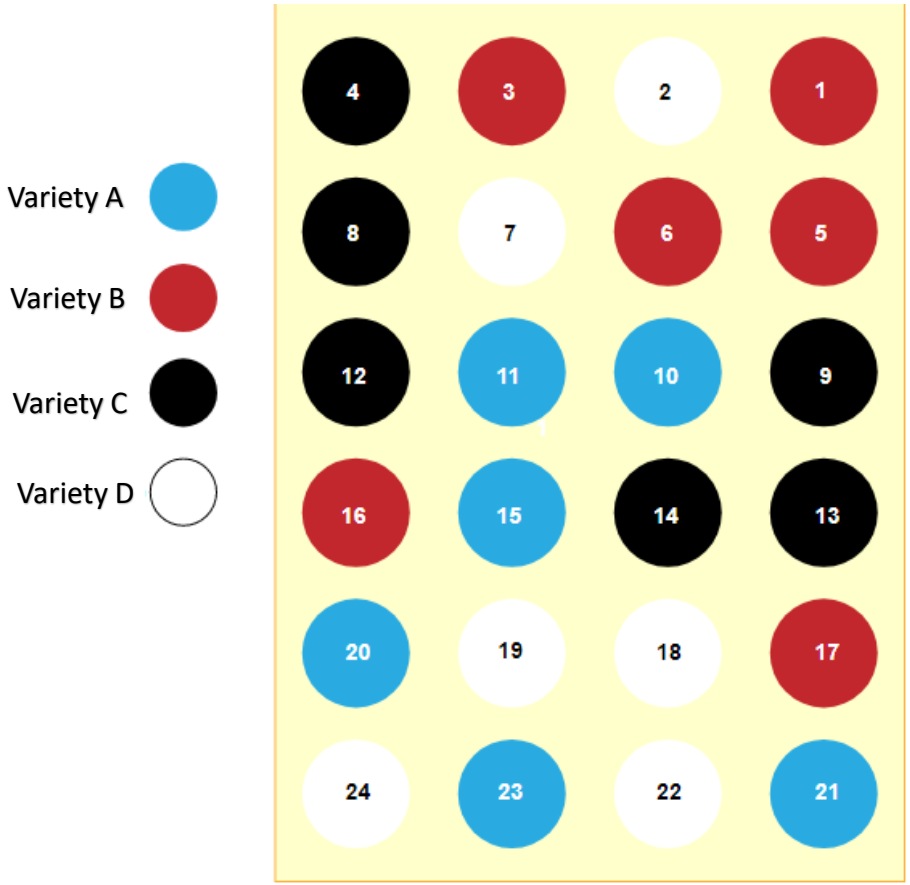
- Large experiments need large areas of land
- They also often have a more complicated blocking structure
- If this large area is quite heterogeneous, but a part of it is more homogeneous, then better results may be obtained from a trial that uses only the homogeneous land

DoE: Sites

- When objective of the trial is to assess potential of different genotypes, the trial will probably be carried out in an “**ideal**” environment (managed so that there is no water stress or competition from weeds)
- When objective is related to response to stress, the choice of sites is crucial, and trials may be repeated over a range of sites and years

