DESIGN AND ANALYSIS OF EXPERIMENTS

Experiment: An experiment is a planned inquiry to obtain new facts or to confirm or deny the results of previous experiments, where such inquiry will aid in an administrative decision such as recommending a variety, a procedure, etc .

Types of Experiments:

1. Preliminary experiment
2. Critical experiment
3. Demonstrational experiment.

In a *preliminary* experiment, the investigator tries out large number of treatments in order to obtain leads for future work. Here most treatments appear once.

In a *critical* experiment the investigator compares responses to different treatments using sufficient observations of the responses.

Demonstrational experiments are performed when an extension worker compare a new treatment or treatments with a standard.

DESIGN OF EXPERIMENTS:

Planning an experiment to obtain appropriate data with respect to any problem under investigation is known as *design of experiments*.

It is a complete sequence of steps taken well in time to insure that appropriate data will be obtained so that it permits an objective analysis of data leading to valid inferences with respect to the stated problem.

Following are the steps used in designing an experiment:

1. Formulation of the objectives of the experiment.
2. Structuring of the dependent and independent variables.
3. The choice of levels of each variable in the experiment.
4. The method of manipulation of the variable on the experimental material.
5. The method of recording and tabulation of data.
6. Mode of analysis of data.
7. The method of drawing sound and valid inferences.

Some basic Concepts:

1. Experimental unit: It is the unit of material to which the application of a treatment is made. The experimental unit may be a plot , a piece of land , a plant , a part of the plant such as leaf, stem or subjects such as human beings, animals, animal parts, etc.

2. Treatment: The treatment is a procedure whose effect is to be measured and compared with other treatments. The treatment may be standard rations, different levels of fertilizers, a spraying schedule, temperature-humidity combinations, etc.

3. Variable: It is a measured characteristic of a unit.

4. Response Variable: It is a variable whose changes we wish to study. It is also called an outcome variable.

5. Explanatory Variable: It is a variable that explains or causes changes in the response variables. An explanatory variable in an experiment is also called a factor.

Example: A fabric researcher is studying the durability of a fabric under repeated washings. Because the durability may depend on the water temperature and the type of cleansing agent used, the researcher decides to investigate the effect of these two factors on durability. Factor A is water temperature and has three levels: hot (145^0 F), warm (100^0F), and cold (50^0F). Factor B is the cleansing agent and also has three levels: regular Tide, low-phosphate Tide, and Ivory liquid . A treatment consists of washing a piece of the fabric (a unit) 50 times in a automatic washer with a specific combination of water temperature and cleansing agent. The response variable is strength after 50 washes, measured by a fabric testing machine that forces a steel ball through the fabric and records the fabric's resistance to breaking.

In this example there are nine treatments. By using them all, the researcher obtains a wealth of information on how temperature alone, washing agent alone, and the two in combination, affect the durability of the fabric .This combination effect is called an *interaction* between cleansing agent and water temperature.

6. Confounded effect: The effects of two variables (explanatory variables or extraneous variables) on a response variable are said to be confounded when they cannot be distinguished from one another.

7. Experimental Error: A characteristic of all experimental material is the variation. Experimental error is a measure of the variation which exists among observations on experimental units treated alike. Variation comes from two main sources:

ANALYSIS OF VARIANCE TECHNIQUE

Meaning: Analysis of variance is a technique of partitioning the total variability present in a set of observations into different components of variabilities produced by different sources – some *known* and others totally *unknown*.

Example: Consider a experiment involving three drugs say A , B and C for comparison. Each of these 3 drugs are applied to 5 subjects. Let the responses to these are as given below:

Reaction time (in hrs.)

	Drug			
Subjects	A	B	C	
1	7	10	14	
2	5	12	16	
3	6	9	13	
4	3	8	12	
5	4	6	10	
Total:	25	45	65	GT =135
Mean:	5	9	13	GM = 9

A quick look at the data indicates that the variations in the reaction times are present both among different drugs and within each drug. This shows that the variations in the reaction times among different subjects given with different drugs are the results of influences from many sources which are large in number and complex in nature. Of these sources, one is a known source namely, "drug factor" as this has been introduced into the experiment as a factor of main interest and the other sources are all unknown and although their influences are unwanted , they still exist. Thus the technique of Analysis of variance tries to isolate the amount of variation attributable to the known and the unknown causes.

