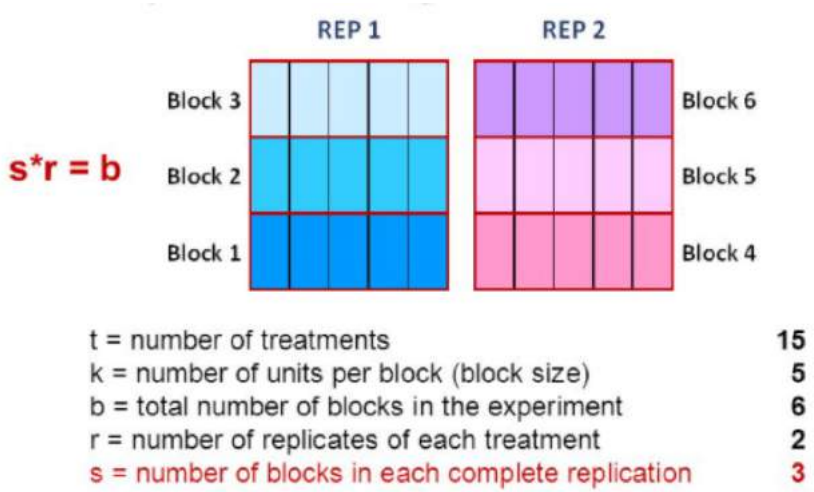
AGR 6322 Advanced Plant Breeding

Fall 2018

Experimental Designs in Plant Breeding



Goals for today

Review of Exp. Designs used in Plant Breeding

Randomization of Experiments

Complete Randomized Design (CRD): Randomize all repetitions (r) of the treatments (t) to all experimental units in a single block

Randomized Complete Block Design (RCBD): Randomize all treatments (t) to the experimental units of a single block (b). Repeat for each block

Split-plot Arrangement (SD): Randomize the whole plot treatments (WPT) as in RCBD, then randomize the subplot (or split-plot) treatment into each WPT independently.

Complete Randomized Design (CRD) - ANOVA

Source of Variation	Degree of Freedom	Sum of Squares	Mean Squares	F-Stat or F-Ratio
Treatment (T)	t – 1	SS(Treat)	$\frac{SS(Treat)}{t-1}$	$\frac{MS(Treat)}{MS(Res)}$
Error or Residual	N – t	SS(Res)	$\frac{SS(Res)}{N-t}$	
Total	N-1			

Treatment effect is tested against the Residual Mean Square "MS(Res)"

Model: $y = \mu + t + e$

Assumptions:

Independent errors, Normality and Homogeneity of Variances

u and *t* are fixed effects

e is the random residual effect $e \sim N(0, \sigma^2)$

Randomized Complete Block Design (RCBD) – ANOVA

Source of Variation	d.f	Sum of Squares	Mean Squares	F-Stat or F-Ratio
Block (b)	b – 1	SS(block)	$\frac{SS(block)}{b-1}$	$\frac{MS(block)}{MS(Res)}$
Treatment (t)	t – 1	SS(Treat)	$\frac{SS(Treat)}{t-1}$	$\frac{MS(Treat)}{MS(Res)}$
Error or Residual	(b–1)x(t–1)	SS(Res)	$\frac{SS(Res)}{(b-1)x(t-1)}$	
Total	N-1			

Block and Treatment effects are tested against the Residual Mean Square "MS(Res)"

Model: $y = \mu + b + t + e$

Assumptions:

Independent errors, Normality and Homogeneity of Variances

u, b and *t* are fixed effects

e is the random residual effect $e \sim N(0, \sigma^2)$

Factorial in a CRD - ANOVA

Source of Variation	Degree of Freedom	Sum of Squares	Mean Squares	F-Stat or F-Ratio
Factor A	a – 1	SS(A)	$\frac{SS(A)}{a-1}$	$\frac{MS(A)}{MS(Res)}$
Factor P	p – 1	SS(P)	$\frac{SS(P)}{p-1}$	$\frac{MS(P)}{MS(Res)}$
Factor AxP	(a-1)(p – 1)	SS(AxP)	$\frac{SS(AxP)}{(a-1)(p-1)}$	$\frac{MS(AxP)}{MS(Res)}$
Error or Residual	N – t	SS(Res)	$\frac{SS(Res)}{N-t}$	
Total	N-1			

