ISSN:3107-4200
SABUJEEMA Agri Newsletter
An International Newsletter for Agriculture and Allied Sciences
Article ID: SAN/02-03/18 Popular Article

Basic Statistical Concepts of Design of Experiments

Dr. Anju A. Bhagat
Assistant Professor of Statistics,
College of Agriculture, Pune -411 005 (MPKV, Rahuri)

Abstract

For comparing of mean performance of two samples/populations we use t-test $i.e.\ \mu_1 = \mu_2$. If we want to compare the mean performance more than two samples/populations then we cannot apply t-test $i.e.\ \mu_1 = \mu_2 = \cdots = \mu_n$

In such situation we can apply Analysis of Variance (ANOVA)

ANOVA is the procedure of splitting the overall variation into different components. Each component is attributed to an identifiable cause or source of variation. The structure of these component parts is determined by the design of experiment.

Assumptions of ANOVA

- Treatment effects of different factors(treatments and environment) are additive.
- The experimental errors are independent.
- The experimental errors are distributed normally with mean 0 and common variance $\boldsymbol{\sigma^2}$.

Design of Experiments

The choice of treatments, the method of assigning treatments to experimental units and arrangement of experimental units in various patterns to suit the requirement of particular problems are combined known as design of experiments.

It was given by **R.A. Fisher**. He developed the planning of agricultural field experiments.

Basic terminologies

- **Experiment:** An experiment is a device of getting an answer to the problem.
- **Experimental unit:** An experimental unit is a unit of material on which application of treatment is done.

Example: A plot, an animal, an insect etc.

- **Experimental material: This** is the name given to the material on which experiment is performed.

Example: A field, soil, seeds etc.

- **Response/dependent variable:** The variable whose change we wish to study is known as response variable.

Example: grain yield.

- **Factor/independent variable:** The variable whose effect on the response we wish to study is known as response variable.

Example: fertilizer, spacing, irrigation schedule, pesticide etc.

- **Degree of freedom (d.f.) :** It is the number of independent comparisons which cane be made between the members of the samples.

i.e. **Total number of observations – Total no. of restrictions**

- **Level of Significance (l.s.): The** accuracy with which the decision of acceptance or rejection of null hypothesis is termed as level of significance.
- **Standard Deviation (S.D.):** The root mean square deviation of each observation from arithmetic mean is called standard deviation.
- **Standard Error (S.E.):** The standard deviation of a sample mean is called as standard error.

$$S.E. = \frac{\sigma}{\sqrt{n}}$$

- **Critical Difference (C.D.) :** It is the minimum difference between two treatment means which would be recognized as statistically significant.

$$C.D. = S.E.(d) X\, t_{error\ d.f.}$$

Where,

$$S.E.(d) = \sqrt{\frac{2EMS}{r}}$$

r= no. of replications

- **Treatment:** In any experimentation certain conditions are compared then conditions are called treatment.

Or

The object of comparison, which an experimenter has to try in the field or laboratory for assessing their values are known as treatments.

Example: variety of crop, different fertilizers, different pesticides, method of seed treats. Cultivation practices etc.

- **Factor:** A set of related treatment.

Example: different doses of nitrogen, dates of sowing,

- **Extraneous factors:** In Agril. experiments, heterogeneity in soil, climatic factors, genetic differences etc. such factors termed as extraneous factors.
- **Experimental Error:** The variation in responses caused by extraneous /uncontrolled factors is termed as experimental error.

A major problem in experiment is that the responses of the experimental units are influenced not only by treatment but also by uncontrolled factors is termed as experimental error.

If the error is small we obtain a better estimate of two differences and if the error is large we obtain a poor estimate of a difference. For proper interpretation of experiment we should have correct estimate of experiment. The amount of information per observation is measured as the reciprocal of the error variable. When error variance is more, the amount of information will be less and vice versa. When information from experiment is more, the precision of experiment is more.

Example: heterogeneity of soil, climatic factors, genetic differences etc.

- **Sum of Squares (S.S.) :** It is mean sum of squares of the deviations from their mean.

If $X_1, X_2, \dots . X_n$ are n observations with their mean X then,

$$\text{Sum of Squares} = \sum (X_i - \overline{X})^2 = \sum X_i^2 - \frac{(\sum X_i)^2}{n}$$

$$\text{Sum of Squares} = \sum X_i^2 - \text{Correction Factor}$$

<u>Basic Principles and it's role in Design of Experiment</u>

The purpose of designing an experiment is to increase precision of the experiment. In order to increase the precision, we try to reduce the experimental error. For reducing the experimental error, we adopt certain techniques. These techniques form the basic principles of experimental designs. There are three basic principles of design of experiment: **replication, randomization and local control.** These are useful for increasing the precision of an experiment and also for drawing valid inferences from experiments.

