COURSE: STA 408
DESIGN & ANALYSIS OF EXPERIMENT

LECTURE DELIVERED BY

SARAH AHMAD

DEPARTMENT OF STATISTICS

VIRTUAL UNIVERSITY OF PAKISTAN



COURSE CONTENTS

- Introduction
- Basic principles of Experimental Design; Randomization, Replication and Local control
- Completely Randomized Design (CRD); Description, Layout, Statistical analysis, advantages and limitations
- The Randomized Complete Block Design (RCBD); Description, Layout, Statistical analysis, advantages and limitations.
- Replication within block, missing observations.
- Efficiency of RCBD and CRD.
- The Latin Square Design; Description, Layout, Statistical analysis, advantages and limitations
- Missing observations in Latin Square Design, efficiency of Latin Square
- Factorial Experiment; Main effects, Interaction effects
- Effect in a 2^2 and 2^3 in Factorial Experiment, advantages and disadvantages.

RECOMMENDED BOOKS

- Boniface, D.R. (1995). "*Experiment Design & Statistical Methods*", Chapman & Hall.
- Chaudhry, S.M. and Kamal, S. (2010). *"Introduction to Statistical Theory"* Part II, Ilmi Kitab Khana, Urdu Bazar, Lahore.
- Clarke, G.M. (1994). "*Statistics & Experimental Design*". Edward Arnold.
- Clarke, G.M., and Kempton, R.E. (1997), "*Introduction to the Design & Analysis of Experiments*", Edward Arnold.
- Das, M.N. and Geri, N.C, (1986). "*Design and Analysis of Experiments*", John Wiley, New York.
- Harold, R. L (1992). "Analysis of Variance in Experimental Design". Springer Verlage
- Mead, R. (1988). "*The Design of Experiments*". Cambridge University Press, Cambridge.
- Montgomery, D. C., Design and analysis of experiments, Edition: 7th ed. Publisher: New Jersey : John Wiley, 2009

COURSE: STA 408
DESIGN & ANALYSIS OF EXPERIMENT

LECTURE NO.1

Planning an experiment to obtain appropriate data and drawing inference out of the data with respect to any problem under investigation is known as ***design and analysis of experiment***.

Why all this?

INTRODUCTION

- **EXPERIMENT**

An experiment deliberately imposes a treatment on a group of objects or subjects in the interest of observing the response.

- **TREATMENTS**

In experiments, a treatment is something that researchers administer to experimental units.

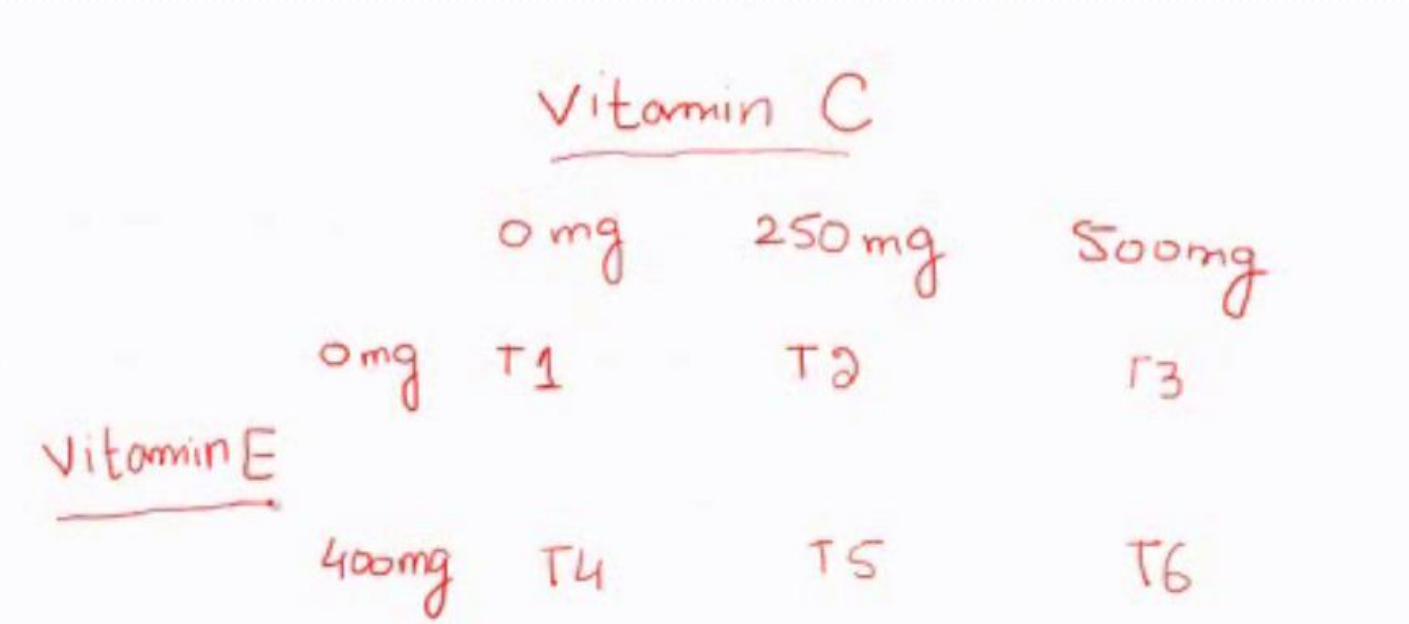
For example, a corn field is divided into four, each part is 'treated' with a different fertilizer to see which produces the most corn.

- **FACTOR**

A factor of an experiment is a controlled independent variable; a variable whose levels are set by the experimenter.

A factor is a general type or category of treatments. Different treatments constitute different levels of a factor.

EXAMPLE



- **EXPERIMENTAL UNIT**

The experimental unit is the basic entity or unit on which the experiment is performed. i.e. It is an object to which the treatment is applied and in which the variable under investigation is measured and analyzed.

For example, experimental unit may be:











IDENTIFY THE EXPERIMENTAL UNIT ?

- A teacher practices the different teaching methods on different groups in her class to see which yields the best results.

- A doctor treats a patient with a skin condition with different creams to see which is most effective.

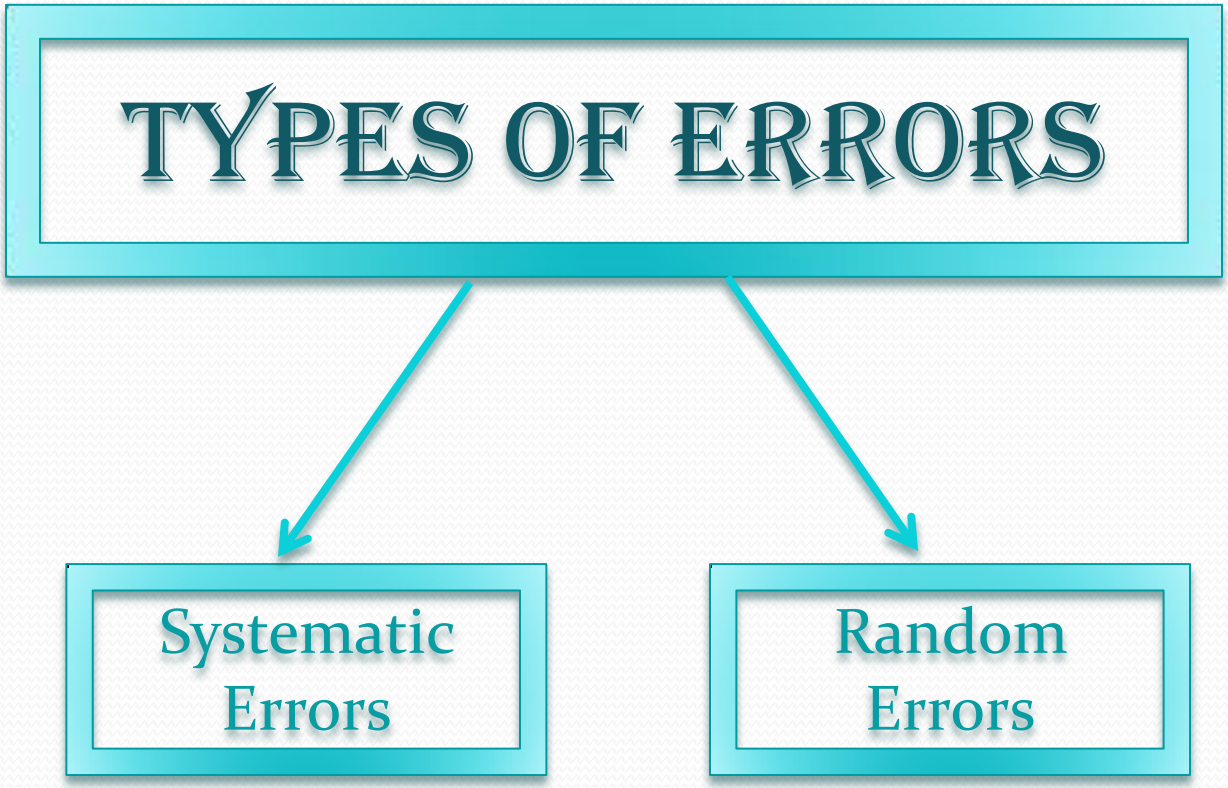
- **EXPERIMENTAL ERROR**

It describes the variation among identically and independently treated experimental units.

The various **origins of experimental error** include:

- ✓ The natural variation among the experimental units.
- ✓ Inability to reproduce the treatment conditions exactly from one unit to another.
- ✓ Interaction of treatments and experimental units.
- ✓ Any other extraneous factors that influence the measured characteristics.

TYPES OF ERRORS



- **EXPERIMENTAL DESIGN**

An experimental design is a plan to collect the data relevant to the problem under investigation. In such a way as to provide a basis for valid and objective inferences about the stated problem.

The plan usually consists of the selection of the treatments , specifications of the experimental layouts, allocation of the treatments and collection of observation for analysis.

- **PURPOSE OF EXPERIMENTAL DESIGN**

  - ✓ Get maximum information for minimum expenditure in the minimum possible time
  - ✓ Helps to reduce the experimental error
  - ✓ To ignore spurious effects, if any. *(FOR DETAIL, SEE THE LECTURE)*
  - ✓ To evaluate and examine the outcomes critically and logically.

STEPS IN DESIGNING AN EXPERIMENT

A statistically designed experiment may consist of the following steps.

- ✓ Statement of the problem
- ✓ Formulation of the hypothesis
- ✓ Choice of response or dependent variable.
- ✓ Selection of the factors to be varied.
- ✓ Choice of levels of these factors i.e. qualitative or quantitative and fixed or random.
- ✓ How factors are to be combined.
- ✓ Number of observations to be taken.

- ✓ Methods of randomization to be used.
- ✓ Mathematical model to describe the experiment.
- ✓ Data collection and processing.
- ✓ Application of statistical techniques to the experiment results.
- ✓ Drawing conclusions with measures of reliability of the estimates of any quantities that are evaluated.
- ✓ Evaluation of whole investigations particularly with other investigations on the same or similar problems.

BASIC PRINCIPLES OF EXPERIMENTAL DESIGN

The basic principles of experimental designs are:

- Randomization
- Replication
- Local Control

These principles make a valid test of significance possible.

RANDOMIZATION

It is a random process of assigning treatments to the experimental units.

Randomization is usually done by using tables of random numbers, by drawing numbered cards from a well shuffled pack of cards, or by drawing numbered balls from a well-shaken container.

PURPOSE OF RANDOMIZATION

- ✓ To remove bias and other sources of extraneous variation, which are not controllable .
- ✓ To form the basis of any valid statistical test.

REPLICATION

Replication is basically a repetition of the basic experiment.

In other words, it is a complete run for all the treatments to be tested in the experiment.

In all experiments, some variation is introduced because of the fact that the experimental units such as individuals or plots of land in agriculture experiments, cannot be physically identical. This type of variation can be removed by using a number of experimental units. We therefore perform the experiment more than once, i.e., we repeat the basic experiment. An individual repetition is called a replicate.

REPLICATION IS USED ?

- ✓ To secure more accurate estimate of the experimental error.
- ✓ To decrease the experimental error and thereby to increase precision.
- ✓ To obtain more precise estimate of the mean effect of a treatment.

LOCAL CONTROL

A term referring to the amount of balancing, blocking and grouping of the experimental units.

Balancing means that treatments should be assigned to the experimental units in such a way that the result is a balanced arrangements of the treatments.

Blocking means that like experimental units should be collected together to form a relatively homogeneous group. A block is also a replicate.

The main purpose of the principle of local control is to increase the efficiency of an experimental design by decreasing the experimental error.

TYPES OF DESIGNS

1- Systematic designs

2- Randomized designs

SYSTEMATIC DESIGNS

In systematic designs treatments are applied to the experimental units by some systematic pattern, i.e. by the choice of the experimenter.

FOR EXAMPLE:

The experimenter wishes to test three treatments and he decides to have four repetitions of each treatment. The arrangements of three treatments over the experimental area could be : *(FOR SOLUTION, LISTEN THE LECTURE)*

RANDOMIZED DESIGNS

In randomized designs , as the treatments are applied randomly therefore the conclusions drawn are supported by statistical tests.

In this way, inferences are applicable in wider range and the random process minimizes the systematic error.

The analysis of variance techniques are also suitable to randomized designs only.

CLASSIFICATIONS OF RANDOMIZED DESIGNS

- **Completely Randomized Design (CRD)**
- **Randomized Complete Block Design (RCBD)**
- **Latin Square Design**

"Detailed discussion will be made in further lectures"

EXAMPLE

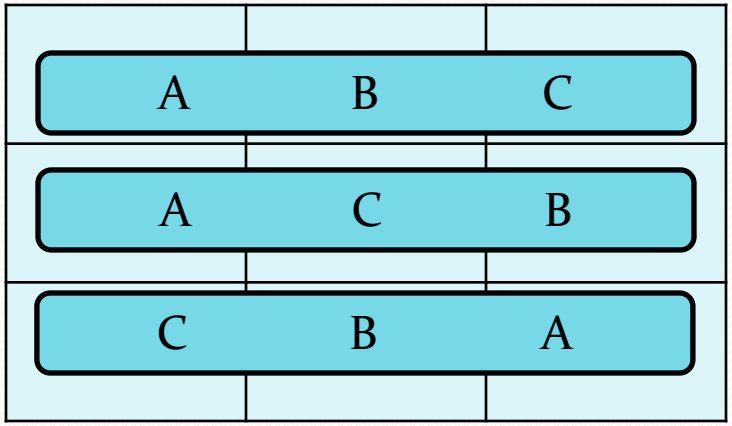
A market gardener wants to test three types of peas, A, B, C, on his land. He has a square plot which he divides into nine equal squares, three to be planted with each type of pea. The problem which he then faces is which square to plant with which type.

- One method is **Completely Randomized Design** which might,

C	A	C
B	A	A
B	B	C

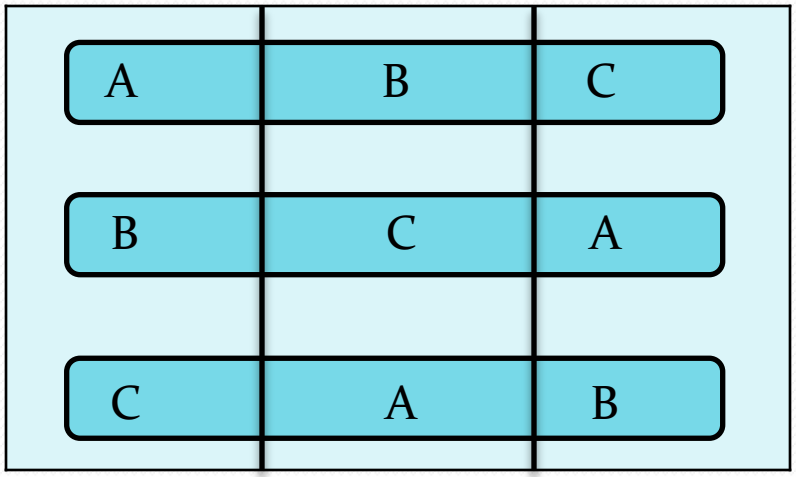
This would be all right if all the plots were equally desirable. If however, there were prevailing north wind so that the northernmost plots were exposed, he might decide to use ,

- **Randomized Complete Block Design**, where each of the types A, B and C is planted once in each west-east block.



If the gardener also felt that the soil to the east was rather better than that to the west, he would use ,

A **Latin Square design**, where each type of pea is planted once in each row (west-east), and once in each column (north-south).



THANK YOU

STA 408: DESIGN AND ANALYSIS OF EXPERIEMENTS

S. M. Muslim Raza

Instructor Statistics

Virtual university of Pakistan



Contents

1. Basic concepts of ANOVA
2. CR Design

Analysis of Variance: A Conceptual Overview

- Analysis of Variance (ANOVA) can be used to test for the equality of three or more population means.
- Data obtained from observational or experimental studies can be used for the analysis.
- We want to use the sample results to test the following hypotheses:

H_0: $\mu_1 = \mu_2 = \mu_3 = \cdot\ \cdot\ \cdot = \mu_k$

H_a: Not all population means are equal

Analysis of Variance: A Conceptual Overview

H_0: $\mu_1 = \mu_2 = \mu_3 = \cdot\ \cdot\ \cdot = \mu_k$

H_a: Not all population means are equal

- If H_0 is rejected, we cannot conclude that *all* population means are different.
- Rejecting H_0 means that at least two population means have different values.

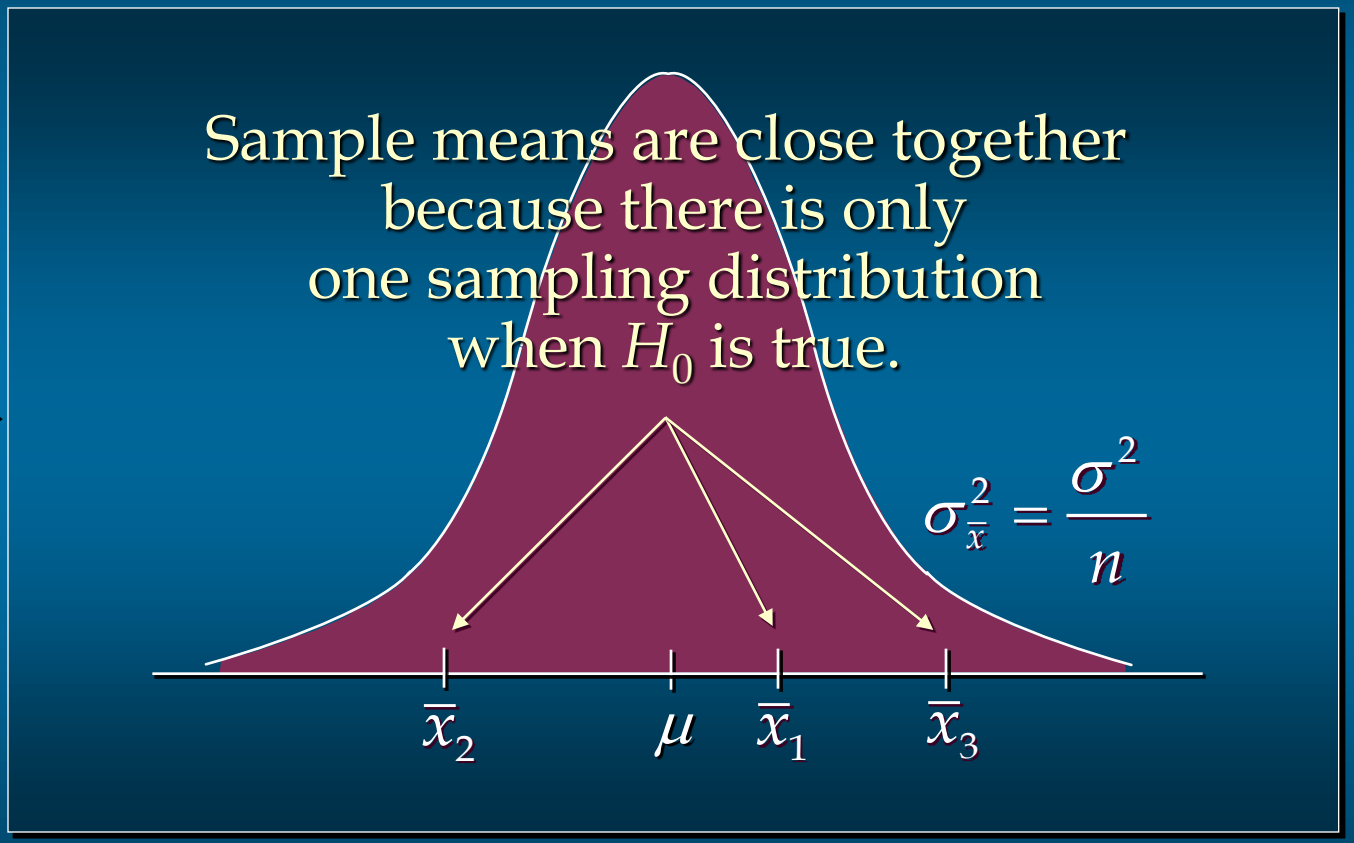
Analysis of Variance: A Conceptual Overview

- Assumptions for Analysis of Variance

  - For each population, the response (dependent) variable is normally distributed.
  - The variance of the response variable, denoted σ^2, is the same for all of the populations.
  - The observations must be independent.