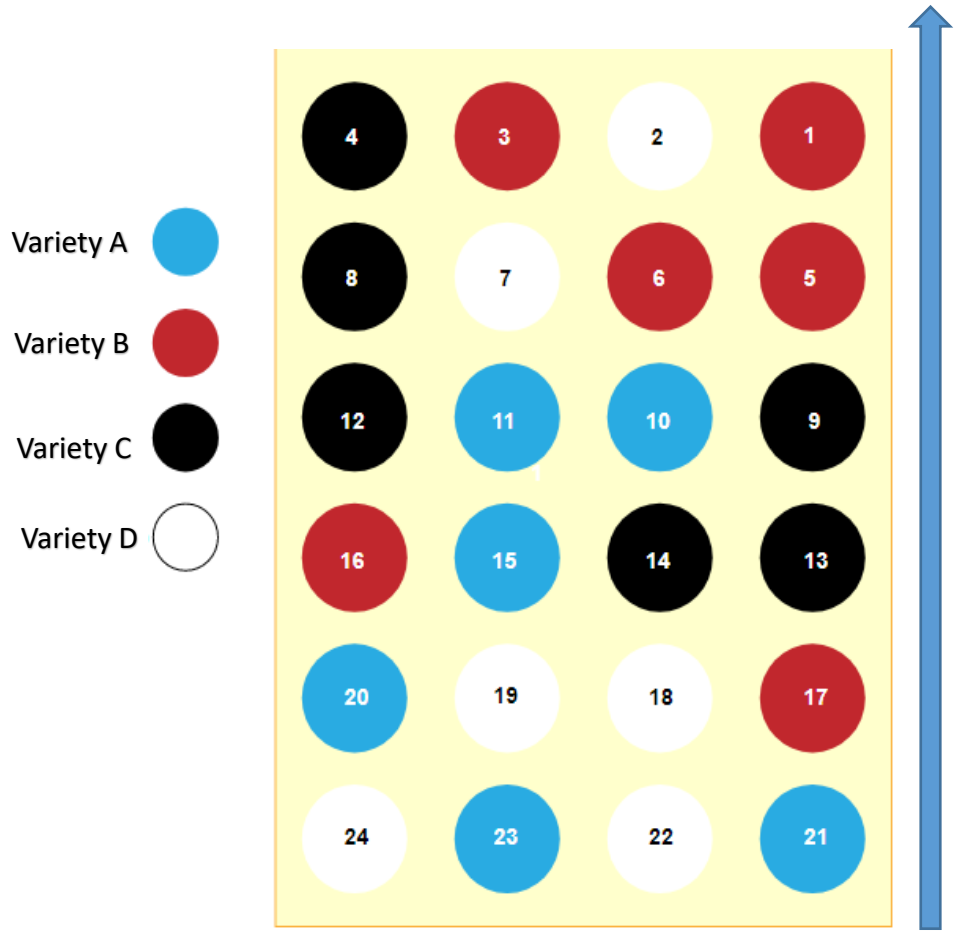
DoE: Assuming Homogeneity

- CRD is perhaps the simplest experimental design in terms of convenience and data analysis.
- Experimental Units (EU) are **homogeneous** – *No basis for grouping.*
- All EUs are randomly divided into treatment groups.
- **Randomization** guarantee the validity of the experiment.
- Each EU is subjected to the unique levels of treatment – Equal probability.
- Without replication it is impossible to treatment effect.



DoE: Assuming Homogeneity

- Sub-samples or duplicate measurements **CANNOT** substitute for replicates – *leads to biased estimate.*
- Replication can be equal or unequal. **For efficiency use equal replicates.**
- # of EUs ***(n)*** is the product of # of treatment groups ***(t)*** & number of replicates ***(r)*** i.e., ***n = tr***
- Very useful with **homogenous Eus.**
- However, the underlying assumption of this design is often **NOT** true due to **inherent heterogeneity** of EUs (field) called **Fertility Gradient.**
- Hence, **Blocking.**



DoE: General Concept

- Let’s take a simple example with just six genotypes: A, B, C, D, E and F.
- There are two main decisions to take:
  - the number of plots to be sown with each genotype, i.e., the number of “replicates”
  - how these replicates of each genotype will be in the field

DoE: General Concept

- Replications, and how they are distributed within experimental layout, are important because they can be used to control the experimental error, due both to
  - variation inherent in the material being tested, and
  - variation in the site where the material is being tested

DoE: Blocking

- The purpose of blocking is to group plots within a part of the field that is as homogeneous as possible
- This enables evaluation of genotypes with greater precision than if the position of the plots were not restricted in this way

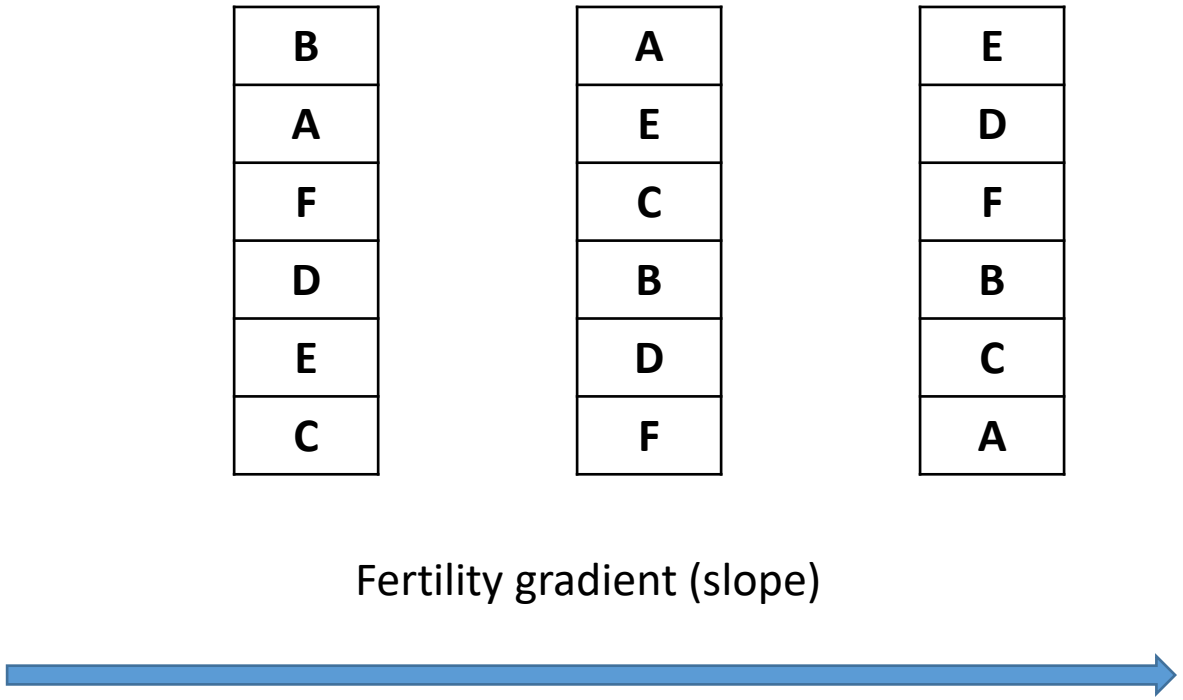
DoE: Blocking

- Example: with 6 varieties A, B, C, D, E, and F we assume that a small experimental field can accommodate 18 plots and that the soil has an inherent fertility gradient that changes smoothly from left to right.
- How to block?

