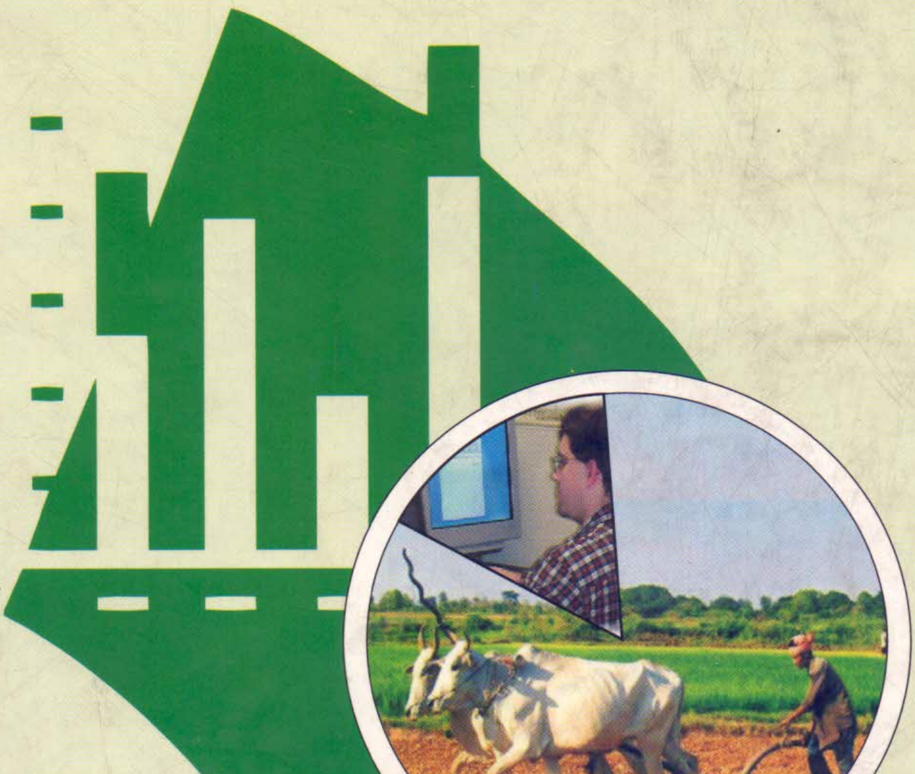
Agriculture and Applied Statistics – II

Pradip KUMAR Sahu
Bidhan Chandra Krishi Viswavidyalaya

Agriculture and Applied Statistics-II



P.K. SAHU • A.K. DAS

CONTENTS

CHAPTER

1 Index Number

1.1. GENERAL IDEA

Quite frequently people use to say that people are saying, 'prices of essential items **have** gone up', 'prices of computers have come down', 'cost of cultivation for agriculture crops **has** gone up', 'agricultural production of the country has been static', 'rupee has been devaluated', 'stock index has fallen' and so on. In each of the statement, there is a tendency to compare the present value of certain phenomenon with its past values. Similarly, we use to say that cost of living in city A is high compared to city B, prices of agricultural commodities in villages are low compared to the cities and so on. In the second case we are comparing a phenomenon in two different places. Thus the inherent tendency of the human being is to compare different phenomena under different situations, viz, over different time periods or over different places. While comparing a phenomenon (may be a group of related variables) over different situations, question arises, how to measure the average change of a phenomenon (group of related variables) over two different situations. **An index number is the measure of the relative changes in level of a phenomenon or variable or group of related variables under two different situations**. Thus the situations may be different time periods, different geographical locations or some other characteristics such as group of people, professions etc. An index number compares the levels of a phenomenon under two different situations. One can compare the cost of living of a certain type of people (say agricultural labourers) in two different places or in two different time periods. Thus if we say the cost of living index of agricultural labourers in the year 2007 is 185 compared to the year 1997; it means to maintain same standard of living by the agricultural labourers in 2007 as per the standard of living during the year 1997 they require 185% of the money in the year 2007. That means two maintain the same standard of living if an agricultural labourer required Rs. 100 in the year 1997, he or she requires Rs. 185 in the year 2007. Similarly, if we say that the cost of living index of the city of Mumbai is 225 with respect to Kolkata then it means if a person can maintain his life with Rs. 100 in Kolkata, he requires Rs. 225 to maintain same standard of living in Mumbai. Thus in the above two examples the cost of living indices (185 and 225) give an idea about the relative costs of livings in two different situations viz. two different time periods (2007 compared to 1997) and two different places (Mumbai with respect to Kolkata) respectively.

