



Virtual University of Pakistan

FEDERAL GOVERNMENT UNIVERSITY

COURSE HANDOUTS

Design and Analysis of Experiment

STA408

Order Your VU Books Now

**Get all Virtual University books in high-quality hard copy,
delivered right to your doorstep. Visit vubookshoppk.com**



Order Now

Contents

1	Experimental Design	1
1	Design and Analysis of Experiment	1
2	Purpose of Experimental Design	2
3	Steps in Designing an Experiment	3
4	Basic Principles of Experimental Design	3
5	Types of Design	4
6	Classifications of Randomized Designs	4
2	Basic Concepts of ANOVA	6
7	An Overview	6
8	Assumptions of ANOVA	6
9	F-test	7
10	Completely Randomized Design (CRD)	7
10.1	Steps for performing ANOVA for CRD	8
10.2	Assumptions of CRD	8
10.3	Advantages of a CRD	8
10.4	Disadvantages	8
10.5	Experimental Layout	8
3	Randomized Complete Block Design	11
11	Main Points	11
11.1	Experimental Layout	11
11.2	Statistical Model	12
11.3	Partitioning of Variations	13
11.4	Advantage and Disadvantages of RBCD	14
4	RCBD with Replications within Blocks	16
12	Introduction	16
12.1	Missing Observations in RCBD	16
13	Efficiency of RCBD Relative to A CRD	19
5	Latin Square Design-I	22
14	Introduction	22
14.1	Layout	22
14.2	Mathematical Model	23
14.3	Statistical Analysis	23
6	Latin Square Design-II	27
15	Latin Square	27

15.1	Advantages and Disadvantages	27
15.2	Missing Observations in Latin Square Design	27
7	Latin Square Designs-II	34
16	Efficiency of Latin Square Designs	34
17	Multiple Comparison Tests	35
17.1	LSD Test	36
8	Factorial Design	43
18	Introduction	43
18.1	What is a Factorial Design?	43
18.2	Advantage and Disadvantages of Factorial Experiment	43
9	2^k and 2^2 Factorial Design	49
19	2^k Factorial Design	49
20	2^2 Factorial Design	49
10	2^3 Factorial Design	53
21	2^3 Factorial Design	53
22	Properties of Sign Table	53

Lecture 1

Experimental Design

Lesson 1: Design and Analysis of Experiment

Definition 1.1

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.

Terminologies

Definition 1.2

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

Definition 1.3

EXPERIMENTAL UNIT: The experimental unit is the basic entity or unit on which the experiment is performed. 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: a single person, animal, plant, manufactured item, or country that belongs to a larger collection of such entities being studied.



Identify the Experimental Units

- ◆ 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.

Definition 1.4

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.

Definition 1.5

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.

Definition 1.6

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.

There are two types of error:

- ◆ Systematic Errors
- ◆ Random Errors

Definition 1.7

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.

Lesson 2: 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.
- ◆ To evaluate and examine the outcomes critically and logically.

Lesson 3: 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.

Lesson 4: Basic Principles of Experimental Design

The basic principles of experimental designs are: (i) Randomization, (ii) Replication, and (iii) Local of 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.

Lesson 5: Types of Design

The types of Design are:

◆ 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: See Example.

◆ 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.

Lesson 6: Classifications of Randomized Designs

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

Example: A market gardener wants to test three types of peas, A , B , and 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, Randomize Complete Block Design.

- ◆ 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, Latin Square Design.

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

A	B	C
---	---	---

A	C	B
---	---	---

C	B	A
---	---	---

A	B	C
B	C	A
C	A	B

Lecture 2

Basic Concepts of ANOVA

Lesson 7: An Overview

Analysis of variance 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 hypothesis:

$$H_o : \mu_1 = \mu_2 = \cdots = \mu_k \quad vs \quad H_1 : \text{Not all population means are equal}$$

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

Lesson 8: Assumptions of ANOVA

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

Note that

- ◆ When H_o is true, the sample means are close together because there is only one sampling distribution of $\bar{X}$.
- ◆ When H_o is false, the sample means come from different sampling distributions and are not as close together.

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 may be denoted by $MSTR$.

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

The numerator in $MSTR$ is the sum of squares due to treatment ($SSTR$) and in denominator is the degrees of freedom associated with $SSTR$.

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 due to error* and may be denoted by *MSE*.

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

The numerator in *MSE* is the sum of squares due to error (*SSE*) and in denominator is the degrees of freedom associated with *SSE*.

Lesson 9: F-test

If the null hypothesis is true and the ANOVA assumptions are valid, the sampling distribution of $\frac{MSTR}{MSE}$ is an F-distribution with *MSTR* df equal to $k - 1$ and *MSE* df equal to $n_T - k$.

If the means of the k populations are not equal, the values of $\frac{MSTR}{MSE}$ will be inflated because *MSTR* overestimates σ^2 .

We will reject H_0 if the resulting value of $\frac{MSTR}{MSE}$ appears to be too large to have been selected at random from the appropriate F-distribution.

ANOVA Table

The ANOVA table decomposes the variance into two components:

- ◆ A between-group component (treatments), and
- ◆ a within-group component (error)

The F-ratio is the ratio of the between-group estimate to the within group estimate.

Lesson 10: Completely Randomized Design (CRD)

Some Definitions

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

A completely randomized 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, that is, 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

- ◆ The experimental units are homogeneous,
- ◆ The experiments are small such as laboratory experiments, and
- ◆ Some experimental units are likely to be destroyed or to fail to respond.

The CRD is used to compare the treatment means:

$$H_o : \mu_1 = \mu_2 = \dots = \mu_k \quad vs \quad H_1 : \text{At least two of the treatment means differ}$$

10.1 Steps for performing ANOVA for CRD

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

10.2 Assumptions of CRD

- ◆ **Normality**
Normality can be checked using Normal probability plots and Kolmogrove test.
- ◆ **Constant Variance**
The homogeneity of variances can be checked using histogram and Levene’s test.

ANOVA Table for CRD

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

10.3 Advantages of a CRD

- ◆ CRD is very flexible design. That is, the 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.

10.4 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.

10.5 Experimental Layout

The layout of an experiment is the actual placement of the treatment to the experimental units. For example, if the number of treatment = 4, number of replication = 3, then experimental units = $4 \times 3 = 12$.

A (Y ₁₁)	D (Y ₄₁)	B (Y ₂₁)
C (Y ₃₁)	B (Y ₂₂)	A (Y ₁₂)
A (Y ₁₃)	D (Y ₄₂)	C (Y ₃₂)
C (Y ₃₃)	B (Y ₂₃)	D (Y ₄₃)

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

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

Question 1: Three fertilizer treatments *A*, *B*, and *C* each applied to seven plots of strawberry plants, resulted in the following weights of crops (lb/plot): Perform the ANOVA to test the

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

hypothesis of no difference in the treatment effects. Use a 0.01 level of significance.

- 1) **Formulation of hypothesis:**
 $H_o : \mu_1 = \mu_2 = \mu_3 \quad vs \quad H_1 : \text{at least two means are not equal}$
- 2) **Level of significance:** $\alpha = 0.05$
- 3) **Test Statistics:** $F = \frac{MS_{treatment}}{MS_{error}}$
Required calculations: Here $T_1 = 143, T_2 = 291, T_3 = 220, T_{..} = 654$.

$$CF = \frac{T_{..}^2}{N}, \quad \text{where } N = nk = 7(3) = 21$$
$$CF = \frac{(654)^2}{21} = 20367.43$$
$$\begin{aligned} \text{Total SS} &= \sum \sum x_{ij}^2 - CF \\ &= 24^2 + 18^2 + 18^2 + \dots + 28^2 - 20367.43 \\ &= 22520 - 20367.43 = 2152.57 \end{aligned}$$
$$\begin{aligned} \text{Treat SS} &= \frac{\sum_{j=1}^k T_j^2}{n} - CF \\ &= \frac{1}{7}[(143)^2 + (291)^2 + (220)^2] - 20367.43 = 1564.43 \end{aligned}$$
$$\begin{aligned} \text{Error SS} &= \text{Total SS} - \text{Treat SS} \\ &= 2152.57 - 1564.43 = 587.14 \end{aligned}$$

- 4) **Critical Region:** Reject H_o if $F_{cal} \geq F_{\alpha(k-1, v-k)} \Rightarrow 24.00 \geq F_{0.05(2, 18)} \Rightarrow 24 \geq 3.55$
We reject H_o

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

Source	SS	df	MS	F
Treatment	1565.43	$k - 1 = 2$	$\frac{1565.43}{2} = 782.72$	$F = \frac{782.72}{32.62} = 24.00$
Error	587.14	$N - k = 18$	$\frac{587.14}{18} = 32.62$	
Total	2152.57	$N - 1 = 20$		

- 5) **Conclusion:** Since the $24 > 3.55$ fall in the critical region, so we reject H_o . So, we conclude that there is a difference between the treatment effects.

□

Lecture 3

Randomized Complete Block Design

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

Lesson 11: Main Points

- ◆ Experimental material is divided into groups or blocks
- ◆ Each block contains a complete set of treatments
- ◆ Treatments are assigned randomly (Randomization is restricted within block)
- ◆ A new randomization is done for each block/ group
- ◆ It has different layouts: 2×3 , 3×5 , etc.
- ◆ The RCBD is perhaps the most frequently used experimental design



Figure 3.1: RCBD Design may Look Like

11.1 Experimental Layout

An experimental layout for such a design using four treatments in three blocks be as follows.

Treatments are randomly allocated to blocks.



Figure 3.2: Agricultural Research Field

Block 1	Block 2	Block 3
A	B	C
B	C	D
C	D	A
D	A	B

11.2 Statistical Model

The linear model for RCBD is

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

where μ = True mean effect, T = effects of treatment, B = Effect of block/ groups, and e = effect of random errors.

