Statistical Data Analysis for Post-Graduate Students Using R Programming Language

Facilitators
Susan Balaba Tumwebaze
Thomas Lapaka Odong
Hellen Namawejje

RUFORUM
Capacity Building in Agriculture

MAKERERE UNIVERSITY
Leveraging 100 years of Excellence
in Building a Transformed Society

Introduction to Research Process

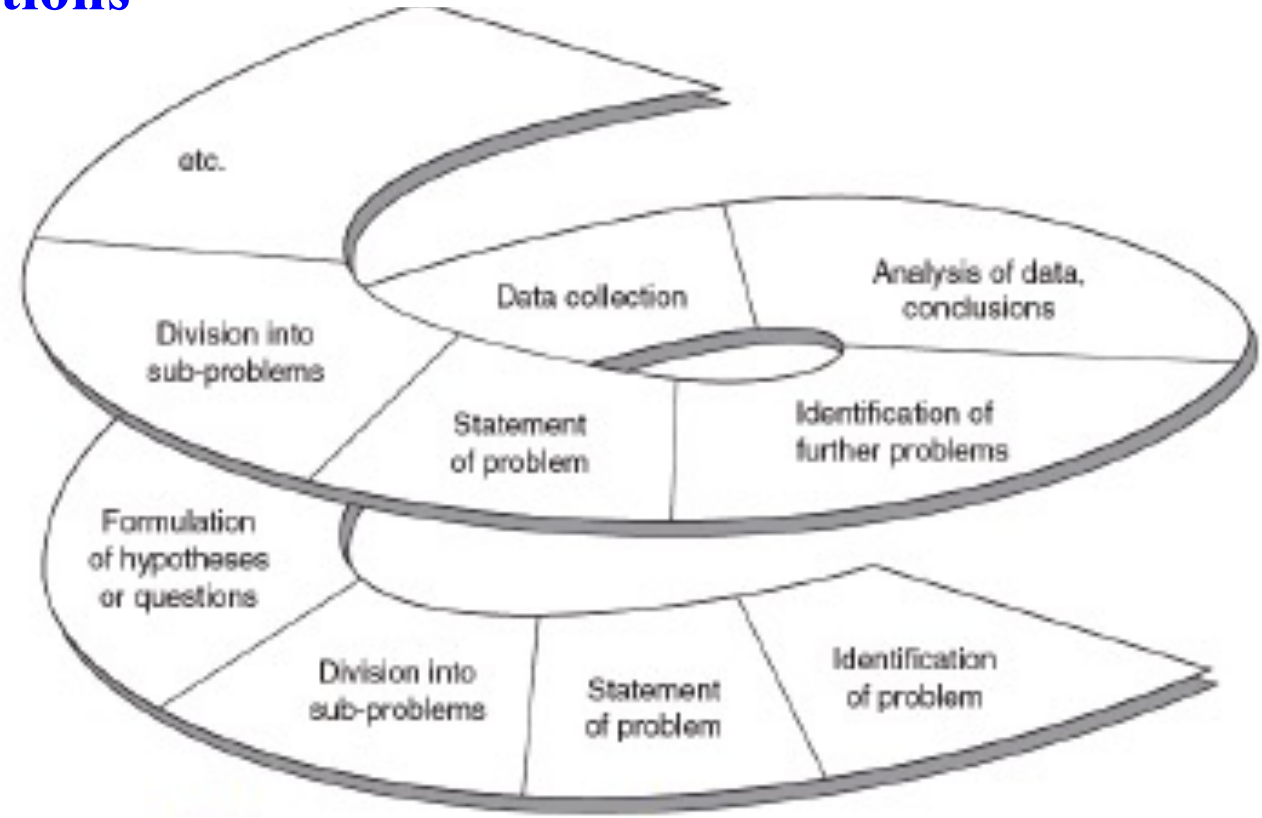
Introduction to Research process

Definition of Research

- Research is a process of **systematic inquiry** that entails collection of data; documentation of critical information; and analysis and interpretation of that data/information, in accordance with suitable methodologies set by specific professional fields and academic disciplines.

Introduction to Research process

Research cyclic – Research generate more problems – Careful observations



http://www.uk.sagepub.com/upm-data/40600_9781849204620.pdf

Introduction to Research process

Process (Biological, social, etc.) generate variations

1) Variation **evoked by the investigator/scientist**, e.g. different animal breeds, plant species, fertilisers, varieties, chemical doses, types of food preservative, etc.
2) Those **not under direct control** of the investigator, e.g. altitude, temperature, soil texture at different experimental sites, sex of animal; etc

Introduction to Research process

Steps in Scientific/agricultural research

- **Problem identification**
  - scientific interest - basic research
  - practical concern - applied/adaptive research
- **Literature review** to identify
  - what is known - pool
  - what is unknown - gap

Introduction to Research process

Steps in Scientific/agricultural research

- **Defining objectives – is** one of the most important stages of research
  - it is the basis for other stages of research.
  - You have to come up with specific objectives in relation to the research problem, which may be:

Introduction to Research process

.

- **Hypothesis**
  - theoretical proposition or a tentative guess at the answer to a research question.
  - determine what kind of data to collect, what factors to include, etc.

- **Data collection**
  - We need data that will provide evidence for or against the hypotheses
  - A researcher usually generates his/her own data by means of an **experiment** or **survey**.
  - *Design* of survey or experiment determines the quality of results. So when we get to experimentation, we shall spend some time on **experimental design**.

Introduction to Research process

- **Data analysis**
  - Takes various forms depending on study objectives.
  - Statistical -estimates, comparisons (e.g. t, F), associations (chi-square) or relationships and optimization (regression), etc.
  - Economics - net benefit ratio, marginal returns, etc.
  - Other (e.g. farmer's assessment/perception)

- **Interpretation**
  - draw logical conclusions
  - make valid inferences
  - new principles or methodologies
  - solution or no solution
  - acknowledge limitations of the research

Introduction to Research process

- **Dissemination of results**
  - project reports, theses
  - publications
- NB: RESEARCH - continuous, dynamic & iterative (cyclic in nature)
  - steps overlap

Introduction to Research process

Roles of objectives

- With a clearly defined objectives, the researcher can focus on the study – provide direction of the investigation
- Avoid collection of data which are not necessary and leave other necessary data
- Organize the study in a clearly defined parts or phases
- Facilitate the development of research methodology – help orient collection, analysis and interpretation of data

Introduction to Research process

FORMULATION OF RESEARCH HYPOTHESES

- Hypotheses are developed by asking *what, why, how, when, where, what if,* ... of different aspects of the causes of the problem. It is the hypotheses that are subjected to testing in a study.
- A clearly stated hypothesis will:
  - aid and determine the **design** of an experiment/survey
  - help identify appropriate **treatments** and **controls**
  - determine the type of **data** to collect, and
  - (probably) suggest a **method of analysis**

Introduction to Research process

FORMULATION OF RESEARCH HYPOTHESES

- Hypothesis is derived from **a research problem** which usually framed as a question
- Hypotheses are developed by asking *what, why, how, when, where, what if*, ... of different aspects of the causes of the problem. It is the hypotheses that are subjected to testing in a study – **Answers to such questions**
- Problem:
  - **Low yield (farmers' problem) and lack information on how to improve yield of Acholi white groundnut variety(Researcher's problem)**

Introduction to Research Process

FORMULATION OF RESEARCH HYPOTHESES

- Problem:
  - **Low yield of Acholi white groundnut variety in Uganda**
- **"Why are farmers in Uganda getting low yield when they grow Acholi white?"**
- **Or "How can we increase Acholi white yield on the farmers' fields?"**

Introduction to Research Process

Formulation of Research Hypothesis

- In developing the hypothesis, you can be influenced by
  - An existing theory,
  - Related research,
  - Personal experience.
- "Your idea": ***Acholi white variety*** has become susceptible to ground rosette virus disease

Introduction to Research Process

FORMULATION OF RESEARCH HYPOTHESES

- Your idea: ***Acholi white variety*** has become susceptible to ground rosette virus disease

Possible hypothesis

"Resistance to groundnut rosette virus disease in Acholi white is broken down due to emergence of new virus strain."

Introduction to Research process

FORMULATION OF RESEARCH HYPOTHESES

- A clearly stated hypothesis will:
  - aid and determine the **design** of an experiment/survey
  - help identify appropriate **treatments** and **controls**
  - determine the type of **data** to collect, and
  - (probably) suggest a **method of analysis**

Introduction to Research process

Some simple rules in research hypothesis formulation:

- Hypotheses must/should be:
  - clearly relate to the problem to be addressed
  - stated in a way that suggests a solution to a problem
  - stated in such a way as to provide direction for research
  - capable of verification or rejection
  - stated as simple as possible

Overview of Experimental Design

Introduction

- Experiments: **Establish cause-effect relationships**
- Principles of experimental design are same irrespective of field of study: Agriculture, biological, ecological, social, economics

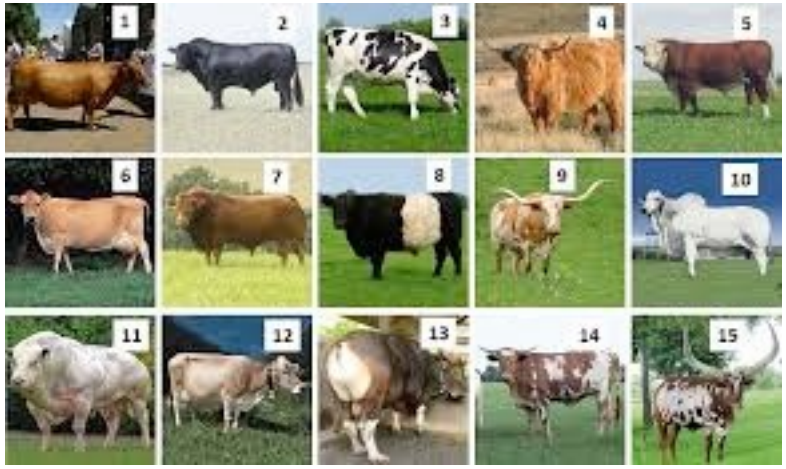
Field Experiment

Laboratory Experiment

Importance of experimental design

- Experimental design has two components:
  - Structure of experimental units (type & nature of exptal material) – **Variation outside the control of the researcher**
  - Structure of treatment set – **Variation under the control of researcher**

Treatment



Experimental material

Experimental Unit, Treatments, Sampling Units

Experimental Unit: Pot

Treatment: Fertilizer, pesticides, varieties, etc.