The above discussion clearly indicate that the main idea behind the study / use / construction of index number is to compare the level of phenomenon / variable / group of variables with respect to either a particular time or a particular place. Thus the comparison or the relative position of a phenomenon in a time period / or place with respect to a base time period or place is of prime importance in analysis of index number. **The time period or the place based on which the level of certain phenomenon / variable / group of variables is compared at certain other time period or for certain other place is called the base period or the base place and**

CHAPTER

2 Time Series Analysis

2.1. Information (data), its variability and the pattern of variability is the major concern in statistical problems. A characteristic may vary over individual objects, over the places, over the times and so on. We refer a set of data as **a cross section data if the observations are pertaining to a particular time period but vary over different individuals, places etc.** A particular commodity, say fertilizer price, at a given point of time over different retail shops constitutes a cross sectional data. On the other hand, a **time series data is observations / information taken or arranged chronologically**. Thus the production of wheat in India over the years, the gross domestic product (GDP) of USA over the years, the intensity of pests in mustard crop over the weeks, industrial production in India over the years, birth and mortality rate over the years in different states of India etc. are the examples of time series data.

Table 1 : Examples of cross section and time section data

Production of milk in different districts of West Bengal in 2005-06

Districts	Production ('000 tones)
Burdwan	480
Birbhum	100
Bankura	173
Midnapore	401
Howrah	135
Hoogly	410
Nadia	488
Murshidabad	301
Malda	152
Jalpaiguri	111
Cooch Behar	131

A. Cross section data

Monthly retail price of sugar

Month	Price (Rs./kg)
January'07	20.00
February'07	19.50
March'07	19.00
April'07	18.50
May'07	18.00
June'07	18.00
July '07	18.00
August'07	17.00
September'07	17.00
October'07	16.00
November'07	16.00
December'07	16.00

B. Time series data

Yearly production wheat in India

Year	Production in million tones
1980-81	36.31
1990-91	55.13
2000-01	69.68
2003-04	72.11
2004-05	68.64
2005-06	69.48

C. Time series data

In the above table A, data pertaining to the production of milk is a cross section data as it relates to production figures of different districts at a particular point of time i.e. the year 2005-06. On the other hand the information given in table B and C are time series data because in

CHAPTER 3 Analysis of Variance

In statistical studies variability and measures of variability are the major focal points of attention. Variability, measured in any character or variables can be partitioned in to a number of components associated with the type, classification and nature of data. **Analysis of variance is a systematic approach towards partitioning the variance of a variable into assignable and non-assignable causes.** Depending upon the nature, type of data and classification of data, analysis of variance is developed for one way classified data, two-way classified data with one observation per cell, two-way classified data with more than one observation per cell etc. Before taking up the analysis of variance in detail, let us discuss about linear model which is mostly being used in analysis of variance.

3.1. LINEAR MODELS

Any observable quantity can be assumed of two parts: (*i*) its true value and (*ii*) the other one is the error part which may be because of chance factor. Thus, if y_i be the i-th observation of a variable y then it can be decomposed as $y_i = \alpha_i + e_i$, where α_i is the true value of the variable y and e_i is the error. This α_i again may be the linear combination of 'm' unknown quantities $\gamma_1, \gamma_2, ..., \gamma_m$; called effects. Thus $\alpha_i = a_{i1}\gamma_1 + a_{i2}\gamma_2 + ... + a_{im}\gamma_m$ where a_{ij} (j = 1,2,……,m) are the constants and takes the values 0 or 1. Thus,

$$y_i = a_{i1}\gamma_1 + a_{i2}\gamma_2 + ... + a_{im}\gamma_m + e_i$$

If a_{ij} = 1 for a particular value of j and for all i then such as γ_j is termed as general mean or general effect. A linear model in which all the γ_j's are unknown constants (known as parameters) is termed as **fixed effect model**. On the contrary, a linear model in which γ_j's are random variables excepting the general mean or general effect is known as **variance component model** or **random effect model.** A linear model in which at least one γ_j is a random variable and at least one γ_j is a constant (other than general effect or general mean) is called a **mixed effect model.**