Data Computations Table

Blocks	Treatments						Block	
	1	2	...	j	...	k	Total	Mean
1	Y_{11}	Y_{12}					$B_{1.}$	$\bar{Y}_{1.}$
2								
$\vdots$								
i								
$\vdots$								
r								
Total	$B_{.1}$							
Means	$\bar{Y}_{.1}$							

11.3 Partitioning of Variations

[Total Variation] = [Variation due to Treatment] + [Variation due to Blocking] + [Variation due to Unknown factors]

Total SS = Treatment SS + Block SS + Error SS

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

SOV	df	SS	MS	F
Treatment	$k - 1$	$SS_{Treat} = r \sum_j (\bar{Y}_{.j} - \bar{Y}_{..})^2$	$MS_{Treat} = \frac{SS_{Treat}}{k-1}$	$F_t = \frac{MS_{Treat}}{MSE}$
Blocks	$r - 1$	$SS_{Block} = k \sum_i (\bar{Y}_{i.} - \bar{Y}_{..})^2$	$MS_{block} = \frac{SSB}{r-1}$	$F_{Block} = \frac{MSB}{MSE}$
Error	$(k - 1)(r - 1)$	$SSE = \text{by subtraction}$	$MS_{Error} = \frac{SS_{Treat}}{(r-1)(k-1)}$	
Total	$rk - 1$	$SS_{Treat} = \sum_i \sum_j (Y_{ij} - \bar{Y}_{..})^2$		

Example 1: Four varieties of wheat are tested in randomized complete block 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			
	V_1	V_2	V_3	V_4
I	2	5	4	1
II	2	3	3	1
III	4	6	6	2
IV	1	4	2	3

Solution:

- Step 1: State the Hypothesis
- $H_o : T_j = 0$ (There is no difference in the means of four varieties of wheat)
- $H_1 : T_j \neq 0$ (At least two treatment means are different from each other)
- $H_o : B_i = 0$ (There is no difference in the means of four varieties of wheat)
- $H_1 : B_i \neq 0$ (At least two varieties are different from each other)

- Step 2: Level of Significance
- $\alpha = 0.05(5\%)$

- Step 3: Test-statistics to be used

$$F = \frac{MS_{Treat}}{MS_{Error}}$$

Required Computations

$$SS_{Total} = \sum \sum Y^2 - \frac{Y_{..}^2}{n} = (2^2 + 5^2 + 4^2 + \dots + 3^2) - \frac{(2 + 5 + \dots + 3)^2}{16} = 40.94$$
$$SS_{Treat} = \sum_{.j} \frac{T_{.j}^2}{n} = \frac{9^2 + 18^2 + 15^2 + 7^2}{4} - \frac{(49)^2}{16} = 19.69$$
$$SS_{Block} = \sum_{i.} \frac{T_{i.}^2}{n} = \frac{12^2 + 9^2 + 18^2 + 10^2}{4} - \frac{(49)^2}{16} = 12.19$$
$$SS_{Error} = SS_{Total} - SS_{Treat} - SS_{Block} = 40.96 - 19.69 - 12.19 = 9.06$$

Blocks	Treatments				Block Total	$\left(\frac{\text{Block}}{\text{Total}}\right)^2$	SS
	V_1	V_2	V_3	V_4			
I	2	5	4	1	12	$12^2 = 144$	$2^2 + 5^2 + 4^2 + 1^2 = 46$
II	2	3	3	1	9	81	23
III	4	6	6	2	18	324	92
IV	1	4	2	3	10	100	30
Varieties Total	9	18	15	7			
$(\text{Varieties Total})^2$	81	324	225	49			
SS	25	86	65	15			

SOV	df	SS	MS	F
Treatment	4-1=3	19.69	$\frac{19.69}{3} = 6.56$	$F_{Treat} = \frac{6.56}{1.01} = 6.50$
Block	4-1=3	12.19	$\frac{12.19}{3} = 4.06$	$F_{Block} = \frac{4.06}{1.01} = 4.02$
Error	15-6=9	9.06	$\frac{9.06}{9} = 1.01$	
Total	16-1	40.96		

Step 4: Critical Region
Critical Value for $F_{Treat} = 3.86$ (as $F_{df_1,df_2,\alpha=0.05} = F_{3,9} = 3.96$)
Critical Value for $F_{Block} = 3.86$ (as $F_{df_1,df_2,\alpha=0.05} = F_{3,9} = 3.96$)

Step 5: Conclusion
The computer value of F_{Treat} is greater than the Critical value of F_{Treat} , So F_{Treat} falls in the critical region. Therefore, we will reject the null hypothesis and conclude that the means of four varieties of wheat are statistically different.
The computer value of F_{Block} is greater than the Critical value of F_{Block} , So F_{Block} falls in the critical region. Therefore, we will reject the null hypothesis and conclude that the Blocking is effective.

□

Question 1: The following is the plan of a field layout testing four varieties, A , B , C , and D of wheat in each of 5 blocks. The plot yields in Kg are also indicated.

□

Block I	Block II	Block III	Block IV	Block V
D 29.3	B 33.0	D 29.8	B 36.8	D 28.8
B 33.3	A 34.0	A 34.3	A 35.00	D 28.8
C 30.8	C 34.3	B 36.3	D 28.0	B 34.5
A 32.3	D 26.0	C 35.3	C 32.3	A 36.5

11.4 Advantage and Disadvantages of RBCD

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.

Note that

- ◆ **Extraneous Variables** are undesirable variables that influence the relationship between the variables that an experimenter is examining.
These variables influence the outcome of an experiment, though they are not the variables that are actually of interest. These variables 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.

Lecture 4

RCBD with Replications within Blocks

Lesson 12: Introduction

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

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

12.1 Missing Observations in RCBD

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

- ◆ Unrooted plants
- ◆ Records lost
- ◆ Died animals
- ◆ Destroyed plants
- ◆ Broken Flask
- ◆ Uncooperative people

Suppose,

Blocks	Treatments				Block Total
	1	2	...	k	
1	x	y_{12}	...	y_{1k}	$B'_{1.} + x$
2	y_{21}	y_{22}	...	y_{2k}	$B_{2.}$
$\vdots$	$\vdots$	$\vdots$	...	$\vdots$	$\vdots$
r	y_{r1}	y_{r2}	...	y_{rk}	$B_{r.}$
Treatment Total	$T'_{.x}$	$T_{.2}$	...	$T_{.k}$	$T'_{..} + x$

We get the following sums of squares, when we perform the analysis of variance,

$$\begin{aligned} SS_{Treat} &= \frac{(T_{.1} + x)^2}{r} + \frac{T_{.2}^2}{r} + \dots + \frac{T_{.r}^2}{r} - \frac{(T'_{..} + x)^2}{rk} \\ SS_{Blocks} &= \frac{(B'_{1.})^2}{k} + \frac{B_{2.}^2}{k} + \dots + \frac{B_{r.}^2}{k} - \frac{(T'_{..} + x)^2}{rk} \\ SS_{Error} &= \text{By subtraction} \\ SS_{Total} &= \sum_{ij} y_{ij}^2 + x^2 - \frac{(T'_{..} + x)^2}{rk} \end{aligned}$$

Thus the error sum of squares is obtained as

$$\begin{aligned} SSE &= x^2 + \frac{(T'_{..} + x)^2}{kr} - \frac{(T'_{.1} + x)^2}{r} - \frac{(B'_{1.} + x)^2}{k} + \text{terms which do not involve } x \\ \frac{\partial SSE}{\partial x} &= 2x + \frac{2(T'_{..} + x)}{kr} - \frac{2(T'_{.1} + x)}{r} - \frac{2(B'_{1.} + x)}{k} \tag{Eq. (1)} \\ \Rightarrow 2x + \frac{2(T'_{..} + x)}{kr} - \frac{2(T'_{.1} + x)}{r} - \frac{2(B'_{1.} + x)}{k} &= 0 \\ \Rightarrow x + \frac{T'_{..}}{kr} + \frac{x}{kr} - \frac{T'_{.1}}{r} - \frac{x}{r} - \frac{B'_{1.}}{k} - \frac{x}{k} &= 0 \\ \Rightarrow x \left[1 + \frac{1}{kr} - \frac{1}{r} - \frac{1}{k} \right] + \frac{T'_{..}}{kr} - \frac{T'_{.1}}{r} - \frac{B'_{1.}}{k} &= 0 \\ \Rightarrow x \left(\frac{kr - 1 - k - r}{kr} \right) &= \frac{rB'_{1.} + kT'_{.1} - T'_{..}}{kr} \\ x &= \frac{rB'_{1.} + kT'_{.1} - T'_{..}}{(k - 1)(r - 1)} \end{aligned}$$

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 observations
- G = Grand total of all observed values

Example 1: The following data, obtained from a randomized complete block design with 3 treatments A, B, C and three blocks, contain one missing observation represented by X . Estimate

		Treatments		
		A	B	C
Blocks	I	5	12	15
	II	7	10	14
	III	8	16	x

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

Solution:

$$x = \frac{rB + kT - G}{(k - 1)(r - 1)}$$
$$r = 3, k = 3, B = 8 + 16 = 24, T = 15 + 14 = 29, G = 87$$
$$x = \frac{3(24) + 3(29) - 87}{(3 - 1)(3 - 1)} = 18$$

Therefore, data table is,

		Treatments		
		A	B	C
Blocks	I	5	12	15
	II	7	10	14
	III	8	16	18

- Step 1: Hypothesis:
 $H_0 : \mu_1 = \mu_2 = \mu_3$
 $H_1 : \text{Not all treatment means are equal}$
 $H_0 : \mu_1 = \mu_2 = \mu_3$
 $H_1 : \text{Not all block means are equal}$

Step 2: Level of significance: $\alpha = 0.05$

Step 3: Test Statistics

$$F_{Treat} = \frac{s^2_{Treat}}{s^2_{Error}} \qquad F_{Block} = \frac{s^2_{Block}}{s^2_{Error}}$$

Calculations:

$$CF = \frac{(T_{..})^2}{rk} = \frac{(105)^2}{9} = 1225$$
$$SS_{Total} = \sum_i \sum_j y_{ij}^2 - CF = 5^2 - 12^2 + \cdots + 16^2 + 18^2 - 1225 = 158$$
$$SS_{Treat} = \sum_j \frac{T_{.j}^2}{r} - CF = 1351 - 1225 = 126$$
$$SS_{Block} = \sum_i B_{i.}^2 - CF = \frac{32^2 + 31^2 + 42^2}{3} - 1225 = 24.7$$
$$SS_{Error} = SS_{Total} - SS_{Treat} - SS_{Block} = 158 - 126 - 24.7 = 7.3$$

- Step 5: Critical Region
Reject the Null hypothesis if $\begin{cases} F_{Treat} \geq F_{0.05,(2,3)=9.55} \\ F_{Block} \geq F_{0.05,(2,3)=9.55} \end{cases}$
For treatment means since $25.9 > 9.55$ Reject H_0
For Block means since $5.08 < 9.55$ Do not Reject H_0

- Step 6: Conclusion
As for treatments calculated F-value all in the critical region, it means that all the treatments means are not equal (or at least two treatment means are not equal). Similarly, for blocks the calculated F-value does not lie in the critical region, so so conclude that all the block means are equal.

