EXPERIMENTAL DESIGNS

Statement

- The experimental designs are planned experiment aimed at obtaining statistically valid results by planning or controlling a number of possible sources of variation from the beginning of an experiment
- They differ from statistical surveys or the analysis of field data where different source of variations cannot be planned before experiment
- They are on use in experiment whose purpose is to compare the effects of different treatments on the experimental units

Precision and Accuracy of an experiment

- The precision is the possibility of obtaining "repeatable" results, i.e., not too dissimilar results between two experiments with the same objective. Precision is linked to the use of experimental units that are representative of a population (sampling error).
- Accuracy is the correlation of experimental observations with the truth. It is obtained by applying the experimental techniques always in the same way (detection errors).

Experimental size

- Which is the minimum no. of replicate to guarantee statistically sound results (or a highly precise experiment)?
- A rule: $n \geq 2\, s^2 (t_{\alpha;\nu}/d_o)^2$
  where:
  s^2: Population Variance of an experimental variable (known or supposed)
  $t_{\alpha;\nu}$: Student t value for a given probability level α and a specific no. of degrees of freedom ν, where $\nu = 2\,(n-1)$
  d_o: desired minimum significant difference between two means

WARNING:
Because n is dependant on ν and vice versa, the formula is iteratively applied up to convergence.

Example

- $S^2=1.5954$
- $d_0=1$
- $\alpha=0.05$

Suppose at first step n=40, i.e.,

- $\nu=2(40-1)=78$
- $t_{0.05;78}=1.984$

$n=[2(1.5954)(1.984)^2]/1^2=12.6$

Considering now n=13, that gives $\nu=24$ & $t_{0.05;24}=2.064$

$n=[2(1.5954)(2.064)^2]/1^2=13.6$

With n=14, that gives $\nu=26$ & $t_{0.05;26}=2.056$

$n=[2(1.5954)(2.056)^2]/1^2=13.5$ (Convergence reached)

Each sample should contain at least 13 replicates

List of experimental designs

1. Completely Randomized Design (CRD)
2. Randomized Block Design
3. Latin Square Design
4. Factorial Design
5. Nested design
6. Split-Plot design (analysis of longitudinal data)

Completely Randomized Design

- Rather than an experimental design is a form of random allocation of the compared treatments to the experimental units.
- For example, an experiment involving one-two or more way ANOVA including interaction (factorial design) can be a completely randomized experiment.
- Example: to 9 Groups of piglets (6 piglets each) are attributed 3 level of Vit B12 in the diet (0, 5 or 10 mg) randomly assigning treatment to the groups

Randomized Block Design

- The randomized block design means to distribute the experimental units within GROUPS or BLOCKS that are similar or homogeneous. E.g., Land parcels physically arranged on the basis of soil characteristics, animals raised in different cages, etc.
- A BLOCK factor is therefore a source of variability which is not of primary interest to the experiment, but is a source to be controlled to obtain a greater statistical accuracy in comparing the different treatments within BLOCK.
- Usually, each block contains a number of experimental units in which all treatments are represented.
- The experimental structure is balanced and includes two factors of variability: one due to the treatments (main factor) and a secondary effects due to the group or block (secondary factor)

Rules in the Randomized Block Design

La general rule is:

- "Block what you can, randomize what you cannot."
- Blocking is used to remove the effects of a few of the most important nuisance variables.
- Randomization is then used to reduce the contaminating effects of the remaining nuisance variables.

Example of Randomized Block Design

4 Varieties of Wheat distributed in 5 blocks dependent from the soil fertility

Fertility – **Fertility -**

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

Fertility + **Fertility +**

Blocks

That reminded:

Total Sum of Squares

$$\sum_{i=1}^{I}\sum_{j=1}^{J}\left(y_{ij}-\bar{y}\right)^2=\sum_{i=1}^{I}\sum_{j=1}^{J}y_{ij}^2-\frac{\left(\sum_{i=1}^{I}\sum_{j=1}^{J}y_{ij}\right)^2}{n}=\sum_{i=1}^{I}\sum_{j=1}^{J}y_{ij}^2-\frac{\bar{y}^2}{n}$$