3.1.1. Assumptions in Analysis Variance

The analysis of variance is based on certain basic assumptions:

(*i*) the effects are additive in nature,

(*ii*) the observations are independent

(*iii*) the variable concerned must be normally distributed

(*iv*) Variances of all populations from which samples have been drawn must be the same. In other words, all samples should be drawn from normal populations having common variance with same or different means.

CHAPTER 4

Basics of Experimental Designs

One of the ingredients of statistics is the information – data. Collection of information (data) can mainly be done through sample survey technique or through conduction of experiments. When a researcher records observation on existing population, taking a sample from the population without interfering it then it may be done through sample survey technique. On the other hand when a researcher designs a set of activities keeping in mind the objective and tries to measure the variations on certain variable / entities / objects keeping other variations as minimum as possible, it is known as experiment. Thus **an experiment is a systematic process or activity which leads to collection of information on certain objects to reply to the objectives that the researcher has already in mind.** For example if one wants to know the average mileage given by a specific brand of motor bike, he can pick up a random sample of certain number of bikes of the specific brand and record the mileage given by the bikes per unit of fuel consumed to get the average mileage given by the brand per unit of the fuel consumed. If one wants to know the physical ability of a certain type of students, he can select a random sample of the specific type of students and record their physical performance to get the average performance as well as the variations. If one wants to know the relationship between the number of tiller per hill and the number of effective tillers per hill (panicle bearing tillers) he can select a good number of hills at random from a field and study the relationship. All these examples are the examples of absolute experiment. Thus, an **absolute experiment is an experiment in which certain absolute values like average, correlation coefficient, median, mode, etc are worked out to describe the population**. In other type of experiment, where **the researcher compares the effects of different objects/entities under consideration are known as comparative experiment**. In a comparative experiment, for example, an experimental design is done in such a way that the experimenter can compare the objects or entities under identical conditions keeping the other sources of variations as minimum as possible. In agriculture, some times we are interested to study the varietal performances of different varieties of certain crop at a particular place, so that a variety could be recommended for the specific crop for that area. Fertilizer trials are taken up to identity the best combinations of fertilizer providing maximum yield. Experiments among the students are conducted to understand the effect of certain type of health drinks on improving the agility of the students. All these are the examples of comparative experiment.

Whenever the objective of the study is comparative in nature, it is generally designed experiment. However, an experiment may also be designed for getting absolute values depending upon the objective of the study. An experiment is mainly designed on the basis of the objective, time given to fulfill the objective, cost allowed to conduct the experiment, availability of the experimental materials, requirement of human resources and above all the feasibility. Sometimes accessibility is also required to be considered before taking up an experiment. Thus the experiment with migrations of birds, average height of the rhododendron trees in North Western Himalayan

CHAPTER 5

Basic Experimental Designs

5.1. COMPLETELY RANDOMIZED DESIGN

Completely randomized design (CRD) is the simplest of all designs where only two principles of design of experiments i.e. replication and randomization have been used. The principle of local control is not used in this design. The basic characteristic of this design is that the whole experimental area (*i*) should be of homogeneous in nature and (*ii*) should be divided into as many number of experimental unit as the sum of the number of replications of all the treatments. Let us, suppose there are five treatments A, B, C, D, E replicated 5, 4, 3, 3, and 5 times respectively then according to this design we require the whole experimental area to be divided in to 20 experimental units of equal size. Thus, completely randomized design is applicable only when the experimental area is homogeneous in nature. Under laboratory condition, where other conditions including the environmental condition are controlled, completely randomized design is the most accepted and widely used design. Let there be t treatments replicated $r_1, r_2, \ldots\ldots r_t$ times respectively. So in total we require an experimental area of $\sum_{i=1}^{t} r_i$ number of homogeneous experimental units of equal size.