□

SOV	df	SS	MS	F
Treatments	$k - 1 =$ $3 - 1 = 2$	$SS_{Treat} = 126$	$MS_{Treat} = \frac{126}{2} = 63$	$F_{Treat} = \frac{MS_{Treat}}{MS_{Error}}$ $= \frac{63}{2.43} = 25.9$
Blocks	$r - 1 =$ $3 - 1 = 2$	$SS_{Block} = 24.7$	$MS_{Block} = \frac{24.7}{2} = 12.35$	$F_{Block} = \frac{MS_{Block}}{MS_{Error}}$ $= \frac{12.35}{2.43} = 5.08$
Error	$(r - 1)(k - 1) - 1 = 3$	$SS_{Error} = 7.3$	$MS_{Error} = \frac{7.3}{3} = 2.43$	
Total	$rk - 1 = 7$	$SS_{Total} = 158$		

SOV	df	SS	MS
Treatments	3	1.1986	0.3995
Blocks	9	0.4074	0.0453
Error	27	0.6247	0.0231
Total	39	2.2309	

Lesson 13: Efficiency of RCBD Relative to A CRD

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 ration of error variances.

$$\begin{aligned}
 RE(RCBD, CRD) &= \frac{\text{Estimated MSE for CRD}}{\text{MSE for the RCBD}} \\
 &= \frac{(r - 1)MS_{Block} + r(k - 1)MS_{Error}}{(rk - 1)MS_{Error}}
 \end{aligned}$$

Example 2: Given the following analysis of variance for RCBD: Estimate the efficiency of this design relative to completely randomized design.

Solution:

$$\begin{aligned}
 RE &= \frac{(9)(0.0453) + (10)(3)(0.0231)}{(39)(0.0231)} \\
 &= \frac{1.1007}{0.9009} = 1.22 = 122\%
 \end{aligned}$$

□

Question 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

Step 1: Hypothesis:

$$H_0 : \mu_A = \mu_B = \cdots = \mu_E$$

H_1 : Not all treatment means are equal

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

H_1 : Not all block means are equal

Step 2: Level of significance: $\alpha = 0.05$

Step 3: Test Statistics

$$F_{Treat} = \frac{s_{Treat}^2}{s_{Error}^2} \quad F_{Block} = \frac{s_{Block}^2}{s_{Error}^2}$$

Calculations:

$$x = \frac{rB + kT - G}{(k-1)(r-1)} = \frac{3(27) + 5(12) - 90}{(5-1)(3-1)} = 6.375$$

variance

Blocks	Treatments				
	A	B	C	D	E
I	5	6	6	7	4
II	6.375	7	7	8	5
III	7	7	7	9	5

$$CF = \frac{(T_{..})^2}{rk} = \frac{(96.375)^2}{15} = 619.209$$

$$SS_{Total} = \sum_i \sum_j y_{ij}^2 - CF = 5^2 - 6^2 + \cdots + 9^2 + 5^2 - 619.209 = 23.432$$

$$SS_{Treat} = \sum_j \frac{T_{.j}^2}{r} - CF = \frac{18.375^2 + 20^2 + 20^2 + 24^2 + 14^2}{3} - 619.209 = 17.338$$

$$SS_{Block} = \sum_i B_i^2 - CF = \frac{28^2 + 33.375^2 + 35^2}{5} - 619.209 = 5.3687$$

$$SS_{Error} = SS_{Total} - SS_{Treat} - SS_{Block} = 23.432 - 17.338 - 5.3687 = 0.725$$

Step 5: Critical Region

Reject the Null hypothesis if $\begin{cases} F_{Treat} \geq F_{0.05, (4,7)=4.12} \\ F_{Block} \geq F_{0.05, (2,7)=4.7374} \end{cases}$

For treatment means since $41.84914 > 4.12$ Reject H_0

For block means since $25.9181 > 4.7374$ Reject H_0

Step 6: Conclusion

As for treatments calculated F-value all in the critical region, it means that all the treatments means are not equal (or at least two treatment means are not equal). Similarly, for blocks the calculated F-value lies in the critical region, so conclude that all the block means are not equal.

□

Question 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.

SOV	df	SS	MS	F
Treatments	$k - 1 =$ $5 - 1 = 4$	$SS_{Treat} = 17.338$	$MS_{Treat} = \frac{17.338}{4}$ $= 4.33437$	$F_{Treat} = \frac{MS_{Treat}}{MS_{Error}}$ $= \frac{4.33437}{0.10357} = 41.84914$
Blocks	$r - 1 =$ $3 - 1 = 2$	$SS_{Block} = 5.3687$	$MS_{Block} = \frac{5.3687}{2}$ $= 2.68437$	$F_{Block} = \frac{MS_{Block}}{MS_{Error}}$ $= \frac{2.68437}{0.10357} = 25.9181$
Error	$(r - 1) \times$ $(k - 1) - 1 = 7$	$SS_{Error} = 0.725$	$MS_{Error} = \frac{0.725}{7}$ $= 0.10357$	
Total	$rk - 1 = 14$	$SS_{Total} = 23.431$		

Sov	df	SS	MS
Treatments	3	134.45	$MS_{Treat} = \frac{134.45}{3} = 44.816$
Blocks	4	21.46	$MS_{Block} = \frac{21.46}{4} = 5.365$
Error	12	26.26	$MS_{Error} = \frac{26.26}{12} = 2.188$
Total	19	182.17	

$$\begin{aligned}
 RE(RCBD, CRD) &= \frac{\text{Estimated MSE for CRD}}{\text{MSE for the RCBD}} \\
 &= \frac{(r - 1)MS_{Block} + r(k - 1)MS_{Error}}{(rk - 1)MS_{Error}} \\
 RE &= \frac{(5 - 1)(5.365) + (5)(4 - 1)(2.188)}{(5 \times 4 - 1)(2.188)} \\
 &= \frac{54.28}{41.572} = 1.305 = 130.5\%
 \end{aligned}$$

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

□

Lecture 5

Latin Square Design-I

Lesson 14: Introduction

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 one” 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 $K \times K$ 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**”.

14.1 Layout

Latin Squares are always constructed by rotations, e.g. in the case of four 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 of a 2×2 Latin Square as,

A	B
B	A

For a 3×3 Latin Square, we have also one Standard Latin Square,

A	B	C
B	C	A
C	A	B

For a 4×4 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 $K \times K$ yields $K!(K - 1)!$ different squares by permuting the rows except first and all the K columns. Accordingly, we may obtain from a square of design.

$$2 \times 2 \Rightarrow 2!(2 - 1)! = 2$$
$$3 \times 3 \Rightarrow 3!(3 - 1)! = 12$$
$$4 \times 4 \Rightarrow 4!(4 - 1)! = 576$$
$$5 \times 5 \Rightarrow 5!(5 - 1)! = 161280$$

14.2 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 $i, j, h = 1, 2, 3, \dots, k$

The μ denote the general mean

- R_i = is the effect of the i th row

C_j = is the effect of the j th column

τ_j = is the effect of the h th treatment

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

14.3 Statistical Analysis

1. Null and Alternative Hypothesis

For *Treatments* ($H_0 : \tau_h = 0$ vs $H_1 : \tau_h \neq 0$)

$H_0 : \mu_1 = \mu_2 = \dots$

H_1 : Not all treatment means are equal

For *Columns* ($H_0 : C_j = 0$ vs $H_1 : C_j \neq 0$)

$H_0 : \mu_1 = \mu_2 = \dots$

H_1 : Not all Column means are equal

For Rows ($H_0 : R_i = 0$ vs $H_1 : R_i \neq 0$)
 $H_0 : \mu_1 = \mu_2 = \dots$
 $H_1 : \text{Not all Row means are equal}$

- 2. **Level of Significance:** $\alpha = 5\%$
- 3. **Test Statistics:**

$$F_{Treat} = \frac{MS_{Treat}}{MS_{Error}}, \quad F_{Column} = \frac{MS_{Column}}{MS_{Error}}, \quad F_{Row} = \frac{MS_{Row}}{MS_{Error}}$$

Calculations

	Rows	Columns						R_i	Means
		1	2	3	4	...	k	Total	
	1	Y_{11A}	Y_{12B}	Y_{13C}	Y_{14D}	...	...	$R_{1..}$	$\bar{Y}_{1..}$
	2	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$R_{2..}$	$\bar{Y}_{2..}$
	3	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$R_{3..}$	$\bar{Y}_{3..}$
	4	Y_{41A}	Y_{42B}	Y_{43C}	Y_{44D}	...	...	$R_{4..}$	$\bar{Y}_{4..}$
	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
	K	Y_{K1A}	Y_{K2B}	Y_{K3C}	Y_{K4D}	...	...	$R_{K..}$	$\bar{Y}_{K..}$
Total C_j		$C_{.1}$	$C_{.2}$	...	...	...	...	$G = T_{..}$	
Means		$\bar{Y}_{.1}$	$\bar{Y}_{.2}$	$\bar{Y}_{.3}$	$\bar{Y}_{.4}$	$\bar{Y}_{.5}$	...	$\bar{Y}_{.K}$	$\bar{Y}_{...}$

Partitioning of Total Sum of Squares

$$SS_{Total} = SS_{Row} + SS_{Column} + SS_{Treat} + SS_{Error}$$
$$SS_{Error} = SS_{Total} - SS_{Row} - SS_{Column} - SS_{Treat}$$

Degrees of Freedom

$$SS_{Total} = n - 1 = k \times k - 1$$
$$SS_{Row} = k - 1$$
$$SS_{Column} = k - 1$$
$$SS_{Treat} = k - 1$$
$$SS_{Error} = k^2 - 1 - 3(k - 1) = (k - 1)(k - 2)$$

Sum of Squares

$$CS = \frac{(T_{...})^2}{n} - \frac{GT^2}{k^2}$$
$$SS_{Total} = \sum_i \sum_j (y_{ij(h)}^2) - CF$$
$$SS_{Row} = \frac{\sum_i R^2}{k} - CF$$
$$SS_{Column} = \frac{\sum_j C_j^2}{k} - CF$$
$$SS_{Treat} = \frac{\sum_h T_h^2}{k} - CF$$

ANOVA Table

SOV	df	SS	MS	F
Treatment	$k - 1$	SS_{Treat}	$MS_{Treat} = \frac{SS_{Treat}}{k-1}$	$F_{Treat} = \frac{MS_{Treat}}{MS_{Error}}$
Row	$k - 1$	SS_{Row}	$MS_{Row} = \frac{SS_{Row}}{k-1}$	$F_{Row} = \frac{MS_{Row}}{MS_{Error}}$
Column	$k - 1$	SS_{Column}	$MS_{column} = \frac{SS_{column}}{k-1}$	$F_{column} = \frac{MS_{column}}{MS_{Error}}$
Error	$(k - 1)(k - 2)$	SS_{Error}	$MS_{Error} = \frac{SS_{Error}}{(k-1)(k-2)}$	
Total	$k^2 - 1$	SS_{Total}		

