	GOVERNMENT COLLEGE(AUTONOMOUS)RAJAHMUNDRY II B.Sc MAJOR & MINOR (Semester-IV) DESIGN AND ANALYSIS OF EXPERIMENTS

Syllabus:

Unit – 1: Analysis of variance (ANOVA)

Concept, Definition and assumptions.ANOVA one way classification – mathematical model, analysis
– with equal and unequal classification. ANOVA two way classification – mathematical model, analysisand problems.

Unit – 2: Completely Randomised Design (CRD)

Definition, terminology, Principles of design of experiments, CRD – Concept, advantages and disadvantages, applications, Layout, Statistical analysis. Critical Differences when hypothesis is significant.

Unit – 3: Randomised Block Design (RBD)

Concept, advantages and disadvantages, applications, Layout, Statistical analysis and Critical Differences. Efficiency of RBD relative to CRD. RBD with one missing value and its analysis, problems.

Unit – 4: Latin Square Design

Concept, advantages and disadvantages, applications, Layout, Statistical analysis and Critical Differences. Efficiency of LSD over RBD and CRD. Estimation of one missing value in LSD and its analysis, problems.

Unit – 5: Factorial experiments

Main effects and interaction effects of 2^2 and 2^3 factorial experiments and their Statistical analysis. Yates procedure to find factorial effect totals

II B.Sc MAJOR & MINOR (Semester-IV)

DESIGN AND ANALYSIS OF EXPERIMENTS

Time: 2 ½ hrs MODEL PAPER Max Marks: 50

SECTION-A

Answer any FIVE questions. 5 X4 =20M

1. Write about assumptions of ANOVA.
2. Write Short note on ANOVA
3. Define completely randomized design (CRD)
4. Explain about fixed effect & random effect model.
5. Describe about applications for RBD.
6. Write about advantages and disadvantages of LSD.
7. Compare the Efficiencies of RBD over LSD
8. Describe Yates procedure to find factorial effect totals.

SECTION-B

Answer any THREE questions 3X10=30M

9. Explain ANOVA two way classification

(OR)

10. One way Classification Model Problem
Four Varieties of Fertilizers have been applied to Five plots each. the yield given below

Varieties	Plots				
	I	II	III	IV	V
1	1.9	2.2	2.6	1.8	2.1
2	2.5	1.9	2.3	2.6	2.2
3	1.7	1.9	2.2	2.0	2.1
4	2.1	1.8	2.5	2.3	2.4

11. Explain about CRD

(OR)

12. Explain missing plot technique is RBD

13. Explain about LSD

(OR)

14. Explain 2^3 factorial experiments

Analysis of Variance

5.1. INTRODUCTION

To test the significance difference between the two samples means of small samples usually conduced by using t-distribution. But if we want test the significance difference of three or more samples means, an alternative procedure is required for testing the hypothesis that all the samples are drawn from the same mean *i.e.* from the same population. The analysis of variance technique provides answer to this problem. The main purpose of analysis of variance is to test the homogeneity of several means.

The term analysis of variance as introduced by Prof. R.A. Fisher in 1920 in solving the problem of analysis of agronomical data. In fact the analysis of variance is a powerful statistical tool which separate the controlled (assignable) and uncontrolled (chance) variations in the data.

5.2. DEFINITION

According to Prof. R.A. Fisher Analysis of variance termed as ANOVA is the "Separation of variance ascribable to one group of causes from the variance ascribable to the group."

By using ANOVA technique, the total variation in the sample data can be expressed as sum of variations due to some specific independent factors (non-negative components). The technique ANOVA is most powerful in estimating the amount of variation due to each independent factors separately. This technique is also helpful in comparing the estimates due to various factors.

5.3. CAUSES OF VARIATION

The variation in any experiment is inherent in nature. The total variation in any numerical data is due to a number of actors or causes which may be categorised mainly into the following two ways :

1. Assignable causes of variation
2. Chance causes of variation.

The variation due to assignable causes can be detected and measured, hence these are called controlled variation. The variation due to chance causes is beyond the control of human hand and they cannot be traced separately, hence these are called uncontrolled variation.

5.4. ASSUMPTIONS OF ANOVA

The technique of ANOVA is based on F-test, for the validity of the F-test in ANOVA, the following assumptions are required.

1. The observations are independent.
2. The parent population from which the sample observations are taken is normal.
3. Various treatments and environmental effects are additive in nature.

5.9. ANOVA —ONE WAY CLASSIFICATION

Let us consider a random variable Y consists of N observations y_{ij} ($i = 1, 2, ..., k, j = 1, 2, ..., n_i$) is divided into k classes of sizes $n_1, n_2, ..., n_k$ respectively on the basis of some criterion such that $\sum_{i=1}^{k} n_i = N$. The one way classified data can be explained in the following table.