Sampling Unit: Each plant

Variables: Plant height, number of combs, leaf length, number of leaves

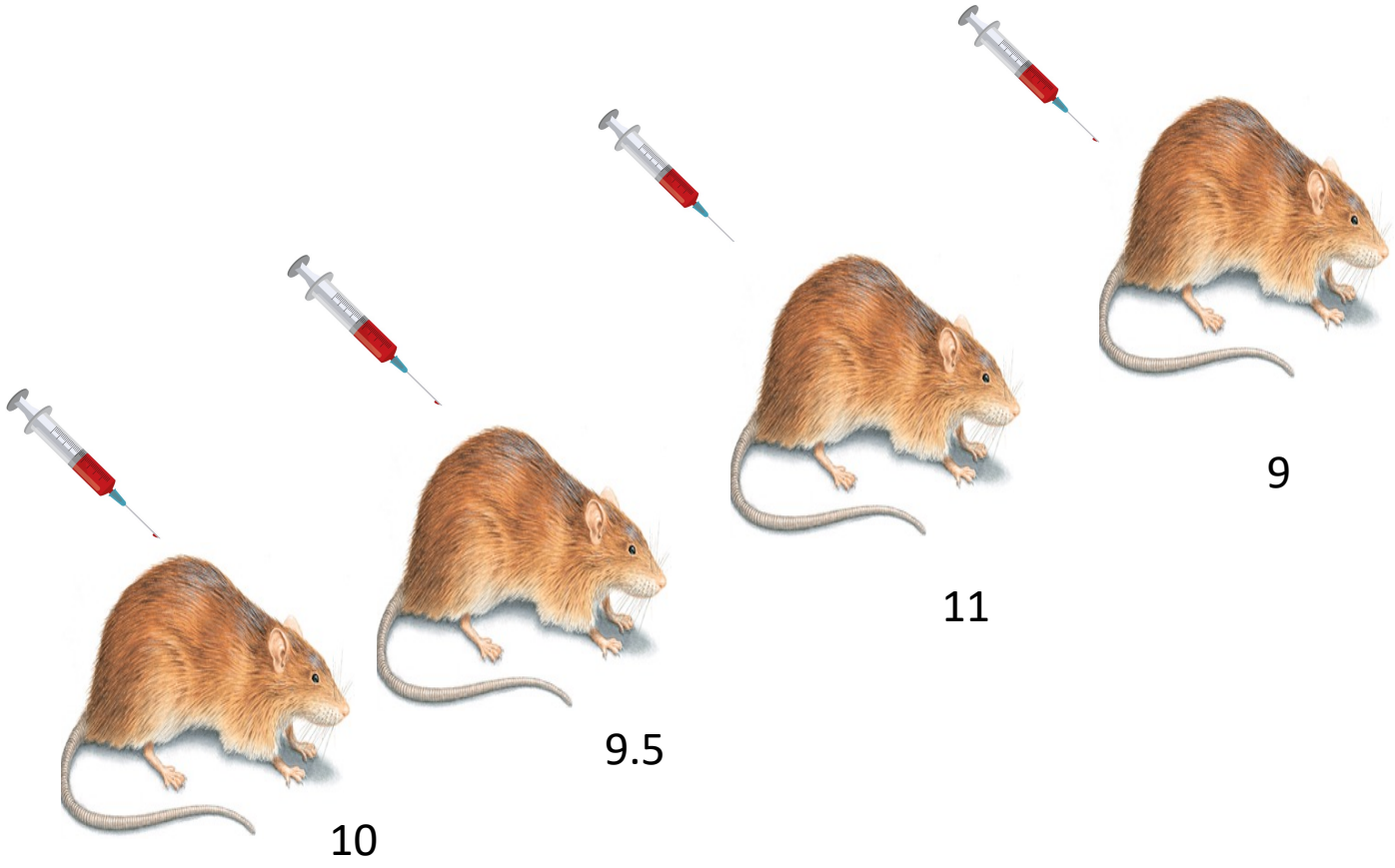
Importance of experimental design

- A good experiment should provide:
  - Unbiased estimate of treatment differences

Comparing the effect two energy drinks: Monster and Red bull

Importance of experimental design

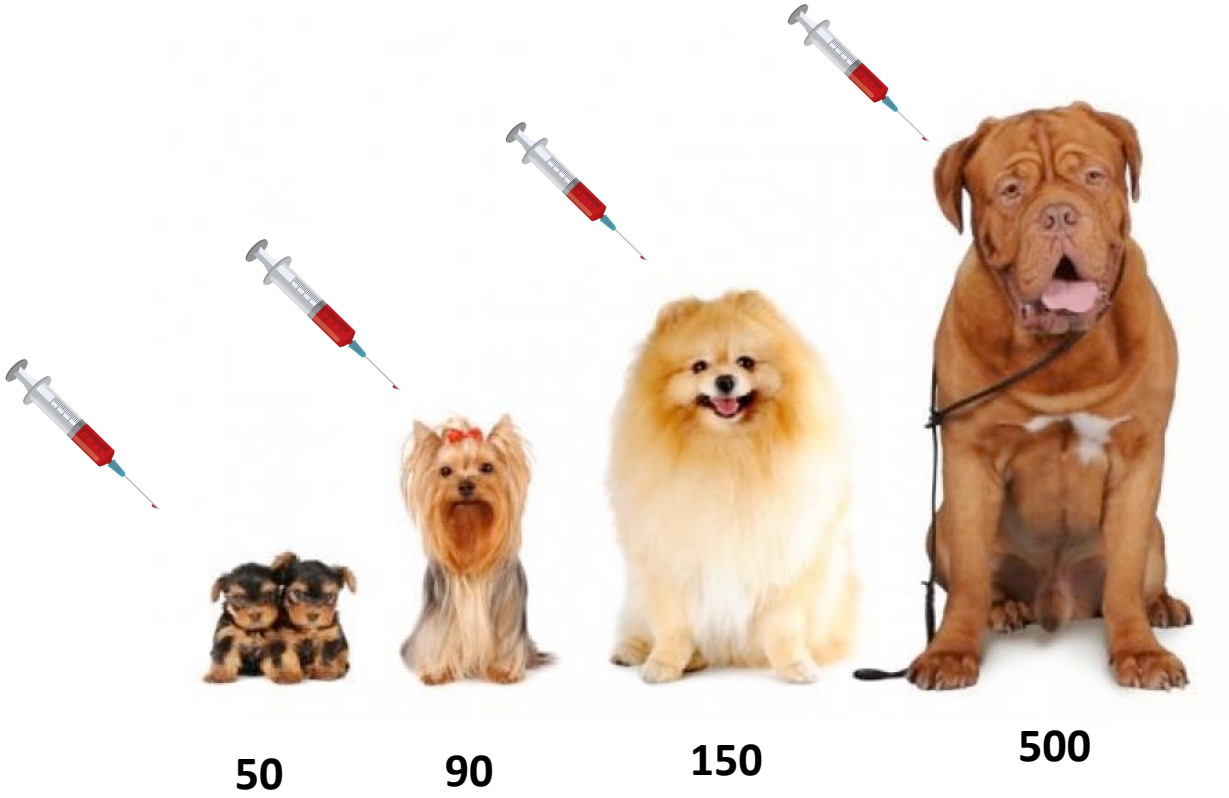
- A good experiment should provide:
  - Unbiased estimate of the experimental error



Variation between similar experimental units receiving the same treatment

Importance of experimental design

- A good experiment should provide:
  - Unbiased estimate of the experimental error



Variation between similar experimental units receiving the same treatment

Basic Principles of Experimental Design

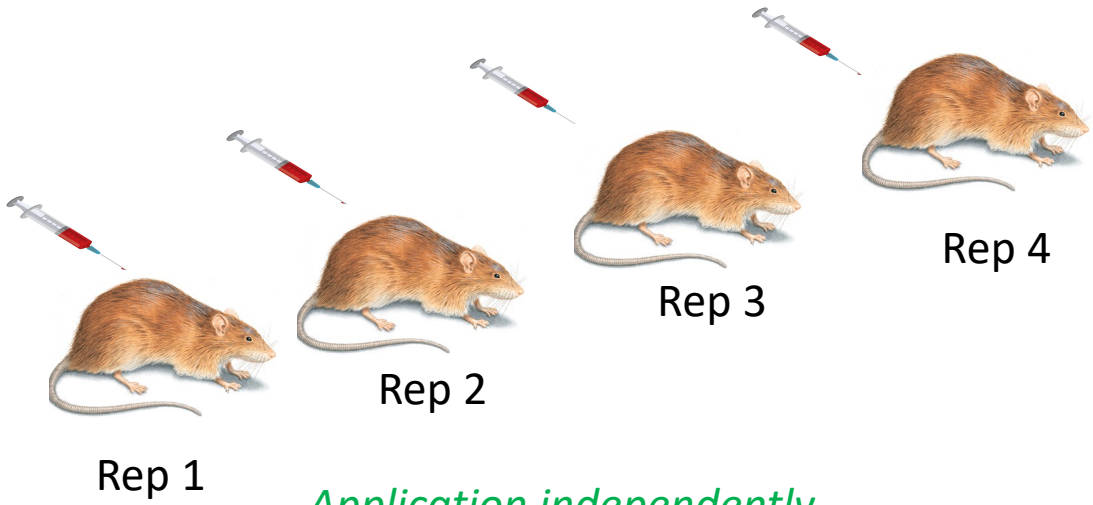
- Basic principles of experimental design are prerequisites of obtaining valid and unbiased estimates of treatment effects and experimental error
- The basic principles of experimental design area:
  - **Replication**
  - **Blocking/Local Control/Restriction**
  - **Randomization**

Replication

What, why, how

What is replication?

- **Replication** (r) is the repetition of a treatment under identical conditions in an experiment.
- Each treatment is ***applied independently*** to two or more experimental units



Application independently

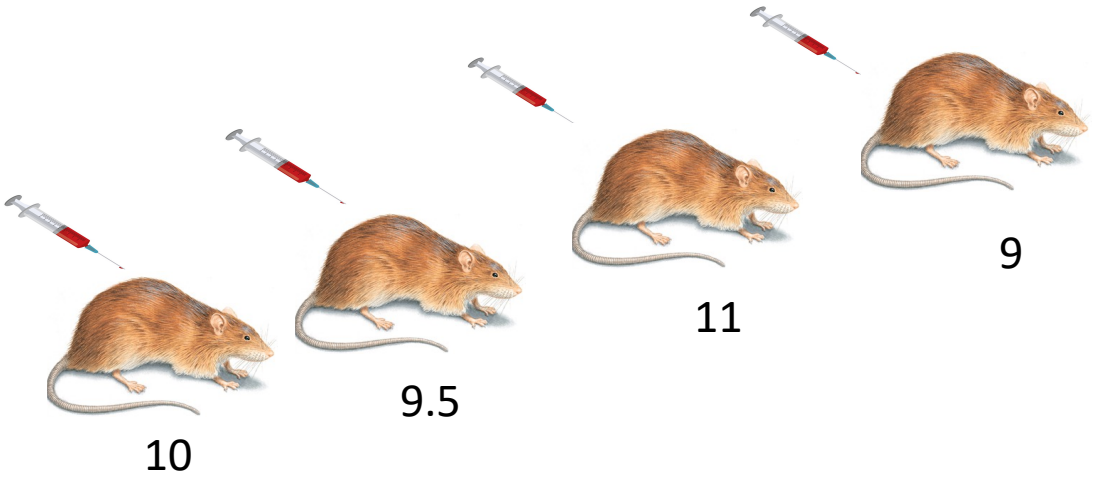
Application not independently

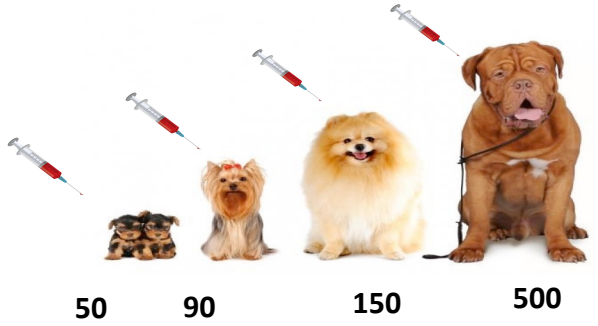
Psuedo replication or subsampling

Why replicate?

Roles of replication

- Permits estimation of experimental error variance





Reaction times for rats

> reaction<-c(10, 9.5, 11, 9)
> var(reaction)
[1] 0.7291667

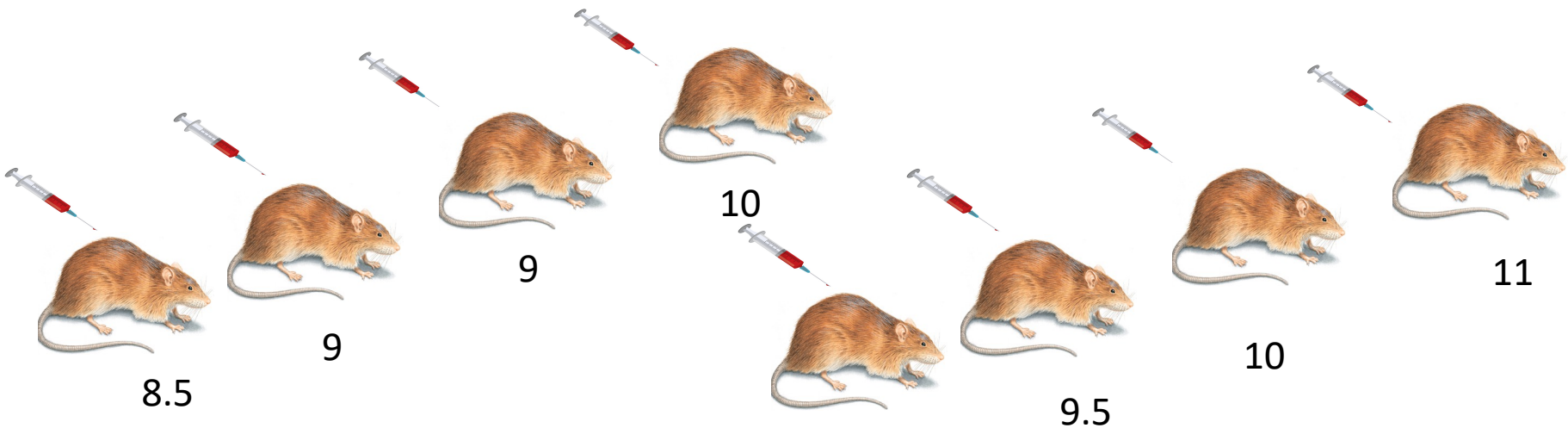
Reaction time for dogs

> reaction2<-c(50,90, 150, 500)
> var(reaction2)
[1] 42358.33

Why replicate?