4. Critical Region

Reject the null hypothesis when

$$F_{Treat} \geq F_{\alpha, (df_{treat}, df_{error})}$$

$$F_{Row} \geq F_{\alpha, (df_{Row}, df_{error})}$$

$$F_{Column} \geq F_{\alpha, (df_{Column}, df_{error})}$$

5. Conclusion

Example 1: 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 test. Is there any

Heights	Districts			
	1	2	3	4
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

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

Solution:

1. Null and Alternative Hypothesis

For *Treatments* ($H_0 : \tau_h = 0$ vs $H_1 : \tau_h \neq 0$)

$H_0 : \mu_1 = \mu_2 = \dots$

H_1 : Not all treatment means are equal

For *Columns* ($H_0 : C_j = 0$ vs $H_1 : C_j \neq 0$)

$H_0 : \mu_1 = \mu_2 = \dots$

H_1 : Not all Column means are equal

For *Rows* ($H_0 : R_i = 0$ vs $H_1 : R_i \neq 0$)

$H_0 : \mu_1 = \mu_2 = \dots$

H_1 : Not all Row means are equal

2. Level of Significance: $\alpha = 5\%$

3. Test Statistics:

$$F_{Treat} = \frac{MS_{Treat}}{MS_{Error}}, \quad F_{Column} = \frac{MS_{Column}}{MS_{Error}}, \quad F_{Row} = \frac{MS_{Row}}{MS_{Error}}$$

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	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	
$\sum_i \sum_j y_{ij(h)}^2$	182.42	85.08	65.62	77.88	411	

SOV	df	SS	MS	F
Treatment	4-1=3	2.3075	$MS_{Treat} = \frac{2.3075}{3} = 0.7692$	$F_{Treat} = \frac{0.7692}{0.2490} = 3.09$
Row	3	3.5275	$MS_{Row} = \frac{3.5275}{3} = 1.1758$	$F_{Row} = \frac{1.1758}{0.2490} = 4.72$
Column	3	17.5475	$MS_{Column} = \frac{17.5475}{3} = 5.8492$	$F_{Column} = \frac{5.8492}{0.2490} = 23.4749$
Error	6	1.4950	$MS_{Error} = \frac{1.4950}{6} = 0.2490$	
Total	15	24.8775		

$$\begin{aligned}
 CF &= \frac{GT^2}{k^2} = \frac{(78)^2}{16} \\
 SS_{Total} &= \sum_i \sum_j y_{ij(h)}^2 - CF \\
 &= 411 - 386.1225 = 24.8775 \\
 SS_{Row} &= \frac{\sum R_i^2}{k} - Cf = \frac{1558.60}{4} - 386.1225 = 3.5275 \\
 SS_{Column} &= \frac{\sum_j C_j^2}{k} - CF = \frac{1614.68}{4} - 386.1225 = 17.5475 \\
 SS_{Treat} &= \frac{\sum T_h^2}{k} - CF = \frac{1553.72}{4} - 386.1225 = 2.3095 \\
 SS_{Error} &= SS_{Total} - SS_{Row} - SS_{Column} - SS_{Treat} \\
 &= 24.8775 - 3.5275 - 17.5475 - 2.3075 = 1.4950
 \end{aligned}$$

4. Critical Region: Reject Null hypothesis if

$$\begin{aligned}
 F_{Treat} &> F_{0.05(3,6)} = 4.76 \\
 F_{Row} &> F_{0.05(3,6)} = 4.76 \\
 F_{Column} &> F_{0.05(3,6)} = 4.76
 \end{aligned}$$

5. 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.

□

Lecture 6

Latin Square Design-II

Lesson 15: Latin Square

15.1 Advantages and 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.

15.2 Missing Observations in Latin Square Design

	Rows	Columns						R_i
		1	2	3	4	...	k	Total
	1	x	Y_{12B}	Y_{13C}	Y_{14D}	...	...	$R_1 + x$
	2	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	R_2
	3	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$R_{3..}$
	4	Y_{41A}	Y_{42B}	Y_{43C}	Y_{44D}	...	...	R_4
	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
	K	Y_{K1A}	Y_{K2B}	Y_{K3C}	Y_{K4D}	...	...	R_K
Total C_j		$C'_1 + x$	C_2	C_3	C_4	...	...	$G' + x$

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

	Treatments				
	<i>A</i>	<i>B</i>	<i>C</i>	...	<i>K</i>
	—	—	—	...	—
	—	—	—	...	—
e get Total	$T'_1 + x$	T_2	T_3	...	T_k

SOV	Sum of Squares
Rows	$SS_{Row} = \frac{(R'_1 + x)^2 + \sum_{i=2}^k R_i^2}{k} - \frac{(G' + x)^2}{k^2}$
Columns	$SS_{Column} = \frac{(C'_1 + x)^2 + \sum_{j=2}^k C_j^2}{k} - \frac{(G' + x)^2}{k^2}$
Treatments	$SS_{Treat} = \frac{(T'_1 + x)^2 + \sum_{h=2}^k T_h^2}{K} - \frac{(G' + x)^2}{k^2}$
Error	$SS_{Error} = SS_{Total} - SS_{Row} - SS_{Column} - SS_{Treat}$
Total	$SS_{Total} = x^2 + \sum_i \sum_j y_{ij}^2 - \frac{(G' + x)^2}{k^2}$

The Error sum of squares can be obtained as

$$\begin{aligned}
 SS_{Error} &= SS_{Total} - SS_{Row} - SS_{Column} - SS_{Treat} \\
 &= x^2 + \sum_i \sum_j y_{ij}^2 - \frac{(G' + x)^2}{k^2} - \frac{(R'_1 + x)^2}{k} - \frac{\sum_{i=2}^k R_i^2}{k} + \frac{(G' + x)^2}{k^2} \\
 &\quad - \frac{(C'_1 + x)^2}{k} - \frac{\sum_{j=2}^k C_j^2}{k} + \frac{(G' + x)^2}{k^2} - \frac{(T'_1 + x)^2}{k} - \frac{\sum_{h=2}^k T_h^2}{k} + \frac{(G' + x)^2}{k^2} \\
 &= x^2 + \frac{2(G' + x)^2}{k^2} - \frac{1}{k} \left[(R'_1 + x)^2 + (C'_1 + x)^2 + (T'_1 + x)^2 \right] + \text{Terms in which } x \text{ is not involved}
 \end{aligned}$$

Differentiate SS_{Error} w.r.t. x

$$\begin{aligned}\frac{\partial SS_{Error}}{\partial x} &= 2x + \frac{(2 \times 2)(G' + x)}{k^2} - \frac{1}{k} [2(R'_1 + x) + 2(G' + x) + 2(T'_1 + x)] \\&\Rightarrow 2x + \frac{4(G' + x)}{k^2} - \frac{2}{k} [(R'_1 + x) + (C'_1 + x) + (T'_1 + x)] = 0 \\&\Rightarrow 2x + \frac{4G'}{k^2} + \frac{4x}{k^2} - \frac{2R'_1}{k} - \frac{2x}{k} - \frac{2C'_1}{k} - \frac{2x}{k} - \frac{2T'_1}{k} - \frac{2x}{k} = 0 \\&\Rightarrow 2x + \frac{4x}{k^2} - \frac{2x}{k} - \frac{2x}{k} - \frac{2x}{k} - \frac{2R'_1}{k} - \frac{2C'_1}{k} - \frac{2T'_1}{k} + \frac{4G'}{k^2} = 0 \\&\Rightarrow 2x \left(1 + \frac{1}{k^2} - \frac{1}{k} - \frac{1}{k} - \frac{1}{k} \right) = \frac{2}{k} (R'_1 + C'_1 + T'_1) - \frac{4G'}{k^2} \\&\Rightarrow x \left(1 + \frac{2}{k^2} - \frac{3}{k} \right) = \frac{2}{2k} (R'_1 + C'_1 + T'_1) - \frac{4G'}{2k^2} \\&\Rightarrow x \left(\frac{k^2 - 3k + 2}{k^2} \right) = \frac{1}{k} (R'_1 + C'_1 + T'_1) - \frac{2G'}{k^2} \\&\Rightarrow x \left(\frac{k^2 - k - 2k + 2}{k^2} \right) = \frac{k(R'_1 + C'_1 + T'_1) - 2G'}{k^2} \\&\Rightarrow x \left(\frac{(k - 1)(k - 2)}{k^2} \right) = \frac{k(R'_1 + C'_1 + T'_1) - 2G'}{k^2} \\&x = \frac{k(R'_1 + C'_1 + T'_1) - 2G'}{k^2} \times \frac{k^2}{(k - 1)(k - 2)} \\&x = \frac{k(R'_1 + C'_1 + T'_1) - 2G'}{(k - 1)(k - 2)}\end{aligned}$$

In general

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

- where
- k = is number of treatments
 - R = is row total
 - C = is column total
 - G = is grand total

Example 1: 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	D 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.