Class	Values of Variables						Totals	Means
1	y_{11}	y_{12}	...	y_{1j}	...	y_{1n_1}	$T_{1.}$	$\bar{y}_{1.}$
2	y_{21}	y_{22}	...	y_{2j}	...	y_{2n_2}	$T_{2.}$	$\bar{y}_{2.}$
⋮	⋮	⋮		⋮		⋮	⋮	⋮
i	y_{i1}	y_{i2}	...	y_{ij}	...	y_{in_i}	$T_{i.}$	$\bar{y}_{i.}$
⋮	⋮	⋮		⋮		⋮	⋮	⋮
k	y_{k1}	y_{k2}	...	y_{kj}	...	y_{kn_k}	$T_{k.}$	$\bar{y}_{k.}$
							G	$\bar{y}_{..}$

G = Grand Total

$\bar{y}_{..}$ = Overall mean.

If $n_1 = n_2 = ... = n_k = n$ (say), then the data is called one way classified data with equal number of observations. Otherwise, the data is known as one way classified data with unequal number of observations.

The total variation in the observations y_{ij} can be divided into the following two components.

1. The variation between the classes (*i.e.* treatments). This is due to assignable causes which can be detected and controlled by human endeavour.
2. The variation within the classes *i.e.*, inherent variation. This is due to chance causes which cannot be control of human hand.

The sources of variation in the data are :

1. Effect of the treatment α_i, $i = 1, 2, ..., k$.
2. Error due to chance named ε_{ij} follows a normal distribution.

Mathematical Model

In one way classification, the linear mathematical model will be

$$\begin{aligned} y_{ij} &= \mu_i + \varepsilon_{ij} \\ &= \mu + (\mu_i - \mu) + \varepsilon_{ij} \\ &= \mu + \alpha_i + \varepsilon_{ij} \end{aligned}$$

Assumptions of the Model

1. All the observations (y_{ij}) are independent.
2. Various effects are additive in nature.
3. ε_{ij} are *i.i.d* N $(0, \sigma_e^2)$

Null Hypothesis

We have to test the equality of the population means *i.e.*, we are testing homogeneity of different treatments.

$\therefore$ Null hypothesis becomes

$H_0 : \mu_1 = \mu_2 = \ldots = \mu_k = \mu$ (say)

i.e., $H_0 : \mu_1 - \mu = \mu_2 - \mu = \ldots = \mu_k - \mu = 0$

i.e., $H_0 : \alpha_1 = \alpha_2 = \ldots = \alpha_k = 0.$

Statistical analysis of the Model

Let us consider

$$\bar{y}_{i.} = \text{mean of the } i\text{th class}$$

$$= \frac{\sum_{j=1}^{n_i} y_{ij}}{n_i}$$

$\bar{y}_{..}$ = Overall mean

$$= \frac{1}{N} \sum_{i=1}^{k} \sum_{j=1}^{n_i} y_{ij} = \frac{1}{N} \sum_{i=1}^{k} n_i \bar{y}_{i.}$$

Degrees of freedom of sum of squares

1. The total sum of squares (TSS) has N – 1 degrees of freedom (d.f.), one d.f. lost since the linear constraint

$$\sum_{i=1}^{k}\sum_{j=1}^{n_i}(y_{ij}-\bar{y}_{..})=0$$

2. The sum of the squares due to treatments (SST) carries $k-1$ d.f.

   Since $\sum_{i=1}^{k} n_i(\bar{y}_{i.}-\bar{y}_{..})=0$

3. The sum of the squares due to error (SSE) carries N – K d.f.

   Since N quantities $\sum_{j=1}^{n_i}(y_{ij}-\bar{y}_{i.})=0$ of k linear constraints.

Mean Sum of Squares (MSS)

Mean sum of squares is sum of the squares divided by its degrees of freedom

1. MSS for treatments is

$$S_T^2 = \frac{SST}{K-1}$$

2. MSS for Error is

$$S_E^2 = \frac{SSE}{N-K}$$

Test a Statistic

By using Cochran's theorem, under null hypothesis, $\frac{SST}{\sigma_e^2}$, $\frac{SSE}{\sigma_e^2}$ are two independent chi-square variates with K – 1 and N – K d.f. respectively. Hence by the definition of F-statistic, under H_0

$$F = \frac{\frac{SST}{\sigma_e^2}\Big/(K-1)}{\frac{SSE}{\sigma_e^2}\Big/(N-k)} \sim F_{(k-1,\, N-k)} = \frac{SST/(k-1)}{SSE/(N-k)}$$

$$F = \frac{S_T^2}{S_E^2} \sim F_{(k-1,\, N-k)}$$

Conclusion :

If calculated value of F is greater than the significant (tabulated) value of F at specified level of significance and at (K – 1, N – K) d.f., we may reject the null hypothesis H_0. Otherwise we may accept H_0.