Fertility gradient

DoE: Blocking

- Blocks perpendicular to the gradient

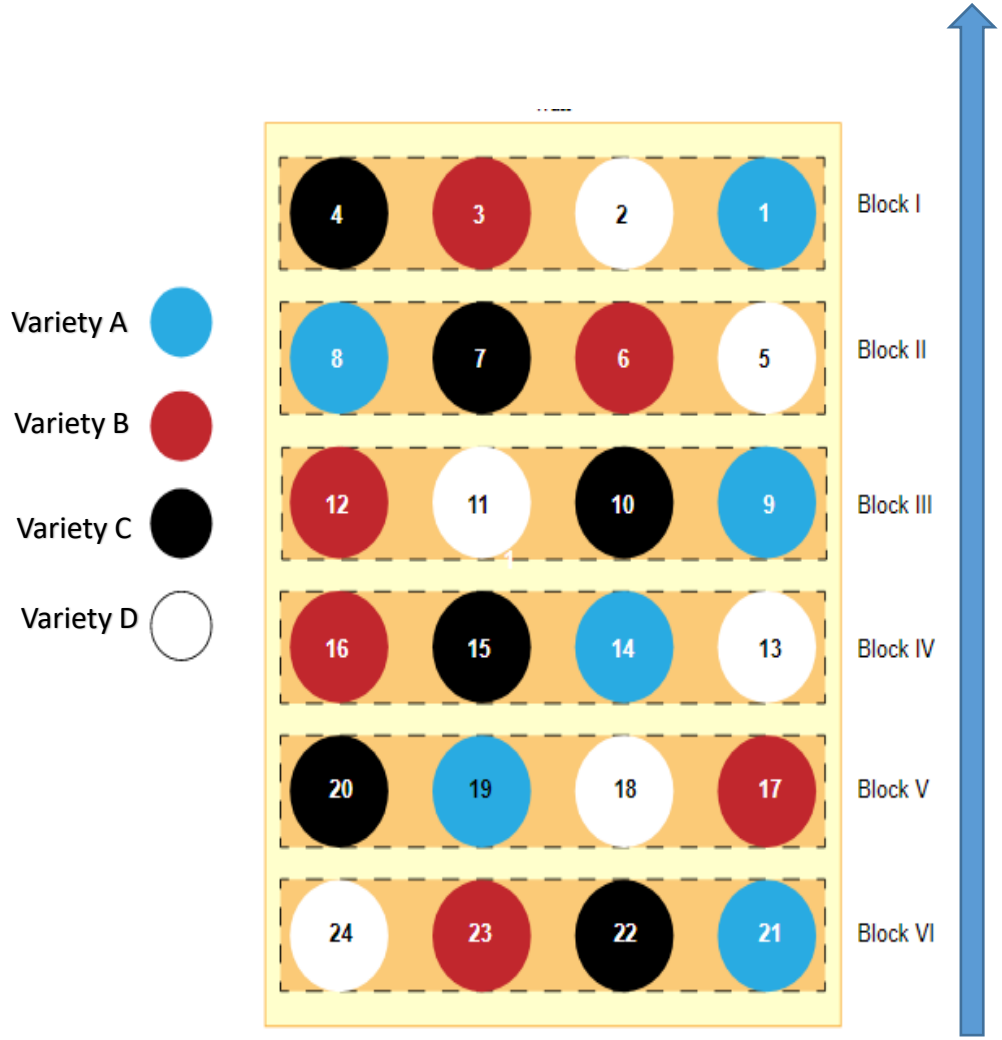


DoE: Blocking

- A total of 18 plots is used, and each genotype has been replicated three times
- The layout is in 3 blocks, with each block containing 6 plots, each block contains one replicate
- This is a very common design called the “**randomized complete block design**”, or **RCBD** for short

