EXPERIMENT IN STAT AND ITS DESIGN

Thursday, November 16, 2023 6:36 PM

RBD / D of Ex

Design of an Experiment

deciding how the obsn should be taken

to answer a particular question

in a valid statistical efficient way

Random data

9862395123

well designed ??

quickest possibly when decision can be taken

Experiment

Experiment

Absolute

(measurement of relationships)

Ex Corrn analysis & Regression.

Comparative

(Compare the effect of 2/more treatments)

Exp Comparison of effect of 2/>2 medicine.

medium.

2020 covid vaccines...

Unit smallest division

Block → Whole experiment into smaller units. → homogeneous units..

Subgrouping

Strata → uniformity → Block

Precision Degree of Uncertainty

$$= \frac{1}{V(\bar{x})} = \frac{1}{\sigma^2/r} = \frac{r}{\sigma^2}$$

2003

$r \to$ treatment replications

$\sigma^2 \to$ variance...

✓

Season
2023

	Team	M	W	L	NRR	Pts	Last 5
1	IND	9	9	0	+2.570	18	✓ ✓ ✓ ✓ ✓
2	SA	9	7	2	+1.261	14	✓ ✗ ✓ ✓ ✓
3	AUS	9	7	2	+0.841	14	✓ ✓ ✓ ✓ ✓
4	NZ	9	5	4	+0.743	10	✓ ✗ ✗ ✗ ✗
5	PAK	9	4	5	-0.199	8	✗ ✓ ✓ ✗ ✗
6	AFG	9	4	5	-0.336	8	✗ ✗ ✓ ✓ ✓
7	ENG	9	3	6	-0.572	6	✓ ✓ ✗ ✗ ✗
8	BAN	9	2	7	-1.087	4	✗ ✓ ✗ ✗ ✗
9	SL	9	2	7	-1.419	4	✗ ✗ ✗ ✗ ✓
10	NED	9	2	7	-1.825	4	✗ ✗ ✗ ✓ ✗

2015

Season
2015

A

	Team	M	W	L	NRR	Pts
1	NZ	6	6	0	+2.564	12
2	AUS	6	4	1	+2.257	9
3	SL	6	4	2	+0.371	8
4	BAN	6	3	2	+0.136	7
5	ENG	6	2	4	-0.753	4
6	AFG	6	1	5	-1.853	2
7	SCO	6	0	6	-2.218	0

B

	Team	M	W	L	NRR	Pts
1	IND	6	6	0	+1.827	12
2	SA	6	4	2	+1.707	8
3	PAK	6	4	2	-0.085	8
4	WI	6	3	3	-0.053	6
5	IRE	6	3	3	-0.933	6
6	ZIM	6	1	5	-0.527	2
7	UAE	6	0	6	-2.032	0

Basic Principles

2015 ✓
2023 ✓

Replication — Repeating >1 any experiment. Are

Randomization — allocation of treatment to the experimental unit in a chance based random manner