Roles of replication

- Increase of replication increases precision of parameter estimate.
  - Variance of treatment mean = $\left(\frac{\sigma^2}{r}\right)$ hence reduce standard error for treatment



Variance with 8 eight rats

```
> reaction3<-c(10, 9.5, 11, 9, 10, 9.5, 11, 8.5)
> var(reaction3)
[1] 0.78125
> var(reaction)/8
[1] 0.09765625
```

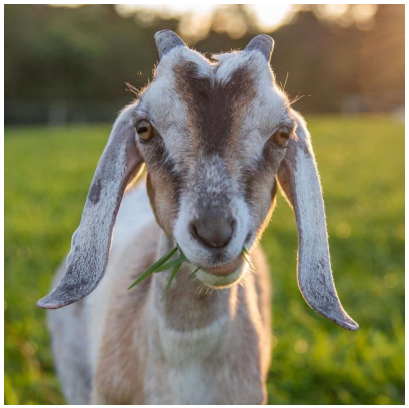
Variance with 4 rats

```
> reaction<-c(10, 9.5, 11, 9)
> var(reaction)
[1] 0.7291667
>var(reaction)/4
[1] 0.1822917
```

Why replicate?

Replication is a form of insurance

- Avoid loosing whole trial

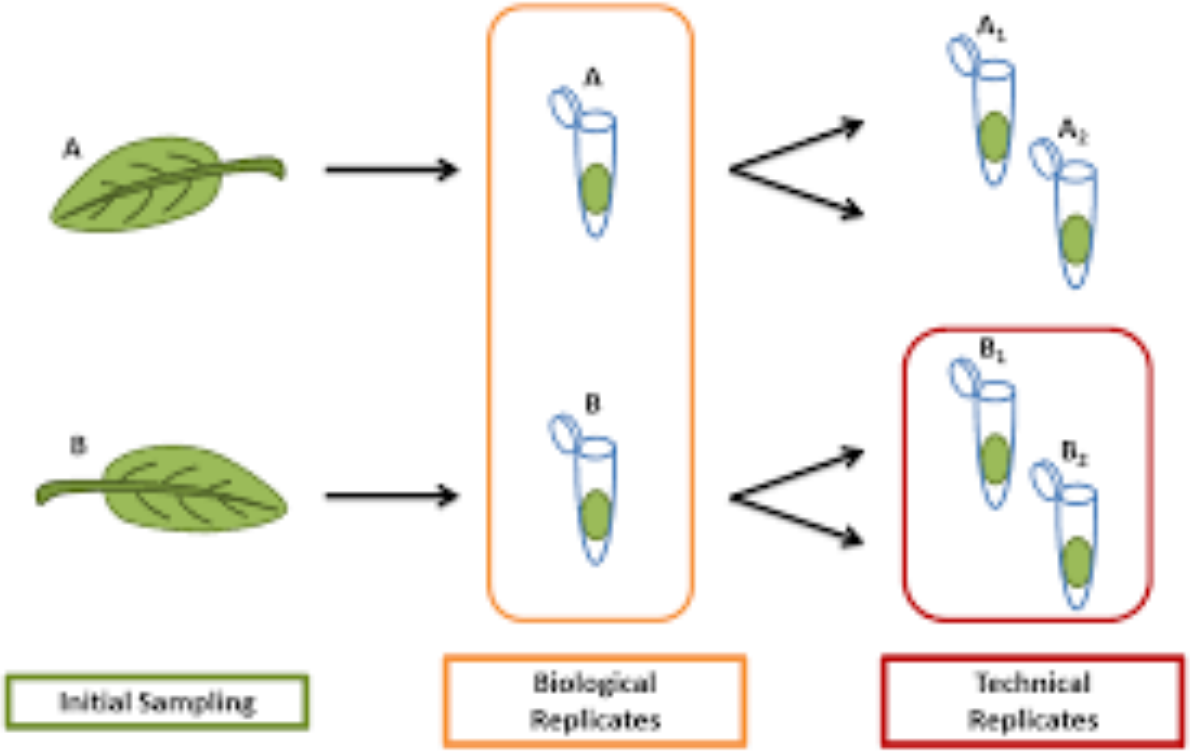


Increase range of validity

- Observe results over a larger sample of farms, soils,...
- Look at context
- How widely do **you** need to test something to be convinced of the results?

How should it be done?

- **Avoid pseudo replications**
- **Beware of Biological versus Technical Replications**



How many replications

Factors that determine the number of replications

- Pattern and magnitude of variability in among the study subject

Which set of potatoes is easy to represent or can be represented by few pieces?

How many replications

Factors that determine the number of replications

- Size of the difference to be detected; the smaller the difference between treatments to be detected, the higher the number of replications

Vs

Vs

How many replications

Factors that determine the number of replications

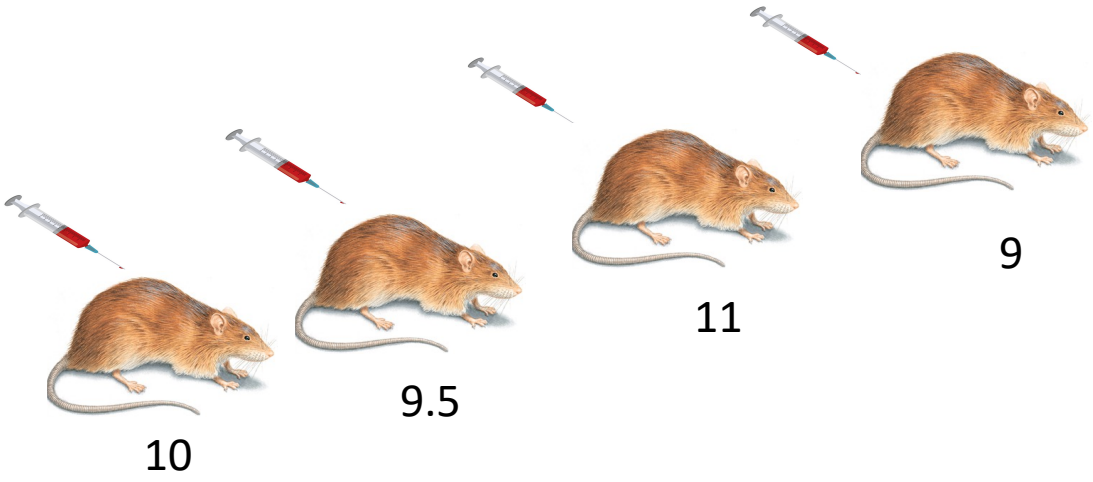
- Required significance/confidence level
- *Available resources for the experiment*

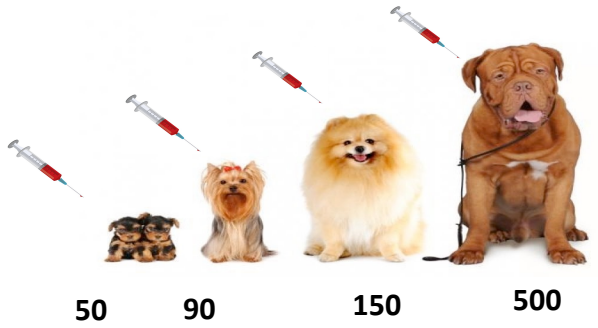
Blocking/Restriction/LC

allowing for variation

Blocking

- Blocking controls/minimizes non-treatment in an experiment
- Non-treatment variation confound effects of treatment
- Increases noise in an experiment and makes it difficult to detect treatment differences or effects





Reaction times for rats

> reaction<-c(10, 9.5, 11, 9)
> var(reaction)
[1] 0.7291667

Reaction time for dogs

> reaction2<-c(50,90, 150, 500)
> var(reaction2)
[1] 42358.33

Blocking

Blocking

- These non-treatment variation needs to be controlled
  - homogeneous material
  - homogeneous management
  - careful measurement
- Blocking can allow us to remove/control the effects of variations in experimental material *(non-treatment variations)*

Blocking

Blocks are - groups of experimental units expected to be similar

Examples: plots with similar soils, same animal breed, etc.

Block 1

Block 2

Block 3

Block 4

Blocking

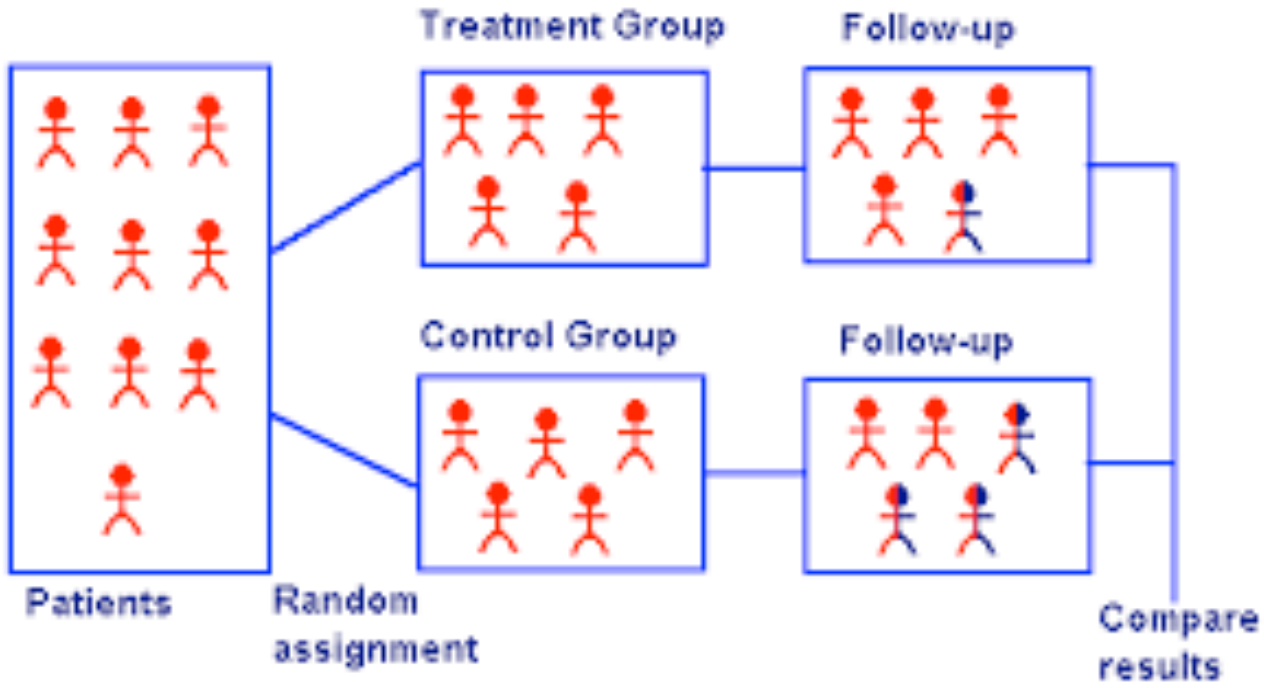
Blocks are - groups of experimental units expected to be similar

Examples: plots with similar soils, same animal breed, etc.

Randomization

Randomization

- Process of assigning “treatments” to the different experimental units (using probability/chance)
- Each treatment has an equal chance to end up in a given subject
- Use explicit process, not just ‘mixed up’



Randomization

- Which plot should receive which treatment?