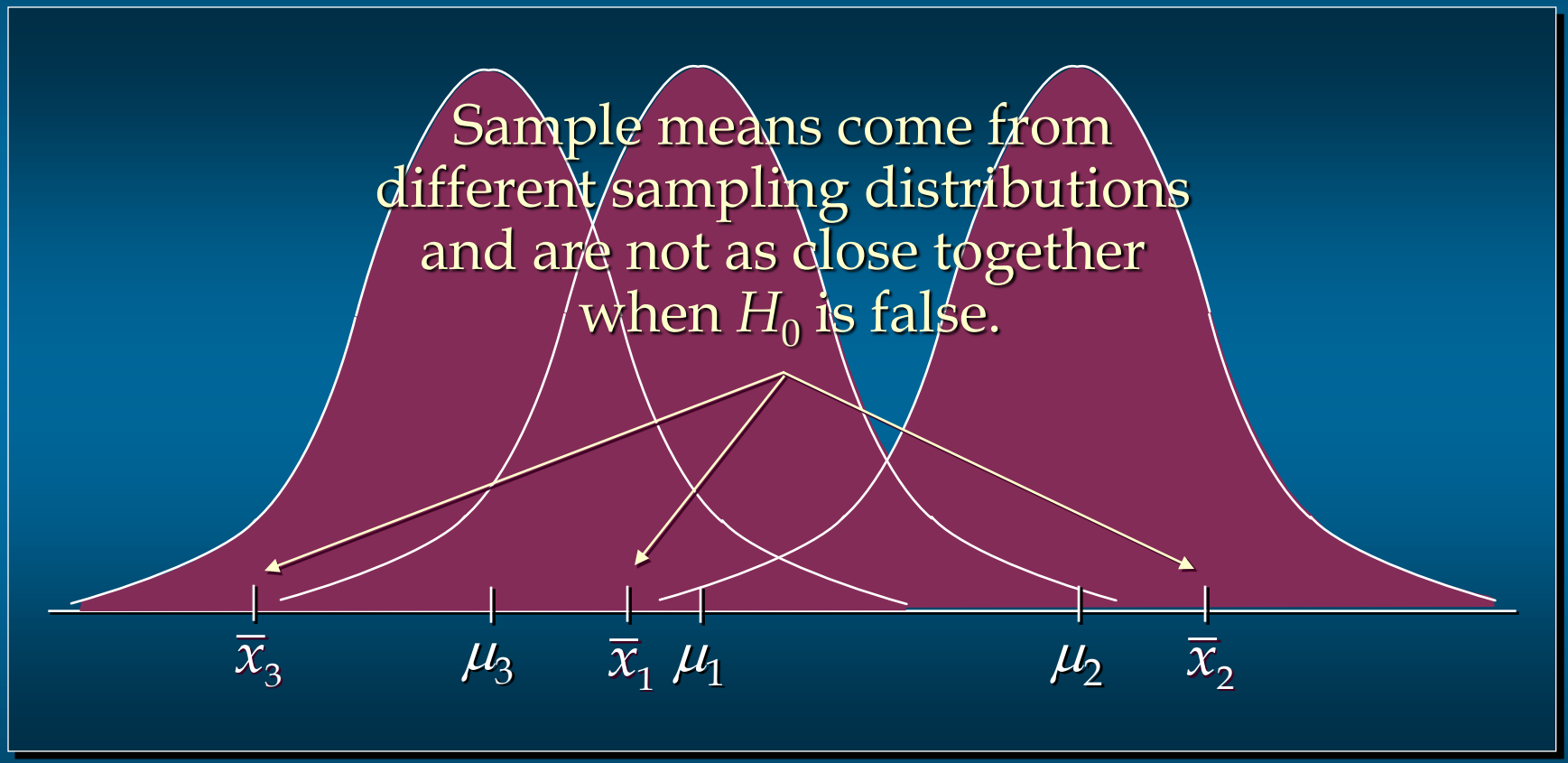
Analysis of Variance: A Conceptual Overview

- Sampling Distribution of $\bar{x}$ Given H_0 is True



Analysis of Variance: A Conceptual Overview

- Sampling Distribution of $\bar{x}$ Given H_0 is False



Analysis of Variance

- Between-Treatments Estimate of Population Variance
- Within-Treatments Estimate of Population Variance
- Comparing the Variance Estimates: The *F* Test
- ANOVA Table

Between-Treatments Estimate of Population Variance σ^2

- The estimate of σ^2 based on the variation of the sample means is called the mean square due to treatments and is denoted by MSTR.

$$\text{MSTR} = \frac{\sum_{j=1}^{k} n_j(\bar{x}_j - \bar{\bar{x}})^2}{k-1}$$

Denominator is the degrees of freedom associated with SSTR

Numerator is called the sum of squares due to treatments (SSTR)

Within-Treatments Estimate of Population Variance σ^2

- The estimate of σ^2 based on the variation of the sample observations within each sample is called the mean square error and is denoted by MSE.

$$\text{MSE} = \frac{\sum_{j=1}^{k}(n_j - 1)s_j^2}{n_T - k}$$

Denominator is the degrees of freedom associated with SSE

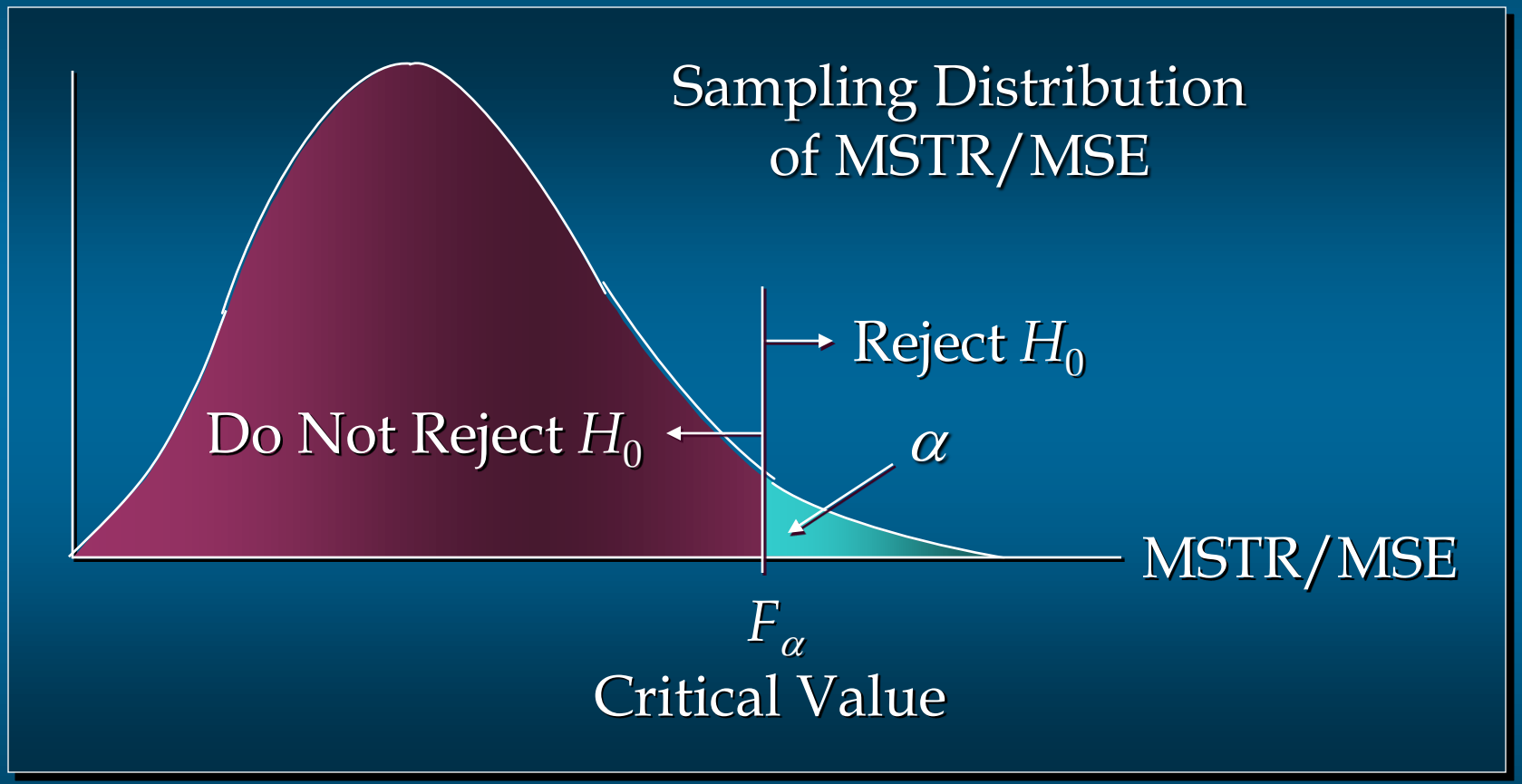
Numerator is called the sum of squares due to error (SSE)

Comparing the Variance Estimates: The *F* Test

- If the null hypothesis is true and the ANOVA assumptions are valid, the sampling distribution of MSTR/MSE is an *F* distribution with MSTR d.f. equal to k - 1 and MSE d.f. equal to n_T - k.
- If the means of the k populations are not equal, the value of MSTR/MSE will be inflated because MSTR overestimates σ^2.
- Hence, we will reject H_0 if the resulting value of MSTR/MSE appears to be too large to have been selected at random from the appropriate *F* distribution.

Comparing the Variance Estimates: The *F* Test

- Sampling Distribution of MSTR/MSE



Test for the Equality of k Population Means

- Hypotheses

H_0: $\mu_1 = \mu_2 = \mu_3 = \cdot\ \cdot\ \cdot = \mu_k$

H_a: Not all population means are equal

- Test Statistic

F = MSTR/MSE

Test for the Equality of k Population Means

- Rejection Rule
  - p-value Approach: Reject H_0 if p-value $\leq \alpha$
  - Critical Value Approach: Reject H_0 if $F \geq F_\alpha$

where the value of F_α is based on an F distribution with k - 1 numerator d.f. and n_T - k denominator d.f.

ANOVA Table

- The ANOVA table decomposes the variance into two components (sources): a between-group component (treatments) and a within-group component (error). The F-ratio, which in this case equals 15.6234, is a ratio of the between-group estimate to the within-group estimate. Since the P-value of the F-test is less than 0.05, there is a statistically significant difference between the means from one level to another at the 95.0% confidence level.

ANOVA Table for hours by brand

Source	*Sum of Squares*	*Df*	*Mean Square*	*F-Ratio*	*P-Value*
Between groups	1.08917	3	0.363057	15.62	0.0000
Within groups	0.557714	24	0.0232381		
Total (Corr.)	1.64689	27			

Completely randomized design (CRD)

- A factor is a variable that the experimenter has selected for investigation (the independent variable).
- A treatment is a level of a factor.
- Experimental units are the objects of interest in the experiment.
- A completely randomized design is an experimental design in which the treatments are randomly assigned to the experimental units.

COMPLETELY RANDOMIZED DESIGN

A completely randomized (CR) design, which is the simplest type of the basic designs, may be defined as a design in which the treatments are assigned to experimental units completely at random, that is the randomization is done without any restrictions. The design is completely flexible, i.e. any number of treatments and any number of units per treatment may be used. Moreover, the number of units per treatment need not be equal. A completely randomized design is considered to be most useful in situations where (i) the experimental units are homogeneous, (ii) the experiments are small such as laboratory experiments and (iii) some experimental units are likely to be destroyed or to fail to respond.

CRD

- When the samples of experimental units for each treatment are random and independent of each other
- Design is used to compare the treatment means:

$$H_0 : \mu_1 = \mu_2 = \ldots \mu_k$$

H_a : *At least two of the treatment means differ*

- The hypotheses are tested by comparing the differences between the treatment means.

- Test statistic is calculated using measures of variability within treatment groups and measures of variability between treatment groups

Steps for Conducting an Analysis of Variance (ANOVA) for a Completely Randomized Design:

- 1- Assure randomness of design, and independence, randomness of samples
- 2- Check normality, equal variance assumptions
- 3- Create ANOVA summary table
- 4- Conduct multiple comparisons for pairs of means as necessary/desired

Assumptions

1- Normality:

You can check on normality using

1- Normal probability plot

2- Kolmogorve test

2- Constant variance:

You can check on homogeneity of variances using

1- Histogram Plot

2- leven's test.

CRD

ANOVA Summary Table for a Completely Randomized Design

Source	*df*	*SS*	*MS*	*F*
Treatments	$k-1$	SST	$MST = \frac{SST}{k-1}$	$\frac{MST}{MSE}$
Error	$n-k$	SSE	$MSE = \frac{SSE}{n-k}$	
Total	$n-1$	$SS(Total)$		

Advantages of a CRD

- Very flexible design (i.e. number of treatments and replicates is only limited by the available number of experimental units).
- Statistical analysis is simple compared to other designs.
- Loss of information due to missing data is small compared to other designs due to the larger number of degrees of freedom for the error source of variation.

Disadvantages

- If experimental units are not homogeneous and you fail to minimize this variation using blocking, there may be a loss of precision.
- Usually the least efficient design unless experimental units are homogeneous.
- Not suited for a large number of treatments.

Experimental Layout:

An example of the experimental layout for a completely randomized design (CR) using four treatments A, B, C and D, each repeated 3 times, is given below:

Example:

An experiment was conducted to compare the yields of three varieties of potatoes. Each variety was assigned at random to equal-size plots, four times. The yields were as follows:

Variety		
A	B	C
23	18	16
26	28	25
20	17	12
17	21	14

Test the hypothesis that the three varieties of potatoes are not different in the yielding capabilities.

Example:

Three fertilizer treatments A, B, C each applied to seven plots of strawberry plants, resulted in the following weights of crops (Ib/plot):

A	24	18	18	29	22	17	15
B	46	39	37	50	44	45	30
C	32	30	26	41	36	28	27

Perform the analysis of variance to test the hypothesis of no difference in the treatment effects. Use a 0.01 level of significance.

Solution:

Formulation of Hypothesis:

$H_0 : \mu_1 = \mu_2 = \mu_3$

$H_1 : At\ least\ two\ means\ are\ not\ equal.$

Level of Significance:

$\alpha = 0.05.$

Test Statistics:

$$F = \frac{MS\ Treatment}{MS\ Error}$$

Calculation:

A	24	18	18	29	22	17	15	**143**
B	46	39	37	50	44	45	30	**291**
C	32	30	26	41	36	28	27	**220**
Total								**654**

Here $T_1 = 143, T_2 = 291, T_3 = 220$

$T_{..} = 654$

Total SS$= \sum\sum X_{ij}^2 - CF$

CF $= T_{..}^2 / N$, where N= nk= 7(3) = 21.

CF$= (654)^2 / 21 = 20367.43$

Total SS$= \sum\sum X_{ij}^2 - CF$

Total SS$= 24^2 + 18^2 + 18^2 + \ldots + 28^2 - 20367.43$

$= 22520 - 20367.43 = 2152.57$

Treat SS$= \dfrac{\sum_{j=1}^{k} T_{.j}^2}{n} - CF$

Treat SS= $\frac{\sum_{j=1} T_{.j}^2}{n} - CF$

$$= \frac{1}{7}\left[(143)^2 + (291)^2 + (220)^2\right] - 20367.43$$

$$= 21932.86 - 20367.43 = 1564.43$$

Error SS= Total SS - Treat SS

$$= 2152.57 - 1564.43 = 587.14$$

Source of Variation	Sum of squares	Degree of freedom	Mean squares	F ratio
Treatment	1565.43	k-1 = 2	=1564.43/2 =782.72	F=782.72/32.62 =24.00
Error	587.14	N – k = 18	=587.14/18 =32.62	
Total	2152.57	N – 1 = 20		

Critical Region:

Reject Ho, if

$$F_{cal} \geq F_{\alpha(k-1,N-k)}$$

$$24.00 \geq F_{0.05(2,18)}$$

$$24.00 > 3.55$$

We reject Ho.

Conclusion:

Since the value $24.00 > 3.55$ fall in the critical region, so we reject Ho or also we conclude that the there is a difference between the treatment effects.

THANK YOU

STA408
Design and Analysis of Experiments
Lesson 3

JAVED IQBAL

Department of Statistics

Virtual University of Pakistan

Randomized Complete Block Design

Testing the effects of two independent variables or factors on one *dependent variable.*

Hypothesis Testing – General Procedure

Procedure Table

Solving Hypothesis-Testing Problems (Traditional Method)

Step 1 State the hypotheses and identify the claim.

Step 2 Find the critical value(s) from the appropriate table

Step 3 Compute the test value.

Step 4 Make the decision to reject or not reject the null hypothesis.

Step 5 Summarize the results.