- **Replication: The repetition of same treatments under investigation in different experimental unit is known as replication.** Replication is necessary in order to get estimate of experimental error variation caused due to uncontrolled factors because the variation in soil fertility is of great that all the treatments do not get the equal chance of experiencing every type environment in the field. Replication increases the precision. If we repeat a single treatment r times, the mean of the treatment will be subjected to standard error.

$$\boldsymbol{S.E.} = \frac{\boldsymbol{\sigma}}{\sqrt{\boldsymbol{r}}}$$

Where,

σ is the standard deviation of indiviual plots and is estimated from experiment.

If r increases the value of standard error (S.E.) decreases. Thus replication of treatment helps in reducing the error in the experiment in addition to provide an estimate of error.

Functions of replication

- To provide estimate of an experimental error.
- To reduce the experimental error.

We can get the required no. of replications from the formula:

$$t = \frac{d}{\sqrt{\frac{2S^2}{r}}}$$

Where,

$$d = |\overline{Y}_1 - \overline{Y}_2| \text{ that is absolute difference between means of treatments.}$$

$$t = \text{table value of } t \text{ for given level of significance and d.f.}$$

$$S^2 = \text{error variance taken from similar earlier expt. conducted.}$$

$$r = \text{no. of replications}$$

$$\therefore\ t^2 = \frac{rd^2}{2S^2}$$

$$\therefore r = \frac{2t^2S^2}{d^2}$$

In practice, it may not be possible to know the value of S^2 hence some approximate rules may be followed to fix the no. of replication. *A satisfactory precision can be obtained by taking no. of replication in such a way that error degree of freedom is not less than 12 because the value of F table do not decreases rapidly after 12 d.f.*

Advantages of replications

- Replications are essential to obtain valid estimate of the experimental error variance.
- The larger no. of replication reduces the standard error of the treatment means. But we cannot take more than certain no. of replication due to management problem and limitation of resources.
- Replications are essential in finding the variation between the treatment means.

- **Randomization: The allocation of the treatments to the different plots by random process is known as randomization of treatment.** When all the treatments have an equal chance of being allocated to different experimental units it is known as randomization. The comparison of treatments will be valid only when they are studied under equal environmental conditions. This is possible by randomization of treatments and not by deliberate allotment. To get valid results in an experiment, sufficient no. of replication should accompany randomization. The procedure of allocating treatments to experimental unit may vary according to the design of experiment. Randomization also helps in removing personal biases.

Advantages of Randomization

- To an unbiased estimate of experimental error.
- To get valid conclusions.

- **Local control:** The principle of making greater homogeneity in experimental units for reducing the experimental error is known as local control. It means that the local control is device for minimizing the experimental error. **Grouping of homogenous experimental units is known as local control.** This is done by dividing the experimental material into homogeneous groups of experimental units. The experimental field is divided into the homogeneous blocks, within block homogeneity and between block heterogeneity.

Example: soil fertility of field affect the growth of plant and yield.

Before conducting the field experiment we prepare the fertility contour map on the basis of fertility gradient. These contour map help to form the block.

Advantages of Local Control

- Reduces experimental error.
- It increases efficiency of design.
- It makes any test of significance more sensitive and powerful.
- A reduction in experimental error consequently helps the investigator to detect the small real difference between treatment means.

Precision of experiment: The reciprocal of variance of a mean is termed as precision.

$$P = \frac{1}{V(\bar{X})} = \frac{1}{\frac{\sigma^2}{r}} = \frac{r}{\sigma^2}$$

OR

Amount of information of a design.

Where, σ^2 error variance per unit

Efficiency of design: Consider the design D_1 and D_2 with error variances per unit σ_1^2 and σ_2^2 and replication r_1 and r_2 respectively.

$$variance\ between\ two\ treatment\ means\ for\ D_1 = \frac{2\sigma_1^2}{r_1}$$

$$variance\ between\ two\ treatment\ means\ for\ D_2 = \frac{2\sigma_2^2}{r_2}$$

Then ratio (efficiency) = E = $\frac{\left(\frac{2\sigma_2^2}{r_2}\right)}{\left(\frac{2\sigma_1^2}{r_1}\right)}$ *is called efficency of* D_1 *w.r.t.* D_2

$$\frac{r_1}{2\sigma_1^2} * \frac{2\sigma_2^2}{r_2} = \frac{r_1}{\sigma_1^2} * \frac{\sigma_2^2}{r_2} = \frac{\boldsymbol{Precision\ of\ D_1}}{\boldsymbol{Precision\ of\ D_2}}$$

is the efficiency of design w.r.t. D_2

In other words, Efficiency of design D_1 w.r.t. design D_2 is the ratio of precision of D_1 to precision of D_2.

When E = 1, both design D_1 and D_2 are said to be equally efficient.