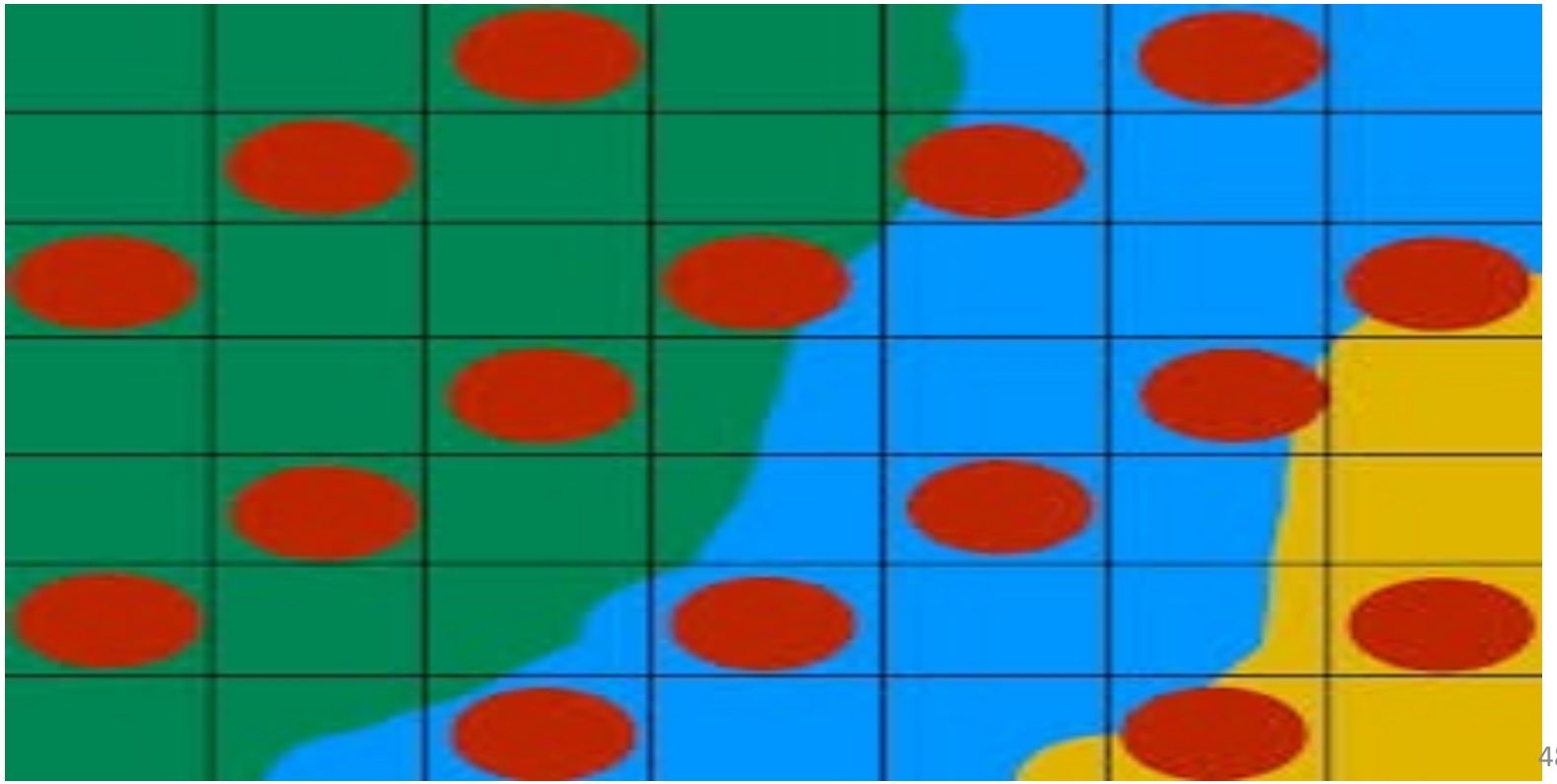
	BLOCKS		
Plots	1	2	3
1	B	C	A
2	D	A	B
3	A	C	D
4	C	B	D

Why do Randomization?

- Avoid favouring some treatments
  - conscious
    - Plots easy to reach
  - Unconscious
    - People tend to take sub-samples close to the centre of a plot of maize
  - avoiding unknown systematic variation
  - “blindness” = medical researcher doesn‘t know which treatment he applies to the patient
- Unbiased estimate of residual variance

Why do Randomization?

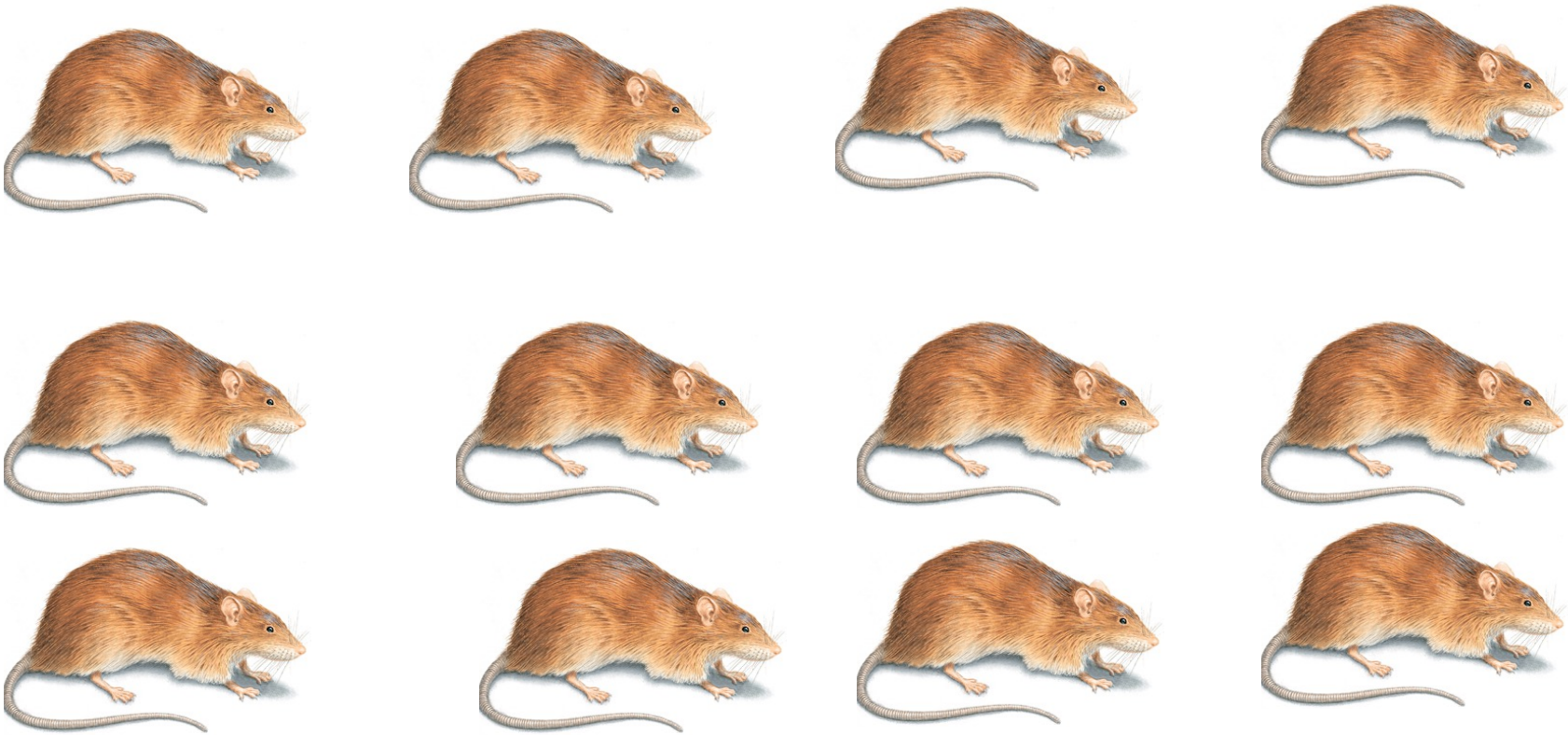
- Avoiding unknown systematic variation
- Unbiased estimate of residual variance



Standard Designs

Completely Randomized Design (CRD)

- **Main Assumption**: **Experimental Units are homogenous**



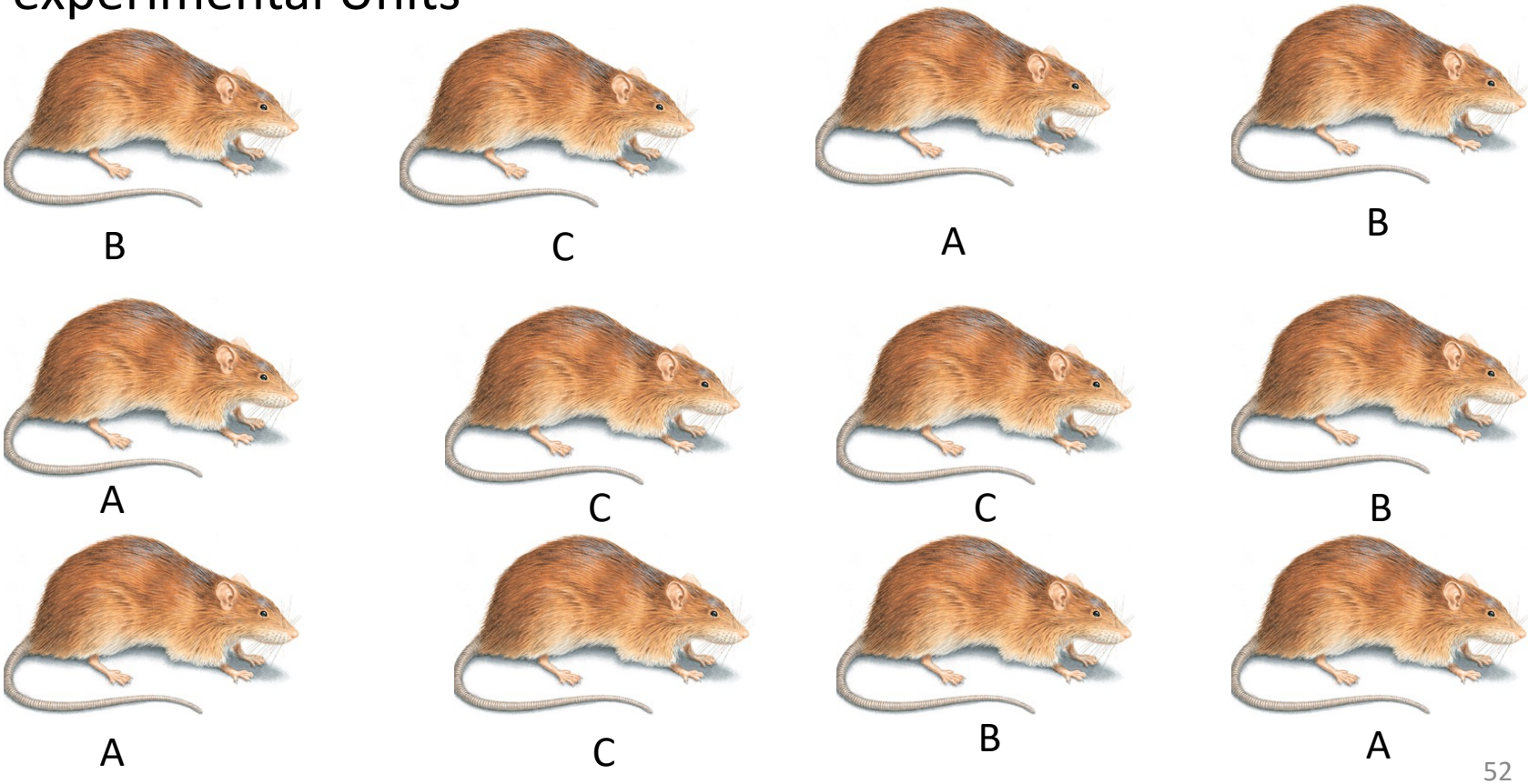
Completely Randomized Design

```
> #randomization for CRD using R
> set.seed(500) #allow for get same randomization
> f <- factor( rep( c("A", "B", "C" ), each = 4)) #creating a vector of factor levels (treatments)
> fac <- sample( f, 12 ) #Randomizes the order of the levels (sample 12 without replacement)
> expt_unit <- 1:12 #
> plan <- data.frame( Rat=expt_unit, Treatment=fac )
> plan
```