$\boxed{9062395723}$

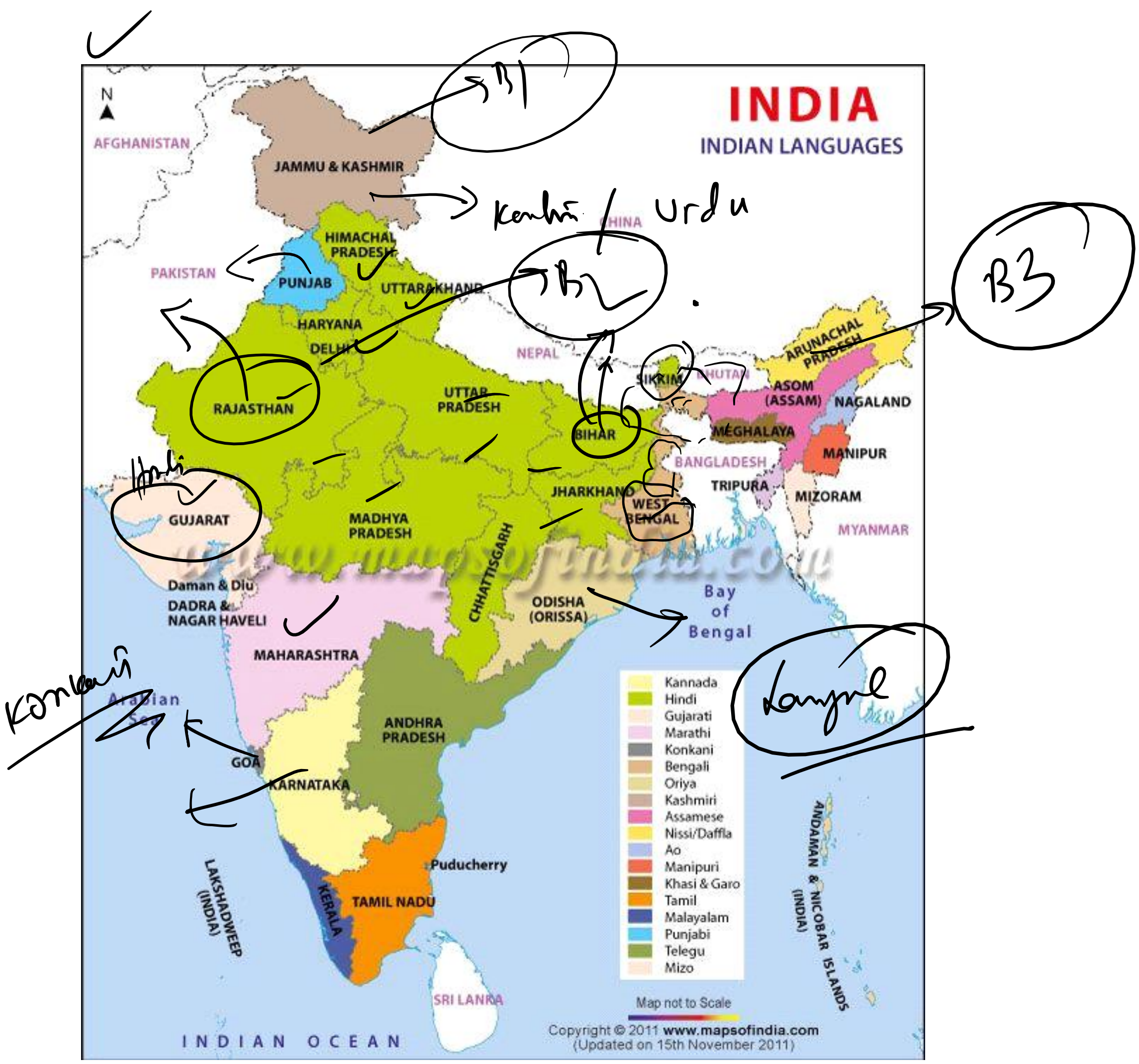
Local Control — experimental error ↓ the homogeneous data into

Local control

experimental error ↓

by dividing the homogeneous data into

Blocks ..



B1

Kashmiri / Urdu

B2

B3

Hindi

Konkani

Language

CRD Complete Randomised Design

Replication

Randomisation

Layout

Let us suppose that we have p treatments and i^{th} treatment is replicated n_i times $(i = 1,2, \dots, p)$.

The layout is given as:

equal prob for all experiments

1 2 3 4 5

ANOVA

S_D^2

Treatments	Replications				Total $(y_{i.})$
1	y_{11}	y_{12}	...	y_{1n_1}	$y_{1.}$
2	y_{21}	y_{22}	y_{33} ...	y_{2n_2}	$y_{2.}$
...	...	...	...	...	...
p	y_{p1}	y_{p2}	...	y_{pn_p}	$y_{p.}$