5.1.1. Randomization and Layout

To facilitate easy understanding we shall demonstrate the layout and randomization procedure in a field experiment conducted in CRD with 5 treatments A, B, C, D, E being replicated 5, 4, 3, 2, 6 times respectively. The steps are given as follows :

(*i*) Total number of experimental unit required is $5 + 4 + 3 + 2 + 6 = 20$. Divide the whole experimental area into 20 experimental units of equal size. For laboratory experiments the experimental units may be test tubes, petri dishes, beakers, pots etc. depending upon the nature of the experiment.

(*ii*) Number the experimental units 1 to 20.

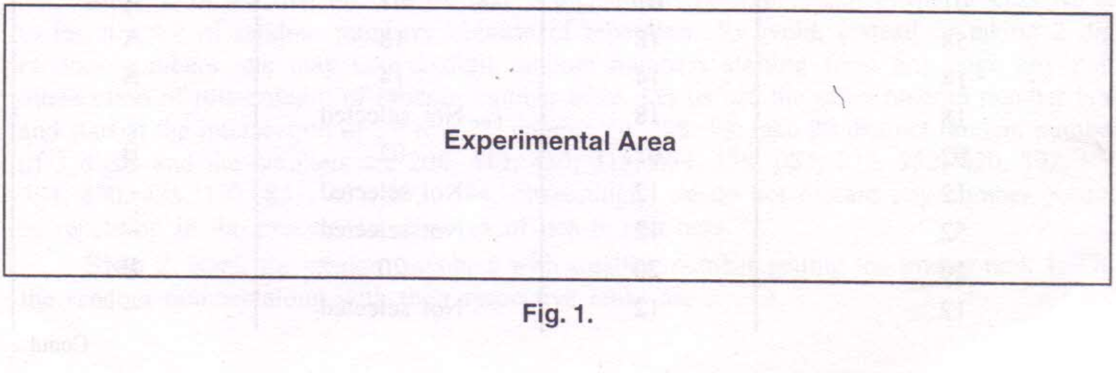


Fig. 1.

CHAPTER 6

Factorial Experiments

In most of the practical situations, instead of conducting experiment for comparing the effects of single set of treatments, more than one set of treatments are required to be tested or compared. In agricultural experiments, particularly the response of different doses /levels of one factor is expected to vary over the different doses or levels of other set of treatments / factors. Thus it is our common experience that a particular variety of wheat or a crop is responding differentially under different irrigation schedule, different rates of fertilizers, different weed management practices, different crop protection practices etc. and an experimenter wants to know all the above. Moreover, if the experimenter wants to know not only the level or the doses of individual factor giving better result but also wants to know the combination of the levels of different factors which is producing the best result. For example, an experimenter may be interested to know not only for the best dose of nitrogen, phosphorus and potassium along with best dose combination of these three essential plant nutrients to get best result in guava. This can be done by designing his experiment in such a way that best levels of all these three factors, N, P, K can be identified along with the combination of N, P K to get best result. Factorial experiments are the methods for inclusions of more than one factor to be compared for their individual effects as well as interaction effects in a single experiment. Thus several factors are investigated for their individual effects and the effects in combination in factorial experiments.

6.1. ADVANTAGES OF FACTORIAL EXPERIMENTS

From the above discussion following advantages of factorial experiments could be noted.

(*i*) Factorial experiments give the opportunity to an experimenter to combine the effects of more than one factor at a time.

(*ii*) Compared to single factor experiments factorial experiments are effective because of the fact that the interaction effects can be worked out from these experiments which is not possible in single factor experiments.

(*iii*) Factorial experiments are not only time saving but also to some extent cost saving also.

(*iv*) The requirement of minimum degrees of freedom for error components in analysis of variance is easy to achieve in factorial experiments compared to single factor experiments; as the number of factors is more in factorial experiment, lesser number of replications compared to single factor experiments may result in minimum required 12 degrees of freedom for the error component in analysis of variance.