	Rat	Treatment
1	1	B
2	2	C
3	3	A
4	4	B
5	5	A
6	6	C
7	7	C
8	8	B
9	9	A
10	10	C
11	11	B
12	12	A

Completely Randomized Design (CRD)

- As a result treatments are assigned completely at random to the experimental Units

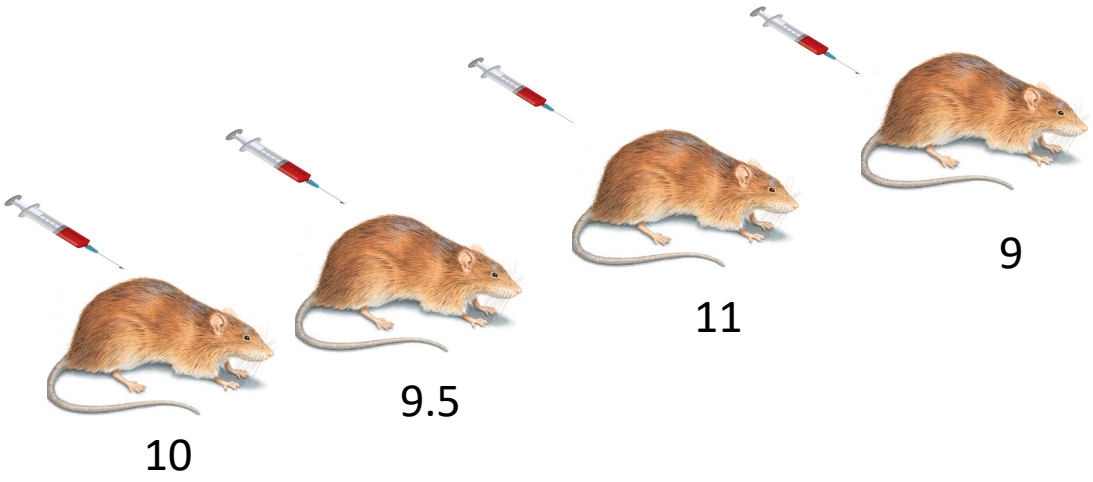


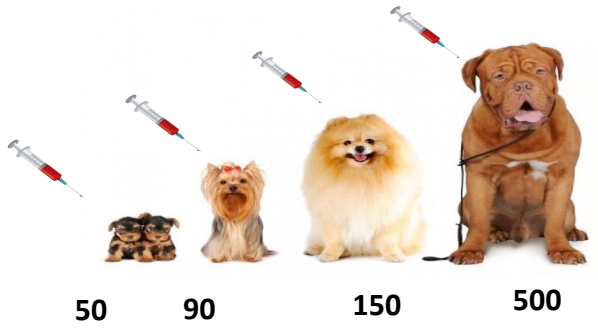
Completely Randomized Design (CRD)

- All experimental units are assumed to be similar **with respect to characteristic that is likely to affect response to treatment**
- Commonly used in laboratory or Screen house experiments
- Application in field experiments is very rare

Completely Randomized Design (CRD)

- The main objection to the CRD is that it is often inefficient.
  - Since randomisation is unrestricted, experimental error includes the entire variation among experimental units except that due to treatments.





Reaction times for rats

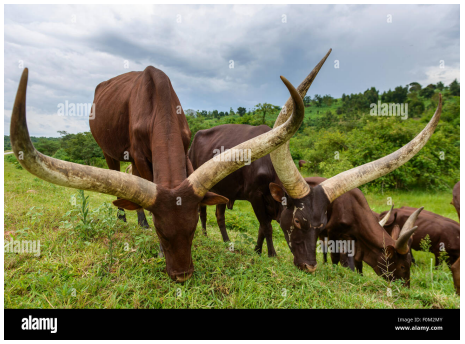
> reaction<-c(10, 9.5, 11, 9)
> var(reaction)
[1] **0.7291667**

Reaction time for dogs

> reaction2<-c(50,90, 150, 500)
> var(reaction2)
[1] 42358.33

Randomized Complete Block Design (RCBD)

- **Experimental material is not always homogenous**



Randomized Complete Block Design (RCBD)

- Experimental material is not homogenous
- **There is a predictable trend in variability in the experimental material e.g. fertility gradient**
- **Experimental units can be grouped into block with homogeneity within block**
- **Knowledge of your experimental material is paramount**

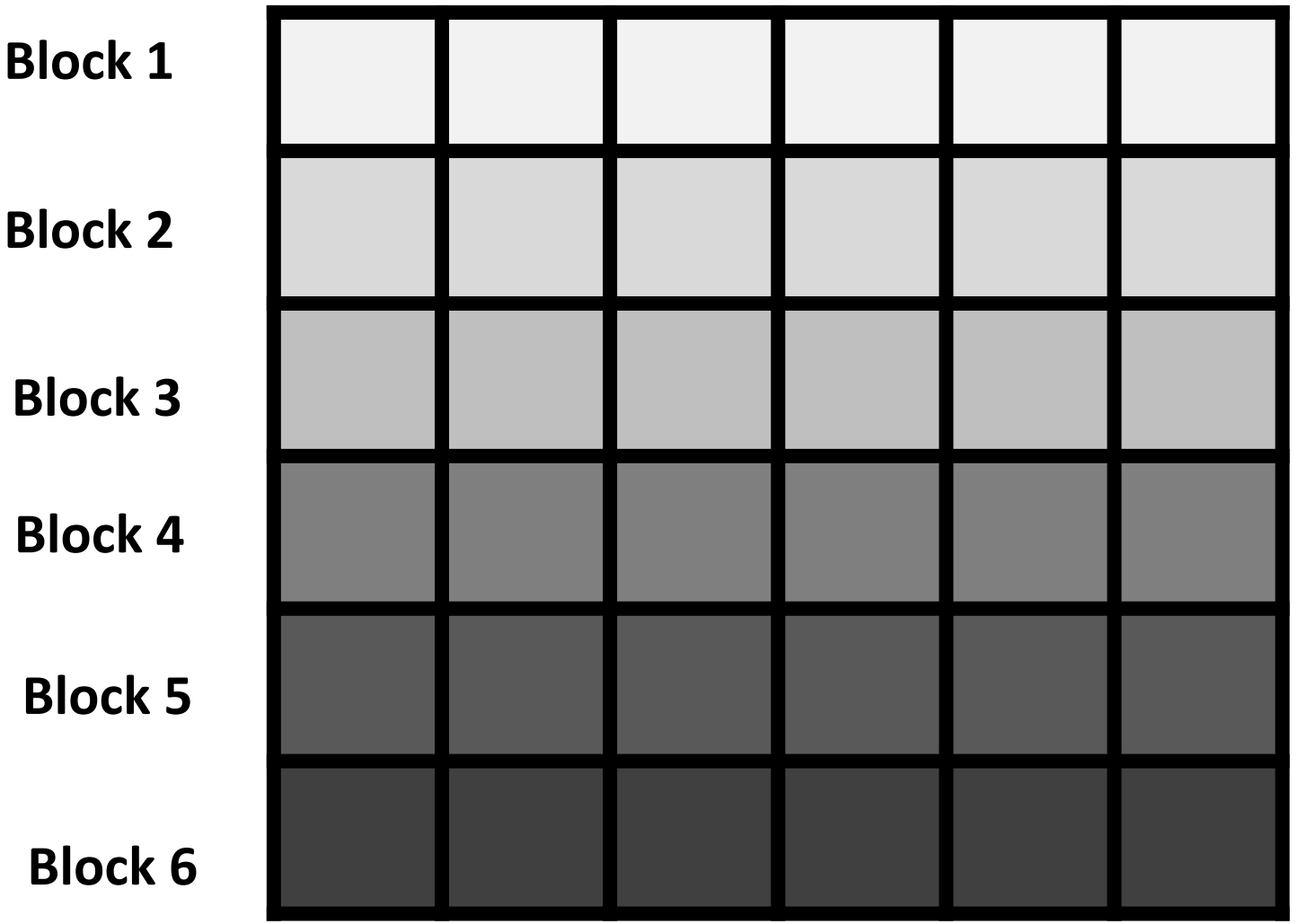
Block 1

Block 2

Block 3

Block 4

Randomized Complete Block Design (RCBD)

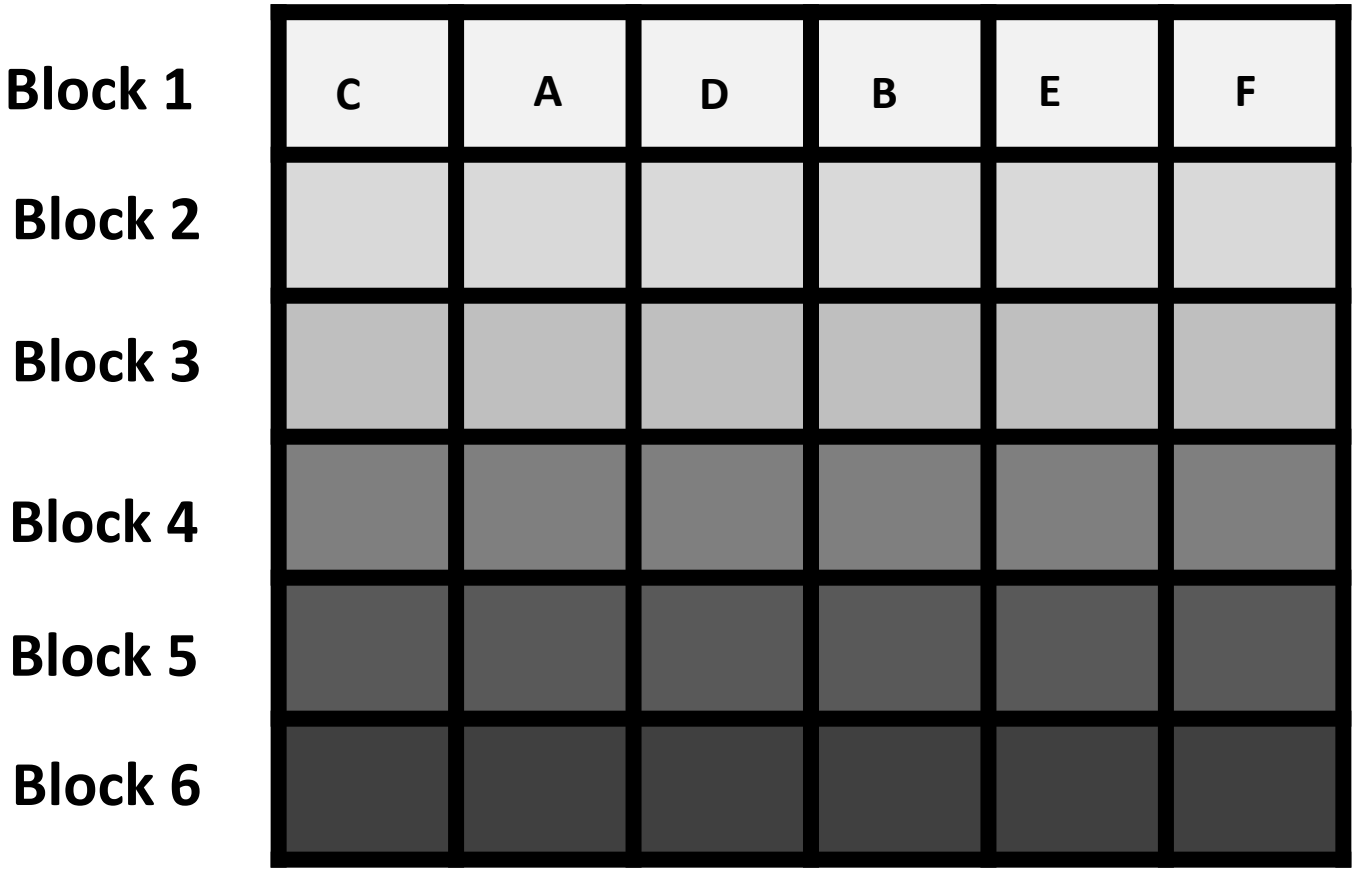


Randomized Complete Block Design

```
> #Randomise complete Block Design - plan
> treat <- c("A","B","C","D", "E", "F")
> b1t <- sample(treat,6) #randomization for block 1
> b2t <- sample(treat,6) #randomization for block 2
> b3t <- sample(treat,6) #randomization for block 3
> b4t <- sample(treat,6) #randomization for block 3
> b5t <- sample(treat,6) #randomization for block 3
> b6t <- sample(treat,6) #randomization for block 3
> treatment<-c(b1t, b2t, b3t) # create a combined vector for treatments in the different blocks
> block <- factor( rep(c("Block 1", "Block2", "Block3","Block4", "Block5", "Block6"),6))
> plot <- rep(1:4,3) #generating labels from plot in each block/category
> plan<-data.frame(Block = block, Plot.Number = plot,treatment=treat)
> plan
```

