**Unit-5**

**Design of Experiment**

**Analysis-of-Variance (ANOVA):**

In this case we say that there is one factor, namely treatment, and the factor is more than two levels and thus more than two samples. In more than 2 sample problem, it will be assumed that there are k samples from k populations. One very common procedure used to deal with testing population means is called the analysis of variance, or ANOVA. The analysis of variance is certainly not a new technique if the reader has followed the material on regression theory. We used the analysis-of-variance approach to partition the total sum of squares into a portion due to regression and a portion due to error.

In other word, when there are only two treatment t-test, or z-test is used to test whether the two sample means are differ significantly. When there is more than two treatments t-test or z-test is not sufficient, so that another test is necessary. In such condition (situation), another test or another technique is used to test the significance difference between means. Such another technique is called analysis of variance or ANOVA.

**Experimental Design:**

Experiment design means planning the experiments in such a way that relevant information should be collected in systematic way for the problem under study, so that efficient information should be drawn.

In estimation and testing for the two-sample case is covered under the important backdrop of the way the experiment is conducted. This falls into the broad category of design of experiments.

It is assumed that the factor levels (treatments) are assigned randomly to the experimental units. The experimental unit is discussed in design of experiment. Simply put, experimental units are the units that provide the heterogeneity that leads to experimental error in a scientific investigation. The random assignment eliminates bias that could result by systematic assignment. The goal is to distribute uniformly among the factor levels the risks brought about by the heterogeneity of the experimental units. A random assignment best simulates the conditions that are assumed by the model. In this Section we discuss blocking in experiments. When comparison between means was accomplished with pairing, that is, the division of the experimental units into homogeneous pairs called blocks. The factor levels or treatments are then assigned randomly within blocks. The purpose of blocking is to reduce the effective experimental error. In this chapter we naturally extend the pairing to larger block sizes, with analysis of variance being the primary analytical tool.

**Objective of Experimental Design:**

The main objectives of experimental design are:

1. To estimate the effects of various treatments.
2. To compare the differences of effects are significant or not.
3. To estimate the interaction effects of various treatments and to compare them
4. To estimate error effects.
5. To control error effects.
6. To give proper interaction of the results.

**Terminology of Experimental Design:**

**Experiment:** Experiment means of getting an answer to the question which is in the experimenter mind or the problem under study. Experiment can be divided into two categories; namely (i) absolute, and (ii) comparative. Absolute experiment consists of determining the absolute value of some characteristic such as correlation coefficients, average of a group of people, and comparative experiment consist of comparing different types of fertilizers, cultivation process, varieties of crops etc.

**Treatment:** Treatments means the inputs. Whose outcomes are to be estimated and compared? The different procedure under comparisons in an experiment is different treatment. Such as, in agriculture experiment, different types of fertilizers, cultivation methods, and different varieties of crops are treatments.

**Experimental units:** The place when different treatments are applied and the effects of treatments are measured. Such as in agricultural experiment, plots are the experimental units, in a hospital, a patient is an experimental unit, in feeding experiment, feeding of cows, a cow is an experimental unit.

**Yield:** The outcomes of the experiment due to the application of treatments in an experimental unit is called yield. It is also called effects. Such as, in agricultural experiment, production of crop on using different fertilizers is yield.

**Block:** The experimental field is divided into relatively homogeneous sub-groups or homogeneous strata or uniform among themselves than field as a whole are called blocks.

**Experimental Error:** The errors which are arises at the time of experiment and cannot be controlled by human hand being is called experimental error.

**Principle of Design of Experiment:**

The basic principles of design of experiments are:

1. Replication.
2. Randomization, and
3. Local control.

**Replication:** Replication means repetition of treatments under investigation. An experimenter resorts to replication in order to average out the influence of the chance factors on different experimental units. A treatment is repeated a large number of times in order to obtain more reliable result than is possible with a single observations. It works in two ways: (i) along with randomization, it provides an estimate of the treatment effect, and (ii) along with local, it provides minimization of error.

**Randomization:** The allocation of treatments to various plots in a random manner is called the randomization. The random process implies that every possible allocation of treatment have the same probability. Thus, if the treatment is replicated a number of times allocated randomly to various plots in a field, we are position to test the significance of observed treatment difference with the help of statistical test such as ANOVA. The main objectives of randomization are: (i) equalization of factor not under control, (ii) eliminates bias in any form, and (iii) provides a basis to estimate the treatment effects.

**Local control:** If the experimental material, say field of agriculture experimentation, is heterogeneous and different treatments are allocated to various unit (plots) at random over the entire (whole) field, the soil heterogeneity will also enter the uncontrolled (chance) factors and thus increase the experimental error as for as practicable without unduly increasing the number of replication or without interfering with the statistical requirement of randomness, so that even smaller differences between treatments can be detected as significant. To reduce experimental error, the whole experimental field is divided into homogeneous groups (blocks) row-wise or column-wise in such a way that variation between the block is maximum and within the block is minimum. This process is known as local control or blocking.