Between Sum of Squares

$$\sum_{i=1}^{I}\left(y_{i.}-\bar{y}\right)^2=\frac{\sum_{i=1}^{I}y_{i.}^2}{I}-\frac{\left(\sum_{i=1}^{I}\sum_{j=1}^{J}y_{ij}\right)^2}{n}=\frac{\sum_{i=1}^{I}y_{i.}^2}{I}-\frac{\bar{y}^2}{n}$$

The Sum of Squares in the ANOVA of a Randomized Block Design will be:

Model: $y_{ij} = \mu + B_i + T_j + e_{ij}$

Block i=1,…,I Treatment j=1,…,J Total N=I*J

$$Y_{i.} = \sum_{j=1}^{J} y_{ij} \qquad Y_{.j} = \sum_{i=1}^{I} y_{ij} \qquad Y_{..} = \sum_{i=1}^{I}\sum_{j=1}^{J} y_{ij}$$

Correction factor: $FC = \dfrac{Y_{..}^2}{N}$

Block (SSB): $\dfrac{\sum_{i=1}^{I} Y_{i.}^2}{J} - FC$

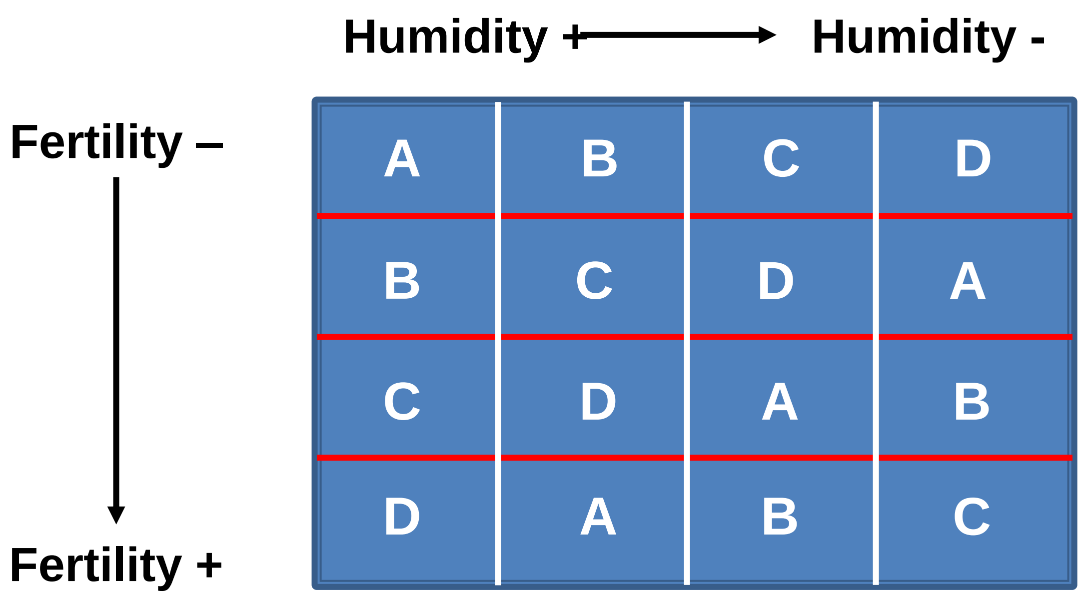
Treatment (SST): $\dfrac{\sum_{j=1}^{J} Y_{.j}^2}{I} - FC$

Latin Square Design

- The Latin Square Design is a special case of "Randomized Block Design" in which the treatments are assigned to experimental units according to a specific pattern following a square matrix;
- In this way each treatment (main factor) is present in each row (secondary factor 1) and in each column (secondary factor 2 or block) only once;
- The number of factors (main and secondary) must be at least 3;
- All factors must have the same number of levels in order to built a square matrix; therefore the number of levels must therefore at least 3.