CHAPTER 7

Statistical Model

7.1. GENERAL IDEA

Starting from our daily life to the development of cryogenic rockets, we frequently use the term model. The term model is defined and used in various ways depending upon the purposes. To an extension personal a model describes the path/strategies for extending a new idea/innovation/technology to its recipients. A plant growth model describes the different growth components or growth parameters during different phases of growth and development. A model may be used to describe the process of development for a community or for an area. Thus, a model can be conceptulised in various ways. As such a model can be presented in the form of a flow chart /diagram/portrait/mathematical equation(s) and so on. In brief, a model can be described as a means or ways of presenting a process. In this section we are concerned about the mathematical models i.e. the models represented only in the form of mathematical formulae/equations.

7.2. TYPES OF MODELS

A mathematical model may be a single equation model or a multiple equations model depending upon the number of equations involved in describing a particular process.

A model can be either **deterministic** or **statistical**. For example, $y = f(x)$ is a deterministic model while $y = f(x, \varepsilon)$ is a statistical model which takes care of the random component.

A statistical model may again be either **linear** or **non linear** in nature. Now the question is **what do we mean by linearity**. We know that an equation, may it be deterministic or statistical, is constituted of three components: the variables, the parameters and the random part. An equation may be linear or non-linear in variables or in parameters. An equation is linear in variables when none of the variables involved in the equation has got power other than unity or zero. Thus, $y = \alpha_0 + \alpha_1 x_1 + \alpha_2 x_2 + \dots\dots\dots + \alpha_k x_k + \varepsilon$ is the example of an equation linear in variables. On the other hand, an equation is linear in parameter when none of the parameter has power other than unity or zero. Thus, the above equation is also a linear equation in parameter.

An equation in which at least one variable / parameter has got power other than unity or zero is termed as non-linear in variable or parameter respectively.

$y = \alpha_0 + \alpha_1 x_1 + \alpha_2 \sqrt{x_2} + \alpha_3 x_3^2 + \dots\dots\dots + \alpha_k x_k^k + \varepsilon$ is an equation non-linear in variables but linear in parameter.

Similarly, the equation $y = \alpha_0 + \alpha_1 x_1 + \sqrt{\alpha_1} x_2 + \alpha_1^2 x_3 + \dots\dots\dots + \alpha_1^k x_k + \varepsilon$ or $y = \dfrac{\alpha}{\alpha + \beta} e^{\alpha + \beta'}$ are non-linear in parameter but linear in variable.

CHAPTER 8 Special Experiments and Designs

8.1. AUGMENTED DESIGN

In many practical situations, experimenters are to conduct experiments where it becomes very difficult to replicate the treatments; mainly because of scarcity of experimental materials. This is more often experienced in breeding experiments, where the material may not be sufficient for even a second replication to facilitate estimation of different effects including experimental error. Experiments with breeders seed materials or germplasms for different crop breeds are found to be constraints by the availability of adequate amount of materials.

Again when a large number of germplasms / lines are required to be evaluated at a time and the process is being constrained by the availability uniform land. It becomes very difficult to accommodate a large number of germplasms (say 100) in a uniform land under CRD setup or in a RBD setup with block which can accommodate such a huge number of germplasms or lines. Thus, making it very difficult to evaluate the germplasms, particularly under RBD set up. We are not considering LSD for its typical characteristics which may not advisable under the present conditions.

Sometimes the experimenters want to compare not only the factorial treatments among themselves but also against a treatment called control, which may be standard package of practice, water spray or zero levels of all the factors under consideration. Thereby, the usual factorial analysis is not possible in these cases, basic design and analysis of the experiments are required to be modified / adjusted accordingly.

To deal with the above situations, experimental designs called **augmented design** have been developed. Federer (1956) has given augmented designs for the first two situations. Suitable adjustment is possible in the analysis of variance to tackle the problem of comparing treatments over the factorial treatments in the form of control treatments.

We know that the quantity of breeders seed is not sometimes sufficient to repeat it to a desired number of times. But for it's suitability it required to be tested with standard comparable and competing materials, which are available in sufficient amount to be repeated desired number of times. So there arises a problem in designing such experiments. Augmented designs may become handy in these situations.