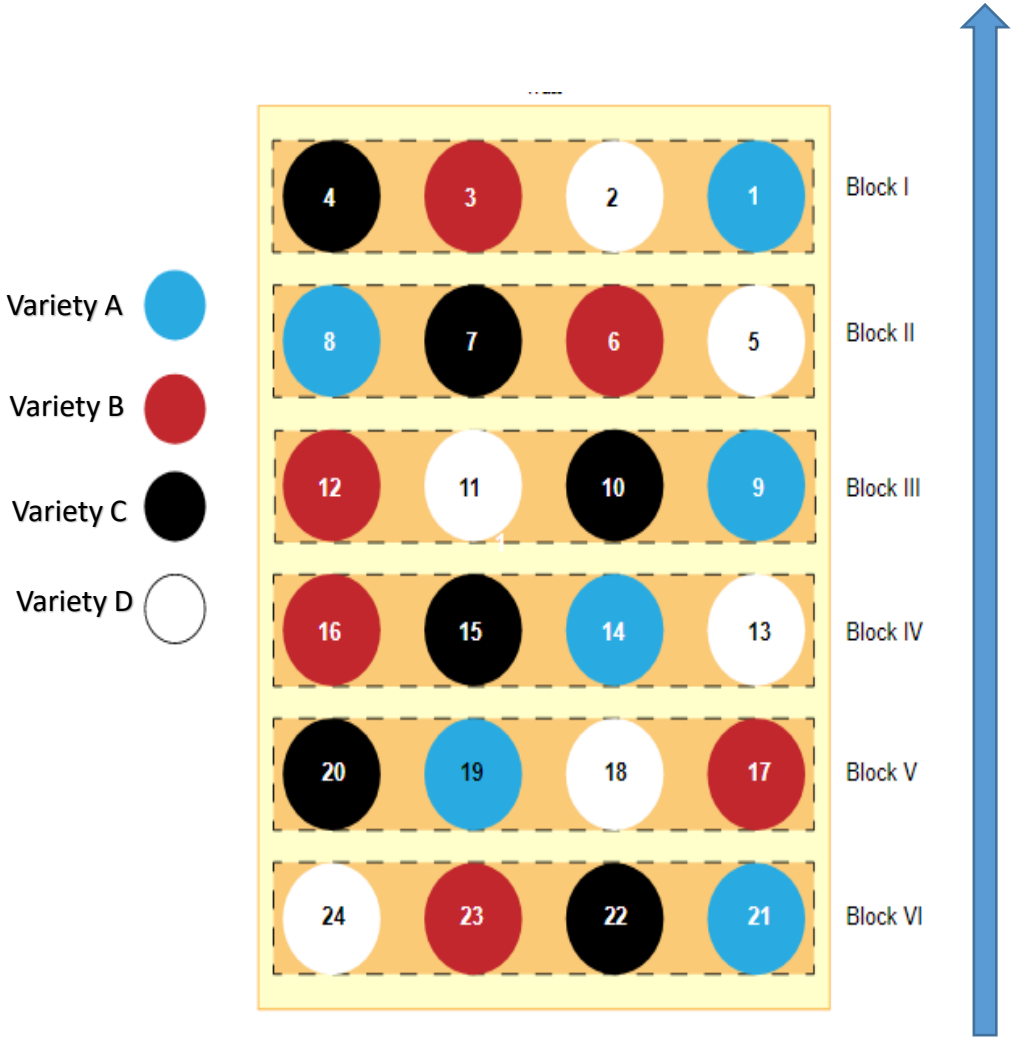
DoE: RCBD Summary

- Experimental Units (EU) are **heterogeneous** – *There is basis for grouping*
- Sub-set of EUs with *homogeneous effects* on response variable makes a **complete block**
- Blocking heterogenous EUs reduces experimental error – *error control*
- Control heterogeneity in one direction
- More EUs per block increases the chance of heterogeneity



DoE: RCBD Summary

- Uniformity trial is used in the formulation of blocks
- Blocking is a feature of the EU's and **NOT** the treatment
- Treatment groups are randomly assigned to EUs in complete blocks
- # of EUs **MUST** be greater than or equal to the # of levels of the treatment factor.
- Sometimes, the assumption behind the usage of this design is not met due to different reasons.
- Hence, **incomplete block**

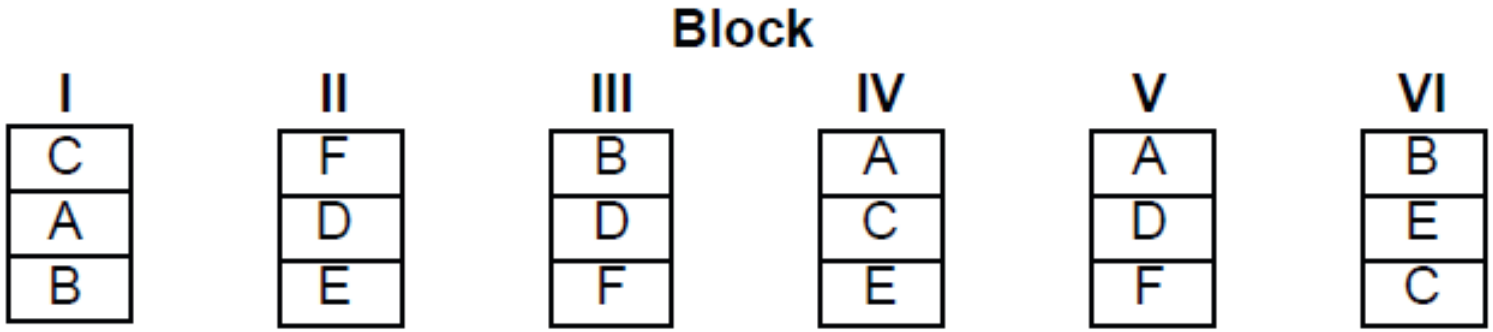


DoE: Blocking

- Sometimes, it is **necessary** to form small blocks with fewer plots per block than the number of genotypes
- This may be due to heterogeneous field conditions, or because there are many genotypes to evaluate

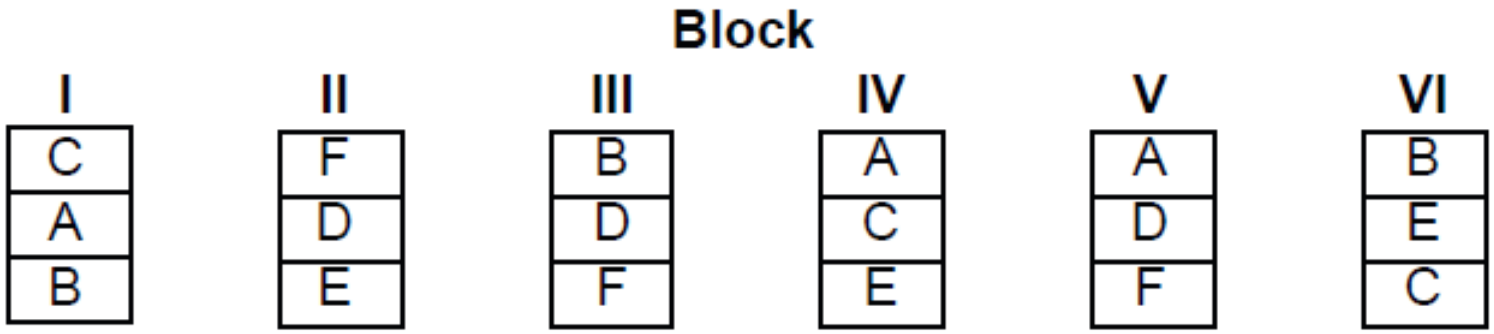
DoE: Blocking

- If the soil fertility in the field were very patchy, then a possible approach to try to preserve soil homogeneity of plots within blocks would be to have the size of blocks from 6 to 3 plots