Why the Latin Square Design

Because sometimes the random distribution of treatments within blocks cannot control adequately other nuisance factors, inflating the residual variance and the statistical test



3x3 Latin Square Design in experiment with animals

Period	Animal		
	1	2	3
I	A	B	C
II	B	C	A
III	C	A	B

3 diets (A,B, & C) given to 3 animals (or groups) in 3 subsequent periods

Features of the Latin Square Design

- A Latin Square Design requires less statistical units than a other experimental design with completely crossed factors to be tested, and it's therefore less expensive.
- Latin Square Designs are more efficient than designs that account for a secondary factor only.
- When a block factor considered is the animal, the Latin Square Design allow the control of the variability due to the animal.
- Interactions between factors are not estimable

Latin Square Design ANOVA

Model: $y_{ij} = \mu + \alpha_t + \beta_i + \gamma_j + \varepsilon_{ij}$

I = 1,..,p j = 1,...,p t = 1,...,p

Correction Factor: $(\Sigma_i\Sigma_j y_{ij})^2/p^2$

Total Sum of Squares: SSY = $\Sigma_i\Sigma_j y^2_{ij}$ - CF

Treatment SS (SST = i,j,t): SST = $\Sigma_t(\Sigma_j y_{ij})^2/p$ – CF

Between rows SS SSR = $\Sigma_i(\Sigma_j y_{ij})^2/p$ – CF

Between Columns SS SSC = $\Sigma_j(\Sigma_j y_{ij})^2/p$ – CF

Error SS: SSE = SSY - SST - SSR - SSC

Factorial Design

- The peculiar aspects of factorial design is the possibility of analyzing not only one ore more factors alone, but also their combination (or interaction)
- Factorial designs allow the analysis of all possible combinations of studied factors

Factorial Design ANOVA

Two factors and their interaction

$$y_{ijk} = \mu + \alpha_i + \beta_j + (\alpha\beta)_{ij} + \varepsilon_{ijk}$$

i = 1, ..., I
j = 1, ..., J
k = 1, ..., K
n = I*J*K

NOTICE THAT:
If there is no interaction between the two factors, the two factors are independent.
If the experimental response of a factor is different within different levels of the other factor, it means that there is interaction and the interaction is significant at the ANOVA.

ANOVA in Factorial Design with 2 factors

Source of variations	Sum of Squares (SS)	d.f.	Mean Squares (MS)	F
Factor A	SSA	I-1	SSA/(I-1)	MSA/MSE
Factor B	SSB	J-1	SSB/(J-1)	MSB/MSE
Interaction (AxB)	SSAB	(I-1)(J-1)	SSAB/(I-1)(J-1)	MSAB/MSE
Error	SSE	IJ(K-1)	SSE/IJ(K-1)	
Total	SSY	n-1		

…..3 Factors Factorial Design

All possible interaction tested:

$$y_{ijkl}=\mu+\alpha_i+\beta_j+\gamma_k+(\alpha\beta)_{ij}+(\alpha\gamma)_{ik}+(\beta\gamma)_{jk}+(\alpha\beta\gamma)_{ijk} + \varepsilon_{ijkl}$$

i = 1, …, I
j = 1, …, J
k = 1, …, K
l = 1,…..,l
n = I*J*K*l

Nested Design

- Nested design are models in which 2 or more factors are analysed, but the levels of a first factor (hierarchically upper factor) are strictly associated to the level of a second factor.
- In this case the second factor is considered hierarchically nested, or subordinated or simply nested into the first factor.

Example of Nested Design

	Sow					
Boar	A	B	C	D	E	F
1	X	X				
2			X	X		
3					X	X

Each level of «Boar» effect (hierarchically upper effect) there are different and unique levels of «sow» (nested effect).

In each combination k repeats are recorded (k piglets, for example).

Basis of ANOVA for Nested Design

- The ANOVA of Nested Design is slightly different form other experimental designs, because of the hierarchy.
- Therefore the mean square of the hierarchically upper factor is tested on the mean square of the nested factor, that is the error line for the first factor.
- This because of the variability of the nested factor is also accounted in the hierarchically upper factor (i.e., the sow variability is include into the boar variability in the previous example).
- Therefore, to estimate a correct F for the ANOVA test, the correct ratio between MS need to be produced.