Take another example, where a particular experiment is required to be conducted under factorial arrangement. If one is interested to know the efficacy of three doses of a particular growth regulator along with four doses of nitrogen in a crop improvement trial, the experiment can be conducted in a factorial randomized block design with $3 \times 4 = 12$ treatment combination. Now if the experimenter wants compare the treatment effects against a treatment where no growth regulator or nitrogen has been used how he or she will do? There are two ways to include zero level in both the growth regulator and nitrogen then making the factorial treatment combinations $4 \times 5 = 20$ (i.e. treatments) and conduct the experiment as per factorial randomized

CHAPTER 9 Variance and Covariance Analysis and Partitioning of Variance and Covariance

Variability is one of the major concerns of statistical theories. As in real life situation we are to deal with a number of variables at a time, co-variability also plays an important role in statistical analyses. In chapter one, we have dealt with various analysis of variance techniques. Analysis of variance is nothing but a technique of partitioning the variability due to different assignable and non-assignable factors. Like that of partitioning of variances into different components one can also partition the co-variabilities among the variables into different components (assignable and non-assignable factors), known as **analysis of covariance.** In the following sections we shall discuss the method of analysis of covariance and later on in this chapter we shall discuss how variance and covariance can farther be partitioned into different components.

9.1. ANALYSIS OF COVARIANCE

Analogous to the sum of squares or the variance $V_x = \frac{1}{n}\sum_{i=1}^{n}\left(x_i - \bar{x}\right)^2$ for any particular variable X, the sum of products or the covariances between any two variables is given as $V_{xy} = \frac{1}{n}\sum_{i=1}^{n}\left(x_i - \bar{x}\right)\left(y_i - \bar{y}\right)$ Thus, in analyses of covariance we are interested in partitioning the co-variability among the variables in to different assignable or non-assignable factor. In real life situation particularly in agricultural field the bio-physical features in any experiment rarely behave independently; rather these are found to be functionally related each other. For example, if we consider a varietal trial where azola, a bio-fertilizer is used as one of the nutrient supplier, in addition to supplying nutrient each also adds profuse amount of organic matter in the field. Now there might be a relation between the amount of organic matter in the field by the population of azola and the yield of paddy. And this functional relationship can be used to adjust the grain yield due to varietal differences. The yield of paddy can be corrected by subtracting the effect of organic matters due to application of azola from the varietal yield. Here the performance, the functional relationship between the yield and the quantum of organic matter is effectively used to extract out the varietal performance from other effects. Thus, in covariance analysis there are two types of variables; the characteristic of main interest and the information on secondary or the auxiliary or the covariates. In covariance analysis any number of covariates can be used to any type of functional relationship between the variables. The simplest among the covariance analyses is one in which only one covariate is used. There are several examples where covariance analysis can effectively be used in augmenting the precession of the experimental results. For example, in yield component analysis of paddy the yield components viz, the number of hills per unit area, number of effective tillers per hill, number of grains per panicles etc. can be used as covariates or concomitant variables. In a study of health drinks on the growth and the physic of

CHAPTER

10 Path Analysis

10.1. GENERAL IDEA

In agriculture and allied fields, including social sciences different response variables are influenced by several other variables. For example, if we consider yield in agricultural crops like paddy, it is quite evident that the yield (y) is the manifestation of effects of different yield components like, number of hills per square meter (x_1), number of tillers per hill (x_2), number of panicle bearing tillers per hill (x_3), length of panicle (x_4), number of grains per panicle (x_5), number of chaffy grains per panicle (x_6), test weight of grains / i.e. thousand grain weight (x_7) etc. All these components are influencing the yield. The yield components are not only individually correlated with yield, but also are correlated among themselves. In socio-economic studies, one can find that Social status of a person is influenced by his/her age, education, family size, income, expenditure, involvement in social programme etc. Social status of the person concerned is the manifestation of all these contributing factors. Similarly, while studying the physical fitness of human being, one considers the body characteristics like height and weight of body, leg length, basal metabolic rate (BMR) etc.