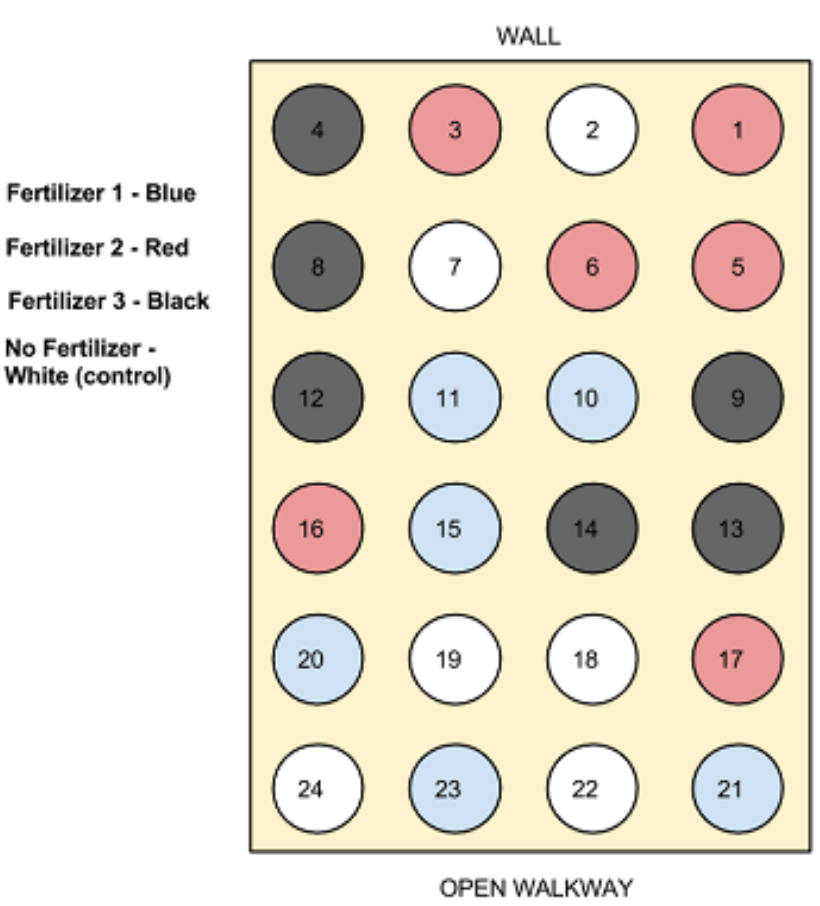
Main Points

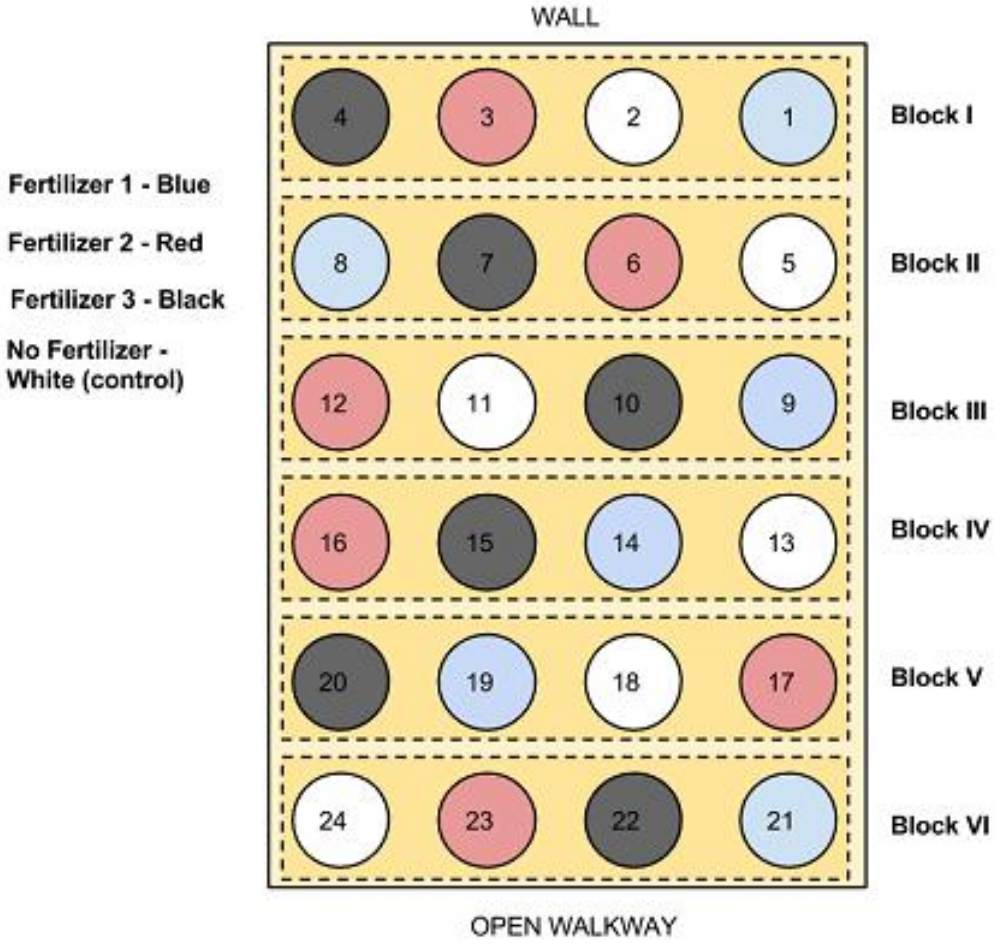
- Experimental material is divided into groups or blocks
  - Why
- Each block contains a complete set of treatments
- Treatments are assigned randomly
  - Randomizations is restricted within block

Main Points

- A new randomization is done for each block/group
- It has different layouts: 2X3, 3X5, etc
- The RCBD is perhaps the most frequently used experimental design

CRD vs RCBD


WALL
Fertilizer 1 - Blue
Fertilizer 2 - Red
Fertilizer 3 - Black
No Fertilizer - White (control)
4 3 2 1
8 7 6 5
12 11 10 9
16 15 14 13
20 19 18 17
24 23 22 21
OPEN WALKWAY


WALL
Fertilizer 1 - Blue
Fertilizer 2 - Red
Fertilizer 3 - Black
No Fertilizer - White (control)
4 3 2 1 Block I
8 7 6 5 Block II
12 11 10 9 Block III
16 15 14 13 Block IV
20 19 18 17 Block V
24 23 22 21 Block VI
OPEN WALKWAY

May look like



The Agricultural Research Field



The Agricultural Research Field



Experimental Layout

Allot treatment randomly



BLOCK - I				
BLOCK - II				
BLOCK - III				

Statistical Model

- The liner model for this design is:

$$Y_{ij} = \mu + B_i + T_j + e_{ij}$$

- Where:

μ = True mean effect

T = Effect of treatment

B = Effect of block/groups

e = effect of random error

Data Computations Table

Block (Groups)	Treatments						Block	
	1	*2*	…	*j*	…	*k*	Total	Mean
1	Y_{11}	Y_{12}					$B_{1\bullet}$	$\bar{Y}_{1\bullet}$
2								
i								
r								
Total	$B_{\bullet 1}$							
Means	$\bar{Y}_{\bullet 1}$							

Partitioning of Variations

$$\begin{bmatrix}\text{Total}\\ \text{Variation}\end{bmatrix}=\begin{bmatrix}\text{Variation due to}\\ \text{Treatment}\end{bmatrix}+\begin{bmatrix}\text{Variation due to}\\ \text{Block/Grouping}\end{bmatrix}+\begin{bmatrix}\text{Variation due to}\\ \text{Unknown factors}\end{bmatrix}$$

Total SS = Treatment SS + Block SS + Error SS

$$\sum_i \sum_j (Y_{ij} - \bar{Y})^2 = r \sum_j (\bar{Y}_{\square j} - \bar{Y}_{\square\square})^2 + k \sum_i (\bar{Y}_{i\square} - \bar{Y}_{\square\square})^2 + \sum_i \sum_j (Y_{ij} - \bar{Y}_{\square j} - \bar{Y}_{i\square} - \bar{Y}_{\square\square})$$

AVOVA Table for RCB

Source of Variation s.o.v	Degree of Freedom d.f	SS = Sum of Squares	MS = Mean Squares	F -test
(Variations due to) **Treatment**	K-1	$SSTr = r\sum_{j}(\bar{Y}_{\square j} - \bar{Y}_{\square\square})^2$	$S_t^2 = \frac{SSTr}{k-1}$	$F_t = \frac{S_t^2}{S_e^2} = \frac{MST}{MSE}$
(Variations due to) **Blocks**	r- 1	$SSB = k\sum_{i}(\bar{Y}_{i\square} - \bar{Y}_{\square\square})^2$	$S_b^2 = \frac{SSB}{r-1}$	$F_b = \frac{S_b^2}{S_e^2} = \frac{MSB}{MSE}$
(Variations due to) **Error**	(k-1) (r-1)	SSE = By Subtraction	$S_e^2 = \frac{SSTr}{(r-1)(k-1)}$	
Total	rk-1	$SST = \sum_{i}\sum_{j}(Y_{ij} - \bar{Y}_{\square\square})^2$	---	

Example

- Four verities of wheat ere tested in randomized complete clock design in four replications/blocks. Yield in kg/plot is shown in the table below.
- Test the hypothesis that there is no difference in the means of four varieties.

Blocks	Varieties of wheat			
	V1	V2	V3	V4
I	2	5	4	1
II	2	3	3	1
III	4	6	6	2
IV	1	4	2	3

Testing Hypothesis – Example

Step 1

Stat the hypothesis.

$H_0 : Tj = 0$ (There is NO difference in the means of four varieties of wheat)
$H_1 : Tj \neq 0$ (There is difference in the means of four varieties)
(Not ALL the four means are equal)

$H_0 : B_i = 0$ (There is NO difference in the means of four varieties of wheat)
$H_1 : B_i \neq 0$ (There is difference in the means of four varieties)
(Not ALL the four means are equal)

Step 2

We set the **level of significance (alpha)** $\boldsymbol{\alpha = 0.05}$ **(5%)**

Testing Hypothesis – Example

Step 3

The test-statistic (formula) to be used is:

$$F = \frac{S_t^2}{S_e^2} = \frac{MS(Treatment)}{MS(Error)}$$

It follows F-distribution with degree of freedom

df = v_1 = from treatment = k-1

df = v_2 = from error = (k-1)(r-1)

Computations

Blocks	Varieties (Treatment)						
	V1	V2	V3	V4	$Block\ Total$	$\left(Block\ Total\right)^2$	Sum of Square of each value
I	2	5	4	1	2+5+4+1 = 12	12*12 = 144	$2^2+5^2+4^2+1^2 = 46$
II	2	3	3	1	9	81	23
III	4	6	6	2	18	324	92
IV	1	4	2	3	10	100	30
$\left(\text{Varities Total}\right)$							
$\left(\text{Varities Total}\right)^2$							
Sum of Square of each value							

Computations

$$\text{SST} = \text{ Sum of Squares of Total } = \Sigma\Sigma Y^2 - \frac{T_{\square\square}}{n}$$

$$= \left(2^2 + 5^2 + 4^2 + \ldots + 3^2\right) - \frac{\left(2+5+4+\ldots+3\right)^2}{16}$$

$$= 191 - \frac{(49)^2}{16} = 40.94$$

$$\text{SSTr} = \text{ Sum of Squares of Treatment(varities)} = -\Sigma\Sigma Y^2 - \frac{T_{\square\square}}{n}$$

$$= \left(2^2 + 5^2 + 4^2 + \ldots + 3^2\right) - \frac{\left(2+5+4+\ldots+3\right)^2}{16}$$

$$= 191 - \frac{(49)^2}{16} = 40.94$$

Computations

$$\text{SSTr} = \text{ Sum of Squares of Treatment(varieties)} = \frac{T^2}{r} - \frac{T_{\square\square}}{n}$$

$$= \frac{679}{4} - \frac{(49)^2}{16} = 169.75 - 150.06 = 19.69$$

$$\text{SSB} = \text{ Sum of Squares of Block} = \frac{B^2}{r} - \frac{T_{\square\square}}{n}$$

$$= \frac{649}{4} - \frac{(49)^2}{16} = 162.25 - 150.06 = 12.19$$

$$\text{SSE} = \text{ Sum of Squares of Error} = SST - SSTr - SSB$$

$$= 40.96 - 19.69 - 12.19 = 9.06$$

ANOVA Table

Source of Variation s.o.v	Degree of Freedom d.f	SS = Sum of Squares	MS = Mean Squares	Computed F -test
(Variations due to) **Treatment**	4-1 = 3	19.69	19.69 / 3 = 6.56	$F_t = \frac{6.56}{1.01} = 6.50$
(Variations due to) **Blocks**	4 – 1 = 3	12.19	12.19 / 3 = 4.06	$F_b = \frac{4.06}{1.01} = 4.02$
(Variations due to) **Error**		9.06	9.06 / 9 = 1.01	
Total	(4*4) – 1 =15	40.96	---	

F – Table Value

Table H *(continued)*

$\alpha = 0.05$

d.f.D.: degrees of freedom, denominator	d.f.N.: degrees of freedom, numerator												
	1	2	3	4	5	6	7	8	9	10	12	15	20
1	161.4	199.5	215.7	224.6	230.2	234.0	236.8	238.9	240.5	241.9	243.9	245.9	248.0
2	18.51	19.00	19.16	19.25	19.30	19.33	19.35	19.37	19.38	19.40	19.41	19.43	19.4
3	10.13	9.55	9.28	9.12	9.01	8.94	8.89	8.85	8.81	8.79	8.74	8.70	8.6
4	7.71	6.94	6.59	6.39	6.26	6.16	6.09	6.04	6.00	5.96	5.91	5.86	5.8
5	6.61	5.79	5.41	5.19	5.05	4.95	4.88	4.82	4.77	4.74	4.68	4.62	4.5
6	5.99	5.14	4.76	4.53	4.39	4.28	4.21	4.15	4.10	4.06	4.00	3.94	3.8
7	5.59	4.74	4.35	4.12	3.97	3.87	3.79	3.73	3.68	3.64	3.57	3.51	3.4
8	5.32	4.46	4.07	3.84	3.69	3.58	3.50	3.44	3.39	3.35	3.28	3.22	3.1
9	5.12	4.26	3.86	3.63	3.48	3.37	3.29	3.23	3.18	3.14	3.07	3.01	2.9
10	4.96	4.10	3.71	3.48	3.33	3.22	3.14	3.07	3.02	2.98	2.91	2.85	2.7
11	4.84	3.98	3.59	3.36	3.20	3.09	3.01	2.95	2.90	2.85	2.79	2.72	2.6
12	4.75	3.89	3.49	3.26	3.11	3.00	2.91	2.85	2.80	2.75	2.69	2.62	2.5
13	4.67	3.81	3.41	3.18	3.03	2.92	2.83	2.77	2.71	2.67	2.60	2.53	2.4
14	4.60	3.74	3.34	3.11	2.96	2.85	2.76	2.70	2.65	2.60	2.53	2.46	2.3
15	4.54	3.68	3.29	3.06	2.90	2.79	2.71	2.64	2.59	2.54	2.48	2.40	2.3
16	4.49	3.63	3.24	3.01	2.85	2.74	2.66	2.59	2.54	2.49	2.42	2.35	2.2
17	4.45	3.59	3.20	2.96	2.81	2.70	2.61	2.55	2.49	2.45	2.38	2.31	2.2
18	4.41	3.55	3.16	2.93	2.77	2.66	2.58	2.51	2.46	2.41	2.34	2.27	2.1
19	4.38	3.52	3.13	2.90	2.74	2.63	2.54	2.48	2.42	2.38	2.31	2.23	2.1
20	4.35	3.49	3.10	2.87	2.71	2.60	2.51	2.45	2.39	2.35	2.28	2.20	2.1
21	4.32	3.47	3.07	2.84	2.68	2.57	2.49	2.42	2.37	2.32	2.25	2.18	2.1
22	4.30	3.44	3.05	2.82	2.66	2.55	2.46	2.40	2.34	2.30	2.23	2.15	2.0
23	4.28	3.42	3.03	2.80	2.64	2.53	2.44	2.37	2.32	2.27	2.20	2.13	2.0
24	4.26	3.40	3.01	2.78	2.62	2.51	2.42	2.36	2.30	2.25	2.18	2.11	2.0
25	4.24	3.39	2.99	2.76	2.60	2.49	2.40	2.34	2.28	2.24	2.16	2.09	2.0
26	4.23	3.37	2.98	2.74	2.59	2.47	2.39	2.32	2.27	2.22	2.15	2.07	1.9
27	4.21	3.35	2.96	2.73	2.57	2.46	2.37	2.31	2.25	2.20	2.13	2.06	1.9

Comparison and Conclusion

Step 5

$$\textit{Calculated Value of F (treatment)} = 6.56$$

$$\text{Critical/Table Value of F} = 3.86$$

$$\Rightarrow 6.56 > 3.86$$

$$\Rightarrow F_t > F(Table/CriticalValue)$$

Conclusion and Summarizing:

Since the computed value of **F (treatment) is GREATER THAN the F Table value, So it** falls in the critical region. We reject our null hypothesis and may conclude that the means of four varieties of wheat are significantly different.

Comparison and Conclusion

Step 5

$$\textit{Calculated Value of F (Block/Group)} = 4.06$$

$$\text{Critical/Table Value of F} = 3.86$$

$$\Rightarrow 4.06 > 3.86$$

$$\Rightarrow F_t > F(Table / CriticalValue)$$

Conclusion and Summarizing:

Since the computed value of **F (Block/group) is GREATER THAN the F Table value, So it** falls in the critical region. We reject our null hypothesis and may conclude that the **BLOCKING or GROUPING was effective.**

Another example

- The following is the plan of a filed layout testing four varieties A, B, C and D of wheat in each of % blocks. The plot yields in Kgs are also indicated.

Block I	
D	29.3
B	33.3
C	30.8
A	32.3

Block II	
B	33.0
A	34.0
C	34.3
D	26.0

Block III	
D	29.8
A	34.3
B	36.3
C	35.3

Block IV	
B	36.8
A	35.00
D	28.0
C	32.3

Block V	
D	28.8
C	35.8
B	34.5
A	36.5

Computations

Blocks	Varieties (Treatment)						
	V1	V2	V3	V4	*Block Total*	$\left(\text{Block Total}\right)^2$	Sum of Square of each value
I							
II							
III							
IV							
$\left(\text{Varities Total}\right)$							
$\left(\text{Varities Total}\right)^2$							
Sum of Square of each value							

ANOVA Table

Source of Variation s.o.v	Degree of Freedom d.f	SS = Sum of Squares	MS = Mean Squares	Computed F -test
(Variations due to) **Treatment**				
(Variations due to) **Blocks**				
(Variations due to) **Error**				
Total				

Advantages

- The source of **extraneous** variation is controlled by grouping the experimental material and hence the estimate of the experimental error is decreased.
- The design is flexible. Any number of replications may be run and any number of treatments may be tested.
- The experiment can be set up easily
- The statistical analysis is simple and straightforward.
- It is easy to adjust the missing observations

Disadvantages

- It control the variability only in one direction
- It is not suitable design when the number of treatments is very large or when the blocks are not homogeneous.

- Extraneous Variables are **undesirable** variables that influence the relationship between the variables that an experimenter is examining.
- Another way to think of this, is that these are variables that influence the outcome of an experiment, though they are not the variables that are actually of interest. These variables are undesirable because they add error to an experiment. A major goal in research design is to decrease or control the influence of extraneous variables as much as possible.

Summary

- Experimental material is divided into groups or blocks
- Each block contains a complete set of treatments
- Treatments are assigned randomly for each block
- Data is organized in ANOVA table
- F-test is used for testing
- Rest of the testing procedure is same.

The END

COURSE: STA 408
DESIGN & ANALYSIS OF EXPERIMENT

LECTURE DELIVERED BY

SARAH AHMAD

DEPARTMENT OF STATISTICS

VIRTUAL UNIVERSITY OF PAKISTAN



COURSE: STA 408
DESIGN & ANALYSIS OF EXPERIMENT

LECTURE NO.4

RCBD WITH REPLICATIONS WITHIN BLOCKS

Let us assume that each treatment is replicated *n* times within each block, so that each block contains *nk* plots. The experiment becomes analogous to two way classifications with each cell containing *n* observations. This leads to new statistical model in the form

$$Y_{ijl} = \mu + B_i + \tau_j + (B\tau)_{ij} + e_{ijl}$$

MISSING OBSERVATIONS IN RCBD

Sometimes it so happens that some experimental units or observations are missing.