In other words, the process of regarding the experimental error by dividing the relatively homogeneous experimental area (field) into homogeneous blocks is known as local control.

**Completely Randomized Design (CRD) or Completely Randomized Block Design (CRBD):**

It is simplest of all the design which is based upon only two principles of design namely replication and randomization. In this design treatments are assigned completely at random manner so that each and every experimental unit has equal chance of receiving any treatment, it is appropriate for the homogeneous experimental material.

**Layout:** Let us consider t treatment in which ith is replicated ri times, i = 1, 2 ……. t, so that there are n experimental units such that n =

Suppose there are four treatments A, B, C, D, each treatments replicated three times. The layout of CRD is

|  |  |  |  |
| --- | --- | --- | --- |
| A | B | C | D |
| C | D | A | B |
| C | D | B | A |

In general number of replication should be made for each treatment except in the case when some treatments are of greater interest than the others.

**Mathematical model:**

Let the linear model be

= +

= μ + ( – μ) +

= μ + +; i = 1, 2 …..…, t; j = 1, 2 ……., r.

Where, = j t h unit of it h treatment.

μ = general mean effect.

= effect due to it h treatment = – μ.

= error due to chance.

**Assumptions:**

1. All the observations are independent.
2. All the observations are drawn from the population having constant variance.
3. All the treatments should be homogeneous as for as possible.
4. Various treatments and environmental effects should be additive in nature.
5. All are independently identical distributed (i.i.d.), N (0, σe2).

**Problem:** To test

H 0T: μ1 = μ2 = ………. = μ t; there is no significance difference between treatments.

H 1T: At least one treatment is different.

**ANOVA Table:**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Source of variation | d f | SS | MSS | F-Ratio |
| 1. Due to treatment 2. Due to error | t – 1  t (r – 1) | SST  SSE | MSST =    MSSE = | Fcal. = |
| Total | r t - 1 | TSS |  |  |

**Critical region:**

The tabulated value of F at α% and [ t – 1, t (r – 1)] d. f. is Fα, [ t – 1, t (r – 1)].

**Decision:**

If Fcal. Fα, [ t – 1, t (r – 1)] ; we accept H0. Otherwise reject H0.

**Computation formula:**

TSS =

= - C.F.

SST = r

=

SSE = TSS – SST

**Advantage of CRD:**

1. It is easy to layout.
2. It allows maximum number of d.f. for MSS due to error which minimize error sum of square.
3. It is simple to statistical analysis due to one way classification.
4. If some observations are missing the analysis still remain simple.

**Disadvantage of CRD:**

1. Principle of local control is not used.
2. It is suitable for small treatment, replication and homogeneous material only.

**Uses of CRD:**

1. It is used in green house.
2. It is used in laboratory etc.

**Example:** Three varieties of wheat are shown in four plots and the following yields were obtained.

|  |  |  |  |
| --- | --- | --- | --- |
| A 8 | C 5 | C 4 | B 5 |
| B 7 | A 4 | B 1 | A 6 |
| C 2 | B 5 | C 3 | A 7 |

Test the significance difference between the yields of varieties.

**Solution:** Here,

**Problem:** To test

H 0T: μ1 = μ2 = μ 3; there is no significance difference between yield of different varieties.

H 1T: At least yield of one variety is different.

Now,

Calculation table is

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Varieties  of wheat |  | | | | Total  Ti. |
| A | 8 | 4 | 6 | 7 | 25 |
| B | 7 | 5 | 1 | 5 | 18 |
| C | 2 | 5 | 4 | 3 | 14 |
|  | | | | | G = 57 |

T = 3, r = 4, N = t r = 3 4 = 12

C.F. = = = 270.75

= 82 + 42 + 62 + 72 + 72 + 52 + 12 + 52 + 22 + 52 + 42 + 32 = 319

TSS = = 319 – 270.75 = 48.25

SST = = = 15.5

SSE = TSS – SST = 48.25 – 15.5 = 32.75

**ANOVA Table:**

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Source of variation | d f | SS | MSS | F-Ratio | F-Table |
| 1. Due to treatment 2. Due to error | 2  9 | 15.5  32.75 | 7.75    3.6389 | Fcal. = 2.13 | F 0.05, (2, 9)  = 4.26 |
| Total | 11 | TSS |  |  |  |

**Critical region:** The tabulated value of F at α = 0.05 is F 0.05, (2, 9) = 4.26

**Decision:** Since, F Cal < F 0.05, (2, 9) ­= 4.26, we accept H0; there in no significance difference between treatments.

**Question:** There are three varieties A, B and C and the following data are

B40 C60 A40 B50

C50 A70 A45 D60

C55 A50 B60 A70

Test the significance of three varieties. [Ans: F cal. = 0.23]

**Question:** Three treatments A, B and C are composed in a completely randomized design with six replications for each. The data are given below.

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| A17 | B10 | A20 | C33 | B23 | B21 |
| B15 | A25 | A17 | C35 | C29 | B29 |
| A33 | C25 | B19 | C37 | A23 | C30 |

Analysis the experimental yield and state your conclusion.