In conventional breeding programme, characters are selected for utilization depending up on the influence/effect of the character concerned in improving the response character/variable. As the variables/characters are not only correlated with the response variable but also among themselves so, the question arises how to separate the direct influence of a particular variable/character on the response variable/character and the effects of the variable/character through other variables/characters because of their interrelationship; possible answer is through the **Path Coefficient Analysis.**

Path coefficient analysis or simply path analysis is special type of multiple regression analysis. It was first developed as a method to decompose correlation coefficients into different components. In a path coefficient analysis there are at least two groups of variables: **independent variables or the exogenous variables and the dependent variables or the endogenous variables.** Path coefficients indicate the direct and indirect effects of the variables on other variable. A path diagram is a flow chart to represent the direction of effects of different variables on the dependent variables.

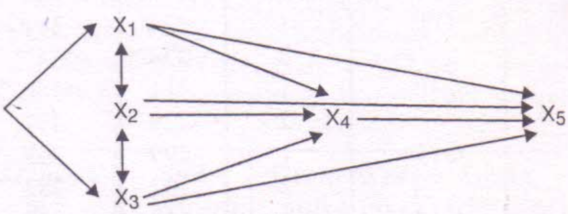


Fig. 1. Path diagram with two dependent and three independent variables

In the above diagram X_1, X_2, X_3 are the exogenous independent variables and X_4, X_5 are the endogenous dependent variables. Here the variable X_5 is not only dependent on X_4 but also

CHAPTER 11 Stability Analysis

It has already been mentioned that manifestation of phenotypic nature of any character is resultant of the genotype, environment and genotype-environment interaction under which an individual is grown. As a result, one of the major task before releasing a genotype / variety for cultivation at farmers field is to judge its performance under different situations. Multi-locational or multi-seasonal or both trials are conducted to understand the adaptability and the performance of the genotypes or the varieties under different situations. A genotype or a variety would be stable one when its performance remains almost same over the different situations. There are different varieties or genotypes which require to study their suitability under different situations like different locations, different input supply situations (intense or low inputs), different seasons, different land situations etc. Multi-situational / multi-locational / multi-seasonal trials are conducted to oversee the consistency in performance of the varieties or genotypes. Different situations / locations / seasons etc. where the trials are conducted can be regarded as environments. The information generated from these trials is analyzed environment wise followed by pooled analysis. If the genotype-environment effects become non-significant there is no problem; one can conclude that the genotypes have performed consistently over the situations. On the other hand, if the genotype-environment interaction effect becomes significant i.e. the performance of different genotypes has varied over the situations; then questions arises which genotype is more stable or consistent among the test genotypes and the situations where the different genotypes can be exploited for their full potential. Thus, the problem of measuring stability comes in to picture.

Most of the measures relating to stability appeared in literature are based on linear regression methodology. However, works of Comstock and Robinson (1952) and Wricke (1962) were based on mean values. Yates and Cochran (1938), Wilkinson and Finlay (1963), Eberhart and Russell (1966), Perkins and Jinks (1968), Freeman and Perkins (1971) etc. have given the different methods of assessing stability. Among the methods, the Eberhart and Russel method, Perkins and Jinks method and Freeman and Perkins method are most widely used methods of estimating stability.

11.1. COMSTOCK AND ROBINSON MODEL

According to this model the phenotypic performance of several genotypes can be looked up on as the additive effects of five major components. Let there be v varieties of paddy to be tested in e environments over the y years for their stability with respect to yield during kharif season. If the experiments are conducted in Randomised Block Designs with three replications then this model assumes simply the three factor factorial RBD model as discussed in chapter four of this book. A stable genotype is one which shows similar trend over the locations and over the years. A genotype is regarded as best, if it happens to be stable as well as the highest yielder among the varieties over the locations and over the years. On the other hand if the varieties, in

CHAPTER 12

Biological Assay

12.1. GENERAL IDEA

In many biological experiments, chemicals like hormones, vitamins, insecticides, fungicides, herbicides etc are applied to living beings like animal, plants, insects, bacterial cultures etc. to record the responses of the chemicals applied. Application of chemicals may cause changes in the living beings which can be a measurable characteristic and the magnitude of the characteristic change mostly depends on the intensity/dose/potency/concentration of the chemical used. This type of biological experiments from which one estimates the nature of the concentrations/potency/intensity/dose of the materials (may be chemicals) by means of reactions that follow its application to a living matter is known as **biological assay** or **simply bio-assay**. The material which is applied on the living beings are known as **stimulus** and the living being which are subjected to the chemical (stimulus) and from which responses are recorded are known as **subject**. The primary objective of a bio-assay experiment is to determine a valid and efficient estimate of the potency of the test preparation.