Model for Nested Design

$$y_{ijk} = \mu + A_i + B_{ij} + e_{ijk}$$

$i = 1,\ldots,I$ $j = 1,\ldots,J$ $k = 1,\ldots,K$ $n = I*J*K$

Sums : $Y_{ij.} = \Sigma_k y_{ijk}$ $Yi.. = \Sigma_j \Sigma_k y_{ijk}$ $Y\ldots = \Sigma_i \Sigma_j \Sigma_k y_{ijk}$

Correction factor : $Y^2_{\ldots}/N$

Total Sum of Squares, SST: $\Sigma_i \Sigma_j \Sigma_k y^2_{ijk} - CF$

Sum of Square for B, SSB: $\Sigma_i \Sigma_j Y^2_{ij.}/K - CF$

Sum of Square for A, SSA: $\Sigma_i \Sigma_j Y^2_{i..}/JK - CF$

Sum of Squares B(A), SSB(A)= SSB – SSA

SST = SSA + SSB(A) + SSE

ANOVA for Nested Designs

Source of variation	Sum of Squares	df	Mean Square	F
Main factor	SSA	I-1	SSA/(I-1)	MSA/MSB(A)
Nested Factor	SSB(A)	I(J-1)	SSB(A)/I(J-1)	MSB(A)/MSE
Errore	SSE	IK	SSE/IK	
Total	SSY	n-1		