DoE: Blocking

- The blocks are now “**incomplete**” as each contains only 3 of the 6 genotypes
- ***Blocks and replicates are no longer equivalent***, since there are still 3 replicates per genotype, but 6 blocks.
- This is called “**incomplete block design**”



DoE: Blocks & Replications

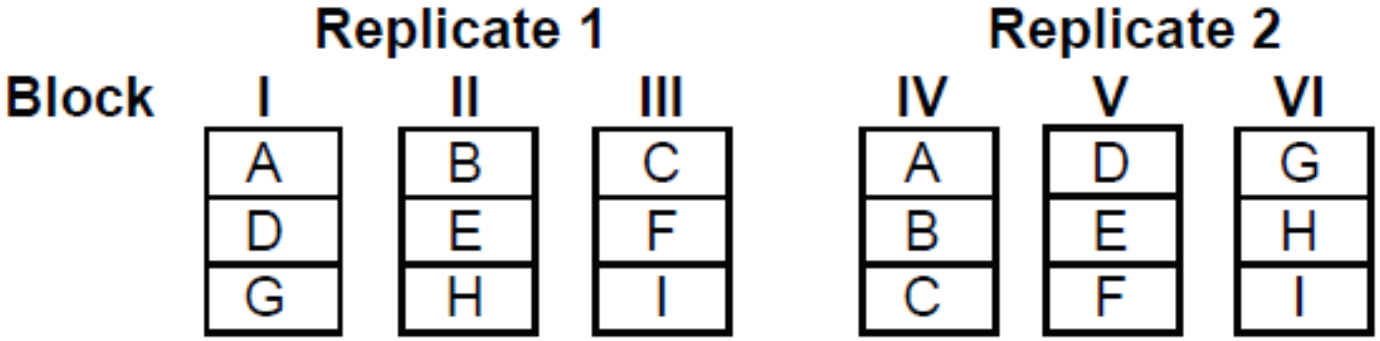
	Replications	**Blocks**
Definition	More than one unit of the same treatment	Group of homogeneous units
Reason	Estimate precision Increase precision but increase size of trial	Increase precision without increasing size of trial
Number	May not be the same for all treatments	Function of the number of treatments

DoE: Lattice & Alpha Designs

- **Lattices** are special cases of **incomplete block designs**
- Let's just consider **square lattices**, where the number of genotypes is a perfect square, for example 9, 16, 25, 144 or 900
- In a square lattice, the block size is fixed as the square root of the number of genotypes.
- Example: with 225 genotypes the blocks would be of 15 plots each

DoE: Lattice & Alpha Designs

- Suppose that 9 genotypes are to be evaluated, each to be replicated twice
- This permits us to use 2 replicates in a 3x3-lattice arrangement, prior to randomization

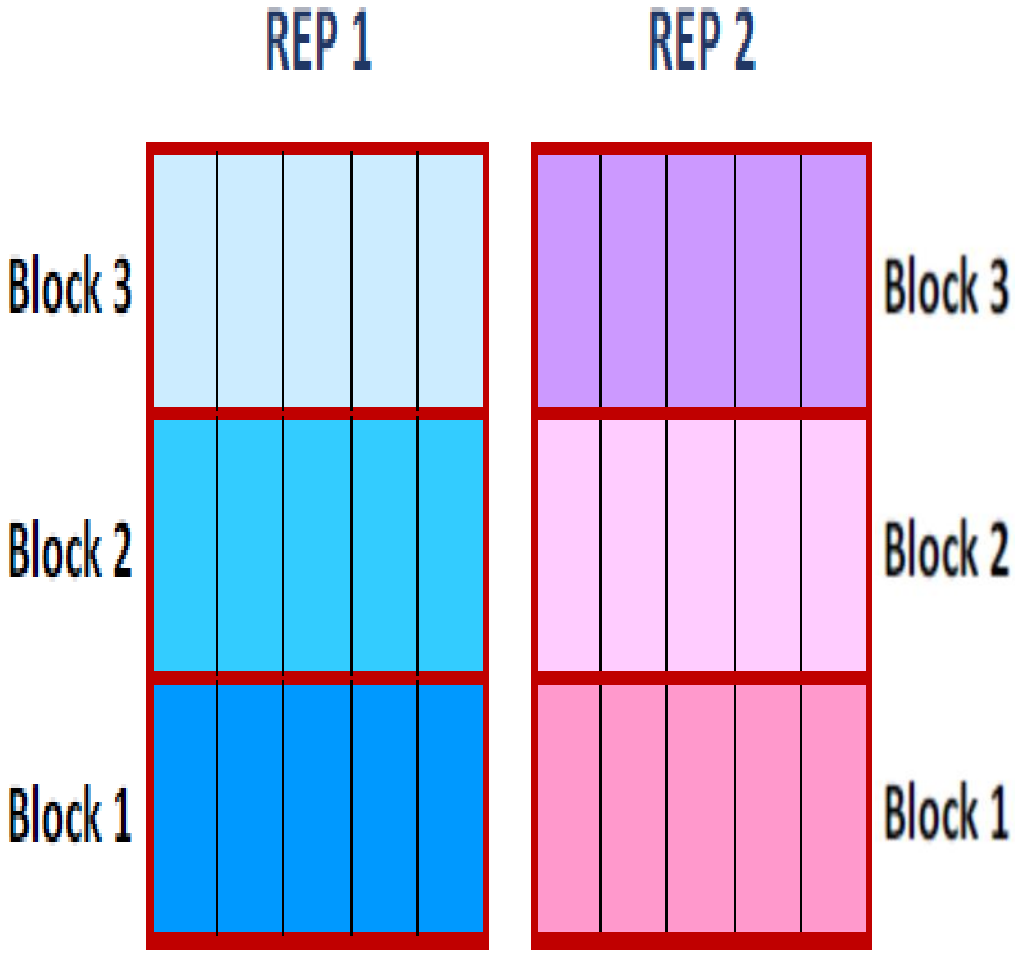


DoE: Lattice & Alpha Designs

- Very large numbers of genotypes can be evaluated by using lattices
- However, lattice lack the flexibility required in many practical situations
- Lattice enforces a rigid block size that may not be appropriate to local field conditions
- The development of a more general class of designs called “**alpha designs**” has removed these restrictions