FOR EXAMPLE



UPROOTED PLANTS



RECORDS LOST



DIED ANIMALS



DESTROYED PLANTS



BROKEN FLASK



UNCOOPERATIVE PEOPLE

STA: 408

Suppose,

		TREATMENTS				Block Total
		1	2	.....	k	
BLOCKS	1	(x)	y_{12}	.....	y_{1k}	$B'_{1.} + x$
	2	y_{21}	y_{22}	.....	y_{2k}	$B_{2.}$
	⋮	⋮	⋮		⋮	⋮
	r	y_{r1}	y_{12}	- - - -	y_{rk}	$B_{r.}$
Treatment Total		$T'_{.1} + x$	$T_{.2}$	- - - -	$T_{.k}$	$T'_{..} + x$

Performing the analysis of variance, we get the following sums of squares:

Source of Variation	Sum of Squares
Treatments	$\frac{(T_{.1}+x)^2}{r} + \frac{T_{.2}^2}{r} + \cdots + \frac{T_{.t}^2}{r} - \frac{(T_{..}'+x)^2}{rK}$
BLOCKS	$\frac{(B_{1.}'+x)^2}{K} + \frac{B_{2.}^2}{K} + \cdots + \frac{B_{b.}^2}{K} - \frac{(T_{..}'+x)^2}{bK}$
ERROR	By Subtraction
Total	$\sum_{ij} y_{ij}^2 + x^2 - \frac{(T_{..}'+x)^2}{rK}$

Thus the error sum of squares is obtained as

$$SSE = x^2 + \frac{(T'_{..} + x)^2}{kr} - \frac{(T'_{.1} + x)^2}{r} - \frac{(B'_{1.} + x)^2}{k}$$

+ terms which do not involve x.

$$\frac{\partial(SSE)}{\partial x} = 2x + \frac{2(T'_{..} + x)}{kr} - \frac{2(T'_{.1} + x)}{r} - \frac{2(B'_{1.} + x)}{k} \quad (1)$$

$$2x + \frac{2(T'_{..} + x)}{kr} - \frac{2(T'_{.1} + x)}{r} - \frac{2(B'_{1.} + x)}{k} = 0$$

$$x + \frac{T'_{..}}{kr} + \frac{x}{kr} - \frac{T'_{.1}}{r} - \frac{x}{r} - \frac{B'_{1.}}{k} - \frac{x}{k} = 0$$

$$\chi\left[1+\frac{1}{Kn}-\frac{1}{n}-\frac{1}{K}\right]+\frac{\bar{T}_{..}'}{Kr}-\frac{\bar{T}_{.1}'}{n}-\frac{B_{1.}'}{K}=0$$

$$\chi\left(\frac{Kr-1-K-r}{Kr}\right)=\frac{rB_{1.}'+K\bar{T}_{.1}'-\bar{T}_{..}'}{Kr}$$

$$\chi=\frac{rB_{1.}'+K\bar{T}_{.1}'-T_{..}'}{(K-1)(r-1)}$$

The formula in general may be written as

$$x = \frac{rB + kT - G}{(k-1)(r-1)}$$

where

r = number of replicates

k = number of treatments

B = sum of remaining values in the blocks with the missing observations

T = sum of the remaining values of the treatment with missing observation, and

G = Grand total of all observed values.

EXAMPLE

The following data, obtained from a randomized complete block design with 3 treatments A, B, C, and 3 blocks, contain one missing observation represented by x.

		TREATMENTS		
		A	B	C
BLOCKS	I	5	12	15
	II	7	10	14
	III	8	16	x

Estimate the missing observation and prepare a table for analysis of variance.

$$x = \frac{rB + kT - G}{(k-1)(r-1)}$$

$r = 3$, $k = 3$, $B = 8 + 16 = 24$, $T = 15 + 14 = 29$

$G = 87$

$$x = \frac{3(24) + 3(29) - 87}{(3-1)(3-1)} = 18$$

$$\boxed{x = 18}$$

		TREATMENTS		
		A	B	C
BLOCKS	I	5	12	15
	II	7	10	14
	III	8	16	18

(1) Hypothesis

$H_0: \mu_{.1} = \mu_{.2} = \mu_{.3}$

H_1: Not all treatments means are equal

$H_0': \mu_{1.} = \mu_{2.} = \mu_{3.}$

H_1': Not all block means are equal

(2) Level of significance

$\alpha = 0.05$

(3) Test Statistic

$F_1 = \frac{s_t^2}{s_e^2}$ → s_t^2: ms for treatments, s_e^2: ms error

$F_2 = \frac{s_b^2}{s_e^2}$ → s_b^2: ms for blocks, s_e^2: ms error

STA: 408

(4) Calculation

$$C.F = \frac{(T..)^2}{n}$$

$$TSS = \sum_i \sum_j y_{ij}^2 - C.F$$

$$n = rK = 9$$

$$TSS = 158$$

$$C.F = \frac{(105)^2}{9} = 1225$$

$$\text{Treat SS} = \frac{\sum_j T_{.j}^2}{r} - C.F = 126$$

$$BSS = \frac{\sum_i B_{i.}^2}{K} - C.F = 24.7$$

$$ESS = TSS - \text{Treat SS} - BSS$$

$$ESS = 7.3$$

ANOVA-TABLE

Source of Variation	Degrees of freedom	Sum of Squares	Mean Square	F-Test
Treatments	$\nu_1 = K-1 = 3-1 = 2$	$126 = \text{Treat SS}$	$s_t^2 = \frac{126}{2} = 63$	$F_1 = \frac{s_t^2}{s_e^2} = \frac{63}{2.43} = 25.9$
Blocks	$\nu_2 = 2$	$24.7 = \text{BSS}$	$s_b^2 = \frac{\text{BSS}}{\nu_2} = 12.35$	$F_2 = \frac{s_b^2}{s_e^2} = \frac{12.35}{2.43} = 5.08$
Error	$\nu_3 = (K-1)(K-1) = 3$	$7.3 = \text{ESS}$	$s_e^2 = 2.43$	
Total	7	$158 = \text{TSS}$	—	

5. Critical Region

Reject Null hypothesis if $\begin{cases} F_1 \geq F_{0.05(2,3)} \\ F_2 \geq F_{0.05(2,3)} \end{cases}$

$25.9 > 9.55 \longrightarrow$ Reject H_0

$5.08 < 9.55 \longrightarrow$ Do not Reject H_0'

6. Conclusion.

As our cal value F_1 fall in critical Region so, all the Treatment means are not equal.

But the F_2 does not lie in reject region. we conclude that all the block means are equal.

EFFICIENCY OF RCBD RELATIVE TO A CR DESIGN

- ✓ For experimental design, efficiency relates to the ability of a design to achieve the objective with minimal expenditure of resource such as time and money.
- ✓ Its relative efficiency is usually expressed in terms of a ratio of error variances.

$$RE(RCB,CR) = \frac{\text{Estimated MSE for CR design}}{\text{MSE for the RCB design}}$$

$$RE(RCB,CR) = \frac{(r-1)s_b^2 + r(k-1)s_e^2}{(rk-1)s_e^2}$$

MS for blocks

MS for error

EXAMPLE

Given the following analysis of variance for RCBD:

Source of Variation	Degrees of freedom	Sum of Squares	Mean Square
Blocks	9	0.4074	0.0453
Treatments	3	1.1986	0.3995
Error	27	0.6247	0.0231
Total	39	2.2309	-

Estimate the efficiency of this design relative to completely randomized design.

Degrees of freedom	Mean Square
= 9 v − 1	0.0453 S_b^2 =
= 3 k − 1	0.3995 S_t^2 =
27	0.0231 S_e^2 =
= 39 vk − 1	-

$$RE = \frac{(9)(0.0453) + (10)(3)(0.0231)}{(39)(0.0231)}$$

$$RE = \frac{1.1007}{0.9009} = 1.22 = 122\%$$

PRACTICE QUESTIONS

1. Determine the missing observation and compute the analysis of variance.

BLOCKS	TREATMENTS				
	A	B	C	D	E
I	5	6	6	7	4
II	Y_{2A}	7	7	8	5
III	7	7	7	9	5

Yates method

$Y_{2A} = ?$

(1) Hypothesis

$H_0 : \mu_{\cdot A} = \mu_{\cdot B} = \mu_{\cdot C} = \mu_{\cdot D} = \mu_{\cdot E}$

H_1: Not all Treatment means are equal

$H_0' : \mu_{1\cdot} = \mu_{2\cdot} = \mu_{3\cdot}$

H_1' : Not all block means are equal

(2) level of significance $\alpha = 0.05$

(3) Test Statistic

$F_1 =$

$F_2 =$

(4) Calculation

$y_{2A} = ?$

ANOVA Table

$[(K-1)(r-1)] - 1$

Sum of Squares

Treatments = 17.6

Blocks = 5.2

Error = 0.8

Total = 23.6

(5) Critical Region

(6) Conclusion

2 Given the following analysis of variance for Randomized Complete Block Design. Fill in the mean squares. Estimate the efficiency of this design relative to Completely Randomized design.

Source of Variation	Degrees of freedom	Sum of Squares	Mean Square
Treatments	3	134.45	$S_t^2 = 134.45/3$
Blocks	4	21.46	
Error	12	26.26	
Total	19	182.17	-

$RE(RCBD, CRD) = \frac{\text{Estimated MSE for CRD}}{\text{MSE for RCBD}} = 1.30 \text{ or } 130\%$

2 Given the following analysis of variance for Randomized Complete Block Design. Fill in the mean squares. Estimate the efficiency of this design relative to Completely Randomized design.

Source of Variation	Degrees of freedom	Sum of Squares	Mean Square
Treatments	3	134.45	$S_t^2 = 134.45/3 =$ 44.82
Blocks	4	21.46	$S_b^2 = 21.46/4 =$ 5.36
Error	12	26.26	$S_e^2 = 26.26/12 =$ 2.19
Total	19	182.17	-

$$RE(RCBD, CRD) = \frac{(r-1)\sigma_b^2 + r(K-1)\sigma_e^2}{(rK-1)\sigma_e^2}$$

$$= \frac{(5-1)(5.36) + (5)(4-1)(2.19)}{[(5\times4)-1](2.19)}$$

$$= \frac{54.29}{41.61} = 1.30 \text{ or } 130\%$$

This indicates that RCBD is 130% as efficient as the CR design.

THANK YOU

COURSE: STA 408
DESIGN & ANALYSIS OF EXPERIMENT

LECTURE DELIVERED BY

SARAH AHMAD

DEPARTMENT OF STATISTICS

VIRTUAL UNIVERSITY OF PAKISTAN



COURSE: STA 408

DESIGN & ANALYSIS OF EXPERIMENT

THE LATIN SQUARE DESIGN

LECTURE NO.5

DESCRIPTION

When the variation is found in two directions, it becomes necessary to remove these two sources of variation simultaneously. This is done by simultaneous blocking of experimental units in two mutually perpendicular directions called rows and columns. Since each row and each column is a complete block, the grouping for a balanced arrangement is performed by imposing the restriction that each treatment must appear " once and only once" in each row and in each column.

If there are “k” treatments then the experimental area will be divided into “k” rows and “k” columns in k x k plots or experimental units. As the experiment is laid out in a square pattern and the treatments are then assigned at random to plots or experimental units. Such a double blocking of experimental units and a corresponding double random assignment of treatments is called “**Latin Square Design**”.

LAYOUTS

Latin squares are always constructed by rotations, e.g. in the case of 4 treatments A, B, C and D, we get

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

A large number of distinct Latin squares can be derived by interchanging rows and columns of a standard Latin square, where a **standard Latin square** is a Latin square in which the treatments in the first row and first column are arranged in alphabetical or numerical order.

A standard Latin square is also called a reduced Latin square.

- There is only one standard square for a 2 x 2 Latin square as,

A	B
B	A

- And for 3 x 3 Latin square we have also one standard square such as,

A	B	C
B	C	A
C	A	B

- But for a 4 x 4 Latin square, there are four possible standard squares as,

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

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

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

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

Each Latin square of k x k yields k! (k-1)! Different squares by permuting the rows except first and all the k columns. According we may obtain from a square of design.

2 x 2 → 2! (2-1)! = 2

3 x 3 → 3! (3-1)! = 12

4 x 4 → 4! (4-1)! = 576

5 x 5 → 5! (5-1)! = 161280

MATHEMATICAL MODEL

In Latin square design, the observation $Y_{ij(h)}$,

$$Y_{ij(h)} = \mu + R_i + C_j + \tau_h + e_{ij(h)}$$

where $i, j, h = 1, 2, 3, \ldots, k.$

The μ denote the general mean

R_i is the effect of the ith row

C_j the effect of the jth column

τ_h the effect of the hth treatment

$e_{ij(h)}$ is the experimental error, that is assumed to be normally and independently distributed with mean zero and common variance σ^2.

1. formulation of Hypothesis
2. level of significance
3. Test Statistic
4. Calculation
5. Critical Region
6. Conclusion

STATISTICAL ANALYSIS

1-Hypothesis

1

H_0 : $\mu_{\cdot 1} = \mu_{\cdot 2} = \mu_{\cdot 3} \ldots$

H_1 : Not all treatment means are equal

3

H_0' : $\mu_{\cdot 1} - \mu_{\cdot 2} = \mu_{\cdot 3} \ldots$

H_1' : Not all column means are equal

2

H_0'' : $\mu_{1\cdot} = \mu_{2\cdot} = \ldots$

H_1'' : Not all row means are equal

2-Level of Significance

$\alpha = ?$

3-Test Statistics

$$F_1 = \frac{s_t^2}{s_e^2}$$

$$F_2 = \frac{s_r^2}{s_e^2}$$

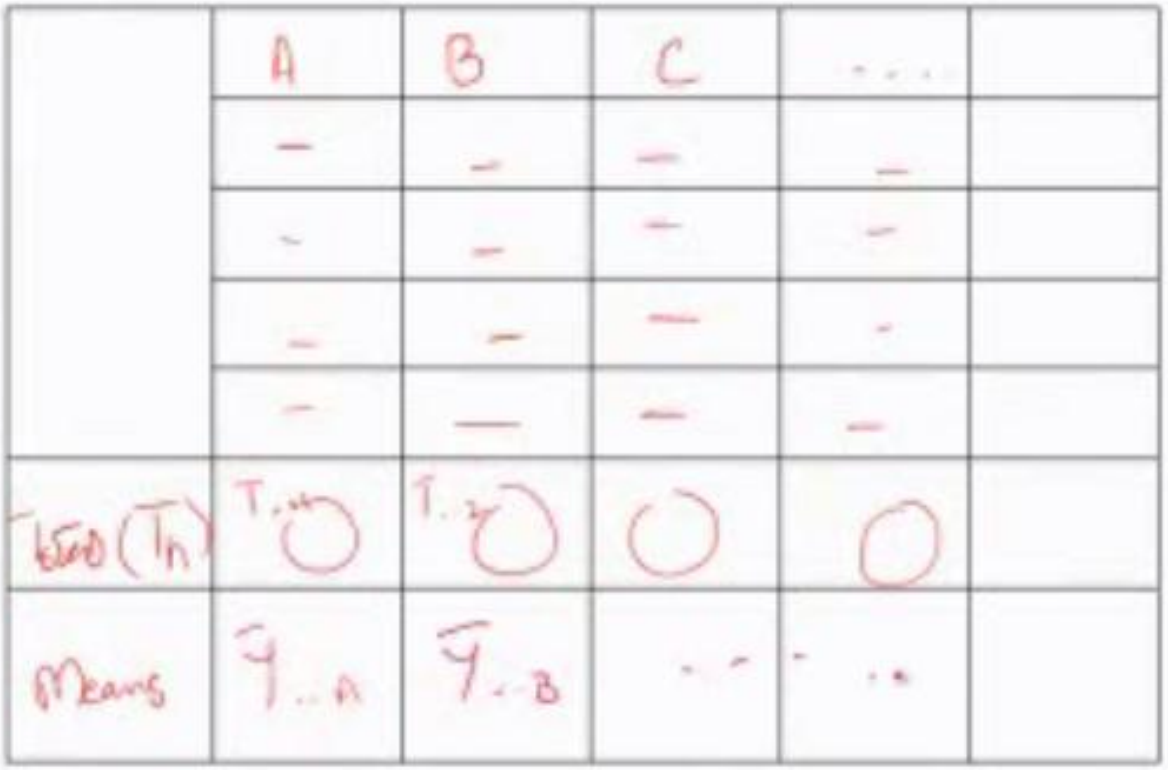
$$F_3 = \frac{s_c^2}{s_e^2}$$

4-Calculation

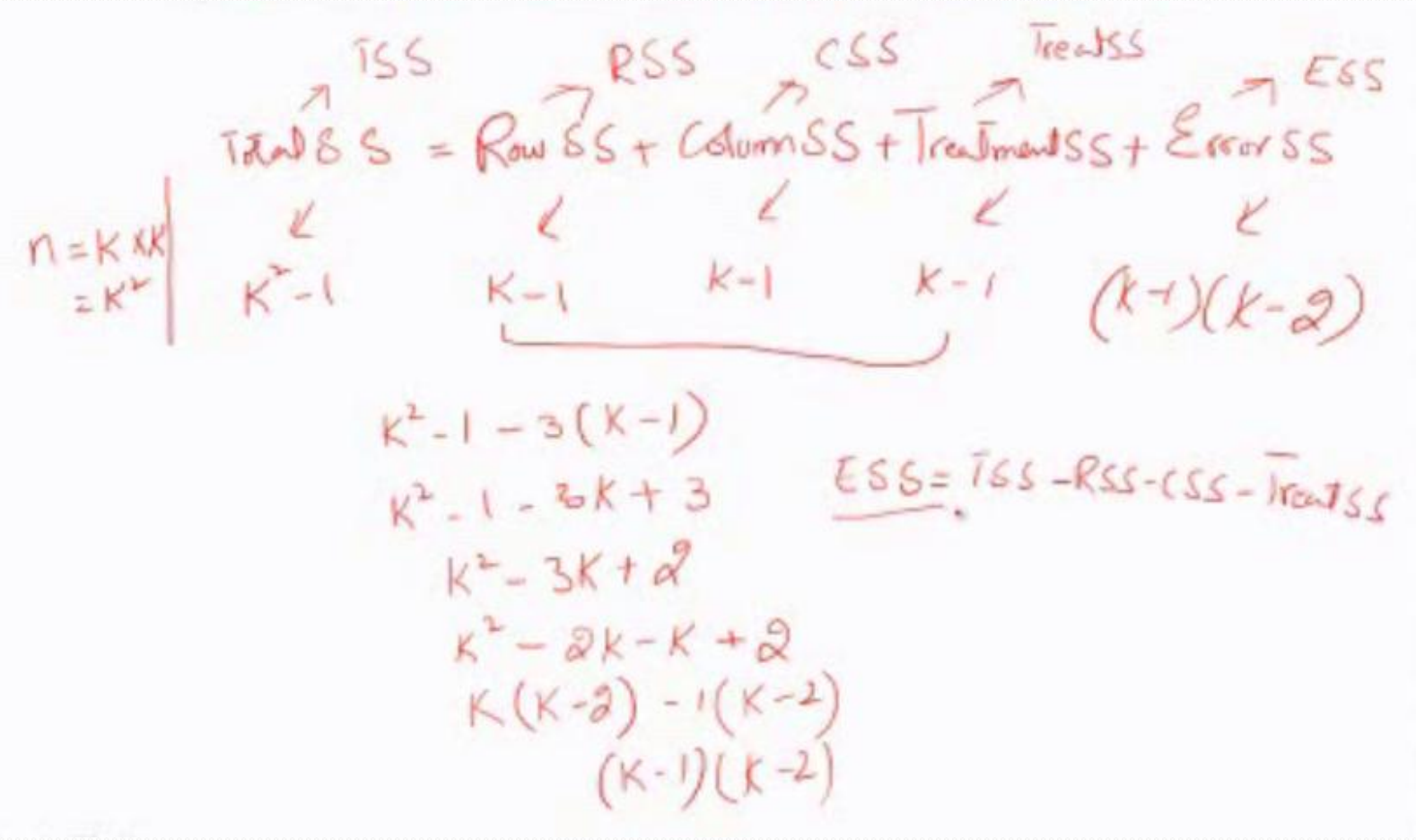
The data obtained from a k x k Latin square design may be arranged in tabular form as,

		COLUMNS						R_i Total	Means
		1	2	3	4	....	K		
ROWS	1	Y_{11A}	Y_{12B}	Y_{13C}	Y_{14D}	....		$R_{1..}$	$\bar{Y}_{1..}$
	2							$R_{2..}$	$\bar{Y}_{2..}$
	3							$R_{3..}$	$\bar{Y}_{3..}$
	4	Y_{41D}	Y_{42A}	Y_{43B}	Y_{44C}				
	K								
Total C_j		$C_{.1.}$	$C_{.2.}$					$G = \bar{Y}_{...}$	
Means		$\bar{Y}_{.1.}$	$\bar{Y}_{.2.}$	$\bar{Y}_{.3.}$					$\bar{Y}_{...}$

COMPUTATIONS FOR THE TREATMENT MEANS



PARTITIONING OF TOTAL SUM OF SQUARES



$$C.F = \frac{(T_{...})^2}{n} = \frac{G^2}{k^2}$$

$$TSS = \sum_i \sum_j \left(Y^2_{ij(h)} \right) - C.F$$

$$RSS = \frac{\sum_i R_i^2}{K} - C.F$$

$$CSS = \frac{\sum_j C_j^2}{K} - C.F$$

$$\text{Treat SS} = \frac{\sum_h T_h^2}{K} - C.F.$$

ANOVA-TABLE for k x k Latin Square Design

Source of Variation	Degrees of freedom	Sum of Squares	Mean Square	F-Test
Between Treatment	$\nu_1 = k-1$	$\text{Treat SS} = \frac{\sum T_h^2}{k} - C.F$	$s_t^2 = \frac{\text{Treat SS}}{\nu_1}$	$F_1 = s_t^2/s_e^2$
B/W Rows	$\nu_2 = k-1$	$RSS = \frac{\sum R_i^2}{k} - C.F$	$s_R^2 = \frac{RSS}{\nu_2}$	$F_2 = s_R^2/s_e^2$
B/W Columns	$\nu_3 = k-1$	$CSS = \frac{\sum C_j^2}{k} - C.F$	$s_C^2 = \frac{CSS}{\nu_3}$	$F_3 = s_C^2/s_e^2$
Within Errors	$\nu_4 = (k-1)(k-2)$	$TSS - \text{Treat SS} - CSS - RSS = ESS$	$s_e^2 = \frac{ESS}{\nu_4}$	
Total	k^2-1	$\sum_i \sum_j (Y_{ij(h)}^2) - CF$	—	

5-Crirtical region

Reject the null hypotheses when

$$F_1 \geq F_{\alpha,(\nu_1,\nu_4)}$$

$$F_2 \geq F_{\alpha,(\nu_2,\nu_4)}$$

$$F_3 \geq F_{\alpha,(\nu_3,\nu_4)}$$

6-Conclusion

EXAMPLE

The atmosphere in 4 different districts of a large town was sampled, the samples being taken at 4 different heights. Four different tests for the presence of a certain chemical were made on the samples. The arrangement is shown in the following table with the percentage by weight of the chemical as determined by the tests. Letter denote the difference tests.