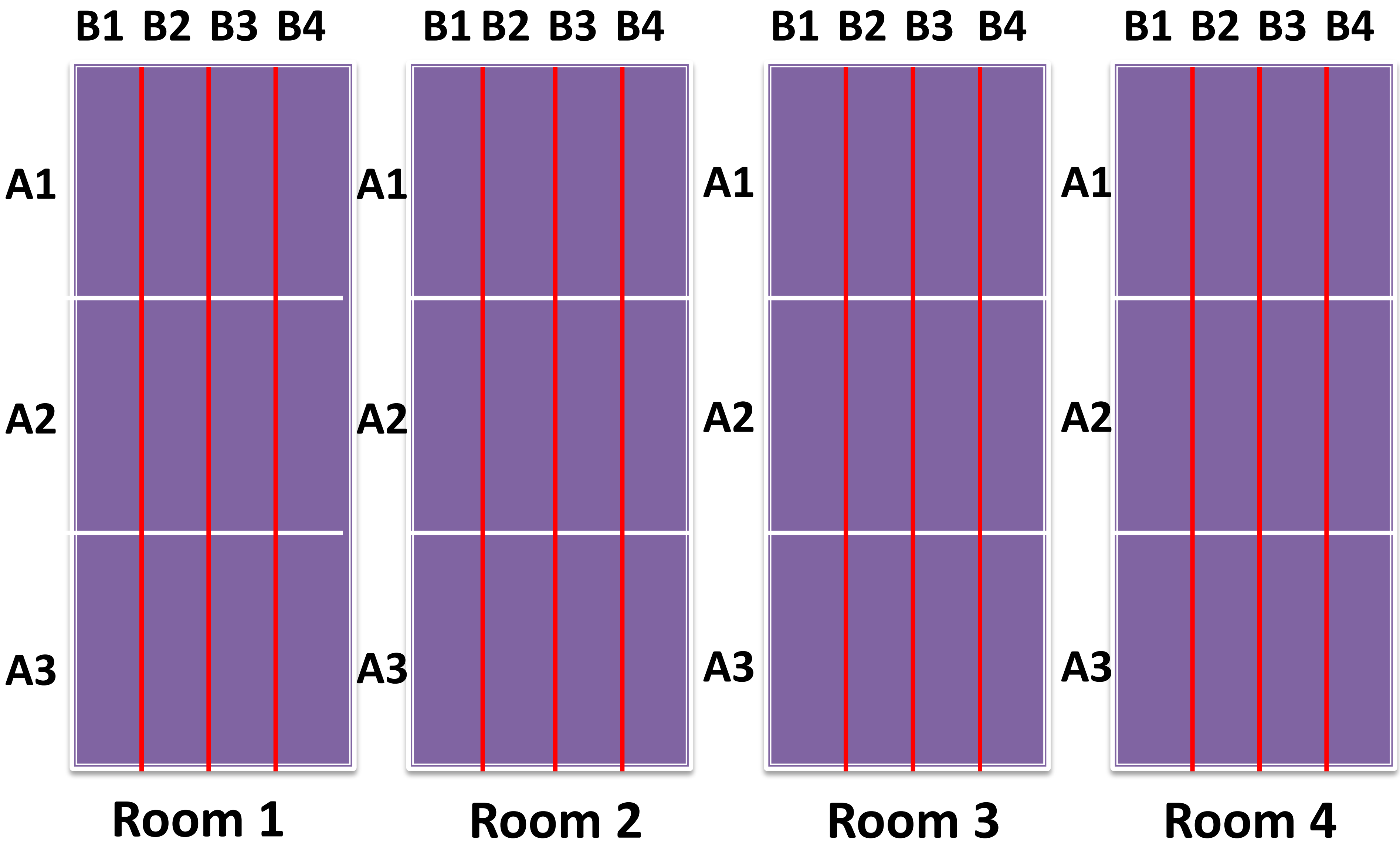
SPLIT-PLOT Designs

- Experimental designs that allow the reduction of blocks by attributing 2 (or more) plots containing each one an experimental treatment (i.e., main and sub-plot or sub-sub-plot in case of a SPLIT-SPLIT-PLOT) and a specific error.
- The experimental units that receive a given level of the first treatment are divided into sub-units receiving all different levels of the second treatment.

SPLIT-PLOT Example

- Average daily gain of broiler chicken fed different diets containing 3 energy levels and 4 protein levels
- Treatment factors:
  - Factor A: Different Energy level in the diet (A_1 A_2 A_3)
  - Factor B: Different Protein level in the diet (B_1 B_2 B_3 B_4)
- One block factor: room
- Each "main plot" (levels of factor A) is divided ("split") in 4 sub-plots containing the different protein level in the diet (levels of factor B).
- In each combination EnergyxProtein contains 4 broiled chickens of the same genetic type.

SPLIT-PLOT



B1 B2 B3 B4
B1 B2 B3 B4
B1 B2 B3 B4
B1 B2 B3 B4
A1
A2
A3
A1
A2
A3
A1
A2
A3
A1
A2
A3
Room 1
Room 2
Room 3
Room 4

SPLIT-PLOT Linear Model

$$y_{ijk} = \mu + \rho_i + \alpha_j + \gamma_{ij} + \beta_k + (\alpha\beta)_{jk} + \varepsilon_{ijk}$$

$i = 1,\ldots,I$ $j = 1,\ldots,J$ $k = 1,\ldots,K$ $N = I*J*K$

ρ_i: Block effect (Room)

α_j: Enery level effect (factor A)

γ_{ij}: Interaction Block*Factor A or main plot error; $\sim N(0, \sigma^2_{ba})$

β_k: Protein level effect (factor B)