Randomized Complete Block Design (RCBD)

- **Treatments are randomized within each block separately**
- **All treatments appear in each block and every block**
- **Each block serve as a replicate (number of blocks = number of replications)**

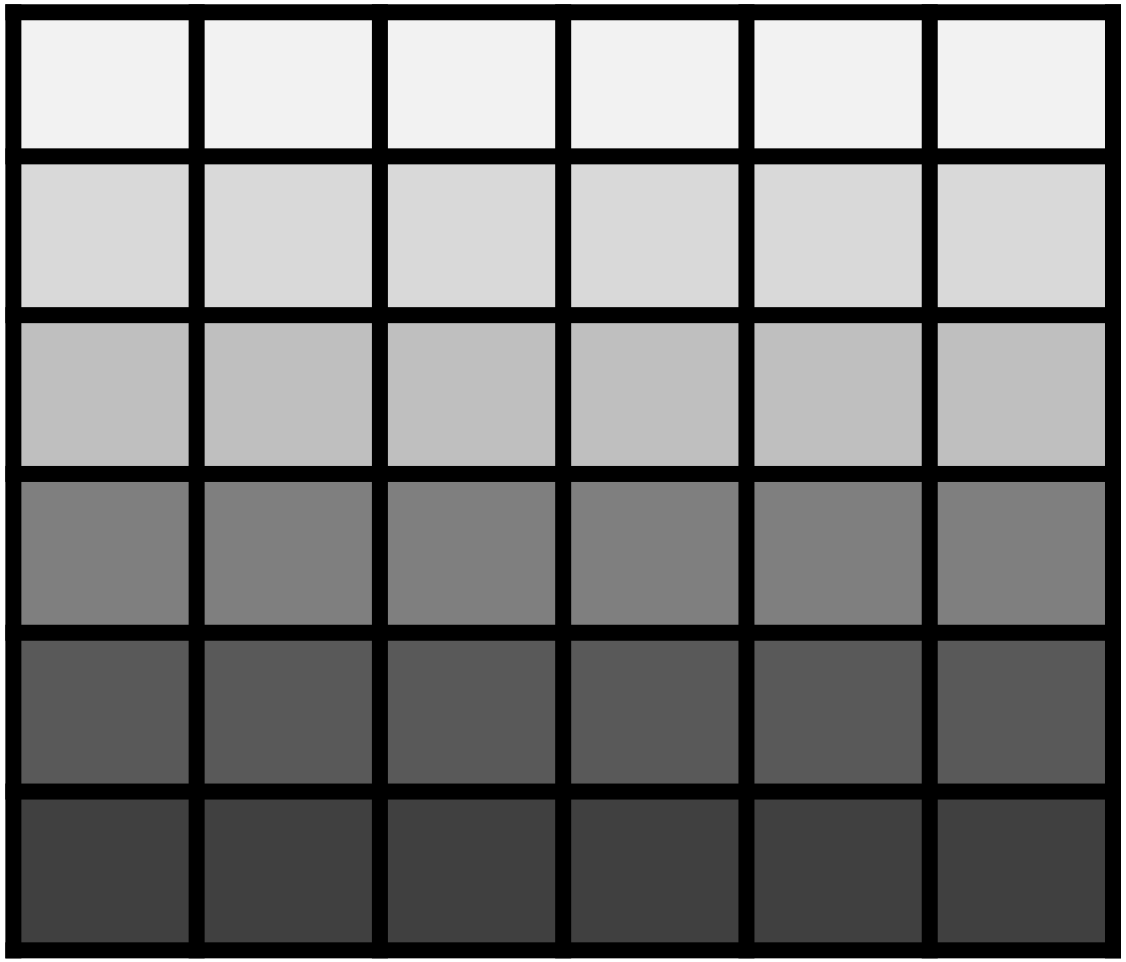


Randomized Complete Block Design (RCBD)

```
> library(agricolae)
> treat<-c(1,2,3,4, 5, 6)
> outdesign <- design.rcbd(treat, 6, seed = 11)
> rcb <- outdesign$book
> levels(rcb$block) <- c("Block 1", "Block 2", "Block 3","Block 4","Block 5","Block 6")
> levels(rcb$treat) <- c("A", "B", "C", "D", "E", "F")
> rcb
  plots   block treat
1   101 Block 1     D
2   102 Block 1     F
3   103 Block 1     B
4   104 Block 1     C
5   105 Block 1     E
6   106 Block 1     A
```

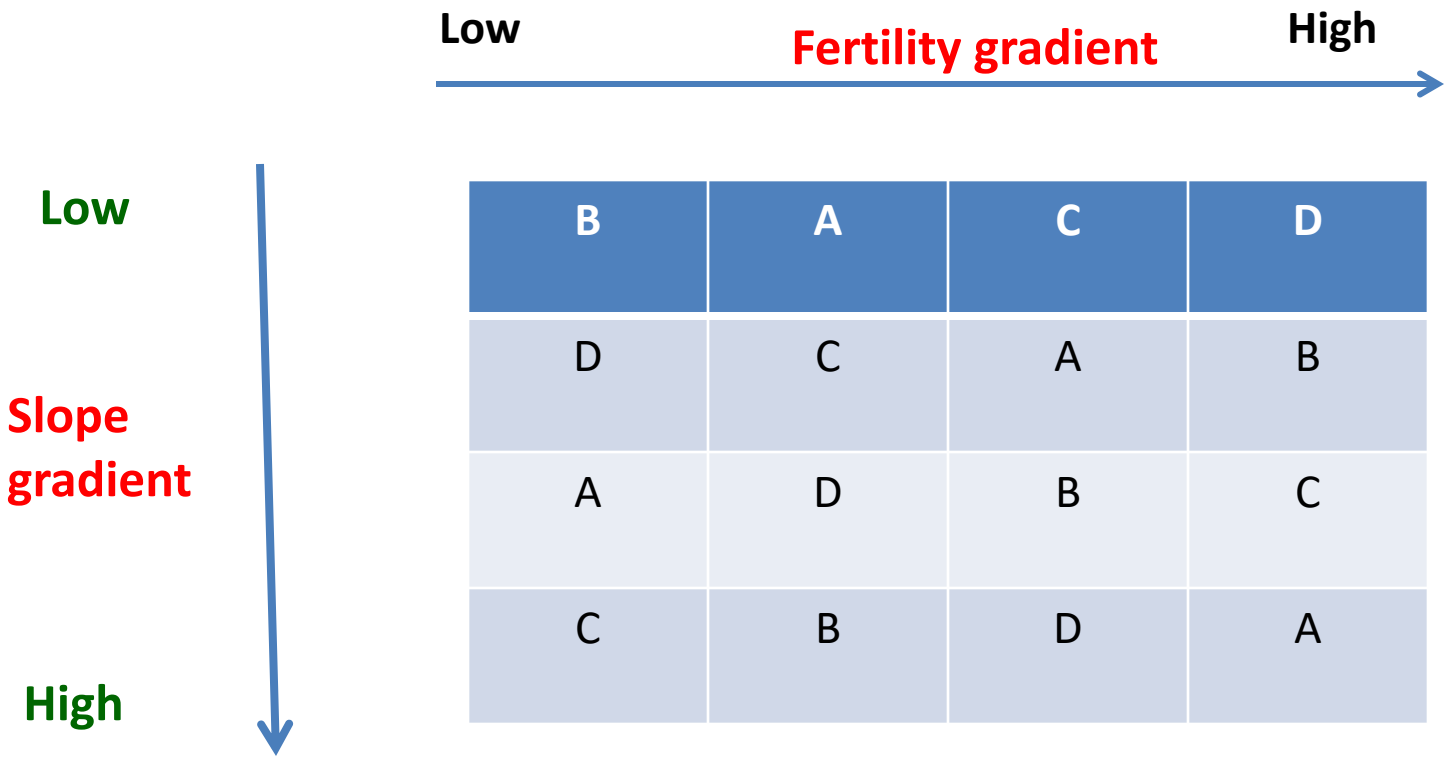
Latin Square Design

- **Only control variation in one direction–Latin Square (2 direction)**



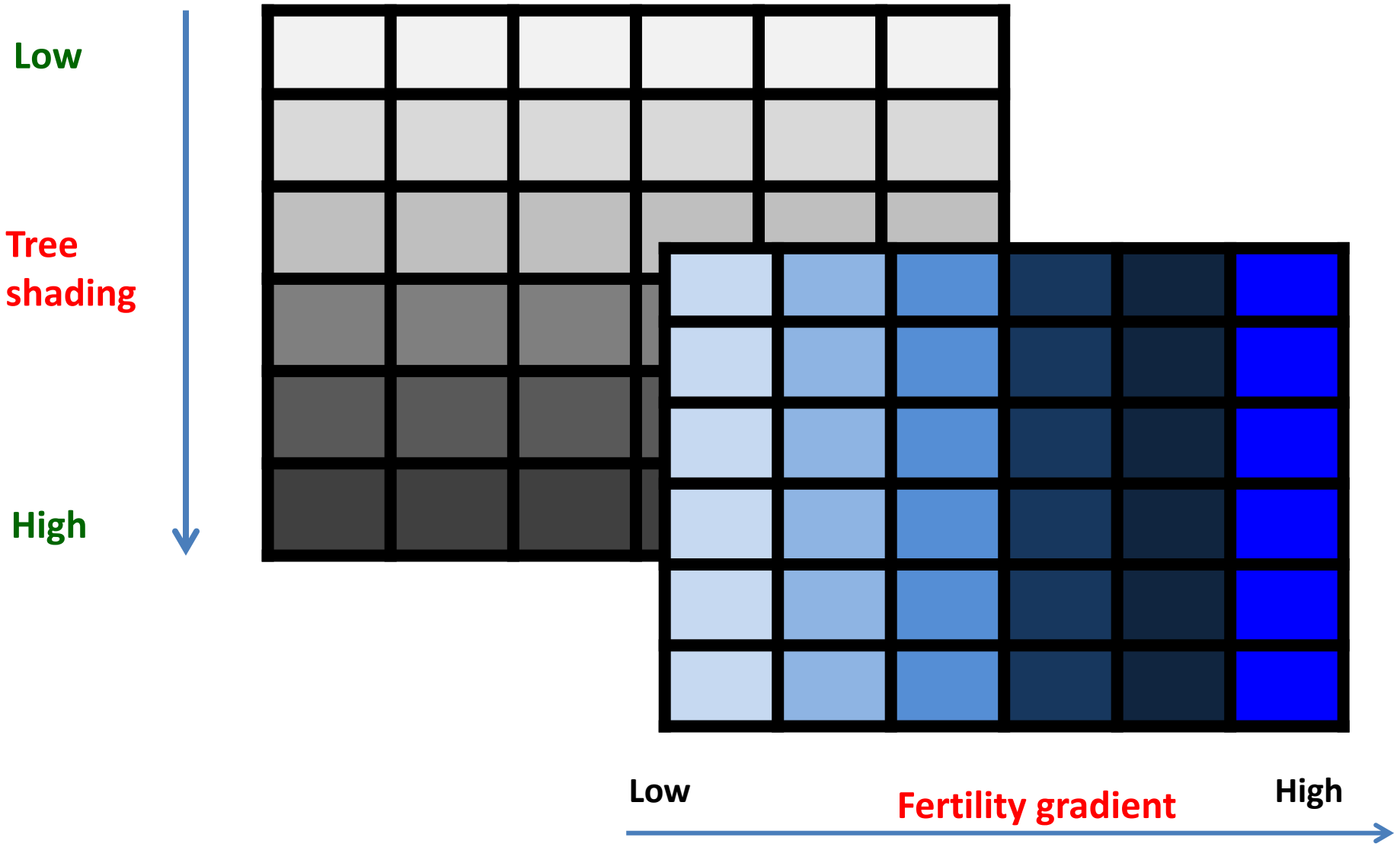
Latin Square design

- In Latin Square, the experimenter is interested controlling two extraneous sources of variability (allows for blocking in two directions).
- Two blocking factors can be look at independently (Two RCBD)