		DISTRICTS			
		1	**2**	**3**	**4**
HEIGHTS	**1**	**A** 8	**B** 5.3	**C** 4.1	**D** 5
	2	**D** 6.8	**A** 4.9	**B** 4.1	**C** 3.2
	3	**B** 6.3	**C** 4.7	**D** 4.0	**A** 5
	4	**C** 5.7	**D** 3.3	**A** 4.0	**B** 4.2

Is there any evidence of significant variation from district to district and between heights in the percentage of the chemical present in the atmosphere..

1-Hypothesis

H_0 : $\mu_{\cdot 1} = \mu_{\cdot 2} = \mu_{\cdot 3} = \mu_{\cdot 4}$

H_1 : Not all treatments means are equal.

H_0' : $\mu_{1\cdot} = \mu_{2\cdot} = \mu_{3\cdot} = \mu_{4\cdot}$

H_1' : Not all row means are equal

H_0'' : $\mu_{\cdot 1} = \mu_{\cdot 2} = \mu_{\cdot 3} = \mu_{\cdot 4}$

H_1'' : Not all column means are equal.

2-Level of Significance

$\alpha = 0.05$

3-Test Statistic

$$F_1 = \frac{s_t^2}{s_e^2}$$ — Tests

$$F_2 = \frac{s_r^2}{s_e^2}$$ — Weights

$$F_3 = \frac{s_c^2}{s_e^2}$$ — Districts

4-Calculations

HEIGHTS	DISTRICTS				R_i	R_i^2
	1	2	3	4		
1	A 8	B 5.3	C 4.1	D 5	22.4	501.76
2	D 6.8	A 4.9	B 4.1	C 3.2	19.0	361.00
3	B 6.3	C 4.7	D 4.0	A 5	20.0	400
4	C 5.7	D 3.3	A 4.0	B 4.2	17.2	295.84
C_j	26.8	18.2	16.2	17.4	78.6	1558.60
C_j^2	718.4	331.24	262.44	302.76	1614.68	—
$\sum\sum r_{ij(ri)}$	182.42	85.48	65.62	77.88	411.00	

	TESTS				
	A	B	C	D	
	8	5.3	4.1	5	
	4.9	4.1	3.2	6.8	
	5	6.3	4.7	4.0	
	4.0	4.2	5.7	3.3	Total
T_n	81.9	19.9	17.7	19.1	78.6
T_n^2	479.61	396.01	313.29	364.81	1553.72

$$TSS = \sum_i \sum_j y_{ij(h)}^2 - C.F \qquad C.F = \frac{G^2}{K^2} = \frac{(78.6)^2}{16}$$

$$= 411.00 - 386.1225 \qquad C.F = 386.1225$$

$$= 24.8775$$

$$RSS = \frac{\sum R_i^2}{K} - C.F = \frac{1558.60}{4} - 386.1225$$

$$RSS = 3.5275$$

Table

$$CSS = \frac{\sum C_j^2}{K} - C.F = \frac{1614.68}{4} - 386.1225$$

$$CSS = 17.5475$$

$$TreatSS = \frac{\sum T_h^2}{K} - C.F = \frac{1553.72}{4} - 386.1225$$

$$TreatSS = 2.3075$$

$$ESS = TSS - RSS - CSS - TreatSS$$

$$ESS = 1.4950$$

ANOVA-TABLE for 4 x 4 Latin Square Design

Source of Variation	Degrees of freedom	Sum of Squares	Mean Square	F-Test
B/W Treatments	$v_1 = 4-1=3$	Treat SS = 2.3075	$s_t^2 =$ 5.8492	$F_1 = 23.47$
B/W Rows	$v_2 = 3$	RSS = 3.5275	$s_r^2 =$ 1.1758	$F_2 = 4.72$
B/W Columns	$v_3 = 3$	CSS = 17.5475	$s_c^2 =$ 0.7692	$F_3 = 3.09$
Error	$v_4 = 6$	1.4950	$s_e^2 =$ 0.2490	
Total	15	24.8725		

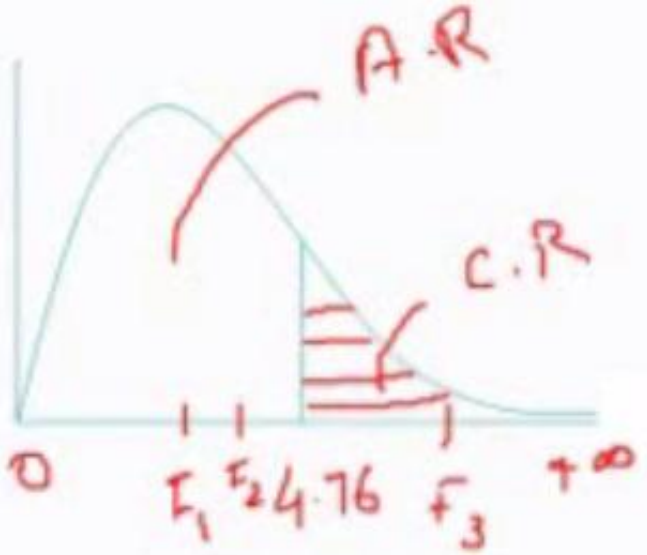
5-Crirtical Region

Reject Null hypothesis if

3.09 $\leftarrow F_1 > F_{0.05(3,6)} = 4.76$

4.72 $\leftarrow F_2 > F_{0.05(3,6)} = 4.76$

23.47 $\leftarrow F_3 > F_{0.05(3,4)} = 4.76$



F-table

6-Conclusion

By comparing the computed values of *F* with the corresponding critical *F*, we conclude that the variation among the districts is significant and the variation among the heights is insignificant. Also, there is no difference between the sensitivity of the test.

THANK YOU

COURSE: STA 408
DESIGN & ANALYSIS OF EXPERIMENT

LECTURE DELIVERED BY

SARAH AHMAD
DEPARTMENT OF STATISTICS
VIRTUAL UNIVERSITY OF PAKISTAN



COURSE: STA 408

DESIGN & ANALYSIS OF EXPERIMENT

LECTURE NO.6

ADVANTAGES & DISADVANTAGES

- ✓ A Latin square design is generally more efficient than RCBD.
- ✓ A Latin square design reduces the error variance by controlling two sources of variation.
- ✓ The analysis of a Latin square design is simple and remains relatively simple with missing observations.

- ✓ A Latin square design is less flexible than RCBD. It is practical only for 5 to 10 treatments. When the number of treatments exceeds 10, the design is seldom used.
- ✓ For a small number of treatments, a Latin Square design does not provide a sufficient number of replicates to give a valid estimate of error.
- ✓ Replication in a Latin square design is costly.
- ✓ In agriculture experimentation, the land requirement is rigid, the actual layout in the field may be laborious and approach to the central most plot becomes difficult.

MISSING OBSERVATIONS IN LATIN SQUARE DESIGN

STA: 408

Suppose,

		COLUMNS						Total
		1	2	3	4	.....	k	
ROWS	1	x	Y_{12B}	Y_{13C}	Y_{14D}	.....	- - -	$R_1' + x$
	2					.....		R_2
	3							R_3
	4	Y_{41D}	Y_{42A}	Y_{43B}	Y_{44C}	- - -	- - -	R_4
	k							
Total		$C_1' + x$	C_2	C_3	C_4	- - -	- - -	$G' + x$

	TREATMENTS				
	A	B	C	.....	*k*
	—	—	—	.....	—
	—	—	—	.....	—
Total	$T_1' + 1$	T_2	T_3	- - - -	T_k

Performing the analysis of variance, we get the following sums of squares:

Source of Variation	Sum of Squares
Rows	$\frac{(R_1'+\gamma)^2+\sum_{i=2}^{K} R_i^2}{K} - \frac{(G'+\gamma)^2}{K^2} = RSS$
Columns	$\frac{(C_1'+\gamma)^2+\sum_{j=2}^{K} C_j^2}{K} - \frac{(G'+\gamma)^2}{K^2} = CSS$
Treatments	$\frac{(T_1'+\gamma)^2+\sum_{k=2}^{K^2} T_k^2}{K} - \frac{(G'+\gamma)^2}{K^2} = TrSS$
Error	By Subtraction
Total	$\gamma^2+\sum_i\sum_j Y_{ij}^2 - \frac{(G'+\gamma)^2}{K^2} = TSS$

Thus the error sum of squares is obtained as

$$ESS = TSS - RSS - CSS - TrSS$$

$$ESS = x^2 + \sum_i \sum_j y_{ij}^2 - \frac{(G'+x)^2}{k^2} - \frac{(R_1'+x)^2}{k} - \frac{\sum_{i=2}^{k} R_i^2}{k} + \frac{(G'+x)^2}{k^2}$$

$$- \frac{(C_1'+x)^2}{k} - \frac{\sum_{j=2}^{k} C_j^2}{k} + \frac{(G'+x)^2}{k^2}$$

$$- \frac{(T_1'+x)^2}{k} - \frac{\sum_{h=2}^{k} T_h^2}{k} + \frac{(G'+x)^2}{k^2}$$

$$ESS = x^2 + \frac{2(G_1' + x)^2}{k^2} - \frac{1}{k}\left[(R_1' + x)^2 + (C_1' + x)^2 + (T_1' + x)^2\right]$$

+ Terms in which x is not involved

Differentiate ESS w.r.t x,

$$\frac{\partial(ESS)}{\partial x} = 2x + \frac{(2\times 2)(G_1' + x)}{k^2} - \frac{1}{k}\left[2(R_1' + x) + 2(G_1' + x) + 2(T_1' + x)\right]$$

$$2x + \frac{4(G_1' + x)}{k^2} - \frac{2}{k}\left[(R_1' + x) + (C_1' + x) + (T_1' + x)\right] = 0$$

$$2x + \frac{4G'}{K^2} + \frac{4x}{K^2} - \frac{2R_1'}{K} - \frac{2x}{K} - \frac{2C_1'}{K} - \frac{2x}{K} - \frac{2\bar{T}_1'}{K} - \frac{2x}{K} = 0$$

$$2x + \frac{4x}{K^2} - \frac{2x}{K} - \frac{2x}{K} - \frac{2x}{K} - \frac{2R_1'}{K} - \frac{2C_1'}{K} - \frac{2\bar{T}_1'}{K} + \frac{4G'}{K^2} = 0$$

$$2x\left(1 + \frac{2}{K^2} - \frac{1}{K} - \frac{1}{K} - \frac{1}{K}\right) = \frac{2}{K}\left(R_1' + C_1' + \bar{T}_1'\right) - \frac{4G'}{K^2}$$

$$x\left(1 + \frac{2}{K^2} - \frac{3}{K}\right) = \frac{2}{2K}\left(R_1' + C_1' + \bar{T}_1'\right) - \frac{4^2}{2}\frac{G'}{K^2}$$

$$\chi\left(\frac{K^2-3K+2}{K^2}\right) = \frac{1}{K}\left(R_1' + C_1' + T_1'\right) - \frac{2G'}{K^2}$$

$$\chi\left(\frac{K^2-K-2K+2}{K^2}\right) = \frac{K\left(R_1' + C_1' + T_1'\right) - 2G'}{K^2}$$

$$\chi\left(\frac{K(K-1)-2(K-1)}{K^2}\right) = \text{''}$$

$$\chi\left\{\frac{(K-1)(K-2)}{K^2}\right\} = \text{''}$$

$$\chi = \frac{K(R_1' + C_1' + T_1') - 2G'}{K^2} \times \frac{K^2}{(K-1)(K-2)}$$

$$\chi = \frac{K(R_1' + C_1' + T_1') - 2G'}{(K-1)(K-2)}$$

$$x = \frac{k(R_i' + C_j' + T_h') - 2G'}{(k-1)(k-2)}$$

no. of treatments
Row total (missing obs)
column total (missing obs)
Grand total
no. of treatments

$$x = \frac{K(R+C+T) - 2G}{(K-1)(K-2)}$$

EXAMPLE

An agriculture experiment was conducted on the Latin square plan to test the effect on yield due to change of treatment (5 kinds) and also to variation of soil in each of two perpendicular directions. The results are set out in the Latin square below, in which letters correspond to treatments, while rows and columns correspond to the two perpendicular directions.

A 7.4	D x	E 5.8	B 12.0	C 14.3
C 11.8	B 6.5	A 8.7	E 7.6	D 7.9
D 10.1	C 17.9	B 9.0	A 8.5	E 7.1
E 8.8	A 10.1	C 15.7	D 11.1	B 7.4
B 11.8	E 8.8	D 14.3	C 18.4	A 10.1

Estimate the missing observation and perform an analysis of variance.

A 7.4	D x	E 5.8	B 12.0	C 14.3	R=39.5
C 11.8	B 6.5	A 8.7	E 7.6	D 7.9	
D 10.1	C 17.9	B 9.0	A 8.5	E 7.1	
E 8.8	A 10.1	C 15.7	D 11.1	B 7.4	
B 11.8	E 8.8	D 14.3	C 18.4	A 10.1	
	C=43.3				G=251.1

Treatment D :	7.9	10.1	11.1	14.3	43.4

$$x = \frac{k(R + C + T) - 2G}{(k-1)(k-2)}$$

$K = 5$

$$x = \frac{(5)(39.5 + 43.3 + 43.4) - 2(251.1)}{(5-1)(5-2)}$$

$$x = \frac{631 - 502.2}{12} = \frac{128.8}{12} = 10.7333$$

$x = 10.7$

① Hypotheses

H_0: All effects on yield due to variation of soil in columns are zero.

H_1: - - - - - - - - are not zero.

H_0': - - - - - - - - rows are zero.

H_1': - - - - - - - - - rows are not equal to zero.

H_0'': All effects on yield due to change in treatments are zero.

H_1'': - - - - - - are not equal to zero.

	COLUMNS						
ROWS	1	2	3	4	5	R_i	R_i^2
1	A 7.4	D 10.7	E 5.8	B 12.0	C 14.3	50.2	2520.04
2	C 11.8	B 6.5	A 8.7	E 7.6	D 7.9	42.5	1806.25
3	D 10.1	C 17.9	B 9.0	A 8.5	E 7.1	52.6	2766.76
4	E 8.8	A 10.1	C 15.7	D 11.1	B 7.4	53.1	2819.615
5	B 11.8	E 8.8	D 14.3	C 18.4	A 10.1	63.4	4019.56
C_j	49.9	54	53.5	57.6	46.8	T... 261.8	13932.22
C_j^2	2490.01	2916	2862.25	3317.76	2190.24	13776.26	
$\sum\sum y_{ij}(n)$	512.69	656.6	641.31	735.78	474.08	3020.46	

	TREATMENTS					
	A	B	C	D	E	
	7.4	12	14.3	10.7	5.8	
	8.7	6.5	11.8	7.9	7.6	
	8.5	9	17.9	10.1	7.1	
	10.1	7.4	15.7	11.1	8.8	
	10.1	11.8	18.4	14.3	8.8	
T_h^2	2007.04	2180.89	6099.61	2926.81	1451.61	14665.96

$$C.F = \frac{(T_{...})^2}{k^2} = \frac{(961.8)^2}{25} = 2741.5696$$

$$TSS = \sum_i \sum_j Y_{ij}^2(h) - C.F = 3020.46 - 2741.5696$$

$$TSS = 278.8904$$

$$CSS = \frac{\sum_j C_j^2}{k} - C.F = \frac{13776.26}{5} - 2741.5696$$

$$CSS = 13.6824$$

$$RSS = \frac{\sum_i R_i^2}{k} - C.F = \frac{13932.22}{5} - 2741.5696$$

$$RSS = 44.8744$$

ANOVA-TABLE

Source of Variation	Degrees of freedom	Sum of Squares	Mean Square	F-Test
Columns	$K-1 = 5-1 = 4$	13.6824	$S_c^2 = 3.4206$	$F_1 = 1.3105$
Rows	4	44.8744	$S_r^2 = 11.2186$	$F_2 = 4.2981$
Treatments	4	191.6224	$S_t^2 = 47.9056$	$F_3 = 18.3539$
Error	$(K-1)(K-2)-1 = 11$	28.7114	$S_e^2 = 2.6101$	
Total	23	278.8904	–	–

(5) Critical Region

Reject null hypothesis if $F_{cal} > F_{tab}$

$F_{0.05(4,11)} = 3.36$



(6) Conclusion

As $F_1 < F_{tab}$, we do not reject H_0

$F_2 > F_{tab}$, H_0' is rejected

$F_3 > F_{tab}$, H_0'' reject

(a) The variation of soil in columns is not significant

(b) There are significant differences in soil in rows.

(c) The effects on yield due to change in treatments are significantly different.

EXAMPLE

The following data represent the yield (in tons per ha) of four cultivars of Khaya arranged in a 4 x 4 Latin square with one missing value. The four cultivars are A (KH4), B (KH2O), C (KH29) and D (KS15).

	COLUMNS			
ROWS	1	2	3	4
I	6.2 (C)	5.1 (D)	3.3 (B)	5.7 (A)
II	5.1 (B)	3.8 (A)	*M* (C)	3.9 (D)
III	5.4 (D)	4.8 (C)	6.9 (A)	4.6 (B)
IV	3.2 (A)	2.9 (B)	5.2 (D)	5.4 (C)

Estimate the missing observation and perform the analysis of variance to test for significant differences among the cultivars.

① Hypothesis

H_0: There is no significant difference in the yield of the cultivars

H_1: There is significant .....