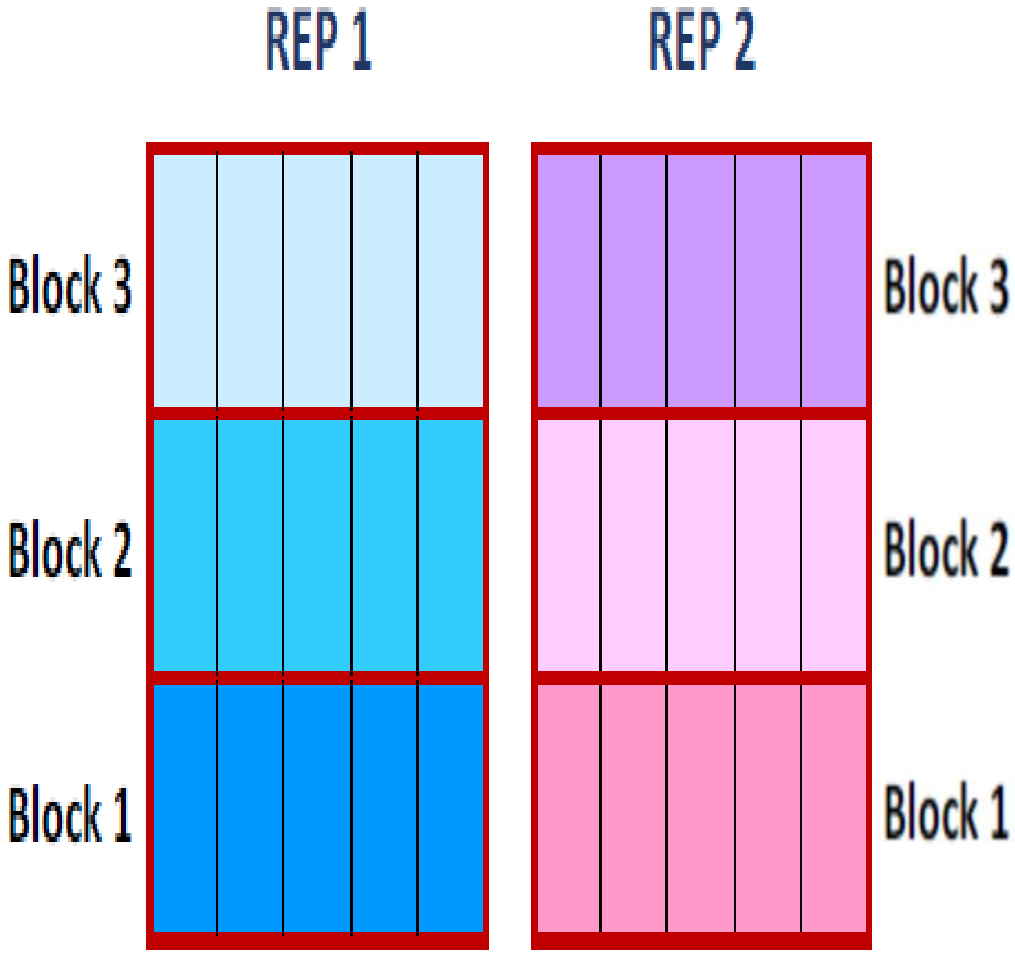
DoE: Lattice & Alpha Designs Summary

- Lattice designs are in the class of ***resolvable*** incomplete block design i.e., the incomplete blocks are in complete replicates.
- Distinguishing features of LD
  - *# of Treatments = t*
  - *# of Replicates = r*
  - *Block size = k*
  - # of blocks in a complete replicate = s
  - *# of blocks* $s * r = b$



DoE: Lattice & Alpha Designs Summary

- Types of LD & their distinguishing features
  - Square Lattice Designs – SLD
    - # of treatment **MUST** be a perfect square (t=k*k)
    - s **MUST** be equal to k (s=k)
    - r = k+1 for complete balance
  - Rectangular Lattice Designs – RLD
    - t = s*(s – 1)
    - k = s - 1
  - Alpha Lattices Designs – ALD
    - t = s*k
    - Wide range of s and k options