All treatment effects are tested against the Residual Mean Square "MS(Res)"

Model: $y = \mu + a + p + a * p + e$

Assumptions:

Independent errors, Normality and Homogeneity of Variances

u, a, p and *a*p* are fixed effects

e is the random residual effect $e \sim N(0, \sigma^2)$

Split-plot Design with Blocks- ANOVA

Source of Variation	Degree of Freedom	Sum of Squares	Mean Squares	F-Stat or F-Ratio
Block (B)	b – 1	SS(Block)	$MS(B)= \frac{SS(Block)}{b-1}$	$\frac{MS(B)}{MS(RWP)}$
Factor A (Whole Plot Factor)	a – 1	SS(A)	$MS(A)= \frac{SS(A)}{a-1}$	$\frac{MS(A)}{MS(RWP)}$
Whole Plot Res (BxA)	(b-1)(a- 1)	SS(BxA)	$MS(RWP)= \frac{SS(BxA)}{(b-1)(a-1)}$	$\frac{MS(RWP)}{MS(RSP)}$
Factor P (Split-plot Factor)	p – 1	SS(P)	$MS(P)= \frac{SS(P)}{p-1}$	$\frac{MS(P)}{MS(RSP)}$
Factor AxP	(a-1)(p – 1)	SS(AxP)	$MS(AxP)= \frac{SS(AxP)}{(a-1)(p-1)}$	$\frac{MS(AxP)}{MS(RSP)}$
Error or Residual	a(p-1)(b-1)	SS(RSP)	$\frac{SS(RSP)}{N-t}$	
Total	N-1			

Factor A tested against Whole Plot Residual (WPR=BxA)

Factor P tested against Split Plot Residual (SPR)

Model: $y = \mu + b + a + a*b + p + a*p + e$

More “complex designs”

We can have a big improvement in the precision of an experiment by shifting from CRD to RCBD.

However, this benefit will only be clear if large pieces of land (or experimental area) have the similar environmental conditions. So we can control them by blocking.

This also works well when we have **one** known source of variation (gradient). Example slope

What happens when there is more than one source of variation?

More “complex designs”

What happens when the variation is in spots across the experimental area. Example soil texture variation

Or when you have too many treatments levels. Example breeding in early generations

There are more experimental designs that you can use, some of them are:

- Latin Square Design
- Incomplete Block Design
- Lattice designs

Rule 1 (when is a factor random?)

- A factor is random
  - when the observed levels are a random sample from a defined population or
  - when it represents a randomization unit (error stratum)
- Otherwise a factor is usually considered as fixed
- When levels of a factor are to be compared (e.g. treatments), the factor is fixed, independently of whether or not it is random by design
- When a factor is random, all effects containing it are random

Example

- Series of experiments with selected set of plant varieties
- Experiments conducted at a sample of sites selected at random from target region
- Want to estimate variety means in target region

$\Rightarrow$ factors: variety and site

$\Rightarrow$ varieties fixed

$\Rightarrow$ sites random

Block factors:

- Randomly selected sampling units (plants, soil specimens, etc.)
- Randomisation units (rows, columns, incomplete blocks, main plots, sub plots, etc.)
- Block units, which are not themselves involved in randomization process (complete blocks, environments, etc.).

Block factors are *innate* to observational units

Treatment factors:

- Selected by experimenter to answer research question
- Levels of treatment factor are randomly allocated to observational units by a defined randomization procedure

Treatments are not *innate* to observational units

Rule 3 (Keep treatment model and block model separate)

- When setting up a full model, it is useful to (at least initially) keep treatment model and block model separete
- The treatment model can be formulated with treatment factors alone
- The block model can be formulated with block factors alone (i.e. without treatment factors!)