② $\alpha = 0.05$

③ $F = \frac{s_t^2}{s_e^2}$

④ Calculation

⑤ Critical Region

⑥ Conclusion

	COLUMNS				
ROWS	1	2	3	4	Total
I	6.2 (C)	5.1 (D)	3.3 (B)	5.7 (A)	
II	5.1 (B)	3.8 (A)	M (C)	3.9 (D)	R = 12.8
III	5.4 (D)	4.8 (C)	6.9 (A)	4.6 (B)	
IV	3.2 (A)	2.9 (B)	5.2 (D)	5.4 (C)	
Total			C = 15.4		G = 71.4

K = 4

Treatment C :	6.2	M	4.8	5.4	16.4

Total

$$\hat{\tau} = \frac{K(R+C+T) - 2G}{(K-1)(K-2)} = \frac{(4)(12.8+15.4+16.4) - 2(71.4)}{(4-1)(4-2)}$$

$$\hat{\tau} = 5.9$$

ROWS	COLUMNS			
	1	2	3	4
I	6.2 (C)	5.1 (D)	3.3 (B)	5.7 (A)
II	5.1 (B)	3.8 (A)	5.9 (C)	3.9 (D)
III	5.4 (D)	4.8 (C)	6.9 (A)	4.6 (B)
IV	3.2 (A)	2.9 (B)	5.2 (D)	5.4 (C)

Sum of Squares

RSS = 3.4675

CSS = 2.9325

TrSS = 5.1825

ESS = 8.115

TSS = 19.6975

C.F = 374.4225


THANK YOU
STA: 408

COURSE: STA 408
DESIGN & ANALYSIS OF EXPERIMENT

LECTURE DELIVERED BY

SARAH AHMAD

DEPARTMENT OF STATISTICS

VIRTUAL UNIVERSITY OF PAKISTAN



COURSE: STA 408
DESIGN & ANALYSIS OF EXPERIMENT

LECTURE NO.7

EFFICIENCY OF LATIN SQUARE DESIGN

- **Relative to RCBD**

When taking rows as blocks

$$\frac{1}{k}\left[\frac{s_c^2}{s_e^2}+(k-1)\right] = RE(LS, RCBD_{rows})$$

ms for columns; MS Error; no of rows

When taking columns as blocks

$$\frac{1}{k}\left[\frac{s_r^2}{s_e^2}+(k-1)\right] = RE(LS, RCBD_{cols})$$

MS rows; MS error

- **Relative to CR Design**

MS rows, MS columns, MS error

$$RE(LS, CR) = \frac{1}{k}\left[\frac{s_r^2 + s_c^2 + (k-1)s_e^2}{(k+1)s_e^2}\right]$$

EXAMPLE

Consider an experiment to investigate the effect of four different diets on milk production of cows. There are four cows in the study. During each lactation period the cows receive a different diet. Assume that there is a washout period between diets so that previous diet does not affect future results. Lactation period and cows are used as blocking variables.

A 4 *x* 4 Latin square design is implemented.

	Cow			
Period	1	2	3	4
1	A=38	B=39	C=45	D=41
2	B=32	C=37	D=38	A=30
3	C=35	D=36	A=37	B=32
4	D=33	A=30	B=35	C=33

=49.06 18.2292 0.8125

The mean square period, mean square cows and mean square error are 49.06, 18.2292 and 0.8125 respectively. Estimate the relative efficiency of Latin square design to RCBD

i) by taking the cows as blocks

ii) by taking lactation periods as blocks

① by taking the columns as blocks

$$Re(LS, RCBD_{cols}) = \frac{1}{k}\left[\frac{\sigma_r^2}{\sigma_e^2} + (k-1)\right]$$

$$= \frac{1}{4}\left[\frac{49.06}{0.8125} + (4-1)\right] = \frac{1}{4}\left[\frac{49.06}{0.8125} + 3\right]$$

$$= 15.8454$$

$$= 15.85$$

(ii) by Taking rows as blocks

$$RE(LS, RCBD_{rows}) = \frac{1}{K}\left[\frac{\sigma_c^2}{\sigma_e^2} + (K-1)\right]$$

$$= \frac{1}{4}\left[\frac{18.2292}{0.8125} + 3\right] = 6.36$$

EXAMPLE

Consider the given ANOVA table for 5 brands of air cleaner in 5x5 Latin Square and find the efficiency of this design relative to CR design.

Source of variation	Degrees of freedom	Sum of squares	Mean Square	F
Rows	4	44512.56	11128.14	3.85
Columns	4	2177.76	544.44	0.19
Treatments	4	58034.16	14508.54	5.02
Error	12	34647.52	2887.29	
Total	24	139372.00		

$$RE(LS, CR) = \frac{\sigma_\beta^2 + \sigma_\tau^2 + (K-1)\sigma_e^2}{(K+1)\sigma_e^2}$$

$$= \frac{11128.14 + 544.44 + (5-1)(2887.29)}{(5+1)(2887.29)} = 1.34 = 134\%$$

That is, approximately 34% more observations per treatment would be required in a completely randomized design to obtain the same precision for estimating the treatment means as with this Latin square design.

MULTIPLE COMPARISON TESTS

If F -test after the analysis of variance rejects our null hypothesis, all we can conclude is that the k population means are not all equal. The conclusion might not be sufficient to satisfy the experimenter, rather he would like to know which means (or set of means) might differ significantly from each other.

For this purpose, several tests based on different view points, have been developed to make comparisons between pair of means. Such tests, used as a follow-up to F - tests are called the **multiple comparison tests**.

MOST COMMONLY USED TESTS

- Fisher's Least Significance Difference (or LSD) test
- The Student Newman-keul's Multiple Range test
- The Duncan's Multiple Range test
- The Scheffe's test
- The Tukey's *T*-method

LSD-TEST

When the null hypothesis of equal means is rejected by *F*-test after the ANOVA, we can test the significance of differences between means of k samples (or treatments) by computing the smallest difference that would be judge significant and to compare the absolute values of all differences of means with it. This smallest difference is called the least significant difference or *LSD* and is given by

$$LSD = t_{\alpha/2,(\nu)}\sqrt{\frac{2(\text{MSE})}{r}}$$

Two means are declared to come from populations with significantly different means, when the absolute value of their difference exceeds the *LSD*.

EXAMPLE

Five fertilizers A, B, C, D, and E were tested by arranging plants in a Latin Square design in a field. The rows and columns in the table are rows and columns in the field. The yields in bushels per plot are shown.

		COLUMNS				
		1	2	3	4	5
ROWS	1	B 4.9	D 6.4	E 3.3	A 9.5	C 11.8
	2	C 9.3	A 4.0	B 6.2	E 5.1	D 5.4
	3	D 7.6	C 15.4	A 6.5	B 6.0	E 4.6
	4	E 5.3	B 7.6	C 13.2	D 8.6	A 4.9
	5	A 9.3	E 6.3	D 11.8	C 15.9	B 7.6

$\alpha < 0.05$

Analyze the data for evidence at the 5 percent level that the mean yields are not equal for the five fertilizers. Also pick up the best fertilizer.

STA: 408

1-Hypothesis

H_0 : The mean yields for all the five fertilizers are equal.

H_1 : Not all mean yields for the fertilizers are equal.

2-Level of significance

$\alpha = 0.05$

3-Test Statistic

$$F = \frac{MS\ for\ fertilizers}{MS\ for\ error} = \frac{s_t^2}{s_e^2}$$

4-Calculations

ROWS	COLUMNS					R_i	R_i^2
	1	2	3	4	5		
1	B 4.9	D 6.4	E 3.3	A 9.5	C 11.8	35.9	1288.81
2	C 9.3	A 4.0	B 6.2	E 5.1	D 5.4	30	900
3	D 7.6	C 15.4	A 6.5	B 6.0	E 4.6	40.1	1608.10
4	E 5.3	B 7.6	C 13.2	D 8.6	A 4.9	39.6	1568.16
5	A 9.3	E 6.3	D 11.8	C 15.9	B 7.6	50.9	2590.81
C_j	36.4	39.7	41.0	45.1	34.3	196.5	7955.79
C_j^2	1324.96	1576.09	1681.00	2034.01	1176.49	7792.55	

$$\sum_i \sum_j y_{ij(h)}^2 = 1829.83$$

STA: 408

	FERTILIZERS				
	A	B	C	D	E
	9.5	4.9	11.8	6.4	3.3
	4.0	6.2	9.3	5.4	5.1
	6.5	6.0	15.4	7.6	4.6
	4.9	7.6	13.2	8.6	5.3
	9.3	7.6	15.9	11.8	6.3
Total (T_h)	34.2	32.3	65.6	39.8	24.6
Means	6.84	6.46	13.12	7.96	4.92

$$TSS = \sum_i \sum_j Y_{ij(lh)}^2 - C.F$$

$$C.F = \frac{G^2}{K^2}$$

$$RSS = \frac{\sum_i R_i^2}{K} - C.F$$

$$CSS = \frac{\sum_j C_j^2}{K} - C.F$$

$$ESS = TSS - RSS - CSS - TrtSS$$

$$TrtSS = \frac{\sum_h T_h^2}{K} - C.F$$

ANOVA-TABLE for 5 x 5 Latin Square Design

Source of Variation	Degrees of freedom	Sum of Squares	Mean Square	*F*-Test
Rows	4	46.67	11.67	
Columns	4	14.02	3.51	
Treatments (Fertilizers)	4	196.61	49.15	$F = \frac{49.15}{2.34} = 21.00$
Error	12	28.04	2.34	
Total	24	285.34		

STA: 408

5-Crirtical region

Reject H_0 if $F \geq F_{0.05(4,12)}$

$F \geq 3.26$

$F_{cal} = 21.00$

$21.00 > 3.26$

F-table

6-Conclusion

Since the computed value of F = 21.00 falls in the critical region, we therefore reject our null hypothesis. Hence we conclude that the data provide sufficient evidence to indicate at the 5% significance level that the mean yields for the fertilizers are not equal.

LEAST SIGNIFICANCE DIFFERENCE (*LSD*)

Now, we compute the Least Significance Difference (LSD) to determine which fertilizer is to be considered the best one.

A	B	C	D	E
$\bar{y}_A$	$\bar{y}_B$	$\bar{y}_C$	$\bar{y}_D$	$\bar{y}_E$
6.84	6.46	13.12	7.96	4.92

$$LSD = t_{\alpha/2,\nu}\sqrt{\frac{2MSE}{R}}$$

$$LSD = t_{0.025,12}\sqrt{\frac{2(2.34)}{5}}$$

$$LSD = 2.18(0.96) = 2.09$$

$^{K}C_2 = {}^{5}C_2 = 10$ pairs

PAIRS			LSD-Value	Decision
$\lvert\bar{Y}_A - \bar{Y}_B\rvert$	$= \lvert 6.84 - 6.46\rvert = 0.38$	$<$	2.09	Insignificant
$\lvert\bar{Y}_A - \bar{Y}_C\rvert$	$= \lvert 6.84 - 13.12\rvert = 6.28$	$>$	2.09	– Significant
$\lvert\bar{Y}_A - \bar{Y}_D\rvert$	$= \lvert 6.84 - 7.96\rvert = 1.12$	$<$	2.09	Insignificant
$\lvert\bar{Y}_A - \bar{Y}_E\rvert$	$= \lvert 6.84 - 4.92\rvert = 1.92$	$<$	2.09	"
$\lvert\bar{Y}_B - \bar{Y}_C\rvert$	$= \lvert 6.46 - 13.12\rvert = 6.66$	$>$	2.09	– Significant
$\lvert\bar{Y}_B - \bar{Y}_D\rvert$	$= \lvert 6.46 - 7.96\rvert = 1.5$	$<$	2.09	Insignificant
$\lvert\bar{Y}_B - \bar{Y}_E\rvert$	$= \lvert 6.46 - 4.92\rvert = 1.54$	$<$	2.09	"
$\lvert\bar{Y}_C - \bar{Y}_D\rvert$	$= \lvert 13.12 - 7.96\rvert = 5.16$	$>$	2.09	– Significant
$\lvert\bar{Y}_C - \bar{Y}_E\rvert$	$= \lvert 13.12 - 4.92\rvert = 8.2$	$>$	2.09	– "
$\lvert\bar{Y}_D - \bar{Y}_E\rvert$	$= \lvert 7.96 - 4.92\rvert = 3.04$	$>$	2.09	– "

STA: 408

EXAMPLE

Perform the analysis of variance on the following data and analyze the treatment means using the LSD test with a 0.05 level of significance.

	TREATMENTS					
BLOCKS	**1**	**2**	**3**	**4**	**5**	**6**
1	1	3	6	4	3	2
2	1	4	4	8	5	1
3	3	6	7	8	4	3
4	2	3	2	3	2	1

1-Hypothesis

$$H_0 : \mu_{.1} = \mu_{.2} = \mu_{.3} = \mu_{.4} = \mu_{.5} = \mu_{.6}$$

H_1 : Not all six treatment means are equal.

2-Level of significance

$\alpha = 0.05$

3-Test Statistic

$$F = \frac{\text{MS for Treatments}}{\text{MS for Error}} = \frac{s_t^2}{s_e^2}$$

4-Calculations

Blocks	Treatments					$B_{i\cdot}$	$B_{i\cdot}^2$
	1	2	3	4	5		
1	1	3	6	4	3		
2	1	4	4	8	5		
3	3	6	7	8	4		
4	2	3	2	3	2		
$T_{\cdot j}$							
$T_{\cdot j}^2$							
Means							
$\sum_i \sum_j y_{ij}^2$							

ANOVA-TABLE

Source of Variation	Degrees of freedom	Sum of Squares	Mean Square	*F*-Test
Between Treatments	5	51.83	10.37	$F = \frac{10.37}{1.57} = 6.61$
Between blocks	3	98.50	9.50	
Error	15	23.50	1.57	
Total	23	103.83	–	

5-Crirtical region

Reject H_0 if $F \geq F_{0.05(5,15)}$ (5: d.f for treatment, 15: for error)

$F_{cal} = 6.61$

$6.61 > 2.90$

$F_{0.05(5,15)} = 2.90$

6-Conclusion

We reject the null hypothesis of equal treatment means as the computed value of $F = 6.61$ exceeds the table value. As the null hypothesis is rejected, using F - test, we therefore apply the *LSD* test to find out which means differ from each other.

The Least Significant Difference (*LSD*) is given by the relation

$$LSD = t_{\alpha/2,(\nu)}\sqrt{\frac{2(\text{MSE})}{r}}$$

$$LSD = t_{0.025,(15)}\sqrt{\frac{2(1.57)}{4}} = 1.89$$

$$LSD = 1.89$$

Pairs				*LSD*-Value	Decision
$\lvert\bar{y}_1 - \bar{y}_2\rvert$	$\lvert 1.75 - 4.00\rvert$	$= 2.25$	$>$	1.89	Significant
$\lvert\bar{y}_1 - \bar{y}_3\rvert$	$\lvert 1.75 - 4.75\rvert$	$= 3$	$>$	1.89	//
$\lvert\bar{y}_1 - \bar{y}_4\rvert$	$\lvert 1.75 - 5.75\rvert$	$= 4$	$>$	1.89	//
$\lvert\bar{y}_1 - \bar{y}_5\rvert$	$\lvert 1.75 - 3.50\rvert$	$= 1.75$	$<$	1.89	Insignificant
$\lvert\bar{y}_1 - \bar{y}_6\rvert$	$\lvert 1.75 - 1.75\rvert$	$= 0$	$<$	1.89	//
$\lvert\bar{y}_2 - \bar{y}_3\rvert$				1.89	
$\lvert\bar{y}_2 - \bar{y}_4\rvert$				1.89	
$\lvert\bar{y}_2 - \bar{y}_5\rvert$				1.89	
$\lvert\bar{y}_2 - \bar{y}_6\rvert$				1.89	
$\lvert\bar{y}_3 - \bar{y}_4\rvert$				1.89	
$\lvert\bar{y}_3 - \bar{y}_5\rvert$				1.89	
$\lvert\bar{y}_3 - \bar{y}_6\rvert$				1.89	
$\lvert\bar{y}_4 - \bar{y}_5\rvert$				1.89	
$\lvert\bar{y}_4 - \bar{y}_6\rvert$				1.89	
$\lvert\bar{y}_5 - \bar{y}_6\rvert$				1.89	

$^6C_2 = 15$ Pairs

EXAMPLE

The effects of four types of graphite coaters on light box readings are to be studied. As these readings might differ from day to day, observations are to be taken on each of the four types every day for three days. The results are:

	Graphite Coater Type			
DAY	***M***	***A***	***K***	***L***
1	4.0	4.8	5.0	4.6
2	4.8	5.0	5.2	4.6
3	4.0	4.8	5.6	5.0

Construct an analysis of variance table. Determine at 0.05 level whether there is significant difference between Coater Types. Apply the *LSD* test to locate significant difference between the various pairs of coater types.

(1) Hypothesis

H_0: There is no significant difference between coater types.

H_1: There is significant " " ".

(2) $\alpha = 0.05$

(3) $F = \dfrac{MS \text{ for Coater types}}{MSE}$

(4) Calculation

S.V	d.f	SS	MS	F
b/w coater types			0.51	
b/w Days			0.10	
Error			0.09	
Total				

$$LSD = 0.60$$

$$\bar{Y}_M = 4.97, \quad \bar{Y}_A = 4.87, \quad \bar{Y}_K = 5.27, \quad \bar{Y} = 4.73$$

STA: 408

THANK YOU

STA: 408

COURSE: STA 408
DESIGN & ANALYSIS OF EXPERIMENT

LECTURE DELIVERED BY

SARAH AHMAD
DEPARTMENT OF STATISTICS
VIRTUAL UNIVERSITY OF PAKISTAN



COURSE: STA 408

DESIGN & ANALYSIS OF EXPERIMENT

FACTORIAL DESIGN

LECTURE NO.8

Experiments are often planned to investigate the effects of say, different rates of fertilizers , different dates of planting, different categories of education, different intensities of a stimulus, etc.

- The independent variables such as fertilizers, planting, education, stimulus, etc. are called ***factors***.

- While the values such as rates, dates, categories or intensities at which a factor is held fixed, are known as ***levels***.

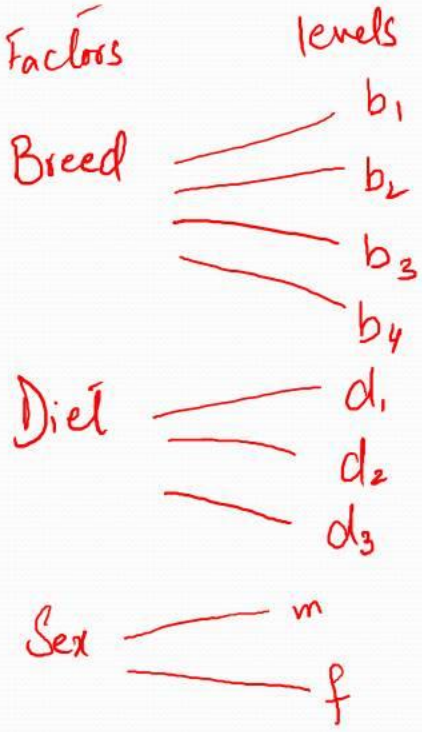
WHAT IS FACTORIAL DESIGN?