Solution:

A 7.4	D x	E 5.8	D 12.0	C 14.3	R = 38.5
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

$$x = \frac{k(R + C + T02G)}{(k - 1)(k - 2)},$$
$$= \frac{5(39.5 + 43.3 + 43.4) - 2(251.1)}{(5 - 1)(5 - 2)} = \frac{631 - 502.2}{12}$$
$$x = 10.7$$

$k = 5$

- Step 1: Hypotheses
- H_0 : All effects on yield due to variation of soil in columns are zero
- H_1 : All effects on yield due to variation of soil in columns are not zero
- H'_0 : All effects on yield due to variation of soil in rows are zero
- H'_1 : All effects on yield due to variation of soil in rows are not zero
- H''_0 : All effects on yield due to change in treatments are zero
- H''_1 : All effects on yield due to change in treatments are not zero

- Step 2: Level of significance: $\alpha = 0.05$
- Step 3: Test Statistics
- Required Calculations

Rows	Columns					R_i	R_i^2
	1	2	3	4	5		
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	$T_{..} = 261.8$	13932.23
C_j^2	2490.01	2916	2862.25	3317.76	2190.24	13776.26	
$\sum_i \sum_j y_{ij(h)}^2$	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	71.	
	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

$$CF \frac{(T_{...})^2}{k^2} = \frac{(261.8)^2}{25} = 2741.5696$$
$$SS_{Total} = \sum_i \sum_j y_{ij(h)}^2 - CF = 3020.46 - 2741.5696 = 278.8904$$
$$SS_{Column} = \frac{\sum_j C_j^2}{k} - CF = \frac{13776.26}{5} - 2741.5696 = 13.6824$$
$$SS_{Row} = \frac{\sum_i R_i^2}{k} - CF = \frac{13932.22}{5} - 2741.5696 = 44.8744$$
$$SS_{Treat} = \frac{\sum_h T_h^2}{k} - CF = \frac{14665.96}{5} - 2741.5696 = 191.6224$$

ANOVA Table

SOV	df	SS	MS	F
Columns	$k - 1 = 4$	$SS_{column} = 13.6824$	$MS_{column} = \frac{13.6824}{4}$	$F_{column} = \frac{3.4206}{2.6101}$
			$= 3.4206$	$= 1.3105$
			$MS_{row} = \frac{44.8744}{4}$	$F_{row} = \frac{11.2186}{2.6101}$
Rows	4	$SS_{row} = 44.8744$	$= 11.2186$	$= 4.2981$
			$MS_{Treat} = \frac{191.6224}{4}$	$F_{Treat} = \frac{47.9056}{2.6101}$
Treatments	4	$SS_{Treat} = 191.6224$	$= 47.9056$	$= 18.3539$
			$MS_{Error} = \frac{28.7114}{11}$	
Error	$(k - 1)(k - 2) - 1$ $= 11$	$SS_{Error} = 28.7114$	$= 2.6101$	
Total	$k^2 - 1 = 23$	$SS_{Total} = 278.8904$		

- Step 4: Critical Region
Reject Null hypothesis if $F_{cal} \geq F_{tab}$
 $F_{0.05(4,11)} = 3.36$
- Step 5: Conclusion:
As $F_{column} < F_{tab}$, we do not reject H_0
The variation of soil in columns is not significant.
 $F_{row} > F_{tab}$, we reject H'_0
There are significant difference in soil in rows.
 $F_{Treat} > R_{tab}$, we reject H''_0
The effects on yield due to change in treatments are significantly different.

□

Example 2: The following data represent the yield (in tons per ha) of found cultivars of Khaya arranged in 4×4 Latin Square with one missing value. The four cultivars are $A(KH4)$, $B(KH20)$, $C(KH29)$, and $D(KH15)$ Estimate the missing observation and perform the analysis of variance to test for significant differences among the cultivars.

Solution:

- Step 1: Hypotheses
 H_0 : All effects on yield due to variation of soil in columns are zero

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)	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)

H_1 : All effects on yield due to variation of soil in columns are not zero

H'_0 : All effects on yield due to variation of soil in rows are zero

H'_1 : All effects on yield due to variation of soil in rows are not zero

H''_0 : All effects on yield due to change in treatments are zero

H''_1 : All effects on yield due to change in treatments are not zero

Step 2: Level of significance: $\alpha = 0.05$

Step 3: Test Statistics

Required Calculations

Rows	Columns				Total
	1	2	3	4	
I	6.2 (C)	5.1 (D)	3.3 (B)	5.7 (A)	R = 12.8
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)	
Total	C = 15.4			G = 71.4	

Treatment C: $6.2 + M + 4.8 + 5.4 = 16.4$

$$\begin{aligned} x &= \frac{K(R + C + T) - 2G}{(k - 1)(k - 2)} \\ &= \frac{(4)(12.8 + 15.4 + 16.4) - 2(71.4)}{(4 - 1)(4 - 2)} = 5.9 \end{aligned}$$

ANOVA Table

2	Columns				Total
	1	2	3	4	
I	6.2 (C)	5.1 (D)	3.3 (B)	5.7 (A)	R = 12.8
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)	
Total	C = 15.4			G = 71.4	

Step 4: Critical Region

Reject Null hypothesis if $F_{cal} \geq F_{tab}$

$F_{0.05(3,5)} = 5.4095$

SOv	df	SS	MS	F
Rows	3	3.4675	1.1558	0.7326
Columns	3	2.9325	0.9775	0.6023
Treatments	3	5.1825	1.7275	1.0643
Error	5	8.115	1.623	
Total	15	19.6875		

Step 5: Conclusion:

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

The variation of cultivars in columns is not significant.

$F_{row} < F_{tab}$, we reject H'_0

There are non significant difference in cultivars in rows.

$F_{Treat} < R_{tab}$, we reject H''_0

The effects on yield due to change in treatments are not significantly different.

□

Lecture 7

Latin Square Designs-II

Lesson 16: Efficiency of Latin Square Designs

The efficiency of Latin Square design can be measured as

- ◆ **Relative to RCBD**
 - ★ When Rows are taken as blocks, that is, $RE(LS, RCBD_{rows})$
$$\frac{1}{k} \left[\frac{MS_{columns}}{MS_{Error}} + (k - 1) \right], \text{ where } k \text{ is number of rows}$$
 - ★ When Columns are taken as blocks, that is, $RE(LS, RCBD_{columns})$
$$\frac{1}{k} \left[\frac{MS_{rows}}{MS_{Error}} + (k - 1) \right], \text{ where } k \text{ is number of columns}$$
- ◆ **Relative to CRD**, that is, $RE(LS, CRD)$
$$\frac{1}{k} \left[\frac{MS_{rows} + MS_{columns} + (k - 1)MS_{Error}}{(k + 1)MS_{Error}} \right]$$

Example 1: 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 of 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 4×4 Latin Square Design is implemented. The mean square period, mean square cows, and mean

Period	Cow			
	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

square error are 49.06, 18.2292, and 0.8125, respectively. Estimate the relative efficiency of Latin Square Design to RCBD,

- ◆ By taking the cows as blocks
- ◆ By taking the lactation period as blocks

Solution:

◆ By Taking the Columns as Blocks

$$\begin{aligned} RE(LS, RCBD_{columns}) &= \frac{1}{k} \left[\frac{MSE_{rows}}{MS_{Error}} + (k - 1) \right] \\ &= \frac{1}{4} \left[\frac{49.06}{0.8125} + (4 - 1) \right] = 15.8454 \end{aligned}$$

◆ By taking the Rows as Blocks

$$\begin{aligned} RE(LS, RCBD_{rows}) &= \frac{1}{k} \left[\frac{MSE_{columns}}{MS_{Error}} + (k - 1) \right] \\ &= \frac{1}{4} \left[\frac{18.2292}{0.8125} + (4 - 1) \right] = 6.36 \end{aligned}$$

□

Example 2: Consider the given ANOVA table for 5-brands of air-cleaner in 5×5 Latin Square and find the efficiency of this design relative to CRD.

SOV	df	SS	MS	F
Row	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		

Solution:

$$\begin{aligned} RE(SL, CRD) &= \frac{MS_{rows} + MS_{columns} + (k - 1)MS_{Error}}{(k + 1)MS_{Error}} \\ &= \frac{11128.14 + 544.44 + (5 - 1)(2887.29)}{(5 + 1)(2887.29)} \\ &= 1.34 \Rightarrow 134 - 100 = 34\% \end{aligned}$$

Results shows that 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. □

Lesson 17: 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 Multiple Comparison Tests are:

- ◆ 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

17.1 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 significance 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

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

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

Example 3: 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.

Rows	Columns				
	1	2	3	4	5
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

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.

Solution:

Step 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

Step 2: Level of significance: $\alpha = 0.05$

Step 3: Test Statistics:

$$F = \frac{\text{MS for fertilizers}}{\text{MS for Error}}$$

Calculations

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

Rows	Columns					R_i	F_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	

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

$$CF = \frac{GT^2}{k^2} =$$

$$SS_{Total} = \sum_i \sum_j y_{ij(h)}^2 - CF = 285.34$$

$$SS_{row} = \frac{\sum_i R_i^2}{k} - CF = 46.67$$

$$SS_{column} = \frac{\sum_j C_j^2}{k} - CF = 14.02$$

$$SS_{Treatment} = \frac{\sum_h T_h^2}{k} - CF = 196.61$$

$$SS_{Error} = SS_{Total} - SS_{row} - SS_{column} - SS_{Treatment} = 28.04$$

ANOVA Table

SOV	df	SS	MS	F
Rows	4	46.67	11.67	
Columns	4	14.02	3.51	
Treatments	4	196.61	49.15	$F = \frac{49.15}{2.34} = 21.00$
Error	12	28.04	2.34	
Total	24	285.34		

- Step 4: Critical Region
- Reject H_0 if $F_{cal} = 21.00 \geq F_{0.05(4,12)} = 3.26$
- Step 5: Conclusion
- Since the computed value of $F_{cal} = 21.00$ falls in the critical region, we therefore, reject our null hypothesis. 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 Significant 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

$$\begin{aligned}
 LSD &= t_{\frac{\alpha}{2}, v} \sqrt{\frac{2MSE}{k}} \\
 &= t_{0.025, 12} \sqrt{\frac{2(2.34)}{5}} = 2.18(0.96) \\
 &= 2.09
 \end{aligned}$$

Pair-wise Differences

Total number of pairs $\binom{k}{2} = \binom{5}{2} = 10$ pairs

Pairs	LSD-value	Decision
$ \bar{Y}_A - \bar{Y}_B $	$= 6.84 - 6.46 = 0.38 < 2.09$	Non-significant
$ \bar{Y}_A - \bar{Y}_C $	$= 6.84 - 13.12 = 6.28 > 2.09$	Significant
$ \bar{Y}_A - \bar{Y}_D $	$= 6.84 - 7.96 = 1.12 < 2.09$	Non-significant
$ \bar{Y}_A - \bar{Y}_E $	$= 6.84 - 4.92 = 1.92 < 2.09$	Non-significant
$ \bar{Y}_B - \bar{Y}_C $	$= 6.46 - 13.12 = 6.66 > 2.09$	Significant
$ \bar{Y}_B - \bar{Y}_D $	$= 6.46 - 7.96 = 1.5 < 2.09$	Non-significant
$ \bar{Y}_B - \bar{Y}_E $	$= 6.46 - 4.92 = 1.54 < 2.09$	Non-significant
$ \bar{Y}_C - \bar{Y}_D $	$= 13.12 - 7.96 = 5.16 > 2.09$	Significant
$ \bar{Y}_C - \bar{Y}_E $	$= 13.12 - 4.92 = 8.2 > 2.09$	Significant
$ \bar{Y}_D - \bar{Y}_E $	$= 7.96 - 4.92 = 3.04 > 2.09$	Significant

□

Example 4: 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.