Rule 4 (effects of block model)

Random effects for randomization units:

- Each randomization unit (every error stratum) has its own effect
- An experimental unit or block unit becomes a randomization unit when levels of a treatment factor are randomly allocated to it
- Crossing of randomization units (error strata) produces further error effects

Fixed effects for block factors, which are not themselves part of the randomization or sampling process

Rule 5 (coding of block factors)

- Every block factor (design factor) is represented by a separate variable
- Sometimes can replace block factor by treatment factor in an effect, but this is not a necessity!
- For clarity it is better to avoid coding of a block factor by a treatment factor

Treatment model:

GEN

Full model

GEN + REP : REP•BLOCK + REP•BLOCK•PLOT

Rule 6 (Interaction between block and treatment factors)

- Usually assume block-treatment additivity (no interaction)
- Check if interaction is to be expected
- Interaction likely when levels of a block factor are very diverse

Latin Square Design

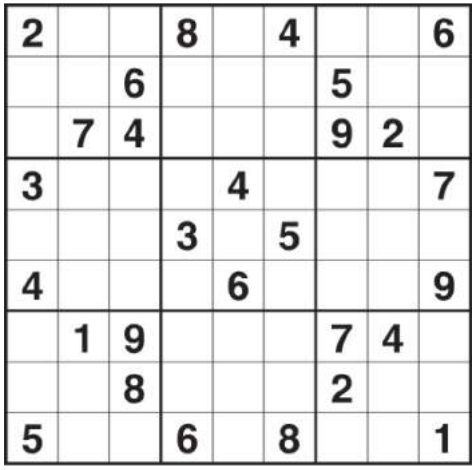
Design that use double blocking to control two gradients simultaneously.

Each treatment occurs once in each row and once in each column. (Randomization constrain)

Randomization: first rows, then columns or first columns, then rows

Example:

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



Latin Square Design

rows = # columns = number of treatments = t

Thus the number of replicates is t and total number of plots is $n=t^2$

Source of Variation	d.f	Sum of Squares	Mean Squares	F-Ratio
Rows (R)	r – 1 or t – 1	SS(rows)	$\frac{SS(rows)}{t-1}$	$\frac{MS(rows)}{MS(Res)}$
Columns (C)	c – 1 or t – 1	SS(cols)	$\frac{SS(cols)}{t-1}$	$\frac{MS(cols)}{MS(Res)}$
Treatment (T)	t – 1	SS(Treat)	$\frac{SS(Treat)}{t-1}$	$\frac{MS(Treat)}{MS(Res)}$
Error or Residual	(t-1) (t-2)	SS(Res)	$\frac{SS(Res)}{(t-1)x(t-2)}$	
Total	N-1			

Model: $y = \mu + r + c + t + e$

Problems and Extensions

Balanced Incomplete Block Design

When the number of treatment is too large or the size of the homogeneous block is smaller than the area needed to fit all treatments.

N= Total number of experimental units (N=txn = mxu)

t = number of treatments

n = number of replicates of each treatment

m = number of incomplete blocks

u = number of experimental units per block [u<t)

λ = number of times each treatment pair occurs together within a block

All treatments occurs together the same number of times (λ) “balanced”. Ensuring all treatments are compared with the same precision.

Balanced Incomplete Block Design

N= Total number of experimental units (N=txn = mxu)

t = number of treatments

n = number of replicates of each treatment

m = number of incomplete blocks

u = number of experimental units per block [u<t)

λ = number of times each treatment pair occurs together within a block

All treatments occurs together the same number of times (λ) "balanced". Ensuring all treatments are compared with the same precision.

Conditions for BIBD $t \times n = m \times u$ and $n(u - 1) = (t - 1) \times \lambda$

Balanced Incomplete Block Design

N= Total number of experimental units (N=txn = mxu)

t = number of treatments

n = number of replicates of each treatment

m = number of incomplete blocks

u = number of experimental units per block [u<t)

λ = number of times each treatment pair occurs together within a block

Conditions for BIBD $t \times n = m \times u$ and $n(u - 1) = (t - 1) \times \lambda$

We can identify the number of Inc. blocks as:

$$m = (n \times t)/u$$

Then we can find out how many times each treatment pair occurs together:

$$\lambda = n(u - 1)/(t - 1)$$

λ must be positive **integer**

Balanced Incomplete Block Design

Example: Seven treatments (t=7), using blocks of size 3 (in A) or 4 (in B), with up to 4 replicates of each treatment (n≤4).