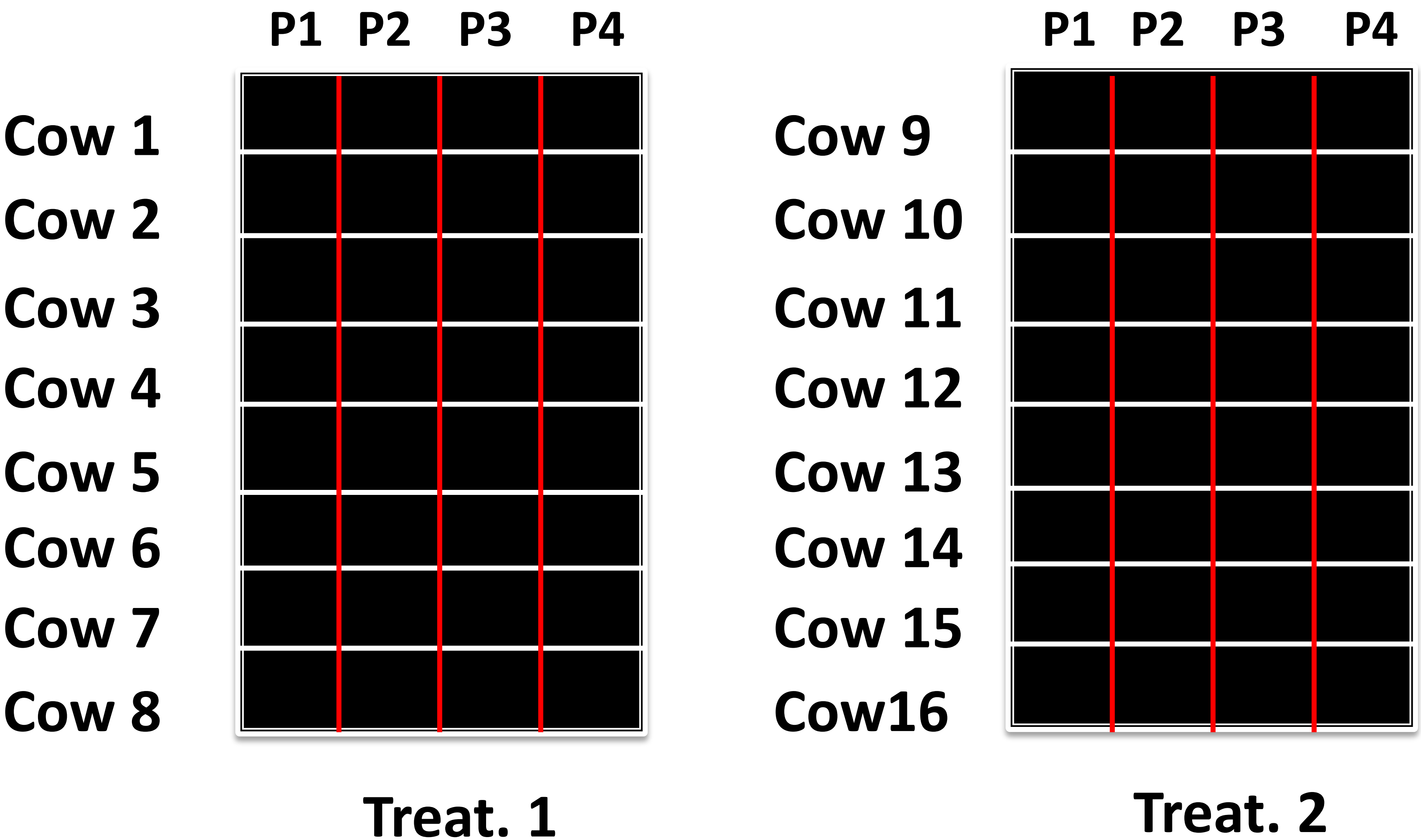
$\alpha\beta_{jk}$: Interaction between factors or treatments (A*B)

ε_{ijk}: sub-plot error $\sim N(0, \sigma^2_e)$

SPLIT –PLOT ANOVA

Source of variation	SS	df	MS	F
Block	SSR	I-1	MSR	MSR/MSE1
Factor A	SSA	J-1	MSA	MSA/MSE1
Error (1st plot)	SSE1	(I-1)(J-1)	MSE(1)	
Factor B	SSB	K-1	MSB	MSB/MSE
Interaction AB	SSAB	(J-1)(K-1)	MSAB	MSAB/MSE
Error (2nd plot)	SSE	J(I-1)(K-1)	MSE	

SPLIT-PLOT: longitudinal data



Spit plot with one Animal effect, i.e., Analysis pf longitudinal data

$$y_{ijk} = \mu + \rho_i + \alpha_{ij} + \beta_k + (\rho\beta)_{jk} + \varepsilon_{ijk}$$

$i = 1,\ldots,I$ $j = 1,\ldots,J$ $k = 1,\ldots,K$ $N = I*J*K$

ρ_i: Treatment effect

α_{ij}: Interaction Animal(treatment): residual or error for the main plot $\sim N(0, \sigma^2_v)$: RANDOM EFFECT

β_k: Time effect

$\rho\beta_{ik}$: Interaction Treatment x Time

ε_{ijk}: residual or error of the sub-plot $\sim N(0, \sigma^2_e)$