Blocks	Treatments					
	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

Solution:

Step 1: Hypothesis

$$H_0 : \mu_1 = \mu_2 = \dots = \mu_6$$

H_1 : Not all six treatments means are equal

Step 2: Level of significance ($\alpha = 0.05$)

Blocks	Treatments						B_i	B_i^2
	1	2	3	4	5	6		
a	1	3	6	4	3	2	19	361
b	1	4	4	8	5	1	23	529
c	3	6	7	8	4	3	31	961
d	2	3	2	3	2	1	13	169
C_j	7	16	19	23	14	7	86	2020
C_j^2	49	256	361	529	196	49	1440	
Mean	1.75	4	4.75	5.75	3.5	1.75		
$\sum_i \sum_j y_{ij(h)}^2 = 412$								

Step 3: Test Statistics,

$$F = \frac{MS_{Treatments}}{MS_{Error}}$$

ANOVA Table

SOV	df	SS	MS	F
Treatments	5	$SS_{Treat} = \frac{1440}{4} - \frac{86^2}{24} = 51.83$	$MS_{Treat} = \frac{51.83}{5} = 10.37$	$F = \frac{10.37}{1.57} = 6.61$
Blocks	3	$SS_{Treat} = \frac{2020}{6} - \frac{86^2}{24} = 28.50$	$MS_{Treat} = \frac{28.50}{3} = 9.50$	$F = \frac{9.50}{1.57} = 6.05$
Error	15	$SS_{Error} = 23.50$	$MS_{Error} = \frac{23.50}{15} = 1.57$	
Total	23	$SS_{Total} = 412 - \frac{86^2}{24} = 103.83$		

Step 4: Critical Region

Reject H_0 : if $F_{cal} \geq F_{0.05(5,15)}$

Step 5: Conclusion

Since $F_{cal} = 6.61 > F_{tab} = 2.90$, we reject the null hypothesis of equal treatment means. As the null hypothesis is rejected, using F -test, we therefore, apply the LSD test to find out which means differ from each other.

Pair-Wise comparison using LSD Test

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

$$\begin{aligned}
 LSD &= t_{\frac{\alpha}{2}, v} \sqrt{\frac{2MSE}{r}} \\
 &= 2.1314 \sqrt{\frac{2(1.57)}{4}} = 1.89
 \end{aligned}$$

Number of pairs possible $\binom{6}{2} = 15$ pairs

□

Pairs		LSD Value	Decision
$ \bar{Y}_1 - \bar{Y}_2 = 1.75 - 4.0 = 2.25$	$>$	1.89	Significant
$ \bar{Y}_1 - \bar{Y}_3 = 1.75 - 4.75 = 3$	$>$	1.89	Significant
$ \bar{Y}_1 - \bar{Y}_4 = 1.75 - 5.75 = 4$	$.$	1.89	Significant
$ \bar{Y}_1 - \bar{Y}_5 = 1.75 - 3.50 = 1.75$	$<$	1.89	Non-significant
$ \bar{Y}_1 - \bar{Y}_6 = 1.75 - 1.75 = 0$	$<$	1.89	Non-significant
$ \bar{Y}_2 - \bar{Y}_3 = 4.0 - 4.75 = 0.75$	$<$	1.89	Non-significant
$ \bar{Y}_2 - \bar{Y}_4 = 4.0 - 5.75 = 1.75$	$<$	1.89	Non-significant
$ \bar{Y}_2 - \bar{Y}_5 = 4.0 - 3.50 = 1.50$	$<$	1.89	Non-significant
$ \bar{Y}_2 - \bar{Y}_6 = 4.0 - 1.75 = 2.25$	$>$	1.89	Significant
$ \bar{Y}_3 - \bar{Y}_4 = 4.75 - 5.75 = 1$	$<$	1.89	Non-significant
$ \bar{Y}_3 - \bar{Y}_5 = 4.75 - 3.50 = 1.75$	$<$	1.89	Non-significant
$ \bar{Y}_3 - \bar{Y}_6 = 4.75 - 1.75 = 3$	$>$	1.89	Significant
$ \bar{Y}_4 - \bar{Y}_5 = 5.75 - 3.50 = 2.75$	$>$	1.89	Significant
$ \bar{Y}_4 - \bar{Y}_6 = 5.75 - 1.75 = 4$	$>$	1.89	Significant
$ \bar{Y}_5 - \bar{Y}_6 = 3.50 - 1.75 = 1.75$	$<$	1.89	Non-significant

Example 5: 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:

Day	Graphite Coater Type			
	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.

Solution:

- Step 1: Hypothesis
- H_0 : There is no significant difference between coater types
- H_1 : There is significant difference between coater types
- Step 2: Level of significance ($\alpha = 0.05$)
- Step 3: Test Statistics $F = \frac{MS_{Treat}}{MS_{Error}}$

Calculations

Day	Graphite Coater Type					
	M	A	K	L		
a	4	4.8	5	4.6	18.4	338.56
b	4.8	5	5.2	4.6	19.6	384.16
c	4	4.8	5.6	5	19.4	376.36
C_j	12.8	14.6	15.8	14.2	57.4	1099.08
C_j^2	163.84	213.16	249.64	201.64	828.28	
$\sum_i \sum_j y_{ij(h)}^2 = 276.84$						
Mean	$\bar{Y}_m = 4.27$	$\bar{Y}_a = 4.87$	$\bar{Y}_k = 5.27$	$\bar{Y}_c = 4.73$		

ANOVA Table

SOV	df	SS	MS	F
Coaters	3	$SS_{Coater} = \frac{828.28}{3} - \frac{57.4^2}{12} = 1.53$	$MS_{coaters} = \frac{1.53}{3} = 0.52$	$F = \frac{0.51}{0.09} = 5.667$
Days	2	$SS_{Days} = \frac{1099.08}{4} - \frac{57.4^2}{12} = 0.2067$	$MS_{Days} = \frac{0.2067}{2} = 0.1033$	$F = \frac{0.1033}{0.09} = 1.148$
Error	6	0.54	0.09	
Total	11	$SS_{Total} = 276.84 - \frac{57.4^2}{12}$		

- Step 4: Critical Value: $F_{0.05,(3,6)} = 4.757$
- Step 5: Decision: $F_{cal} > F_{tab}$, H_1 is accepted.

LSD Test

$$LSD = t_{\frac{\alpha}{2},v} = \sqrt{\frac{2MSE}{r}}$$
$$= t_{\frac{0.05}{2},6} = 2.446\sqrt{\frac{2 \times 0.09}{3}} = 0.60$$

Pair-wise Comparison using LSD test

Number of possible pairs = $\binom{4}{2} = 6$ pairs

Pairs		LSD Value	Decision
$ \bar{Y}_m - \bar{Y}_a = 4.27 - 4.87 = 0.60$	=	0.60	Significant
$ \bar{Y}_m - \bar{Y}_k = 4.27 - 5.27 = 1.0$	>	0.60	Significant
$ \bar{Y}_m - \bar{Y}_c = 4.27 - 4.73 = 0.460$	<	0.60	Non-significant
$ \bar{Y}_a - \bar{Y}_k = 4.87 - 5.27 = 0.40$	<	0.60	Non-significant
$ \bar{Y}_a - \bar{Y}_c = 4.87 - 4.73 = 0.140$	<	0.60	Non-significant
$ \bar{Y}_k - \bar{Y}_c = 5.27 - 4.73 = 0.540$	<	0.60	Non-significant

□

Lecture 8

Factorial Design

Lesson 18: Introduction

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

The independent variable such as fertilizers, planting, education, and 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**.

18.1 What is a Factorial Design?

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 from the effect of other factors. The main effect is the effect of the factors 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. 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 interaction and so on.

18.2 Advantage and Disadvantages of Factorial Experiment

Advantages:

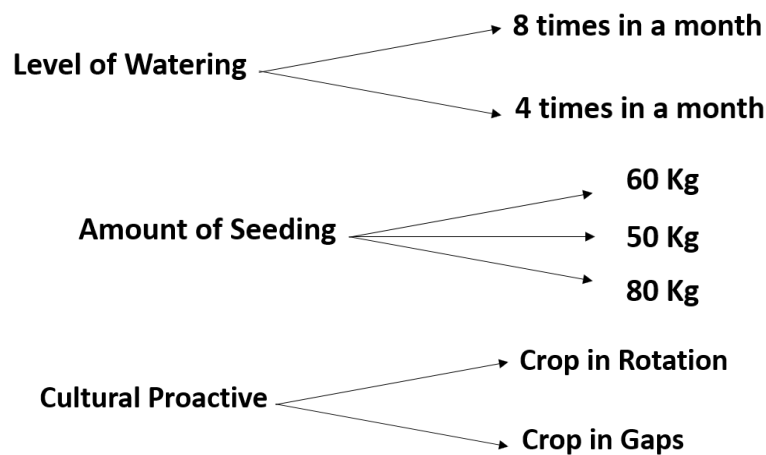
- ◆ A factorial experiment is usually economical.

- ◆ 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.

Disadvantages:

- ◆ A factorial experiment requires an excessive amount of experimentation when there are several factors at several levels. For example, for 8 factors, each factor at 2 level, there will be 256 combinations. Similarly, for 7 factors each at 3 level, there will be 2187 combination.
- ◆ A large number combinations when used decrease 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.

See the following diagrams.

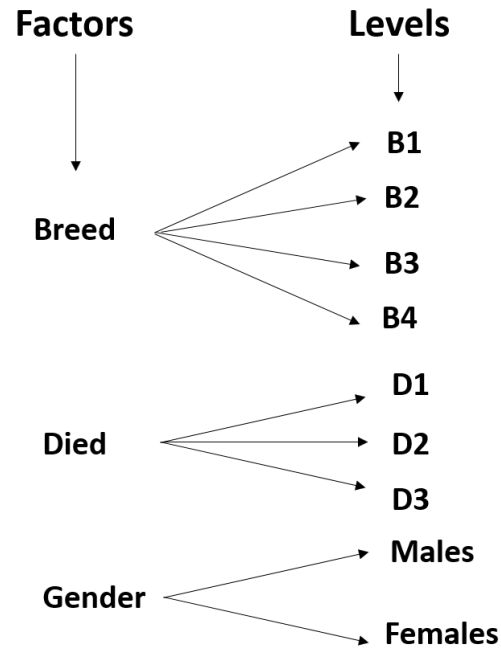


The treatment combinations are

		Diet (D)					
		d_1		d_2		d_2	
		m	f	m	f	m	f
Breed	b_1	b_1md_1	b_1fd_1	b_1md_2	b_1fd_2	b_1md_3	b_1fd_3
	b_2	b_2md_1	b_2fd_1	b_2md_2	b_2fd_2	b_2md_3	b_2fd_3
	b_3	b_3md_1	b_3fd_1	b_3md_2	b_3fd_2	b_3md_3	b_3fd_3
	b_4	b_4md_1	b_4fd_1	b_4md_2	b_4fd_2	b_4md_3	b_4fd_3

Example 1: A soft drink company is studying the effect of percent carbonation (A), the speed (C), and operating pressure (B) in the filler on the volume of the carbonated drink pack. In each bottle, 3 levels of carbonated, 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 determined randomly. The coded data are given in the table.