Computational procedure:

Step (1): Obtain the treatment totals, grand total, treatment means and the grand mean.

$$T_1 = 25 , T_2 = 45 , T_3 = 65 , GT = 135$$

$$\overline{X}_1 = 5 , \overline{X}_2 = 9, \overline{X}_3 = 13 , GM = 9$$

Step (2): Obtain the correction factor (C.F) as

$$C.F = \frac{(GT)^2}{n} = \frac{(135)^2}{15} = 1215$$

where n is the total no. of observations

Step(3): Find the total sum of squares (SSQ_T) as

$$SSQ_T = (y_{11}^2 + y_{12}^2 + \ldots + y_{ij}^2 - C.F$$

$$= (7^2 + 5^2 + \ldots + 10^2) - 1215 = 210 .$$

Step (4): Compute the sum of squares due to treatment (Drugs) as

$$SSQ_t = \frac{T_1^2}{r} + \frac{T_2^2}{r} + \frac{T_3^2}{r} - C.F = \frac{(25)^2}{5} + \frac{(45)^2}{5} + \frac{(65)^2}{5} - 1215 = 160.$$

Step(5) : Compute the error sum of squares (ESS) as

$$ESS = TSS - SSQ_t = 210 - 160 = 50$$

ANOVA TABLE

Source of variation	Degrees of freedom(df)	Sum of squares (SS)	Mean sum of squares (MSS)	F- Ratio	F- Table
Between drugs	2	160	80	19.184**	6.93
Within drugs (error)	12	50	4.17		
Total	14	210			

Since the observed F – value 19. 184 for 2 and 12 df exceeds the table value 6.93 at 1% level of significance, we may conclude that the mean reaction times of different drugs are statistically significant at 1 % level of significance.

NOTE: As there is only one condition involved in the above example (ie. Comparing different drugs), the analysis is termed as One-way analysis of variance.

Some assumptions made in the analysis of variance are:

1. The experimental errors are uncorrelated.
2. Various components of variations are additive.
3. Though the treatment means differ significantly, the experimental errors must have common variance.
4. Sample observations are all drawn independently from populations having normal distributions.

ONE-WAY ANALYSIS OF VARIANCE TECHNIQUE

Test Procedure:

1. Assume that the observations are drawn from a normally distributed population with equal variances.
2. Null hypothesis (H_0): There is no significant difference among several means.

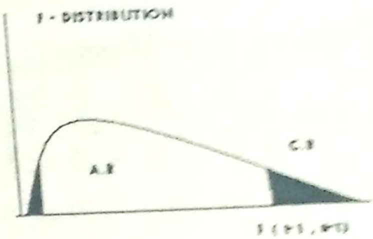
   That is, H_0 : $\mu_1 = \mu_2 = \mu_3$

   Alternative hypothesis(H_1): There is significant difference among several means.

   That is, H_1: Atleast one is different.
3. Level of significance (α): 5% or 1% level
4. Test statistic: The appropriate test statistic is Variance ratio given by

$$F = \frac{Mean\ square\ between\ treatment}{Mean\ square\ within\ treatment}$$

$$which\ \ follows\ \ F-distributon\ \ with\ \ (t-1\,,n-t)\,df$$



5. Critical region: The critical value of F can be referred from the F-distribution tables at (t-1 , n-t) df. and at given level of significance.
6. Decision rule: If the computed value of F lies in the critical region then the null hypothesis H_0 is rejected. On the other hand if it lies in the acceptance region then the null hypothesis H_0 is accepted.
7. Conclusion: Draw appropriate conclusion based on the decision .