B	A	C	D
D	C	A	B
A	D	B	C
C	B	D	A

Latin Square Design



Latin Square Design

	Machine 1	*Machine 2*	*Machine 3*	*Machine 4*
Operator 1	*A*	*B*	*C*	*D*
Operator 2	*C*	*A*	*D*	*B*
Operator 3	*B*	*D*	*A*	*C*
Operator 4	*D*	*C*	*B*	*A*

Latin Square Design

WEEK 1	WEEK 2	WEEK 3	WEEK 4
Control	Low dose	Medium dose	High dose
Low dose	Medium dose	High dose	Control
Medium dose	High dose	Control	Low dose
High dose	Control	Low dose	Medium dose

Latin Square design

- Laid out in Rows and Columns in this way is called a LATIN SQUARE: .i.e., each treatment occurs once and only once in each row block and column block. Thus:
- **Number of Replications = Number of Rows = Number of columns = Number of treatments**
  - The total number of plots is the square of the number of treatments
  - If you can block on two (perpendicular) sources of variation (rows x columns) you can reduce experimental error when compared to the RCBD

Latin Square Design

- Square requirement: Limitation for large experiments
- Inefficient for small experiments: Few error degrees of freedom
- Several incomplete Latin Square designs exist (Youden Design)

Latin Square Design

Generating Latin Square in R

```
> Nutrition <- c("Nutrition 1", "Nutrition 2", "Nutrition 3", "Nutrition 4", "Nutrition 5")
> outdesign <- design.lsd( Nutrition, seed = 23)
> lsd <- outdesign$book
> levels(lsd$row) <- c("Week 1", "Week 2", "Week 3", "Week 4", "Week 5")
> levels(lsd$col) <- c("Animal 1", "Animal 2", "Animal 3","Animal 4", "Animal 5")
> head(lsd)
```

Latin Square Design

```
> head(lsd)
  plots    row      col          Nutrition
1   101 Week 1 Animal 1 Nutrition 1
2   102 Week 1 Animal 2 Nutrition 4
3   103 Week 1 Animal 3 Nutrition 2
4   104 Week 1 Animal 4 Nutrition 3
5   105 Week 1 Animal 5 Nutrition 5
6   201 Week 2 Animal 1 Nutrition 4
```

Split-Plot “Design”

- Split- plot is used with **a factorial treatment structure**
- One factor require larger area for application – used mainly for logistics reasons
- Can be used with CRD, RCBD or Latin Square – it is more of treatment arrangement than design
- Involves split a larger experimental unit onto which the first factor is applied into the smaller units onto which the second factor is applied

Split-Plot “Design”

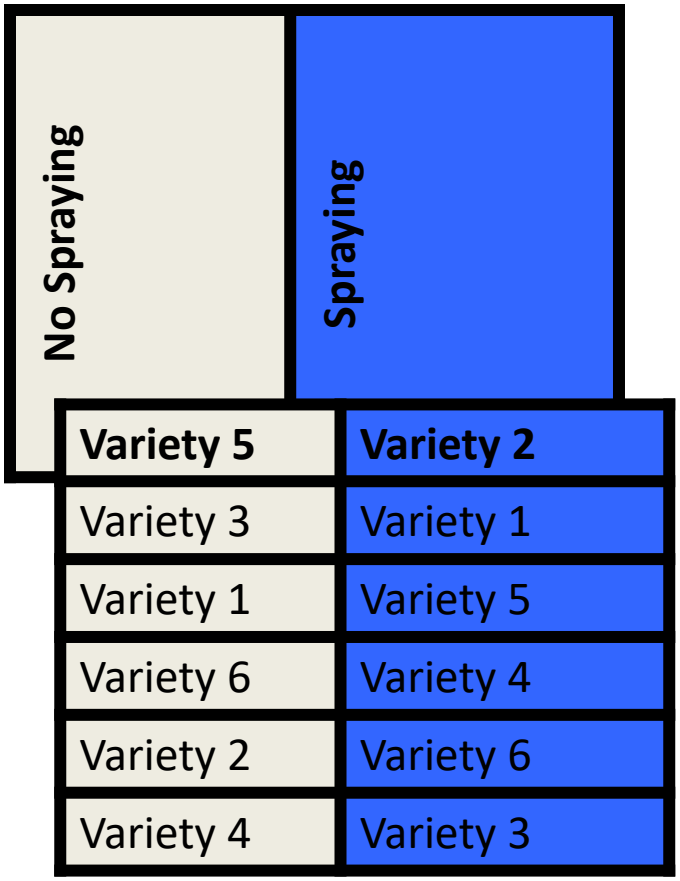
- One factor require larger area for application – used mainly for logistics reasons

Tray A

Tray B

Split-Plot “Design”

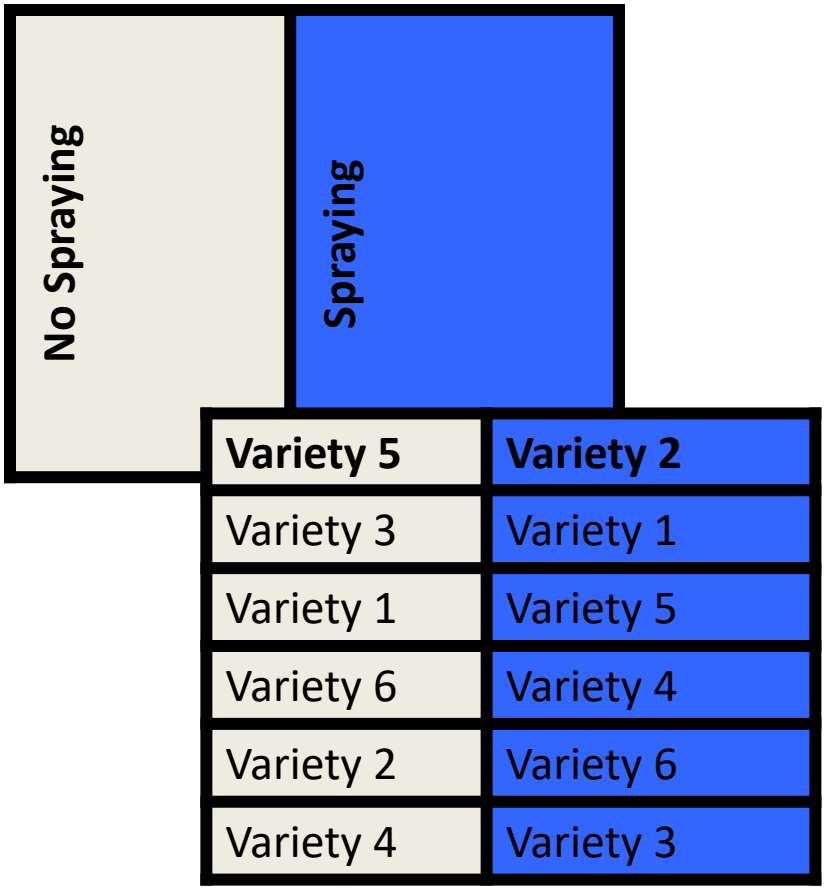
Block 1



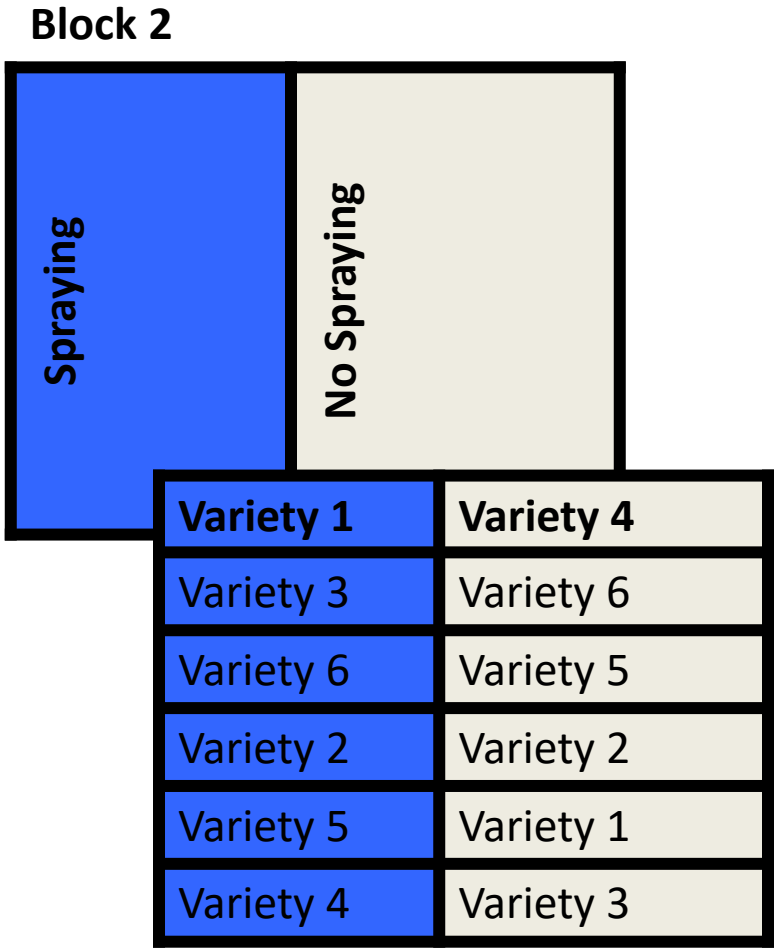
- Used with ***factorial*** treatment structure
- One factor require larger area for application – used mainly for logistics reasons
- Can be used with CRD, RCBD or Latin Square – it is more of treatment arrangement than design
- Involves split a larger experimental unit onto which the first factor is applied into the smaller units onto which the second factor is applied

Split-Plot “Design”

Block 1

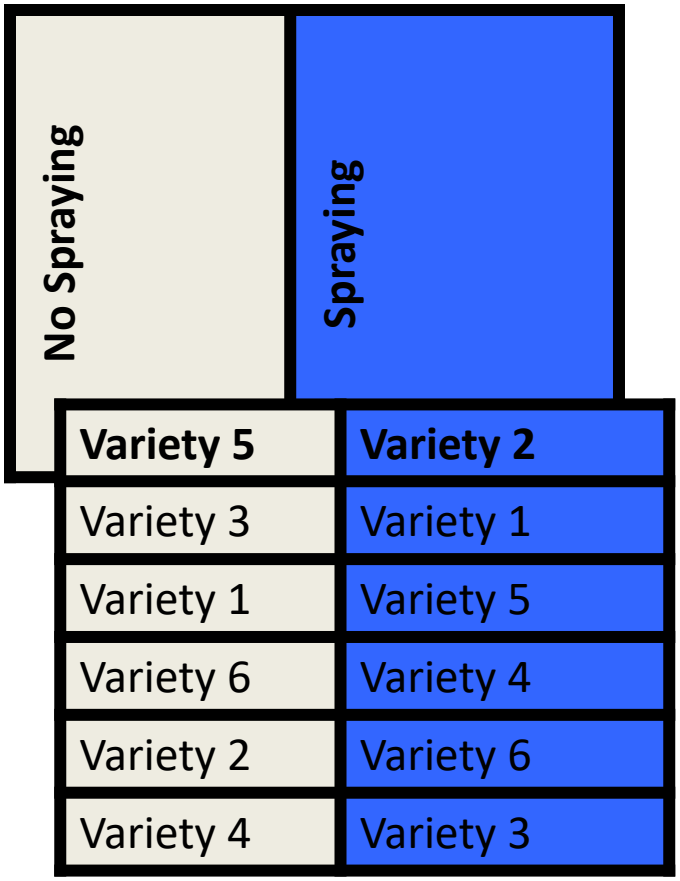


Block 2



Split-Plot “Design”