ANOVA TABLE

The entire statistical analysis of the given data can be easily understand with the help simple presentation of the table is known as analysis of variance (ANOVA) table.

ANOVA Table for One Way Classified Data

Source of Variation	*Sum of Squares*	*d.f.*	*Mean Sum of squares*	*F-ratio (Variance ratio)*
Treatments	SST	$k-1$	$S_T^2 = \frac{SST}{k-1}$	$F = \frac{S_T^2}{S_E^2} \sim F_{(k-1, N-k)}$
Error	SSE	$N-k$	$S_E^2 = \frac{SSE}{N-k}$	
Total	TSS	$N-1$		

5.11. ANOVA—TWO WAY CLASSIFICATION (WITH ONE OBSERVATION PER CELL)

Let us consider a random variable Y consists of N observations y_{ij} ($i = 1, 2, ..., k$; $j = 1, 2, ..., h$) is divided according to two factors *viz.*, treatments and varieties. Let us suppose that N observations are divided into h classes (due to varieties) and each class consists of k observations with the effect of k treatments such that $N = kh$. The values of the variable Y can be explained in the following $k \times h$ two-way table.

Varieties → *Treatments* ↓	*1*	*2*	...	*j*	...	*h*	*Totals*	*Means*
1	y_{11}	y_{12}	...	y_{1j}	...	y_{1h}	$T_{1.}$	$\bar{y}_{1.}$
2	y_{21}	y_{22}	...	y_{2j}	...	y_{2h}	$T_{2.}$	$\bar{y}_{2.}$
⋮	⋮	⋮		⋮		⋮	⋮	⋮
i	y_{i1}	y_{i2}	...	y_{ij}	...	y_{ih}	$T_{i.}$	$\bar{y}_{i.}$
⋮	⋮	⋮		⋮		⋮	⋮	⋮
k	y_{k1}	y_{k2}	...	y_{kj}	...	y_{kh}	$T_{k.}$	$\bar{y}_{k.}$
Total	$T_{.1}$	$T_{.2}$	...	$T_{.j}$	...	$T_{.h}$	G	$\bar{y}_{..}$
Means	$\bar{y}_{.1}$	$\bar{y}_{.2}$	...	$\bar{y}_{.j}$	⋮	$\bar{y}_{.h}$	$\bar{y}_{..}$	

G = Grand total

$\bar{y}_{..}$ = Overall mean.

The source of variation in the data are

1. Effect due to the ith treatment α_i
2. Effect due to the jth variety β_j
3. Error effect due to chance ε_{ij} follows normal distribution.

Mathematical Model

In the two way classification, the linear mathematical model will be

$$y_{ij} = \mu + \alpha_i + \beta_j + \varepsilon_{ij}, \qquad i = 1, 2, \ldots, k$$
$$j = 1, 2, \ldots, h$$

Assumptions of the model

1. All the observations (y_{ij}) are independent.
2. Various effects are additive in nature.
3. ε_{ij} are *i.i.d.* N $(0, \sigma_e^2)$

Null Hypothesis

We have to set up null hypothesis for the treatments and varieties. Separately as their effects are homogeneous.

(1) For treatments :

H_T: All the treatment effects do not differ significantly.

i.e. $H: \alpha_1 = \alpha_2 = \ldots = \alpha_k = 0$

(2) For varieties :

H_V: All the varieties effects do not differ significantly.

i.e. $H_V: \beta_1 = \beta_2 = \ldots = \beta_h = 0$

Statistical Analysis of the Model

Let us consider

$$\bar{y}_{i.} = \text{Mean yield of } i\text{th treatment}$$

$$= \frac{1}{h} \sum_{j=1}^{h} y_{ij}$$

$$\bar{y}_{.j} = \text{Mean yield of } j\text{th variety}$$

$$= \frac{1}{k} \sum_{i=1}^{k} y_{ij}$$

$$\bar{y}_{..} = \text{Overall mean}$$

$$= \frac{1}{N} \sum_{i=1}^{k} \sum_{j=1}^{h} y_{ij} = \frac{1}{hk} \sum_{i=1}^{k} \sum_{j=1}^{h} y_{ij}$$