Comparison of treatment means in ANOVA:

Whenever, the treatment means are found to be significantly different, it is necessary to find out which pair of means are significantly different and which ones are not. In order to verify this, we use *Least Significant Difference Test* (lsd_t) based on Student's "t" statistic given by

$lsd_t = t_\alpha \cdot SE(\bar{X})$

where t_α *is the table value of* t *at error df.*

and $SE(\bar{X}) = \sqrt{\frac{2\ EMSS}{r}}$ *is the standard error of difference between two means , EMSS is the Error means square in ANOVA and r is the no.of observations per treatment*

For the above example,

1. Standard error of mean SE_M is

$$SE_M = \sqrt{\frac{EMSS}{n}} = \sqrt{\frac{4.17}{15}} = 0.5272$$

2. Standard error of difference between two means Se_d is

$$SE_d = \sqrt{\frac{2\ EMSS}{r}} = \sqrt{\frac{2\ X\ 4.17}{5}} = 1.2915$$

$$Lsd_t = t_{\alpha\ (\text{error df})}\ SE_d$$

$$Lsd_t = 3.055\ X\ 1.2915 = 3.9455$$

NOTE: The least significant difference (Lsd_t) is also known as Critical difference (CD).

A Comparison of Means:

1. $|\bar{y}_1 - \bar{y}_2| = |5-9| = 4 > 3.9455$, *Significant*
2. $|\bar{y}_1 - \bar{y}_3| = |5-13| = 8 > 3.9455$, *Significant*
3. $|\bar{y}_2 - \bar{y}_3| = |9-13| = 4 > 3.9455$, *Significant*

TWO-WAY ANALYSIS OF VARIANCE TECHNIQUE

(With single observation per cell)

In this situation we consider every observation to be from two known sources of variation one orthogonal to the other. The two sources are considered as two conditions leading to two way classification of data as shown below:

Condition 1	Condition 2					Total
	1	2	3	....	m	
1	y_{11}	y_{12}	y_{13}	....	y_{1m}	R_1
2	y_{21}	y_{22}	y_{23}	....	y_{2m}	R_2
.	...	...	...	....	...	...
.	...	...	...	....	...	...
n	y_{n1}	y_{n2}	...	....	y_{mn}	R_n
Total	C_1	C_2			C_m	G.T

For instance, consider two factors namely, variety and season on which the yield of a wheat crop is recorded .Now each observation is subjected to the influence of three sources namely, varietal effect, seasonal effect and effect due to unknown source. Here, the analysis of variance technique helps in partitioning the total variability into each of the above components.

Example: Let us assume 4 varieties of wheat tried in three seasons. The yield in kg/plot are as shown below:

Seasons	Varieties					
	V_1	V_2	V_3	V_4	Total	Mean
S_1	8	4	11	5	28	7
S_2	10	11	13	14	48	12
S_3	6	6	12	8	32	8
Total	24	21	36	27	GT= 108	
Mean	8	7	12	9		GM=9

Computational Method:

1. Correction factor (CF) = $\frac{(GT)^2}{N} = \frac{(108)^2}{12} = 972$

2. S. S due to Varieties (SSQv) =

$$\left[\frac{V_1^2+V_2^2+V_3^2+V_4^2}{s}\right]-CF=\left[\frac{(24)^2+(21)^2+(36)^2+(27)^2}{3}\right]-972 = 42$$

3. S. S due to Seasons (SSQs) =

$$\left[\frac{S_1^2+S_2^2+S_3^2}{v}\right]-CF=\left[\frac{(28)^2+(48)^2+(32)^2}{4}\right]-972 = 56$$

4. Total S. S (TSS) = $\left[y_{11}^2+y_{12}^2+...+y_{ij}^2\right]-CF = \left[8^2+4^2+...+8^2\right]-972=120$

5. Error S. S (ESS) = TSS - SSQv – SSQs = 120 – 56 – 42 = 22

ANOVA Table

Source	df	SS	MSQ	F Ratio	F Table
Variety	3	42	14	3.82^{NS}	4.76
Season	2	56	28	7.63*	5.14
Error	6	22	3.67		
Total	11	120			

From the above analysis we conclude that there is no significant difference between varietal means where as there is significant difference among seasonal means.