Audio Lecture

Level of watering — 4 times in a month
— 8 times in a month

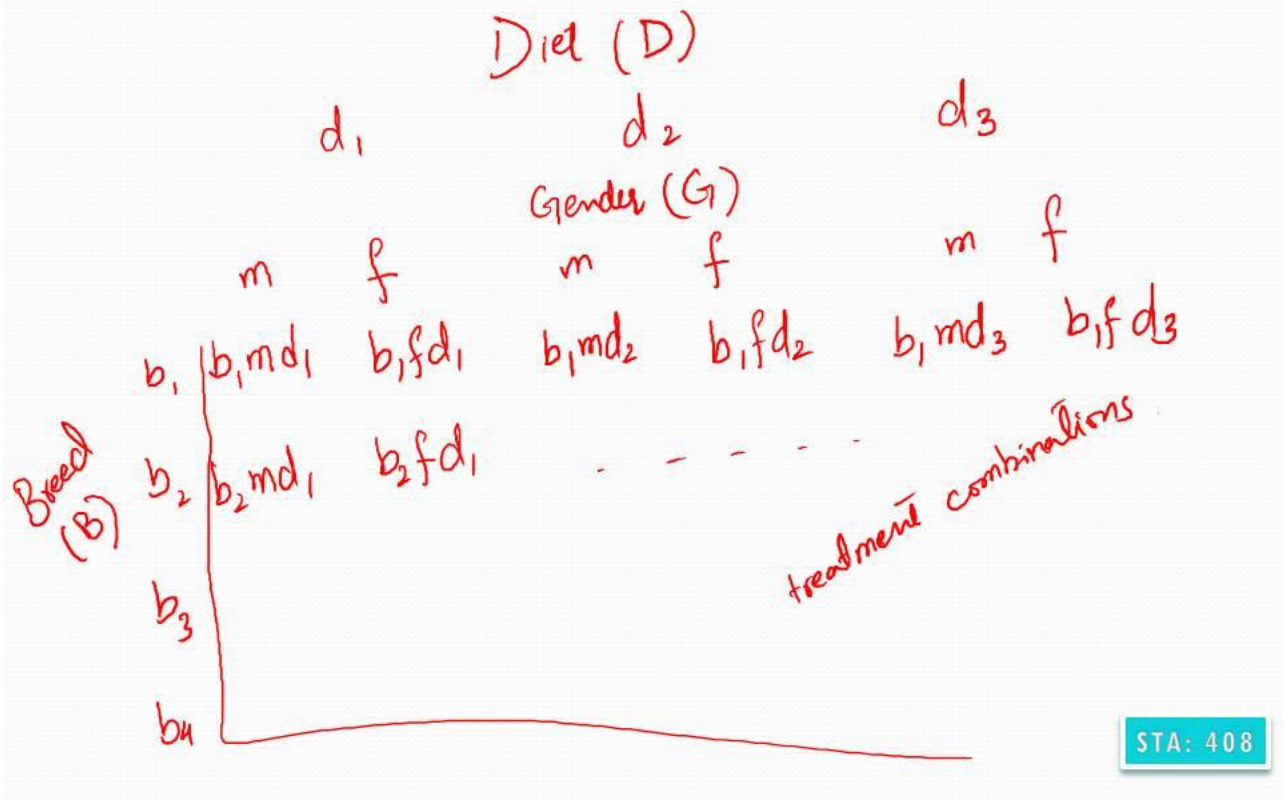
Amount of Seeding — 60 kg
— 50 kg
— 80 kg

Cultural Practice — crop in rotation
— crop in gaps



Diets (D)

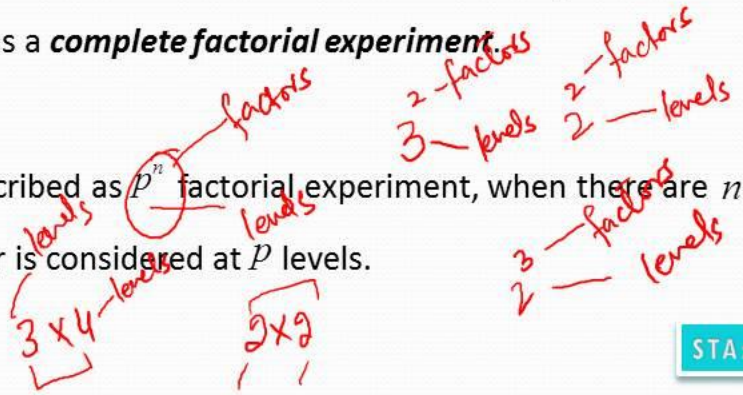
d_1 d_2 d_3


Diet (D)
d_1
d_2
d_3
Gender (G)
m
f
m
f
m
f
Breed (B)
b_1
b_2
b_3
b_4
b_1md_1
b_1fd_1
b_1md_2
b_1fd_2
b_1md_3
b_1fd_3
b_2md_1
b_2fd_1
treatment combinations

- It is customary to represent the factors by capital letters A, B, C etc. and levels in small letters a, b, c, etc.

- An experiment is called a ***factorial experiment*** if the treatments consist of all possible combinations of several levels of several factors. When each treatment combination is used the same number of times, the factorial experiment is known as a ***complete factorial experiment***.

- The experiment is described as p^n factorial experiment, when there are n factors and each factor is considered at p levels.



MAIN EFFECTS

A main effect of a factor is defined as a measure of the average change in effect produced by changing the level of the factor. It is measured independently of other factors and is the effect of the factor only. A main effect is sometimes regarded as an interaction of zero order.

INTERACTION EFFECTS

Factors are said to be interact, when they are not independent. But interaction in a factorial experiment, is a measure of the extent to which the effect of changing the levels of one or more factors depends on the levels of the other factors. Interactions between two factors are referred to as first order interactions, those concerning three factors, as second order interactions and so on.

A, B, C AB 1st AC BC ABC 2nd

A B C

ADVANTAGES & DISADVANTAGES OF FACTORIAL EXPERIMENT

STA: 408

- A factorial experiment is usually economical.
- All the experimental units are used in computing the main effects and interactions.
- The use of all treatment combinations makes the experiment more efficient and comprehensive.
- The interaction effects are easily estimated and tested through the usual analysis of variance.
- The experiment yields unbiased estimates of effects, which are of wider applicability.

- A factorial experiment requires an excessive amount of experimentation when there are several factors at several levels.

8 factors, each at 2 levels
256 combinations
7 factors, each at 3 levels
2187

fractional Replication

- A large number of combinations when used, cause decrease in the efficiency of the experiment. The experiment may be reduced to a manageable size by confounding some effects considered of little practical consequence.

- The experiment set up and the resulting statistical analysis are more complex.

EXAMPLE

A soft drink company is studying the effect of percent carbonation (A), the speed (C) and operating pressure (B) on the filler on the volume of the carbonated drink pack. In each bottle, 3 levels of carbonation, 2 levels of pressure and 2 levels of speed are selected and a factorial experiment with 2 replicates is conducted. The order in which 24 observations are taken is determine randomly. The coded data are given in the table.

		Operating pressure (B)			
		25 PSI		30 PSI	
		Line Speed (C)			
		200	250	200	250
Percent Carbonation (A)	10%	-3 -1	-1 0	-1 0	1 1
	12%	0 1	2 1	2 3	6 5
	14%	5 4	7 6	7 9	10 11

i → A → a
j → B → b
k → C → c
l → no of replicates
n

- HYPOTHESIS

H_0 : Effect of main effects is insignificant.

H_0' : " 1st order interaction is ".

H_0'' : " 2nd order interaction is ".

or

H_b : Effect of factor A is insignificant.

B

C

AB

⋮

ABC

		Operating pressure (B)				
		25 PSI		30 PSI		
		Line Speed (C)				
		200	250	200	250	$\Sigma y_i \ldots$
Percent Carbonation (A)	10%	-3 -1 (-4)	-1 0 (-1)	-1 0 (-1)	1 1 (2)	-4
	12%	0 1 (1)	2 1 (3)	2 3 (5)	6 5 (11)	20
	14%	5 4 (9)	7 6 (13)	7 9 (16)	10 11 (21)	54
$\Sigma y_j \ldots$		6	15	20	34	75

STA: 408

$$C.F = (Y_{\cdots})^2 / abcn \qquad abcn = 3 \times 2 \times 2 \times 2 = 24$$

$Y_{\cdots}$ ↓ Grand Total

$$C.F = (75)^2/24 = 234.375$$

$$TSS = \sum_i \sum_j \sum_k \sum_l Y_{ijkl}^2 - C.F$$

$$= (-4)^2 + (1)^2 + (9)^2 + (-1)^2 + (3)^2 + \cdots (21)^2 - 234.375$$

$$TSS = 1118 - 234.375 = 336.625$$

$$SSA = \frac{\sum_i y_{i\ldots}^2}{bcn} - C.F$$

$$SSA = \frac{(-4)^2 + (20)^2 + (59)^2}{2\times2\times2} - 234.375 = 252.75$$

$$SSB = \frac{\sum_j y_{\cdot j\cdot}^2}{acn} - C.F$$

$$SSB = \frac{(6+15)^2 + (20+34)^2}{3\times2\times2} - 234.375 = 45.375$$

$$SSC = \frac{\sum_k y_{\cdot\cdot k}^2}{abn} - C.F$$

$$= \frac{(6+20)^2 + (15+34)^2}{3\times2\times2} - C.F$$

$SSC = 22.042$

Sub Table (A x B)

		Operating pressure (B)	
		25 PSI	30 PSI
Percent Carbonation (A)	10%	-5	1
	12%	4	16
	14%	22	37

STA: 408

$$TSS(A\times B) = \frac{\sum\sum y_{ij}^2}{cn} - C.F$$

$$TSS(A\times B) = \frac{(-5)^2 + (1)^2 + \cdots + (37)^2}{2\times 2} - 234.375$$

$$TSS(A\times B) = 303.375$$

$$SSAB = TSSAB - SSA - SSB$$

$$= 303.375 - 252.75 - 45.375$$

$$SSAB = 5.25$$

Sub Table (A x C)

		Line Speed (C)	
		200	250
Percent Carbonation (A)	10%	5	1
	12%	6	14
	14%	25	34

$$TSSAC = \frac{\sum\sum y_{i.k.}^2}{bn} - C.F$$

$$TSSAC = 509.75 - 234.375$$

$$TSSAC = 275.375$$

$$SSAC = TSSAC - SSA - SSC$$

$$SSAC = 0.583$$

Sub Table (B x C)

		Line Speed (C)	
		200	250
Operating pressure (B)	25 PSI	6	15
	30 PSI	20	34

STA: 408

$a = 3$
$n = 2$

$$TSSBC = \frac{\sum\sum Y_{.jx.}^2}{jk} - C.F$$

$$an$$

$$TSSBC = \frac{(6)^2 + \cdots + (34)^2}{3x2} - 234.375$$

$$TSSBC = 68.46$$

$$SSBC = TSSBC - SSB - SSC$$

$$SSBC = 1.041$$

Sub Table (A x B x C)

		Operating pressure (B)			
		25 PSI		30 PSI	
		Line Speed (C)			
		200	250	200	250
Percent Carbonation (A)	10%	-4	-1	-1	2
	12%	1	3	5	11
	14%	9	13	16	21

STA: 408

$$TSS(ABC) = \frac{\sum_i \sum_j \sum_k y_{ijk}^2}{n} - C.F$$

$$= \frac{(-4)^2 + (11)^2 + \cdots + (21)^2}{2} - 234.375$$

$$TSS\ ABC = 328.125$$

$$SSABC = TSS\ ABC - SSA - SSB - SSC - SSAB - SSAC - SSBC$$

$$SSABC = 1.084$$

$$\text{Error SS} = TSS - TSS(ABC)$$

$$= 336.625 - 328.125 = 8.5$$

Source of Variation	Degrees of Freedom
Factor A	$a-1$
Factor B	$b-1$
Factor C	$c-1$
AB	$(a-1)(b-1)$
AC	$(a-1)(c-1)$
BC	$(b-1)(c-1)$
ABC	$(a-1)(b-1)(c-1)$
Error	$abc(n-1)$
Total	$abcn-1$

ANOVA TABLE

S.V	d.f	SS	MS	F_{cal}	F_{tab}	Decision
Main Effects						
A	2	252.75	126.375	178.412	3.89	Sig
B	1	45.375	45.375	64.059	4.75	Sig
C	1	22.042	22.042	31.118	4.75	Sig
1st Order Interaction						
AB	2	5.25	2.625	3.706	3.89	Insig
AC	2	0.583	0.292	0.412	3.89	//
BC	1	1.041	1.042	1.471	4.75	//
2nd Order Interaction						
ABC	2	1.084	0.542	0.765	3.89	Insig
Error	12	8.5	0.708			
Total	23	336.625				

PRACTICE QUESTION

The quality control department of a fabric finishing plant is studying the effect of several factors on the dying of cotton-synthetic cloth used to manufacture men's shirts. Two operators, three cycle times and two temperatures were selected, and a factorial experiment with 2 replicates is conducted. The results are as follows. Analyze the data and draw conclusions.

$i \to C \to c = 3$

$j \to R \to r = 2$

$K \to P \to p = 2$

$l \to n = 2$

$i = 1, 2, \ldots c$

$j = 1, 2, \ldots r$

$K = 1, 2, \ldots p$

$l = 1, 2, \ldots n$

		TEMPERATURE (R)			
		300°C		350°C	
		OPERATOR (P)			
		1	2	1	2
CYCLE TIME (C)	40	23, 24 (47)	30, 28 (58)	24, 23 (47)	38, 36 (74)
	50	36, 35 (71)	34, 38 (72)	37, 39 (76)	34, 38 (72)
	60	28, 24 (52)	35, 35 (70)	26, 29 (55)	36, 37 (73)

C.F, TSS, SSR, SSC, SSP

SSCR, SSCP, SSRP

SSCRP, ESS

C — C-1 = 3-1 = 2

R — r-1 = 2-1 = 1

P — p-1 = 2-1 = 1

CR (C-1)(r-1)

CP - - -

C, R, P } Main Effects

CR, CP, RP } 1st order interactions

CRP → 2nd order interaction

THANK YOU

COURSE: STA 408
DESIGN & ANALYSIS OF EXPERIMENT

LECTURE DELIVERED BY

SARAH AHMAD

DEPARTMENT OF STATISTICS

VIRTUAL UNIVERSITY OF PAKISTAN



COURSE: STA 408

DESIGN & ANALYSIS OF EXPERIMENT

2^K AND 2^2 FACTORIAL DESIGN

LECTURE NO.9

2^K FACTORIAL DESIGN

The *K* factors each at only two levels. These levels may be quantitative, such as two values of temperature, pressure, or time; or they may be qualitative, such as two machines, two operators, the "high" and "low" levels of a factor, or perhaps the presence and absence of a factor. A complete replicate of such a design requires 2x2x2.....x2 = 2^k observations and is called a 2^k factorial design.

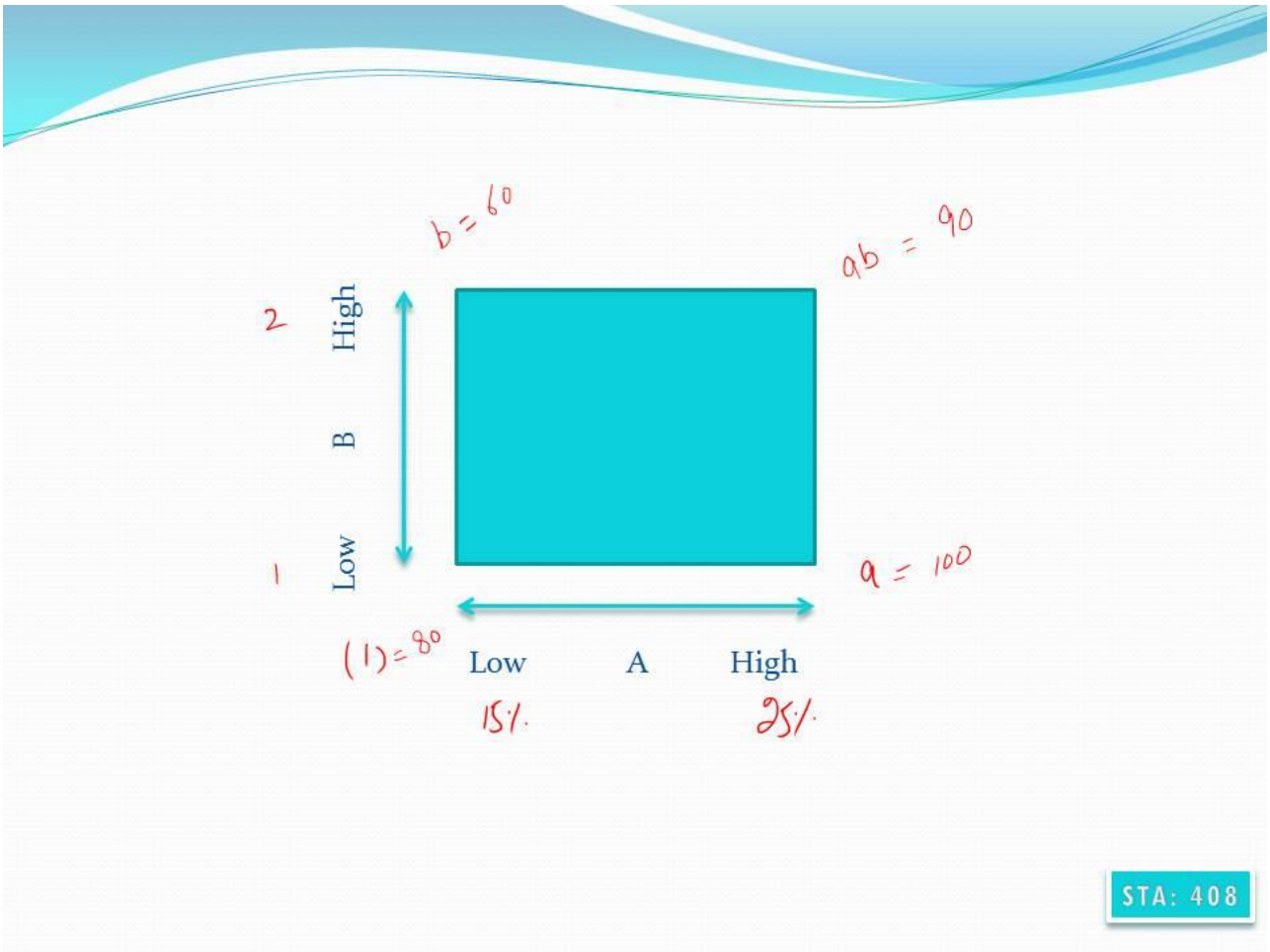
2^2 FACTORIAL DESIGN

The simplest case is one in which two factors are of interest and each factor is said at only 2 levels. This design is called a 2^2 factorial design. The levels of the factors may be arbitrarily called low and high or presence and absence.

EXAMPLE

Consider an investigation into the effect of the concentration of the reactant and the amount of the catalyst on the conversion (yield) in a chemical process. The objective of the experiment was to determine if adjustments to either of these two factors would increase the yield. Let the reactant concentration be factor A and let the two levels of interest be 15 and 25 percent. The catalyst is factor B, with the high level denoting the use of 2 pounds of the catalyst and the low level denoting the use of only 1 pound. The experiment is replicated three times, so there are 12 runs. The order in which the runs are made is random so this is a completely randomized experiment. The data are obtained as follows:

			REPLICATE			
			1	2	3	TOTAL
TREATMENT COMBINATIONS	A low, B low	(1)	28	25	27	**80**
	A high, B low	a	36	32	32	**100**
	A low, B high	b	18	19	23	**60**
	A high, B high	ab	31	30	29	**90**


b = 60
ab = 90
2
High
B
Low
1
a = 100
(1) = 80
Low
A
High
15%
25%
STA: 408

SIGN TABLE

		EFFECTS			
		M	**A**	**B**	**AB**
TREATMENT COMBINATIONS	(1)	+	−	−	+
	a	+	+	−	−
	b	+	−	+	−
	ab	+	+	+	+

$C.F = \frac{(330)^2}{12} = 9075$

EFFECTS

	EFFECTS			
	M	A	B	AB
(1)	+	-	-	+
a	+	+	-	-
b	+	-	+	-
ab	+	+	+	+

Effect M

$C.F = \frac{[\text{contrast } M]^2}{2^k n}$

$M = \frac{1}{2^k n}[\text{Contrast of } M]$

$2^k n \rightarrow$ no. of replicates

$M = \frac{1}{2^2(3)}[(1) + a + b + ab]$

$M = 27.5$

$M = \frac{1}{4 \times 3}[80 + 100 + 60 + 90] = \frac{1}{12} \times 330 = 27.5$

Effect A

$$A = \frac{1}{2^{k-1}n}[\text{contrast } A] = \frac{1}{2^{2-1}(3)}[-(1) + a - b + ab]$$

	EFFECTS			
	M	A	B	AB
(1)	+	-	-	+
a	+	+	-	-
b	+	-	+	-
ab	+	+	+	+

$$A = \frac{1}{2 \times 3}[-80 + 100 - 60 + 90]$$

$$A = \frac{1}{6}[50] = 8.33 \quad \text{——— effect A}$$

$$SSA = \frac{[\text{contrast } A]^2}{2^k n} = \frac{(50)^2}{2^2(3)} = \frac{(50)^2}{12} = 208.33$$

Effect B

$$B = \frac{1}{2^{k-1} \cdot n} \left[-(1) - a + b + ab \right]$$

$$B = \frac{1}{2(3)} \left[-80 - 100 + 60 + 90 \right] = \frac{-30}{6}$$

$$\boxed{B = -5}$$

$$SSB = \frac{[\text{contrast } B]^2}{2^k n} = \frac{(-30)^2}{12} = 75$$

	EFFECTS			
	M	A	B	AB
(1)	+	-	-	+
a	+	+	-	-
b	+	-	+	-
ab	+	+	+	+

Effect AB

	EFFECTS			
	M	A	B	AB
(1)	+	-	-	+
a	+	+	-	-
b	+	-	+	-
ab	+	+	+	+

$$AB = \frac{1}{2^{k-1}n}\left[(1) - a - b + ab\right]$$

$$AB = \frac{1}{2(3)}\left[80 - 100 - 60 + 90\right]$$

$$AB = \frac{1}{6}(10) = 1.667 = \underline{1.67}$$

$$SSAB = \frac{[\text{contrast } AB]^2}{g^k \cdot n} = \frac{(10)^2}{12} = 8.33$$

$$TSS = \sum_{ijk} y_{ijk}^2 - C.F$$

$$TSS = (28)^2 + (25)^2 + \cdots\cdots + (29)^2 - 9075$$

$$TSS = 9398 - 9075 = 323$$

$$ESS = TSS - SSA - SSB - SSAB$$

$$ESS = 323 - 208.33 - 75 - 8.33$$

$$ESS = 31.34$$

$$\boxed{SSE = 31.34}$$

$F_{cal} \geq F_{\alpha(v_1, v_E)}$ (Tab)

ANOVA TABLE						
S.O.V	d.f	SS	MS	F_{cal}	F_{tab}	Decision
Main Effects		SSA =				
A	$a-1 = 2-1 = 1$	208.33	208.33	$F_A = 53.15$	> 5.32	Sig
B	$b-1 = 2-1 = 1$	SSB = 75	75	$F_B = 19.13$	> 5.32	Sig
1st Order Interaction						
AB	$(a-1)(b-1) = 1 \times 1 = 1$	SSAB = 8.33	8.33	$F_{AB} = 2.13$	< 5.32	Insig
Error	$ab(n-1) = 2 \times 2 \times (2) = 8$	SSE = 31.34	3.92			
Total	$abn - 1 = (2 \times 2 \times 3) - 1 = 11$					

① H_0: A is insig H_1: A is sig ✓

② H_0: B " ✓ H_1: B " ✓

③ H_0: AB " ✓ H_1: AB "

STA: 408

EXAMPLE

A bacteriologist is interested in the effects of two different culture media and two different times on the growth of a particular virus. He or she performs six replicates of a 2^2 design, making the runs in random order. Analyze the bacterial growth data that follow and draw appropriate conclusions.