Block 1



- The split can proceed as long as it is required or feasible
- Randomization for the levels of different factors are done separately
- First randomly assign **spraying** to the main plot and then assign varieties to the sub-plots
- – ***implication is that different experimental errors are used for testing different factors***

Split-Plot Design

```
library(agricolae)

treatment1<-c("Spray", "No Spray")

treatment2<-c("Variety 1","Variety 2", "Variety 3","Variety 4","Variety 5","Variety 6")


design.split(treatment1, treatment2,r=3, design="rcbd",serie = 2,

        seed = 10, kinds = "Super-Duper", first=TRUE,randomization=TRUE)
         #treatment1 = treatment in the plot
         #treatment2 = treatment in the subplot
        #r = replications or block (3)
        #seed = randomization seed for repeat
```

Incomplete Block Designs

Introduction to IBD

- **Importance of blocking**
  - ❖ Decrease experimental error and provide more precise comparison of treatments - increase precision
  - ❖ Be able to make comparisons under more uniform conditions (within blocks)

Introduction to IBD

- **Challenges faced in blocking**
  - ❖ Large number of treatments require larger blocks
  - ❖ As blocks get larger, the conditions become more heterogeneous
    - precision decreases

If We have 9 treatments we may need to borrow the neighbour puppies

Introduction to IBD

- **Challenges faced in blocking**
  - ❖ Blocks based on natural grouping of experimental units may have fewer experimental units compare the number of treatments

Introduction to IBD

- **Challenges faced in blocking**
  - ❖ Sometimes it is therefore necessary to block experimental units into groups smaller than complete replicate of all treatments with RCBD or Latin Square designs
  - ❖ i.e. plots are grouped into blocks that are not large enough to contain all treatments (entries)

Introduction to IBD

Incomplete block designs may be classified into based on

1) Number of blocking factors (One vs two Blocking factors)

2) Precision of comparison of treatments (Balanced vs

Partially Balanced)

Introduction to IBD

Classification based number of blocking factors

- Randomized Incomplete Block Design – one blocking factor
- In complete Latin Square – two blocking factors

Introduction to IBD

Classification based on precision of treatment comparison

- **Balanced Incomplete Block design**: Each pair of treatment occur equal number of times in the same block (eq)
- **Partially balanced Incomplete Block design** – pairs of treatments don't occur in equal frequency

Balanced Incomplete Block Design

- All treatments are equally replicated
- Each pair of treatments in the same block equal number of times
- All pairs are compared with the same precision even though differences between blocks may be large

Balanced Incomplete Block Design

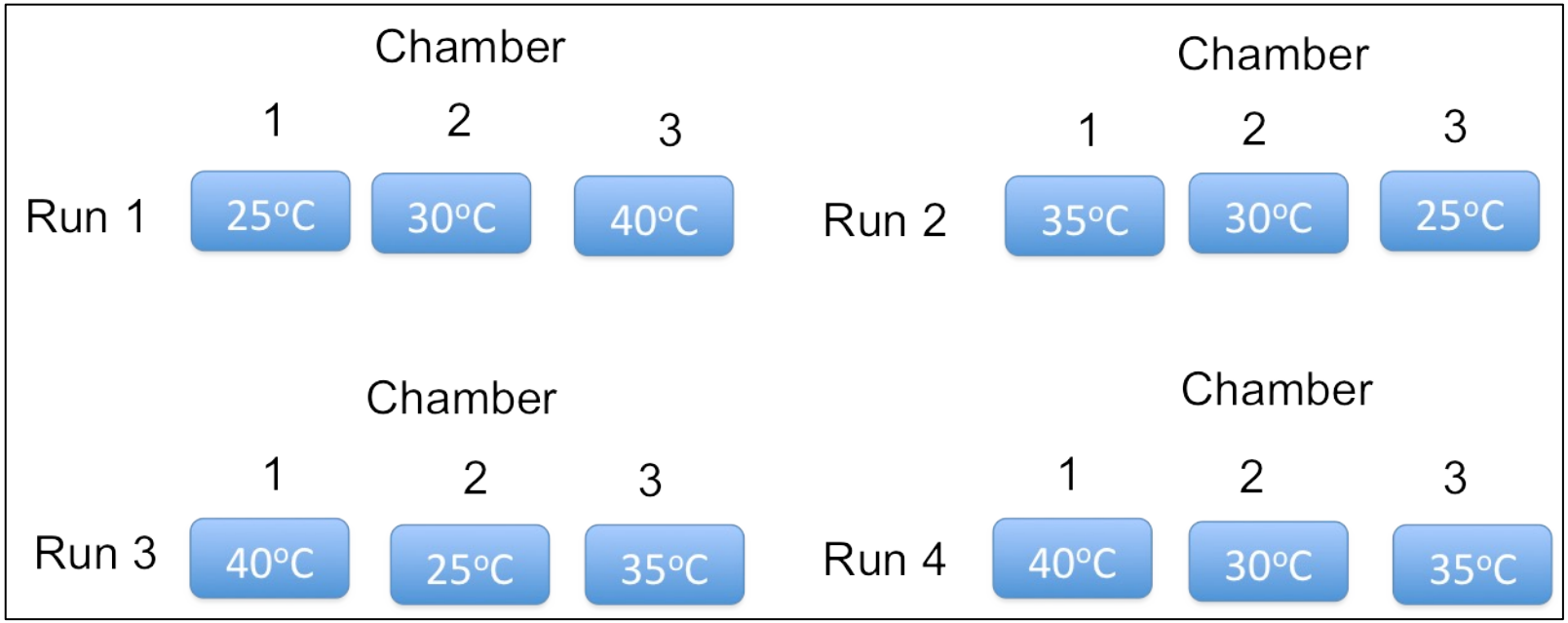
- **Balanced Incomplete Block Designs**
  - ❖ Can balance any number of treatments and any size of block but...**treatments** and **block size** determine the number of replications required for balance
  - ❖ Often the minimum number of replications required for balance is too large to be practical

Balanced Incomplete Block Design

- Example BIB: Tomato Seed Germination Experiment (Kuehl)
  - Effect of temperature inhibition on seed germination
  - **Treatment**: Four temperature (25^{o} C, 30^{o} C, 35^{o} and 40^{o} C). Seed subjected to a constant temperature in a controlled environment chamber
  - **Design**: A single chamber constitute an experimental unit and each run of the experiment constitute a block
  - One complete block/replication would require 4 chambers but only three chambers were available to the scientist **- Incomplete Block**

Balanced Incomplete Block Design

- Example BIB: Tomato Seed Germination Experiment (Kuehl)



Balanced Incomplete Block Design

- Designing BIB using agricolae
- library(agricolae)

  design.bib(trt, k, r=NULL, serie = 2, seed = 0, kinds = "Super-Duper", maxRep=20,randomization=TRUE)

Balanced Incomplete Block Design

#BIB

```
trt<-c("25C", "30C", "35C", "40C")

design.bib(trt, k=3, r=NULL, serie = 2, seed = 30, kinds =     "Super-Duper",

maxRep=20,randomization=TRUE)
```

#trt= number of treatments, k= block size, r= number of rep

(calculate)

Balanced Incomplete Block Design

#BIB

```
trt<-c("25C", "30C", "35C", "40C")
design.bib(trt, k=3, r=NULL, serie = 2, seed = 30, kinds =  "Super-Duper",
maxRep=20,randomization=TRUE)
```

```
$book
   plots block  trt
1    101     1  40C
2    102     1  35C
3    103     1  30C
4    201     2  25C
5    202     2  30C
6    203     2  35C
7    301     3  40C
8    302     3  35C
9    303     3  25C
10   401     4  30C
11   402     4  25C
12   403     4  40C
```

Partially balanced incomplete block designs

- BIB designs can not be constructed for every experimental situation requiring incomplete block
- **Required number of replications may become prohibitive**
- Partially balanced incomplete designs requiring much less replication can be constructed

Partially balanced incomplete block designs

- Different treatment pairs occur in the same blocks an unequal number of times or some treatment pairs never occur together in the same block
- Mean comparisons have differing levels of precision

Partially balanced incomplete block designs

Partially Balanced Incomplete Block Design with six treatments in Blocks of size 4 (Kuel page 323)

Block 1	1	4	2	5
Block 2	2	5	3	6
Block 3	3	6	1	4

- Pairs (1,4), (2,5) and (3,6) occur together in 3 blocks
- All other treatments occur together in only one block
- Different pairs are compared with different levels of precision
- For application ensure that there is reasonable precision for all pair-wise comparison

Row-Column Designs for Two Blocking Criteria

- Latin Square is a complete Block Design used to control variation in two directions
- Latin Square may be impractical because large number of experimental units required $N= t^2$
- There are incomplete block designs derived from Latin Squares
  - Row-Column Designs – either rows or columns or both as incomplete blocks
  - Youden Squares – two or more rows omitted from the Latin Square

Row-Column Designs for Two Blocking Criteria

- The designs are arranged in p rows and q column of experimental units

4 X7 Row-Column Balanced Incomplete Block Design (Kuel page 321)

	Automobiles						
Position	(1)	(2)	(3)	(4)	(5)	(6)	(7)
(1)	3	4	5	6	7	1	2
(2)	5	6	7	1	2	3	4
(3)	6	7	1	2	3	4	5
(4)	7	1	2	3	4	5	6

Row-Column design with seven treatments evaluated on the four car tire positions (Row) and 7 automobiles (Column).

Position – Complete block (all seven treatment)

Automobile – Incomplete Block (only take 4 tires – test only 4 out of 7)

Row-Column Designs for Two Blocking Criteria

Youden Squares

- Are incomplete Latin Squares with two or more rows from the Latin squares omitted.

RESOLVABLE and CYCLIC INCOMPLETE BLOCK DESIGNS

Resolvable incomplete block designs

- **Resolvable designs** - blocks are grouped so that each group of blocks constitute one complete replication of the treatments
  - Useful for trial management - replication-by-replication basis
  - Field operations can be conducted in stages (planting, weeding, data collection, harvest)
  - Complete replicates can be lost without losing the whole experiment
  - If you have two or more complete replications, you can analyze as a RCBD if the blocking turns out to be ineffective

Resolvable incomplete block designs

- Resolvable designs are arranged in ***r*** replicate groups of ***s*** blocks with ***k*** units per blocks
- The number of treatments is a multiple of the number of unit per block $t=sk$
- The total number of blocks satisfies the relationship

$$b = rs \geq t + r - 1$$

for balanced resolvable incomplete block designs

Alpha designs

- Resolvable designs developed by Patterson and Williams, 1976
  - Number of varieties and replications were fixed by statutory requirements
  - Large number of varieties necessitate incomplete block designs
  - Resolvable design were required for proper management in the field
  - Existing resolvable designs did not always accommodate the number varieties and block sizes required by the trials

Alpha designs

- Alpha designs have no limitation on block size
- except for constraint that the number of treatments t is a multiple of block size k, so that t=sk to have resolvable design with equal block sizes
  - Alpha designs are available for many (*r,k,s*) combinations where *r* is the number of replicates, *k* is the block size and *s* is the number of blocks per replicate (the number of treatments *t*=*ks*).

Alpha designs

- Alpha designs have no limitation on block size
  - Efficient alpha designs exist for some combinations for which conventional lattices do not exist. Can also accommodate unequal block sizes.

Alpha designs

```
#alpha lattice design
        trt<-1:100
        t<-length(trt)
        k<-5 #number of blocks
        s<-t/k  # number of blocks that constitute a complete replicate
        r<-2 #number of replications
outdesign<-design.alpha(trt, k=5, r=2, serie = 2, seed = 0, kinds = "Super-Duper",randomization=TRUE)
class(outdesign) #finding out the type of object
str(outdesign) #finding out the structure of the object outdesign
book<-outdesign$book # picking one element from our list called book
plots<-book[,1] #creating an object called plot showing experimental units/plots
dim(plots)<-c(k,s,r) #showing plot labels with respect to k, s, r
for (i in 1:r) print(t(plots[,,i]))
outdesign$sketch #show a field sketch
```