Solution:



		Operating Pressure (B)			
		25 psi		30 psi	
		Line Speed (C)			
		200	250	200	250
% Carbonated (A)	10%	-3	-1	-1	1
		-1	0	0	1
	12%	0	2	2	6
		1	1	3	5
	14%	5	7	7	10
		4	6	9	11
% Carbonated (A)					

H_0 : Effect of main effect (A, B, C) is non-significant.

H'_0 : Effect of first order interaction (AB, AC, BC) is non-significant.

H''_0 : Effect of second order interaction (ABC) is non-significant

Let $i \rightarrow A \rightarrow a$, $j \rightarrow B \rightarrow b$, $k \rightarrow C \rightarrow c$, and $l \rightarrow$ Number of replications (n)

		Operating Pressure (B)				$Y_{i...}$
		25 psi		30 psi		
		Line Speed (C)				
		200	250	200	250	
% Carbonated (A)	10%	-3	-1	-1	1	-4
		-1	0	0	1	
	20%	0	2	2	6	20
		1	1	3	5	
	14%	5	7	7	10	59
		4	6	9	11	
$B \times C$ total	$Y_{.jk.}$	6	15	20	34	$Y_{...} = 75$

$$CF = \frac{(Y_{...})^2}{abcn} = \frac{75^2}{3 \times 2 \times 2 \times 2} = \frac{5625}{24} = 234.375$$

$$SS_C = SS_{Speed} = \frac{\sum_{k=1}^c Y_{..k}^2}{abn} - CF = \frac{(6 + 20)^2 + (15_34)^2}{3 \times 2 \times 2} - 234.375 = 22.042$$

$$SS_{Total} = \sum_{i=1}^a \sum_{j=1}^b \sum_{k=1}^c \sum_{l=1}^n Y_{ijkl}^2 - CF$$
$$= (-3)^2 + (-1)^2 + 0^2 + 1^2 + \dots + 10^2 + 11^2 - 234.375 = 336.625$$

$$SS_A = SS_{carb} = \frac{\sum_{i=1}^a Y_{i...}^2}{bcn} - CF$$
$$= \frac{(-4)^2 + (20)^2 + (59)^2}{2 \times 2 \times 2} - 234.375 = 252.750$$

$$SS_B = SS_{Pressure} = \frac{\sum_{j=1}^b Y_{.j..}^2}{acn} - CF = \frac{(21)^2 + (54)^2}{3 \times 2 \times 2} - 234.375 = 45.375$$

$$SS_{AB} = \frac{\sum_{i=1}^2 \sum_{j=1}^b Y_{ij..}^2}{cn} - CF - SS_A - SS_B$$
$$= \frac{(-5)^2 + 1^2 + \dots + 37^2}{4} - 234.375 - 252.750 - 45.375 = 5.250$$

Table of <i>A × B</i>		
<i>A</i>	<i>B</i>	
	25psi	30psi
10	-5	1
12	4	16
14	22	37

where Total $SS_{AB} = \frac{\sum_{i=1}^2 \sum_{j=1}^b Y_{ij..}^2}{cn} - CF$

$$SS_{AC} = \frac{\sum_{i=1}^a \sum_{k=1}^c Y_{i.k.}^2}{bn} - CF - SS_A - SS_C$$
$$= \frac{(-5)^2 + 1^2 + \dots + 34^2}{4} - 234.375 - 252.750 - 22.042 = 0.583$$

Table of <i>A × C</i>		
<i>A</i>	<i>C</i>	
	200	250
10	-5	1
12	6	14
14	25	34

where Total $SS_{AC} = \frac{\sum_{i=1}^a \sum_{k=1}^c Y_{i.k.}^2}{bn} - CF$

$$SS_{BC} = \frac{\sum_{j=1}^b \sum_{k=1}^c Y_{jk}}{an} - CF - SS_B - SS_C$$

$$= \frac{6^2 + 15^2 + \dots + 34^2}{6} - 234.375 - 45.375 - 22.042 = 1.042$$

Table of $B \times C$		
B	C	
	200	250
25 psi	6	15
30 psi	20	34

where Total $SS_{BC} = \frac{\sum_{j=1}^b \sum_{k=1}^c Y_{jk}}{an} - CF$

$$SS_{ABC} = \frac{\sum_{i=1}^a \sum_{j=1}^b \sum_{k=1}^c Y_{ijk}}{n} - CF - SS_A - SS_B$$

$$- SS_C - SS_{AB} - SS_{AC} - SS_{BC}$$

$$= \frac{-4^2 + -1^2 + -1^2 + \dots + 27^2}{6} - 234.375 - 252.75$$

$$- 45.375 - 22.042 - 5.25 - 0.583 - 1.042 = 1.083$$

Table of $A \times B \times C$				
A	Operating Pressure B			
	25 psi		30 psi	
	Line Speed C			
	200	250	200	250
10%	-4	-1	-1	2
12%	1	3	5	11
14%	9	13	16	21

where Total $SS_{ABC} = \frac{\sum_{i=1}^a \sum_{j=1}^b \sum_{k=1}^c Y_{ijk}}{n} - CF$

$$SS_{error} = SS_{total} - SS_A - SS_B - SS_C - SS_{AB} - SS_{AC} - SS_{BC} - SS_{ABC} = 8.5$$

□

	SOV	df	SS	MS	F
main effects	A	$a - 1 = 2$	252.750	126.375	178.412*
	B	$b - 1 = 1$	45.375	45.375	64.059*
	C	$c - 1 = 1$	22.042	22.042	31.118*
1st order interaction	AB	$(a - 1)(b - 1) = 2$	5.250	2.625	3.706**
	AC	$(a - 1)(c - 1) = 2$	0.583	0.292	0.412
	BC	$(b - 1)(c - 1) = 1$	1.042	1.042	1.471
2nd order interaction	ABC	$(a - 1)(b - 1)(c - 1) = 2$	1.083	0.542	0.765
Error	$abc(n - 1) = 12$	8.5	0.708		
Total	$abcn - 1 = 23$	336.625			

Question 1: The quality control department of a fabric finishing plant is studying the effect of several factors on the dyeing 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.

		Temperature (R)			
		300°C		350°C	
		Operator (P)			
		1	2	1	2
Cycle Time C	40	23	30	24	38
		24	28	23	36
	50	36	34	37	34
		35	38	39	38
	60	28	35	26	36
		24	35	29	37

Let $i \rightarrow C \rightarrow c = 3$, $j \rightarrow R \rightarrow r = 2$, $k \rightarrow P \rightarrow p = 2$, and $l \rightarrow n = 2$

Main Effects are: C, R, P .

First order Interactions are: CR, CP , and RP .

Second order Interaction is CRP

□

Lecture 9

2^k and 2^2 Factorial Design

Lesson 19: 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 $2 \times 2 \times 2 \cdots \times 2 = 2^k$ observations and is called a 2^k factorial design.

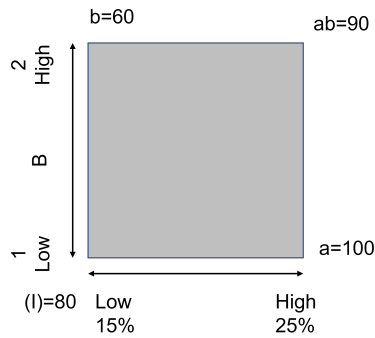
Lesson 20: 2^2 Factorial Design

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 or absence.

Example 1: 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:

Treat. Comb.		Replicate			Total
		1	2	3	
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

Solution:



Sign Table				
Treat. Comb.	Effects			
	<i>M</i>	<i>A</i>	<i>B</i>	<i>AB</i>
(1)	+	-	-	+
<i>a</i>	+	+	-	-
<i>b</i>	+	-	+	-
<i>ab</i>	+	+	+	+

Effects

$$\begin{aligned}
 CF &= \frac{(\text{Contrast } M)^2}{2^k n} - \frac{(330)^2}{12} = 9075 \\
 M &= \frac{1}{2^k n} [\text{Contrast of } M] = \frac{1}{2^2(3)} [(I) + a + b + ab] \\
 &= \frac{1}{12} [80 + 100 + 60 + 90] = 27.5 \\
 A &= \frac{1}{2^{k-1} n} [\text{Contrast } A] = \frac{1}{2^{2-1}(3)} [-(I) + a - b + ab] \\
 &= \frac{1}{2 \times 3} [-80 + 100 - 60 + 90] = 8.33 \\
 SS_A &= \frac{[\text{Contrast } A]^2}{2^k n} = \frac{(50)^2}{2^2(3)} = \frac{(50)^2}{12} = 208.33 \\
 B &= \frac{1}{2^{k-1} n} [-(I) - a + b + ab] = \frac{1}{2(3)} [-80 - 100 + 60 + 90] = -5 \\
 SS_B &= \frac{[\text{Contrast } B]^2}{2^k n} = \frac{(-30)^2}{12} = 75 \\
 AB &= \frac{1}{2^{k-1} n} [(I) - a - b + ab] = \frac{1}{2(3)} [80 - 100 - 60 + 90] = 1.667 \\
 SS_{AB} &= \frac{[\text{Contrast } AB]^2}{2^k n} = \frac{(10)^2}{12} = 8.33 \\
 SS_{Total} &= \sum_{ijk} Y_{ijk}^2 - CF = (28)^2 + (25)^2 + \dots + (29)^2 - 9075 = 323 \\
 SS_{Error} &= SS_{Total} - SS_A - SS_B - SS_{AB} = 323 - 208.33 - 75 - 8.33 = 31.34
 \end{aligned}$$

$H_0 : A$ is non-significant vs $H_1 : A$ is significant

$H_0 : B$ is non-significant vs $H_1 : B$ is significant

$H_0 : AB$ is non-significant vs $H_1 : AB$ is significant

ANOVA Table

SOV	df	SS	Ms	F_{cal}	F_{tab}	Decision
Mean Effects						
<i>A</i>	$a - 1 = 1$	208.33	208.33	$F_A = 53.15$	> 5.32	Significance
<i>B</i>	$b - 1 = 1$	75	75	$F_B = 19.13$	> 5.32	Significance
First Order Interaction						
<i>AB</i>	$(a - 1)(b - 1) = 1$	8.33	8.33	$F_{AB} = 2.13$	< 5.32	Non-Significant
Error	$ab(n - 1) = 8$	31.34	3.92			
Total	$abn - 1 = 11$	323				

Example 2: 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.