If E > 1, Design D_1 is more efficient than design D_2.

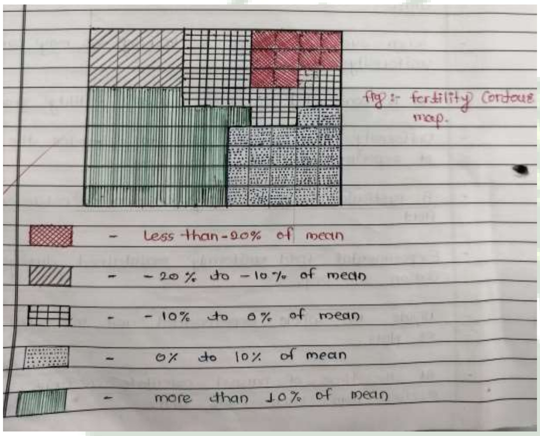
If E < 1, Design D_1 is less efficient than design D_2.

An efficient design has greater ability of detecting the difference of treatment effect. The efficiency is increased by decreasing error variation per unit and by increasing no. of replication.

ISSN:3107-4200
SABUJEEMA
Agri Newsletter
An International Newsletter for Agriculture and Allied Sciences

❖ Uniformity Trials

In order to achieve local control we must have knowledge of the experimental material. We may effect suitable grouping of the experimental units. When such knowledge is lacking, we may carry out uniformity trial. It is constructed to know the soil fertility gradient. Uniformity trial is also used to determine the size of experimental unit. A particular single variety is sown on entire experimental field. Experimental field uniformly maintained during the season of crop. Divide the whole experimental unit into basic units or plots. At the time of harvest calculate average yield received from each basic unit. Find out the mean yield of entire experimental unit. Find out the percentage mean to overall yield of basic unit. Percentage are taken in such a way that we will get at least 5 to 8 groups. The similar units are joined by lines to produce contour map such map is known as fertility contour map. Using this map homogeneous expt. Units can be grouped as blocks/replications.



Soil fertility contour map

❖ Transformation of data

▪ Normality

We know that many statistical tests are valid only when data follows normal distribution .In some experimental situations the distribution of errors may not follows exactly normal as assumed under ANOVA. It may be skewed or follows Poisson distribution or it may follows binomial distribution. If we have count data and occurrence of probability is very less then data follows Poisson distribution.

Example: No. of insects/plant, no. of weed/plot, no. of infected plant/plot, no. of insect caught in a trap etc. such type of data are not follows the Normal distribution.

When the individual observation is in proportions or in percentage then data follows Binomial distribution.

Example: germination percentage of seed material, percentage mortality of insects, percent disease incidence (PDI) etc.

In order to normalize non-normal distribution into normal form we use three transformation of data as:

1. **Square root transformation**
2. **Angular/Arcsine transformation**
3. **Logarithmic transformation**

- **Square root transformation: To normalize poisson distribution**

If original observation Y is converted to a new value by taking its square root, it is known as square root transformation.

It is used in case of count data.

When some of observed counts are numerically small, say less than 10, the more appropriate transformation is $\sqrt{Y + 0.5}$ the transformation of type $\sqrt{Y} + \sqrt{Y + 1}$ is also used.

The square root transformation is used when the observations follows poisson distribution.

Ex: no. of insects/plant, no. of weed/plot, no. of infected plant/plot, no. of insect caught in a trap etc.

- **Angular/Arcsine transformation: To normalize Binomial distribution**

In case of proportions derived from count data, the observed proportion p can be changed to a new form $\sin^{-1}\sqrt{p}$. This type of transformation is known as angular or arcsine transformation. It is also known as inverse sine transformation.

Angular transformation is not applicable to percentage data which are not derived from count data.

Example: percentage of marks, percentage of profit, percentage of protein in rice, infection index etc. can not be subjected to angular transformation.

The angular transformation is not good when $p = \frac{0}{n} = 0$ or $p = \frac{n}{n} = 1$

The transformation is improved by replacing $\frac{0}{n}$ with $\frac{1}{4n}$ and $\frac{n}{n}$ with $1-\frac{1}{4n}$

Where, n is the total number of units under observation.

When all observed proportions lies between 30% to 70%, angular transformation need not be used.

Angular transformation is used to normalize the binomial distribution, when observed proportions are in the range of 0 to 30% or 70 to 100%.

- **Logarithmic transformation**

When original Y is converted to logY, then conversion is known as log transformation. Log to any base can be used. Log to the base 10 is generally easiest. If observed value is 0 then a constant value preferably 1 is added to avoid negative logarithms. When such constant is added, it is added to all the observations. It is used in normalizing the positively skewed distribution. When observed Y consist of index such as index no. of bacterial population then we use logarithmic transformation.

Example: Bacterial growth on petri plate.