1 way ANOVA model

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

$i = 1, 2 \dots n$

$j = 1, 2 \dots n_i$

$y_{ij} \to$ jth obs^n of ith treatment

μ = General mean effect

$\alpha_i \to$ effect of ith treatment

ϵ_{ij} = Error

13 → SA

9 → AUS

9

22 − 9 ⇒ 13

2003 ×

2023

208 → SA

1975 48 years → 5

Model assumption $\sum_{i=1}^{p} \alpha_i = 0$ are iid $N(0, \sigma^2)$

ANOVA Table for CRD

Source of variation	Degree of Freedom (d.f.)	Sum of squares (SS)	Mean sum of squares (MSS)	F- ratio
Treatment	$p-1$	SSE	$MST = \frac{SST}{p-1}$	$F_T = \frac{MST}{MSE}$
Error	$n-p$	SSE	$MSE = \frac{SSE}{n-p}$	
Total	$n-1$	TSS		

eq − unu = df

$\boxed{df \geq 0}$

$\cancel{p} - 1 + n - \cancel{p}$

where

- $TSS = \sum_{i=1}^{p}\sum_{j=1}^{n_i} y_{ij}^2 - \frac{G^2}{n}$; $SST = \sum_{i=1}^{p} \frac{y_{i.}^2}{n_i} - \frac{G^2}{n}$
- $SSE = TSS - SST$
- here $y_{i.}$ = total for the i^{th} treatment.
- $G = \sum_{i=1}^{p}\sum_{j=1}^{n_i} y_{ij}$ = Grand total

$H_0 \rightarrow$

$\alpha_1 = \alpha_2 = \cdots = \alpha_p = 0$

or

$\mu_1 = \mu_2 = \cdots = \mu_p$

H_1: effects of treatments are not same

$F_T = \frac{MST}{MSE}$

Advantages of CRD

The layout of design is easy.

Very useful to conduct small experiments.

The no. of replications need not to be same for each treatment.

The CRD provides maximum d.f. for estimation of error variance, which increases the sensitivity or the precision of the experiments when the number of treatments are small.

The statistical analysis remains simple even if some or all observations for any treatment are rejected or lost or missing for some purely accidental reasons.

CRD results in the maximum use of the experimental units since all the experimental material can be used.

Disadvantages of CRD

If experimental materials are not homogenous, the design suffers from the disadvantage of being inherently less informative than other more sophisticated designs.

Not suited for a large number of treatments. (very large) $n \geq 100$.

Uses of CRD

- Under conditions where the experimental material is homogenous e.g. in physics, chemistry and biological experiment for some green house studies.
- CRD may be used in a chemical or baking experiment where the experimental units are the part of the thoroughly mixed chemical or powder.

A company is considering three different covers for boxes of a brand of cereal. Box type A has picture of a sports hero eating the cereal, type B has a picture of a child eating the cereal, and type C has a picture of a bowl of the cereal. The company wants to determine which cereal box type provides for the most sales. Eighteen test markets were selected by the company and each box type was randomly assigned to six markets. The number of boxes of this cereal sold per 10000 population in a specified period is recorded for each test market. The data are as follows:

18 ← 3 × 6

H_0: no difference

H_1: Difference is there..

Type A	52.4	47.8	52.4	51.3	50	52.1
Type B	50.1	45.2	46	46.5	47.4	46.2
Type C	49.2	48.3	49	47.2	48.6	48.2

$p = 3$ $\quad n_i = 6 \;\forall\; i = 1,2,3$

$n = 18$ $\quad$ Let $n_i = 9$

Test whether there is any significant difference in the mean of three means.

$$TSS = \sum_{i=1}^{k}\sum_{j=1}^{n_i} y_{ij}^2 - G^2/n = (52.4^2 + 47.8^2 + \dots + 48.2^2) - \frac{877.9^2}{18} = 85.40$$

$$SST = \frac{1}{6}\left(306^2 + 281.4^2 + 290.5^2\right) - \frac{877.9^2}{18} = 57.57$$

$$SSE = TSS - SST = 85.40 - 57.57 = 33.83$$

ANOVA Table

Source of variation	d.f.	Sum of squares	Mean sum of squares	F- ratio
Treatment	3-1=2	51.57	25.79	11.41
Error	18-3=15	33.83	2.26	
Total	17	85.40		

Calc value > Table value

H_0 X

- *The tabulated value* ($F_{(2,15)}(0.05)$) = 3.86.
- *Hence F-ratio > tabulated (11.41 > 3.86), so, we reject null hypothesis.*
- *Thus, there is a significant difference between the means of three treatments at the 5% level of significance.*

RBD

Larger number of treatments

Variance Comparison.