Scenarios:

	A			B			
	Unit1	Unit2	Unit 3	Unit 1	Unit2	Unit3	Unit4
Block1	1	5	7	4	1	2	7
Block2	6	4	7	7	6	5	2
Block3	2	5	4	3	6	2	4
Block4	2	7	3	3	4	5	7
Block5	4	3	1	6	5	1	4
Block6	5	3	6	6	3	7	1
Block7	2	6	1	3	5	1	2

Randomize with "*crossdes" or "ibd*" R packages for example

Augmented Design

- Objective: to compare existing treatment to new treatments, when constraints prevent replication of the new treatments
  - Existing treatment = control = “checks” = **used for calculating the error**
  - **You need enough number and replicates of checks to provide good error estimates**
  - New treatments = replication constrained by cost or number of available units

- Examples:
  - Limited number of seeds in early evaluation trials of new varieties/crosses
  - Large number of fertilizer types being evaluated

Augmented Design

- Checks (control): c
- Blocks: r
- New treatments (entries, for example): n

- Error: minimum of 10
  - (c-1)(r-1)
- How many plots?
  - (c*r) + n

- Important: no need to have the same number of plots per block, but the more equal the better

Augmented Design

- Example: I have 48 new entries (new treatment) I would like to test against my checks (controls). I will use 4 checks in 5 blocks.
- Checks (control): 4
- Blocks: 5
- New treatments (example): 48

- Error: (4-1)(5-1) = 3*4 = 12
- How many plots? (4*5) + 48 = 20 + 48 = 68
- How many plots per block? 68/5 = 13.6 ~ 14
  - 4 blocks with 14 plots; last block with 12 plots (4 checks and 8 new varieties)
- Randomization- randomly assign the checks in each block; then randomly assign new entries to plots.

Augmented Design

Block 1	A	6	32	1	D	3	38	25	26	9	B	10	12	C
Block 2	48	20	5	29	D	34	A	C	40	4	B	46	14	35
Block 3	C	15	11	31	24	B	28	D	18	A	43	47	41	27
Block 4	42	2	44	8	A	B	19	33	13	D	17	37	C	30
Block 5	7	36	16	B	23	A	45	21	39	D	C	22		

Augmented Design

$$Y_{ij} = \mu + \beta_i + c + \tau_{k(i)} + \varepsilon_{ij}$$

Mean + blocks + checks + new entries + error

- New entries could be fixed or random; typically considered random
- Need to adjust means for block effects
- Run ANOVA with checks as a RCBD
- Calculate SE and use it for LSD

Lattice Design

- Incomplete block design
- Used for large number of treatments (to control for the variation within block)
- Number of plots per block is smaller than the total number of treatments
- Number of treatments, t, is a perfect square, $t=k^2$
- Block size, k , is the square root of t, or $k = \sqrt{t}$
- Each rep has a complete set of treatments (balanced lattice)
- Number of reps, r=k+1
- Each rep has k blocks, each containing k treatments
- Number of blocks, b = k*r = k(k+1)
- Total number of observations = t*r = k*b
- Every pair of treatments occurs together in the same block exactly once $\lambda = r(k-1)/(t-1) = 1$
- Degree of precision for comparing means is the same for all pairs

Barreto, H. J., et al. "The alpha lattice design in plant breeding and agronomy: Generation and analysis." *Drought-and Low N-Tolerant Maize*(1996): 544.
http://www2.hawaii.edu/~halina/603/IncBlock.pdf

Lattice Design

1	2	3	4	5	6	1	2	3	4	5	6	1	2	3	4	5	6
11	21	23	13	17	6	8	24	12	5	2	19	11	2	17	12	21	3
4	10	14	3	15	12	20	15	11	9	18	7	1	15	18	13	22	5
5	20	16	19	7	24	14	3	21	10	13	6	14	9	4	10	16	20
22	2	18	8	1	9	4	23	17	1	22	16	19	8	6	23	24	7

plot

incomplete block

Block model:

REP/BLOCK/PLOT = REP : REP•BLOCK + REP•BLOCK•PLOT

- Piepho, H. P., A. Büchse, and B. Truberg. "On the use of multiple lattice designs and α-designs in plant breeding trials." *Plant breeding* 125.5 (2006): 523-528.