$2^2 = 4$ Treatment combinations

	Culture Medium			
Time	1		2	
12	21	22	25	26
	23	28	24	25
	20	26	29	27
18	37	39	31	34
	38	38	29	33
	35	36	30	35

TREATMENT COMBINATIONS	REPLICATES					
	1	2	3	4	5	6
(1)	21	22	23	28	20	26
m	37	39	38	38	35	36
c	25	26	24	25	29	27
mc	31	34	29	33	30	35

M / Time	Culture Medium 1 (low)		2 (high)	
12 (low)	21	22	25	26
	23	28	24	25
	20	26	29	27
18 (high)	37	39	31	34
	38	38	29	33
	35	36	30	35

C

STA: 408

		REPLICATES						
		1	2	3	4	5	6	TOTAL
TREATMENT COMBINATIONS	(1)	21	22	23	28	20	26	140
	m	37	39	38	38	35	36	223
	c	25	26	24	25	29	27	156
	mc	31	34	29	33	30	35	192

(1) Hypothesis

H_0: The main effect M is insignificant.

H_1: " " significant

H_0': The effect of C is insignificant.

H_1': " " significant.

H_0'': The first order interaction effect MC is insignificant.

H_1'': " " significant.

(2) level of significance

(3) Test Statistic

$$F = \frac{MS\,Effect}{MSE}$$

(4) Calculations

Sign table

STA: 408

n - no. of replicates
m - levels of factor M
c - levels of factor C

ANOVA TABLE

Source of Variation	Degrees of Freedom	Sum of Squares	Mean Square	*F*-Ratio
M	$m-1 = 2-1 = 1$	SSM	$MS(M) = SSM/m-1$	$F_M = \frac{MS(M)}{MSE}$
C	$c-1 = 2-1 = 1$	SS C	$MS(C) = SSC/c-1$	$F_C = \frac{MSC}{MSE}$
MC	$(m-1)(c-1) = 1$	SSMC	$MS(MC) = SSMC/(m-1)(c-1)$	$F_{MC} = \frac{MS(MC)}{MSE}$
Error	$mc(n-1)$	SSE	$MSE = SSE/mc(n-1)$	
Total	$mcn-1$	SST		

EXAMPLE

Find out the effect of *A*, *B* and *AB* by using the data given below. Also, make a column of signs *AB*.

A	B	Treatment combination	REPLICATES			
			1	2	3	4
-	-	(1)	18.2	18.9	12.9	14.4
+	-	a	27.2	24.0	22.4	22.5
-	+	b	15.7	14.5	15.1	14.2
+	+	ab	41.0	43.9	36.3	39.9

A	B	AB	Treatment combination	REPLICATES				
				1	2	3	4	Total
-	-	+	(1)	18.2	18.9	12.9	14.4	**64.4**
+	-	−	*a*	27.2	24.0	22.4	22.5	**96.1**
-	+	−	*b*	15.7	14.5	15.1	14.2	**59.5**
+	+	+	*ab*	41.0	43.9	36.3	39.9	**161.1**

(1) = 64.4 b = 59.5

a = 96.1 ab = 161.1

Effect of A

$$A = \frac{1}{2^{k-1}(n)}\left[-(1) + a - b + ab\right]$$

$$A = \frac{1}{2^{2-1}(4)}\left\{-64.4 + 96.1 - 59.5 + 161.1\right\}$$

$$A = 133.3/8 = 16.6625$$

Effect of B

$$B = \frac{1}{2^{k-1}(n)}\left[-(1) - a + b + ab\right]$$

$$B = \frac{1}{2(4)}\left[-64.4 - 96.1 + 59.5 + 161.1\right]$$

$$B = 60.1/8 = 7.5125$$

Effect of AB

$$AB = \frac{1}{2^{k-1}(n)}\left[(1) - a - b + ab\right]$$

$$AB = \frac{1}{2(4)}\left\{64.4 - 96.1 - 59.5 + 161.1\right\}$$

$$AB = \frac{69.9}{8} = 8.7375$$

THANK YOU

STA: 408

COURSE: STA 408
DESIGN & ANALYSIS OF EXPERIMENT

LECTURE DELIVERED BY

SARAH AHMAD
DEPARTMENT OF STATISTICS
VIRTUAL UNIVERSITY OF PAKISTAN



COURSE: STA 408

DESIGN & ANALYSIS OF EXPERIMENT

2^3 FACTORIAL DESIGN

LECTURE NO.10

2^3 FACTORIAL DESIGN

In this case, three factors each at two levels, are of interest. The design is called 2^3 factorial design, and the eight treatment combinations are possible in this design.

SIGN TABLE

TREATMENT COMBINATIONS	EFFECTS							
	M	**A**	**B**	**AB**	**C**	**AC**	**BC**	**ABC**
(1)	+	−	−	+	−	+	+	−
a	+	+	−	−	−	−	+	+
b	+	−	+	−	−	+	−	+
ab	+	+	+	+	−	−	−	−
c	+	−	−	+	+	−	−	+
ac	+	+	−	−	+	+	−	−
bc	+	−	+	−	+	−	+	−
abc	+	+	+	+	+	+	+	+

STA: 408

PROPERTIES OF SIGN TABLE

- Except for the first column, every column has an equal number of "+" and "-" signs.

A − + − + − + − +

B − − + + − − + +

- The sum of products of signs is in any two columns is zero that is any two columns are orthogonal.

(−1)(−1) (+1)(−1) (−1)(+1) (+1)(+1)

- If we multiply any two columns of the sign table another column is achieved i.e. A and B are ,multiplied and get AB.

A and B are

- Column 1 multiplied by another column leaves column unchanged that is column 1 is an identity column.

MODE RULE

Let ABC x BC

So, it must be equal to A.

$ABC \times BC = A$

$AB^2C^2 = AB^0C^0 = A$

$\therefore B^0C^0 = 1$

$$2\overline{)2}\ 1,\quad 2,\quad 0$$

EXAMPLE

A 2^3 factorial design was used to develop a nitride etch process on a single wafer plasma etching tool. The design factors are the gap between the electrodes, the gas flow (C_2F_6 is used as the reactant gas), and the RF power applied to the cathode. Each factor is run at two levels and the design is replicated twice. The response variable is the etch rate for silicon nitride. The etch rate data are shown as

ETCH RATE				
		REPLICATES		
		1	2	Total
TREATMENT COMBINATIONS	(1)	604	550	1154
	a	650	669	1319
	b	601	633	1234
	ab	635	642	1277
	c	1052	1037	2089
	ac	868	749	1617
	bc	1063	1075	2138
	abc	860	729	1589

EFFECTS

Effect of A

$$A = \frac{1}{2^{k-1}n}[\text{contrast } A]$$

$$A = \frac{1}{2^{3-1}(2)}\left[-(1)+a-b+ab-c+ac-bc+abc\right]$$

TREATMENT COMBINATIONS	EFFECTS							
	M	A	B	AB	C	AC	BC	ABC
(1)	+	-	-	+	-	+	+	-
a	+	+	-	-	-	-	+	+
b	+	-	+	-	-	+	-	+
ab	+	+	+	+	-	-	-	-
c	+	-	-	+	+	-	-	+
ac	+	+	-	-	+	+	-	-
bc	+	-	+	-	+	-	+	-
abc	+	+	+	+	+	+	+	+

$$A = \frac{1}{4\times 2}[-1154+1319-1234+1277-2089+1617-2138+1589]$$

$$A = \frac{1}{8}[-813] = -101.625$$

$$SSA = \frac{(-813)^2}{16} = 41310.5625$$

Effect of B

$$B = \frac{1}{2^2(2)}\left[b + ab + bc + abc - (1) - a - c - ac\right]$$

$$B = \frac{1}{8}\left[1234 + 1277 + 2138 + 1589 - 1154 - 1319 - 2089 - 1617\right]$$

$$B = \frac{1}{8}\left[59\right] = 7.375$$

$$SSB = \frac{(59)^2}{16} = 217.5625$$

TREATMENT COMBINATIONS	EFFECTS							
	M	A	B	AB	C	AC	BC	ABC
(1)	+	-	-	+	-	+	+	-
a	+	+	-	-	-	-	+	+
b	+	-	+	-	-	+	-	+
ab	+	+	+	+	-	-	-	-
c	+	-	-	+	+	-	-	+
ac	+	+	-	-	+	+	-	-
bc	+	-	+	-	+	-	+	-
abc	+	+	+	+	+	+	+	+

Effect of C

$$C = \frac{1}{8}[c + ac + bc + abc - (1) - a - b - ab]$$

$$C = \frac{1}{8}\{2089 + 1617 + 2138 + 1589 - 1134 - 1319 - 1234 - 1277\}$$

$$C = \frac{1}{8}[2449] = 306.125$$

$$SSC = \frac{(2449)^2}{16} = 374850.0625$$

TREATMENT COMBINATIONS	EFFECTS							
	M	A	B	AB	C	AC	BC	ABC
(1)	+	-	-	+	-	+	+	-
a	+	+	-	-	-	-	+	+
b	+	-	+	-	-	+	-	+
ab	+	+	+	+	-	-	-	-
c	+	-	-	+	+	-	-	+
ac	+	+	-	-	+	+	-	-
bc	+	-	+	-	+	-	+	-
abc	+	+	+	+	+	+	+	+

Effect of AB

$$AB = \frac{1}{8}[ab - a - b + (1) + abc - bc - ac + c]$$

$$AB = \frac{1}{8}[1277 - 1319 - 1234 + 1154 + 1589 - 2138 - 1617 + 2089]$$

$$AB = \frac{1}{8}(-199) = -24.875$$

$$SSAB = \frac{(-199)^2}{16} = 2475.0625$$

TREATMENT COMBINATIONS	EFFECTS							
	M	A	B	AB	C	AC	BC	ABC
(1)	+	-	-	+	-	+	+	-
a	+	+	-	-	-	-	+	+
b	+	-	+	-	-	+	-	+
ab	+	+	+	+	-	-	-	-
c	+	-	-	+	+	-	-	+
ac	+	+	-	-	+	+	-	-
bc	+	-	+	-	+	-	+	-
abc	+	+	+	+	+	+	+	+

Effect of AC

$$AC = \frac{1}{8}[(1) - a + b - ab - c + ac - bc + abc]$$

$$AC = \frac{1}{8}[1184 - 1319 + 1234 - 1277 - 2089 + 1617 - 2138 + 1589]$$

$$AC = \frac{1}{8}[-1299] = -153.625$$

$$SSAC = \frac{(-1299)^2}{16} = 94402.5625$$

TREATMENT COMBINATIONS	EFFECTS							
	M	A	B	AB	C	AC	BC	ABC
(1)	+	-	-	+	-	+	+	-
a	+	+	-	-	-	-	+	+
b	+	-	+	-	-	+	-	+
ab	+	+	+	+	-	-	-	-
c	+	-	-	+	+	-	-	+
ac	+	+	-	-	+	+	-	-
bc	+	-	+	-	+	-	+	-
abc	+	+	+	+	+	+	+	+

Effect of BC

TREATMENT COMBINATIONS	EFFECTS							
	M	A	B	AB	C	AC	BC	ABC
(1)	+	-	-	+	-	+	+	-
a	+	+	-	-	-	-	+	+
b	+	-	+	-	-	+	-	+
ab	+	+	+	+	-	-	-	-
c	+	-	-	+	+	-	-	+
ac	+	+	-	-	+	+	-	-
bc	+	-	+	-	+	-	+	-
abc	+	+	+	+	+	+	+	+

$$BC = \frac{1}{8}[(1) + a - b - ab - c - ac + bc + abc]$$

$$BC = \frac{1}{8}[1154 + 1319 - 1234 - 1277 - 2089 - 1617 + 2138 + 1589]$$

$$BC = \frac{1}{8}(-17) = -2.125$$

$$SSBC = \frac{(-17)^2}{16} = 18.0625$$

Effect of ABC

$$ABC = \frac{1}{8}[abc - bc - ac + c - ab + b + a - (1)]$$

$$= \frac{1}{8}[1589 - 2138 - 1617 + 2089 - 1277 + 1234 + 1319 - 1154]$$

$$ABC = 5.625 = \frac{1}{8}(45)$$

$$SSABC = \frac{(45)^2}{16} = 126.5625$$

		EFFECTS							
		M	A	B	AB	C	AC	BC	ABC
TREATMENT COMBINATIONS	(1)	+	-	-	+	-	+	+	-
	a	+	+	-	-	-	-	+	+
	b	+	-	+	-	-	+	-	+
	ab	+	+	+	+	-	-	-	-
	c	+	-	-	+	+	-	-	+
	ac	+	+	-	-	+	+	-	-
	bc	+	-	+	-	+	-	+	-
	abc	+	+	+	+	+	+	+	+

TSS = 531420.9375

ESS = 18090.50

TREATMENT COMBINATIONS	EFFECTS							
	M	A	B	AB	C	AC	BC	ABC
(1)	+	-	-	+	-	+	+	-
a	+	+	-	-	-	-	+	+
b	+	-	+	-	-	+	-	+
ab	+	+	+	+	-	-	-	-
c	+	-	-	+	+	-	-	+
ac	+	+	-	-	+	+	-	-
bc	+	-	+	-	+	-	+	-
abc	+	+	+	+	+	+	+	+

ANOVA - Table

S.O.V	d.f	SS	MS	F_{cal}	F_{tab}	Decision
Main Effects						
A	$a-1=1$	41310.5625 →		$F_A = 18.34$	>5.32	Sig
B	$b-1=1$	217.5625 →		$F_B = 0.10$	<5.32	Insig
C	$c-1=1$	374850.0625 →		$F_C = 166.41$	>5.32	Sig
1st Order Interaction						
AB	$(a-1)(b-1)$ 1	2475.0625 →		$F_{AB} = 1.10$	<5.32	Insig
AC	$(a-1)(c-1)$ 1	94402.5625 →		$F_{AC} = 41.91$	>5.32	Sig
BC	$(b-1)(c-1)$ 1	18.0625 →		$F_{BC} = 0.01$	<5.32	Insig
2nd Order Interaction						
ABC	$(a-1)(b-1)(c-1)$ 1	126.5625 →		$F_{ABC} = 0.06$	>5.32	Insig
Error	$ab(n-1)$ $=8$	18020.5000	2252.5625			
Total	$abcn-1$ $(2\times2\times2\times2)-1=15$	531420.9375				

STA: 408

EXAMPLE

An engineer is interested in the effects of cutting speed (A), tool geometry (B) and cutting angle (C) on the life (in hours) of a machine tool. Two levels of each factor are chosen, and three replicates of a 2^3 factorial design are run. Compute the signs of AC and ABC, and hence find the sum of squares of effects of C, AC and ABC by using the following results:

		REPLICATES			Sign Table		
		I	II	III	A	B	C
TREATMENT COMBINATIONS	(1)	22	31	25	-	-	-
	a	32	43	29	+	-	-
	b	35	34	50	-	+	-
	ab	55	47	46	+	+	-
	c	44	45	38	-	-	+
	ac	40	37	36	+	-	+
	bc	60	50	54	-	+	+
	abc	39	41	47	+	+	+

		REPLICATES				Sign Table				
		I	II	III	Total	A	B	C	AB	ABC
TREATMENT COMBINATIONS	(1)	22	31	25	78	-	-	-	+	−
	a	32	43	29	104	+	-	-	−	+
	b	35	34	50	119	-	+	-	−	+
	ab	55	47	46	148	+	+	-	+	−
	c	44	45	38	127	-	-	+	+	+
	ac	40	37	36	113	+	-	+	−	−
	bc	60	50	54	164	-	+	+	−	−
	abc	39	41	47	197	+	+	+	+	+

SUM OF SQUARES OF C

Effect of C

$$C = \frac{1}{2^{k-1}(n)} [c + ac + bc + abc - (1) - a - b - ab]$$

$$C = \frac{1}{2^{3-1}(3)} [127 + 113 + 164 + 127 - 78 - 104 - 119 - 148]$$

$$C = \frac{1}{4 \times 3} [82] = \frac{1}{12}(82) = 6.8333$$

$2^k(n)$

$$SSC = \frac{(82)^2}{24} = 280.1667$$

SUM OF SQUARES OF *AC*

Effect of AC

$$AC = \frac{1}{2^{k-1}(n)} \left[(1) - a + b - ab - c + ac - bc + abc \right]$$

$$AC = \frac{1}{2^{3-1}(3)} \left[78 - 104 + 119 - 148 - 127 + 113 - 164 + 127 \right]$$

$$AC = \frac{1}{4 \times 3} [-208] = \frac{-208}{12} = -17.3333$$

$$SSAC = \frac{(208)^2}{24} = 1802.6667$$

SUM OF SQUARES OF *ABC*

Effect of ABC

$$ABC = \frac{1}{2^{k-1}(n)}\left[abc - bc - ac + c - ab + b + a - (1)\right]$$

$$ABC = \frac{1}{12}\left[127 - 164 - 113 + 127 - 148 + 119 + 104 - 78\right]$$

$$ABC = \frac{1}{12}(-26) = -2.1667$$

$$SSABC = \frac{(26)^2}{24} = 28.1667$$



EXERCISE

The following data were obtained from a 2^3 factorial experiment repeated 3 times

	Treatment Combinations							
Replication	(1)	*a*	*b*	*ab*	*c*	*ac*	*bc*	*abc*
1	**12**	**15**	**24**	**23**	**17**	**16**	**24**	**28**
2	**19**	**20**	**16**	**17**	**25**	**19**	**23**	**25**
3	**10**	**16**	**17**	**27**	**21**	**19**	**29**	**20**

Evaluate the sum of squares for all factorial effects.

THANK YOU