1 directional → Blocks..

Layout ← RBD

Let us suppose that we have p treatments and q blocks.

The layout is as follows:

Treatments \ Blocks	1	2	...	q	Total ($y_{i.}$)
1	y_{11}	y_{12}	...	y_{1q}	$y_{1.}$
2	y_{21}	y_{22}	...	y_{2q}	$y_{2.}$
...	...	...	...	...	...
p	y_{p1}	y_{p2}	...	y_{pq}	$y_{p.}$
Total ($y_{.j}$)	$y_{.1}$	$y_{.2}$	...	$y_{.q}$	$y_{..}$

$\Sigma\, y_1 + y_2 + \dots\, y_p$

Model

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

Model $y_{ij} = \mu + \alpha_i + \beta_j + \epsilon_{ij}$

$\beta_j \to$ effect of jth block

$\alpha_i \to$ " " ith treatment

ANOVA Table for RBD

Source of variation	Degree of Freedom (d.f.)	Sum of squares (SS)	Mean sum of squares (MS)	F- ratio
Treatment	$p-1$ 3	SST	$MST = \frac{SST}{p-1}$	$F_T = \frac{MST}{MSE}$
Blocks	$q-1$ 4	SSB	$MSB = \frac{SSB}{q-1}$	$F_T = \frac{MSB}{MSE}$
Error	$(p-1)(q-1)$ 12	SSE	$MSE = \frac{SSE}{(p-1)(q-1)}$	
Total	$n-1$ 19	TSS		

2 way Anova

where

$TSS = \sum_{i=1}^{p}\sum_{j=1}^{q} y_{ij}^2 - \frac{G^2}{n}$; $SST = \frac{1}{q}\sum_{i=1}^{p} y_{i.}^2 - \frac{G^2}{n}$

$SSB = \frac{1}{p}\sum_{j=1}^{q} y_{.j}^2 - \frac{G^2}{n}$; $SSE = TSS - SST - SSB$

$y_{i.}$ = total for the i^{th} treatment

$y_{.j}$ = total for the j^{th} block.

$G = \sum_{i=1}^{p}\sum_{j=1}^{q} y_{ij} = y$ =Grand total

Treatment $H_{0T}: \alpha_1 = \alpha_2 = \cdots = \alpha_p = 0$

Block $H_{0B}: \beta_1 = \beta_2 = \cdots = \beta_q = 0$

Same yields mean of varieties $H_{0A}: \mu_{1A} = \mu_{2A} = \mu_{3A} = \mu_{4A}$

H_{0B}: Yields of fields.

A seed company performs an experiment to compare four varieties of rice. Five fields are available for the study and each field is subdivided

A seed company performs an experiment to compare four varieties of rice. Five fields are available for the study and each field is subdivided into four plots of equal size. Each variety is randomly assigned to a plot, and the yield in bushels is recorded as follows:

		Field				
		1	2	3	4	5
Variety of rice	1	45	37	41	48	32
	2	47	41	38	46	37
	3	53	47	50	56	45
	4	38	32	40	43	29

Test whether there is any significant difference in the yield mean according to variety of rice and fields..

HOB: Yield of fields.

$p = 4$, $q = 5$

$n = pq = 20$

HW

LSD → Latin Square Design

RBD → that reduces the residual Error in an experiment

↓

Removing the variability due to a controllable variable

35/2 245/2

210

Model

$$y_{ijk} = \mu + \alpha_i + \beta_j + \gamma_k + \epsilon_{ijk}$$

(CRD → RDB → LSD)

CRD → RDB → LSD

Only treatment → CRD

Treat + Block → RBD

Treat + Row + Column → LSD