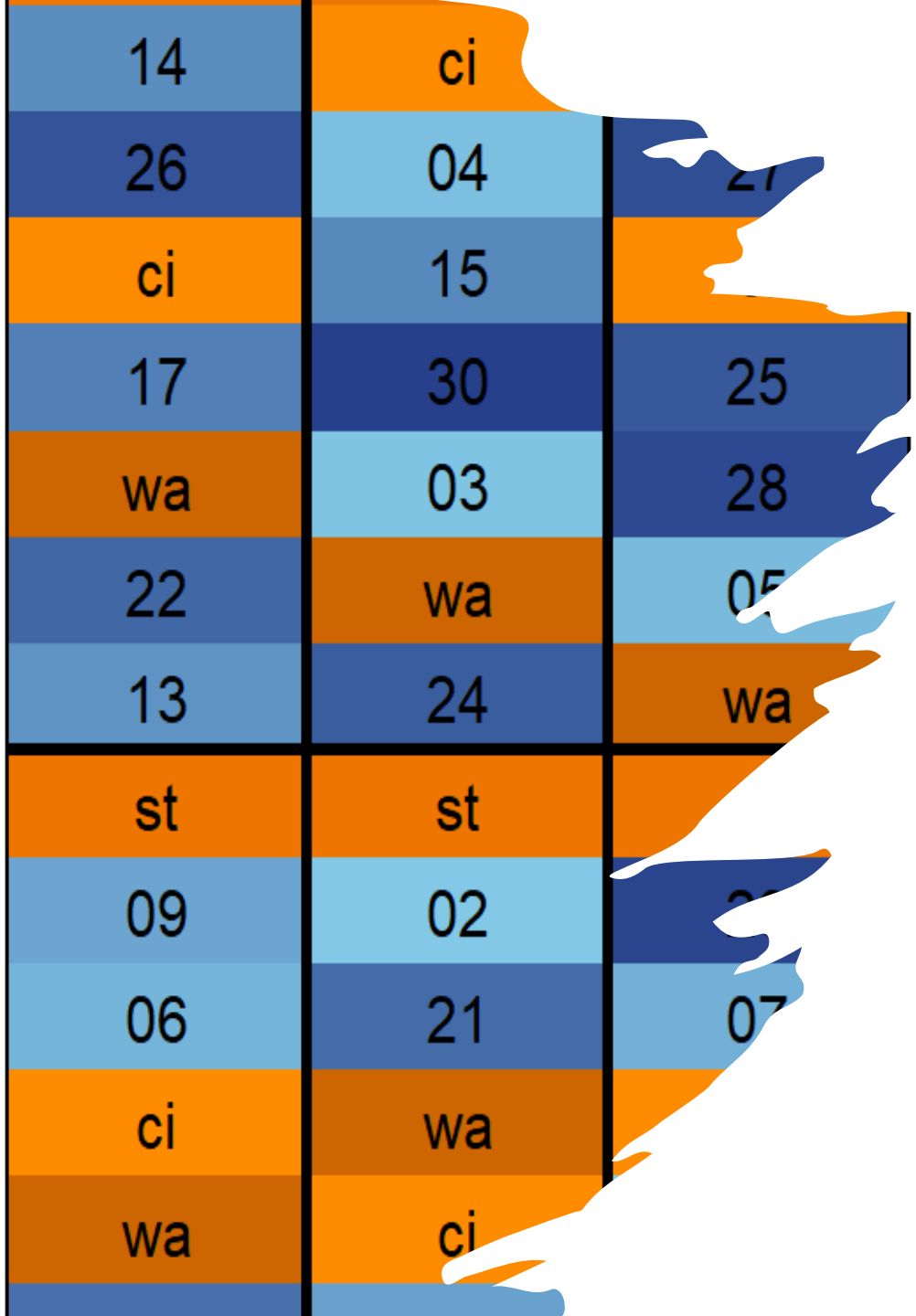


DoE: Practice 1 (15 minutes)

- An experimenter has five (5) genotypes to be tested with 3 different levels of Nitrogen (No treatment, 50%, and 100% of the recommendation).

- A breeder has 35 genotypes and wishes to replicate his trial 3 times.

- Please recommend a design for each of the case and generate the design



DoE: Augmented Designs

- Where there is little seed for some of the genotypes, there may be just a single replicate of some of these genotypes.

- **Augmented designs** are appropriate for early generation stages when hundreds or even thousands of genotypes are being evaluated in the same experiment, using a limited amount of sowing material, perhaps enough for one replicate only

DoE: Augmented Designs

- An augmented design is any standard design in control treatments augmented with additional (new or test) treatments in complete or incomplete blocks in one-way heterogeneity setting
- A survey of the literature reveals that generally these experiments are conducted using an augmented randomized complete block design