Comparison of seasonal means:

As seasons are significantly different, we compare the seasonal means by lsd_t test as

$$SE_M = \sqrt{\frac{EMSS}{vs}} = \sqrt{\frac{3.67}{(4)(3)}} = 0.5530$$

$$SE_{d(Varieties)} = \sqrt{\frac{2EMSS}{s}} = \sqrt{\frac{2X\,3.67}{3}} = 1.5641$$

$$SE_{d(seasons)} = \sqrt{\frac{2EMSS}{v}} = \sqrt{\frac{2X3.67}{4}} = 1.3546$$

$$lsd_{t(seasons)} = t_\alpha \, . \, SE_{d(seasons)} = 2.074X\,1.3546 = 2.8094$$

Multiple comparisons:

1. S_1 *vs* $S_2 = |7-12| = 5 > 2.8094$ *significant*
2. S_1 *vs* $S_3 = |7-8| = 1 < 2.8094$ *not significant*
3. S_2 *vs* $S_3 = |12-8| = 4 > 2.8094$ *significant*

TWO-WAY ANALYSIS OF VARIANCE

(With more than one observation per cell)

Suppose in a fertilizer trial involving two fertilizers say, Nitrogen (N) at three levels and Phosphate (P) at two levels are tried. For each combination of two fertilizers let two observations are recorded from two similar plots. The yield of a crop in kg/plot be as shown below:

		N levels							
		30 lbs		60lbs		90lbs		Total	Mean
P levels	30 lbs	6	8	10	14	19	21	78	13
	60lbs	9	13	14	18	28	32	114	19
	Total	36		56		100		192	16
	Mean	9		14		25			

$$Step(1):\ CF = \frac{(GT)^2}{N} = \frac{(192)^2}{12} = 3072$$

$$Step(2):\ Total\ SS\ (TSS) = [\,6^2 + 8^2 + \ldots + 32^2] - 3072 = 704$$

$$Step(3):\ SS\ between\ Nitrogen\ (SSQ_N) = \left[\frac{(36)^2 + (56)^2 + (100)^2}{4}\right] - 3072 = 536$$

$$Step(4):\ SS\ between\ Phosphate(SSQ_P) = \left[\frac{(78)^2 + (114)^2}{6}\right] - 3072 = 108$$

$$Step(5):\ SS\ between\ Cell\ totals\ (SSQ_C) = \left[\frac{(14)^2 + (24)^2 + \ldots + (60)^2}{2}\right] - 3072 = 668$$

$$Step(6):\ Interaction\ SS\ (SSQ_{NP}) = SSQ_C - SSQ_N - SSQ_P = 668 - 536 - 108 = 24$$

$$Step(7):\ Error\ SS\ (ESS) = TSS - SSQ_N - SSQ_P - SSQ_{NP} = 704 - 536 - 108 - 24 = 36$$

ANOVA Table

Source	df	SS	MSQ	F Ratio	F Table
Nitrogen (N)	2	536	268	44.67**	10.92
Phosphate (P)	1	108	108	18.00**	13.74
Interaction (NXP)	2	24	12	2.00^{NS}	5.14
Error	6	36	6		
Total	11	704			

The ANOVA table indicates that there is significant difference between Nitrogen and Phosphate levels where as no significant difference was observed in the interaction. The linear additive model indicates four different effects apart from the general mean. That is,

Any Observation = (General mean) + (Row effect) + (Column effect) + (Interaction effect) + (Error effect)

Symbolically, $Y_{ijk} = \mu + R_i + C_j + I_{ij} + \varepsilon_{ijk}$

The standard errors are given by

$$SE_M = \sqrt{\frac{EMSS}{npr}} = \sqrt{\frac{6}{12}} = 0.7071$$

$$SE_{d(Nitrogen)} = \sqrt{\frac{2EMSS}{pr}} = \sqrt{\frac{2X6}{4}} = 1.7320$$

$$SE_{d(Phosphate)} = \sqrt{\frac{2EMSS}{nr}} = \sqrt{\frac{2X6}{6}} = 1.4142$$

$$SE_{d(NXP)} = \sqrt{\frac{2EMSS}{r}} = \sqrt{\frac{2X6}{2}} = 2.4494$$

$$lsd_{t(Nitrogen)} = t_\alpha \,.\, SE_{d(Nitrogen)} = 1.943 X 1.7320 = 3.3652$$

$$lsd_{t(Phosphate)} = t_\alpha \,.\, SE_{d(Phosphate)} = 1.943 X 1.4142 = 2.7477$$