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

Solution:

		Replicates						Total
		1	2	3	4	5	6	
Treat. Comb	(I)	21	22	23	28	20	26	140
	m	37	39	38	38	35	36	223
	c	25	26	24	25	29	26	156
	mc	31	34	29	33	30	35	192

Step 1: Hypothesis

H_0 : The main effect M is non-significant vs H_1 : The main effect is significant

H'_0 : The effect of C is non-significant vs H'_1 : The effect of C is significant

H''_0 : The first order interaction effect MC is non-significant vs H''_1 : The first order interaction effect MC is significant

Step 2: Level of significant ($\alpha = 0.05$)

Step 3: Test Statistics $F = \frac{MS_{effect}}{MSE}$

Let n is number of replicates, m is the levels of factor M and c is levels of factor C .

SOV	df	SS	MS	F
M	$m - 1 = 1$	SS_M	$MS_M = \frac{SS_M}{m-1}$	$F_M = \frac{MS_M}{MSE}$
C	$c - 1 = 1$	SS_C	$MS_C = \frac{SS_C}{c-1}$	$F_C = \frac{MS_C}{MSE}$
MC	$mc(c - 1)$	SS_{MC}	$MS_{MC} = \frac{SS_{MC}}{(m-1)(c-1)}$	$F_{MC} = \frac{MS_{MC}}{MSE}$
Error	$mc(n - 1)$	SS_E	$MS_E = \frac{SS_E}{m(n-1)}$	
Total	$mcn - 1$	SS_{Total}		

□

Example 3: Find out the effect of A , B , and AB by using the data below. Also make a column of signs AB .

Solution:

<i>A</i>	<i>B</i>	Treat. comb.	Replicates			
			1	2	3	4
–	–	(<i>I</i>)	18.2	18.9	12.9	14.4
+	–	<i>a</i>	27.2	24.0	22.4	22.5
–	+	<i>b</i>	15.7	14.5	15.1	14.2
+	+	<i>ab</i>	41.0	43.9	36.3	39.9

<i>A</i>	<i>B</i>	<i>AB</i>	Treat. comb.	Replicates				Total
				1	2	3	4	
–	–	+	(<i>I</i>)	18.2	18.9	12.9	14.4	64.4
+	–	–	<i>a</i>	27.2	24.0	22.4	22.5	96.1
–	+	–	<i>b</i>	15.7	14.5	15.1	14.2	59.5
+	+	+	<i>ab</i>	41.0	43.9	36.3	39.9	161.1

Effects

$$A = \frac{1}{2^{k-1}n} [-(1) + a - b + ab] = \frac{1}{2^{2-1}(4)} [-64.4 + 96.1 - 59.5 + 161.1] = 16.6625$$

$$B = \frac{1}{2^{k-1}(n)} [-(I) - a + b + ab] = \frac{1}{2(4)} [-64.4 - 96.1 + 59.5 + 161.1] = 7.5125$$

$$AB = \frac{1}{2^{k-1}(n)} [(I) - a - b + ab] = \frac{1}{2(4)} [64.4 - 96.1 - 59.5 + 161.1] = 8.7375$$

□

Lecture 10

2³ Factorial Design

Lesson 21: 2³ Factorial Design

In this case, three factors each at two levels, are of interest. The design is called 2³ factorial design, and the eight treatment combinations are possible in this design.

		Effects							
		<i>M</i>	<i>A</i>	<i>B</i>	<i>AB</i>	<i>C</i>	<i>AC</i>	<i>BC</i>	<i>ABC</i>
Treatment Combinations	(<i>I</i>)	+	−	−	+	−	+	+	−
	<i>a</i>	+	+	−	−	−	−	+	+
	<i>b</i>	+	−	+	−	−	+	−	+
	<i>ab</i>	+	+	+	+	−	−	−	−
	<i>c</i>	+	−	−	+	+	−	−	+
	<i>ac</i>	+	+	−	−	+	+	−	−
	<i>bc</i>	+	−	+	−	+	−	+	−
	<i>abc</i>	+	+	+	+	+	+	+	+

Lesson 22: 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 in any two columns is zero that is any two columns are orthogonal. $(-1)(-1) + (+1)(-1) + (-1)(+1) + (+1)(+1) = 0$
- ◆ If we multiply any two columns of the sign table another column is achieved, that is, *A* and *B* are multiplied and get *AB*.
- ◆ Column 1 multiplied by another column leaves column unchanged that is column 1 is an identity column.
- ◆ $ABC \times BC = A \Rightarrow AB^2C^2 = AB^2C^2 = A \therefore B^2C^2 = 1$

Example 1: A 2³ 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*₂*F*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		Total
		1	2	
Treatment Combinations	(I)	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

Solution:

Effects

$$A = \frac{1}{2^{k-1}n} [\text{Contrast A}] = \frac{1}{2^{3-1}(2)} [-(I) + a - b + ab - c + ac - bc + abc]$$

$$= \frac{1}{4 \times 2} [-1154 + 1319 - 1234 + 1277 - 2089 + 1617 - 2138 + 1589] = \frac{1}{8} [-813] = -101.625$$

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

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

$$= \frac{1}{8} [1234 + 1277 + 2138 + 1589 - 1154 - 1319 - 2089 - 1617] = \frac{1}{8} [59] = 7.375$$

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

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

$$= \frac{1}{8} [2089 + 1617 + 2138 + 1589 - 1154 - 1319 - 1234 - 1277] = \frac{1}{8} [2449] = 306.125$$

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

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

$$= \frac{1}{8} [1277 - 1319 - 1234 + 1154 + 1589 - 2138 - 1617 + 2089] = \frac{1}{8} (-199) = -24.875$$

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

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

$$= \frac{1}{8} [1154 - 1319 + 1234 - 1277 - 2089 + 1617 - 2138 + 1589] = \frac{1}{8} [-1229] = -153.625$$

$$\begin{aligned}
SS_{AC} &= \frac{(-1229)^2}{16} = 94402.5625 \\
BC &= \frac{1}{8}[(I) + a - b - ab - c - ac + bc + abc] \\
&= \frac{1}{8}[1154 + 1319 - 1234 - 1277 - 2089 - 1617 + 2138 + 1589] = \frac{1}{8}(-17) = -2.125 \\
SS_{BC} &= \frac{(-17)^2}{16} = 18.0625 \\
ABC &= \frac{1}{8}[abc - bc - ac + c - ab + b + a - (I)] \\
&= \frac{1}{8}[1589 - 2138 - 1617 + 2089 - 1277 + 1234 + 1319 - 1154] = \frac{1}{8}(45) = 5.625 \\
SS_{ABC} &= \frac{(45)^2}{16} = 126.5625 \\
SS_{Total} &= \sum_i \sum_j Y_{ij}^2 - \frac{(GT)^2}{N} = 604^2 + 550^2 + \dots + 860^2 + 729^2 - \frac{(12417)^2}{16} = 531420.9375 \\
SS_{Error} &= 531420.9375 - 41310.5625 - 217.5625 - 374850.0625 - 2475.0625 - 94402.5625 \\
&\quad - 18.0625 - 126.5625 = 18020.5
\end{aligned}$$

ANOVA Table

□

SOV	df	SS	MS	F_{cal}	F_{tab}	Decision
Main Effects						
<i>A</i>	$a - 1 = 1$	41310.5625	41310.5625	$F_A = 18.34$	> 5.32	Significant
<i>B</i>	$b - 1 = 1$	217.5625	217.5625	$F_B = 0.10$	< 5.32	Non-Significant
<i>C</i>	$c - 1 = 1$	374850.0625	374850.0625	$F_C = 166.49$	> 5.32	Significant
First order Interaction						
<i>AB</i>	$(a - 1)(b - 1) = 1$	2475.0625	2475.0625	$F_{AB} = 1.10$	< 5.32	Non-Significant
<i>AC</i>	$(a - 1)(c - 1) = 1$	94402.5625	94402.5625	$F_{AC} = 41.9$	> 5.32	Significant
<i>BC</i>	$(b - 1)(c - 1) = 1$	18.0625	18.0625	$F_{BC} = 0.01$	< 5.32	Non-Significant
Second Order Interaction						
<i>ABC</i>	$(a - 1)(b - 1)(c - 1) = 1$	126.5625	126.5625	$F_{ABC} = 0.06$	< 5.32	Non-Significant
Error	$abc(n - 1) = 8$	18020.5	2252.5625			
Total	$abcn - 1 = 15$	531420.9375				

Example 2: 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:

Solution:

		Replicates			Sign Table		
		I	II	III	A	B	C
Treatment Combinations	(I)	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			Total	Sign Table				
		I	II	III		A	B	C	AB	ABC
Treatment Combinations	(I)	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	127	+	+	+	+	+

Effects

$$C = \frac{1}{2^{k-1}n}[c + ac + bc + abc - (I) - a - b - ab]$$
$$= \frac{1}{2^{3-1}(3)}[127 + 113 + 164 + 127 - 78 - 104 - 119 - 148] = \frac{1}{4 \times 3}[82] = 6.8333$$
$$SS_C = \frac{(82)^2}{24} = 280.1667$$
$$AC = \frac{1}{2^{k-1}(n)}[(I) - a + b - ab - c + ac - bc + abc]$$
$$= \frac{1}{2^{3-1}(3)}[78 - 104 + 119 - 148 - 127 + 113 - 164 + 127]$$
$$= \frac{1}{4 \times 3}[-208] = -\frac{208}{12} = -17.3333$$
$$SS_{AC} = \frac{(-208)^2}{24} = 1802.6667$$
$$ABC = \frac{1}{2^{k-1}(n)}[abc - bc - ac + c - ab + b + c - (I)]$$
$$= \frac{1}{12}[127 - 164 - 113 + 127 - 148 + 119 + 104 - 78] = \frac{1}{12}(-26) = -2.1667$$
$$SS_{ABC} = \frac{(26)^2}{24} = 28.1667$$

□

Example 3: The following data were obtained from a 2^3 factorial experiment repeated 3 times. Evaluate the sum of squares for all factorial effects.

Replication	Treatment Combinations							
	(I)	<i>a</i>	<i>b</i>	<i>ab</i>	<i>c</i>	<i>ac</i>	<i>bc</i>	<i>abc</i>
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

Solution:

□