Lattice Design

	Lattice I					Lattice II					Lattice III					Lattice IV					Lattice V				
Replicate 1																									
Block	1	2	3	4	5	1	2	3	4	5	1	2	3	4	5	1	2	3	4	5	1	2	3	4	5
	5	14	17	4	13	C3	29	C2	26	21	49	50	51	60	58	61	79	75	77	C1	95	C2	83	90	96
	C4	8	9	20	18	38	C1	23	32	C5	41	C4	47	48	44	66	80	67	76	70	88	81	100	97	82
	16	2	C2	10	11	27	37	30	24	34	57	53	C5	C2	46	64	C2	73	C4	71	91	92	C4	94	C1
	6	19	15	C3	C5	28	25	31	40	33	C1	42	45	53	54	C3	69	C5	62	78	C5	93	85	84	98
	3	C1	1	12	7	35	36	39	C4	32	52	56	59	55	C3	72	68	63	74	65	89	87	99	C3	86
Replicate 2																									
Block	1	2	3	4	5	1	2	3	4	5	1	2	3	4	5	1	2	3	4	5	1	2	3	4	5
	6	C1	2	3	9	24	C3	C1	32	C4	54	47	41	C1	52	68	C4	78	63	70	91	82	94	89	84
	C5	10	7	C3	19	37	22	26	35	38	C2	57	C4	51	44	71	79	61	74	C5	87	99	85	100	96
	15	13	16	11	12	C5	40	27	C2	31	53	C3	59	56	48	C3	65	75	72	64	C4	81	88	93	C2
	4	17	20	1	18	39	30	23	21	33	45	43	55	58	50	73	66	76	C1	77	86	90	C1	C3	83
	14	5	C2	8	C4	28	25	34	29	36	49	42	46	60	C5	62	67	C2	69	80	97	C5	92	98	95

An α-lattice design with two replicates and block size five for 100 entries and five checks (C1–C5)

- Piepho, H. P., A. Büchse, and B. Truberg. "On the use of multiple lattice designs and α-designs in plant breeding trials." *Plant breeding* 125.5 (2006): 523-528.

Lattice Design

Factors

Design factors:

LAT = lattice
REP = replicate
BLK = incomplete block
PLT = plot

Treatment factors:

GEN = genotypes

Block model:

LAT/REP/BLK/PLT

Treatment model:

GEN

- Piepho, H. P., A. Büchse, and B. Truberg. "On the use of multiple lattice designs and α-designs in plant breeding trials." *Plant breeding* 125.5 (2006): 523-528.

ANOVA - balanced lattice design

Block	Rep I	Block	Rep II	Block	Rep III	Block	Rep IV
(1)	1 2 3	(4)	1 4 7	(7)	1 5 9	(10)	1 8 6
(2)	4 5 6	(5)	2 5 8	(8)	7 2 6	(11)	4 2 9
(3)	7 8 9	(6)	3 6 9	(9)	4 8 3	(12)	7 5 3

- Number of treatments, $t = k^2 = 9$
- Block size, $k = \sqrt{t} = 3$
- Number of reps, $r = k+1 = 4$
- Number of blocks, $b = k*r = k(k+1) = 12$
- Number of observations, $tr = kb = 36$

Source	df	
Total	$k^2(k+1)-1$	35
Treatment	$k^2 -1$	8
Rep	k	3
Block:Rep	k^2-1	8
Error	$(k-1)(k^2 – 1)$	16

Barreto, H. J., et al. "The alpha lattice design in plant breeding and agronomy: Generation and analysis." *Drought-and Low N-Tolerant Maize*(1996): 544.
http://www2.hawaii.edu/~halina/603/IncBlock.pdf