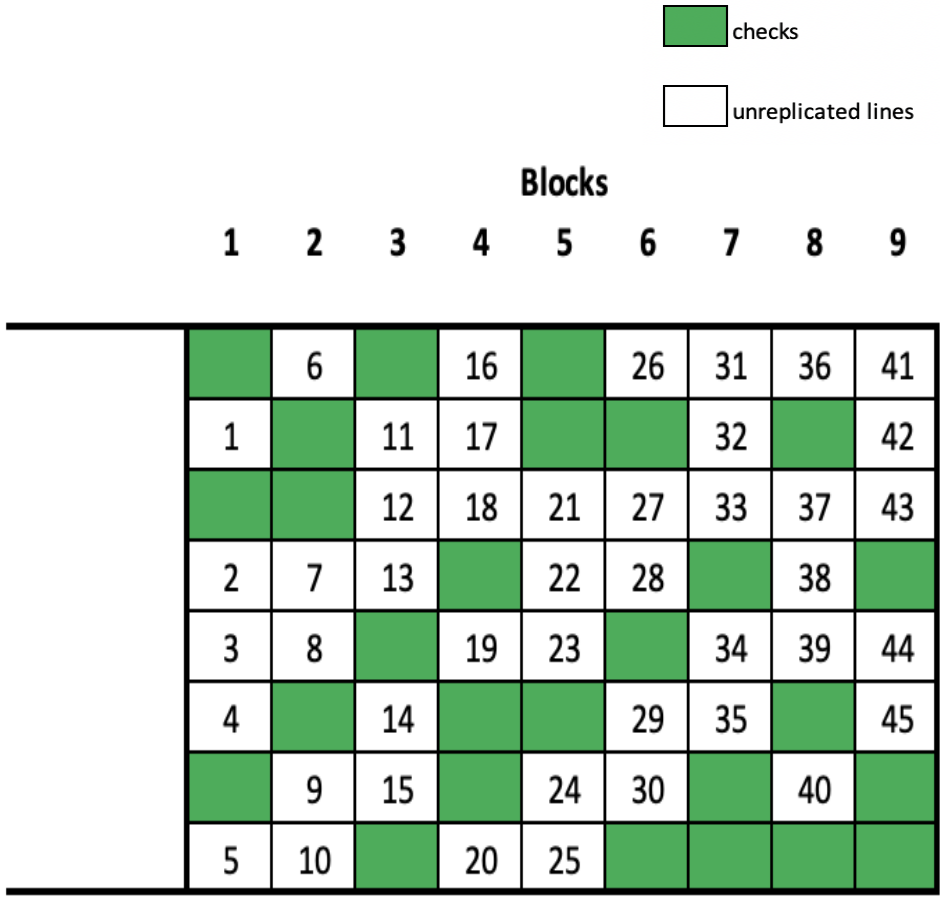
DoE: Augmented Designs

- In an **augmented RCB design**, the basic design plan is to divide the experimental area into several blocks.
- A few check varieties are replicated in each block, while test genotypes are assigned to the remaining plots in each block
- The test genotypes are not replicated but are assigned at random throughout the blocks
- Genotypes do not have replicates; precision is lower than in replicated designs

DoE: Augmented Designs

Example:

- 45 unreplicated lines
- 3 checks
- 9 blocks
- #plots: (45+3*9) = 72 plots
- replicated checks to enable the estimation of block effects and error variances



DoE: Augmented Designs

- The total number of blocks is determined by the number of **check varieties (*c*)** used in the trial.
- As a result, the minimum number of blocks (***b***) must be:

$$b = [12/(c-1)] + 1$$

- For example, with three checks then:

$$b = [12/(3-1)] + 1 = 7$$

DoE: Augmented Designs Summary

- Controls (check varieties) are replicated in a standard experimental design
- New or test treatments (genotypes) are not replicated, or have fewer replicates than the checks i.e., they augment the standard design.
- Control heterogeneity in one direction
- Simplest case is Augmented Randomized Complete Block Design – ARCBD
  - Checks occur once in every block
  - New or test treatment occurs once in the treatment

Block 1	Block 2	Block 3
st	ci	st
14	st	18
26	04	27
ci	15	ci
17	30	25
wa	03	28
22	wa	05
13	24	wa
st	wa	ci
09	02	29
06	21	07
ci	st	01
wa	10	wa
20	ci	12
11	08	st
23	16	19
Block 4	**Block 5**	**Block 6**

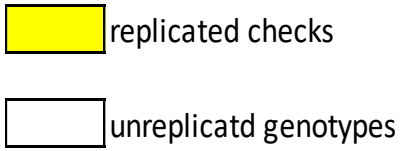
replicated checks

unreplicatd genotypes

DoE: Augmented Designs Summary

- Plant breeding use case
  - *Limited seeds – test treatments*
  - *Land & other resources are limited*
  - *It's difficult to keep homogeneous blocks many genotypes.*
  - Test promising genotypes in as many environments as possible
- Few drawbacks
  - Significant resources are spent on producing & processing of control plots
  - Experimental error has a limited number of degrees of freedom, which limits the power to detect changes across treatments.
  - Unreplicated experiments are inherently imprecise, regardless of the design

Block 1	Block 2	Block 3
st	ci	st
14	st	18
26	04	27
ci	15	ci
17	30	25
wa	03	28
22	wa	05
13	24	wa
st	wa	ci
09	02	29
06	21	07
ci	st	01
wa	10	wa
20	ci	12
11	08	st
23	16	19
Block 4	**Block 5**	**Block 6**



DoE: Practice 2 (15 minutes)

- A plant breeder has 53 genotypes, 3 of which are standard checks to be replicated in each block.
- Please recommend a design and generate the design

DoE: P-rep Designs

- In EGVT, augmented designs (unreplicated trials) using control plots where seed is limited, and full replication is not feasible are popular
- An alternative is to use **p**artially **rep**licated (**p–rep**) designs, where a proportion of the test genotypes are replicated

DoE: P-rep Designs

Example:

- 45 lines:
  - 10 lines replicated twice (20%-25%
  - 35 unreplicated lines
- 5 blocks
- 3 checks in each of the 5 blocks
- #plots: 10*2+35+3*5 = 70 plots

	Blocks				
	1	2	3	4	5
			14	23	
		6			
		7			
		8	15	24	31
	1		16		32
	2		17		
		9	18	25	
		10		26	
	3	11	19	27	33
	4			28	
		12	20	29	
			21		34
	5		22	30	
		13			35

checks

replicated lines

unreplicated lines

DoE: Sparse Testing

- Sparse testing represents a promising approach to expand the number of lines and or locations
- Test each genotype only in a subset of locations and predict missing information
- The main goal is to test more genotypes or the same number of genotypes in more locations (or a combination of the two) at the same cost
- Designs: alpha, row-columns, p-rep, augmented

DoE: Spare Testing

Testing	More genotypes	More locations
Full	500 genotypes in 4 locations	500 genotypes in 4 locations
Sparse	1000 genotypes, each in 2 out of 4 locations	500 genotypes, each in 4 out of 8 locations

DoE: Spare Testing

- Increased selection gain due to testing of more entries and/or locations at same costs
- Relative efficiency depends on relationship of genotypes and correlations among locations
- Logistic challenges:
  - seed dispatching to locations,
  - statistical analysis, etc.

Thank You For Listening