Degrees of freedom of sum of squares

TSS carries N – 1 or $hk - 1$ d.f. since subject to one linear constraint $\sum_{i=1}^{k} \sum_{j=1}^{h} (y_{ij} - \bar{y}_{..}) = 0$

SST carries $k - 1$ d.f. since subject to one linear constraint $\sum_{i=1}^{k} (\bar{y}_{i.} - \bar{y}_{..}) = 0$

SSV carries $h - 1$ d.f. since $\sum_{j=1}^{h} (\bar{y}_{.j} - \bar{y}_{..}) = 0$

$$\begin{aligned} \text{d.f. of SSE} &= \text{d.f. of TSS} - \text{d.f. of SST} - \text{d.f. of SSE} \\ &= hk - 1 - (k-1) - (h-1) \\ &= (h-1)(k-1) \end{aligned}$$

Mean of Sum of Squares

1. MSS for treatment is

$$S_T^2 = \frac{SST}{k-1}$$

2. MSS for varieties is

$$S_V^2 = \frac{SSV}{h-1}$$

3. MSS for error is

$$S_E^2 = \frac{SSE}{(h-1)(k-1)}$$

Test a Statistic (Variance Ratio)

By using Cochran's theorem, under null hypothesis (H_T and H_V), $\frac{SST}{\sigma_e^2}$, $\frac{SSV}{\sigma_e^2}$ and $\frac{SSE}{\sigma_e^2}$ are independent chi-square variates with $k-1$, $h-1$ and $(k-1)(h-1)$ d.f. respectively. By the definition of F-statistic, under H_T and H_V,

(1) For Treatment

$$F_T = \frac{\frac{SST}{\sigma_e^2} \Big/ (k-1)}{\frac{SSE}{\sigma_e^2} \Big/ (h-1)(k-1)} \sim F_{(k-1,\,(h-1)(k-1))}$$

$$= \frac{SST/(k-1)}{SSE/(h-1)(k-1)} = \frac{S_T^2}{S_E^2}$$

(2) For Varieties

$$F_V = \frac{\frac{SSV}{\sigma_e^2} \Big/ (h-1)}{\frac{SSE}{\sigma_e^2} \Big/ (h-1)(k-1)} \sim F_{(h-1,\,(h-1)(k-1))}$$

$$= \frac{SSV/(h-1)}{SSE/(h-1)(k-1)} = \frac{S_V^2}{S_E^2}$$

Conclusion

If calculated value of F_T or F_V is greater than the significant value of F at specified level of significance at $(k-1, (h-1)(k-1))$ or $((h-1), (h-1)(k-1))$ d.f. then we may reject null hypothesis. Otherwise we may accept null hypothesis (H_T or H_V).

ANOVA Table for Two-Way Classified Data

Source of Variation	*Sum of Squares*	*d.f.*	*Mean Sum of squares*	*F-ratio (Variance ratio)*
Treatments	SST	$k-1$	$S_T^2 = \dfrac{SST}{k-1}$	$F_T = \dfrac{S_T^2}{S_E^2} \sim F_{(k-1,\,(h-1)(k-1))}$
Varieties	SSV	$h-1$	$S_V^2 = \dfrac{SSV}{h-1}$	$F_V = \dfrac{S_V^2}{S_E^2} \sim F_{(h-1,\,(h-1)(k-1)}$
Error	SSE	$(h-1)(k-1)$	$S_E^2 = \dfrac{SSE}{(h-1)(k-1)}$	
Total	TSS	$hk-1$		

5.10. CRITICAL DIFFERENCE (C.D.)

If there is a significant difference in between the treatments (*i.e.*, if H_T is rejected), then we would be interested to find out which pair of treatments differ significantly. For this, instead of calculating student's *t* for different pairs of treatment means, we calculate the least significant difference at the given level of significance. This least difference is known as the critical difference (C.D.).

C.D. at α level of significance is given by

$$\text{C.D.} = (\text{S.E. of difference between two treatment means}) \times (t_{\alpha\%} \text{ for error d.f.})$$

C.D. $(\bar{y}_{i.} - \bar{y}_{.j})$ = C.D. = (S.E. of difference of means) $\times\ t_{\alpha\%}$ for error d.f.

$$= \sigma_e \sqrt{\frac{1}{n_i} + \frac{1}{n_j}} \times t_{\alpha\%} \text{ for error d.f.}$$

S_E^2 provides an unbiased estimator for σ_e^2 and if each treatment replicated n times *i.e.*,

$$n_i = n,\ i = 1, 2, \ldots, k, \text{ then}$$

$$\text{C.D.} = S_E \sqrt{\frac{2}{n}} \times t_{\alpha\%} \text{ for error d.f.}$$