In bio-assays the responses may be of two different types : **(*a*) qualitative and (*b*) quantitative**. In qualitative type of responses, the response may be mortality of insects in insecticidal experiment, death of bacterial cells in bacterial culture etc. Thus, the qualitative respond may be of the form "either-or" i.e. the subject will respond or will not respond. For example, an experiment with different doses of a particular insecticide, the response may be the death of the insects. The qualitative response also known as **quantal response** may be of the type "**all or nothing**". On the other hand, in quantitative response, the responses can be measured. For example, change in the weight of the whole subject or some particular portion of the subject etc.

12.2. POTENCY

The potency of a test preparation is the ratio of two equally effective doses of the test preparation and the standard preparation. The responses of selected doses of both the standard preparation and the test preparation are recorded in bio-assay experiment and the doses at which the responses are equal for both the preparation are noted.

12.3. TOLERANCE

The occurrence or non-occurrence of quantal response depends on the intensity of the stimulus applied to the subjects under control condition. Under certain level of intensity there will be no response and above which response could be recorded. The intensity at which the response

CHAPTER 13 Multivariate Analysis

Statistical procedures may broadly be classified in to two groups : univariate analysis and multivariate analysis. In univariate analyses, one variable is considered at a time on the other hand in multivariate analyses more than one variable is considered at a time. The simplest case of multivariate analysis is the bivariate analysis, in which two variables are considered together. In fact, the variables what we consider in the filed of agriculture, economics, anthropology, sociology, psychology, management etc. tend to vary together and as such multivariate analysis will be useful in analysis of the system as a whole. Univariate analysis may throw lights on the characteristics (including statistical properties) of a particular variable but are inadequate in explaining the relationship, interdependence, relative importance of different variables considered together.

Let us take the example of yield component analysis. Yield is the ultimate manifestation of several yield components. If we consider the yield in paddy, then we know that the yield depends on number of hills per unit area, number of tillers per hill, number of effective tiller (panicle bearing tiller) per hill, length of the panicle, number of grains per panicle, test weight (1000 grain weight) of grains. These yield components of paddy are not independent of each other, rather these are correlated, interdependent to each other. These also vary in their importance towards ultimate yield of paddy. Univariate analysis can throw light separately on each of the character but to analyse the system as a whole taking due consideration of their interdependence, relationships and importance, multivariate analysis is the possible option. Several examples in business, economics, management, medical science and other disciplines can also be put forwarded where multivariate analysis is very useful.

As multivariate analysis analyses the system as a whole, having number of variables at a time taking in to consideration of their interdependence, relationships, importance etc., it is complicated compared to univariate analysis. But with the advent of computer technology and statistical softwares the use of multivariate technique is gaining momentum in every sphere. As such the area and coverage of multivariate analysis is huge one and it is not possible to include all these here in this chapter. In this chapter we shall try to provide an outline of some of the multivariate techniques which can be used in different agriculture and allied fields. **The details of analytical procedure is beyond the scope of this book, in this chapter we shall take examples to present the results (analysed using various computer softwares) of various multivariate analyses to demonstrate the usefulness of the multivariate techniques. Interested readers may find useful references in the reference section of this book.**

13.1. CLASSIFICATION OF MULTIVARIATE ANALYSIS

J N Sheth (1971) has classified multivariate analysis in to two broad groups: a) dependence methods and b) interdependence method. In the first group of analysis there must be some dependent variables along with the independent variables on the other hand in the second group

CHAPTER 14

Computer and Statistics

Development of statistics goes side by side with the development of human civilization. Because of the need of the society and inquisitiveness of human beings, science has developed by leaps and bounds. Newer and newer disciplines of science have come up and in each discipline of science the need and usefulness of statistics have been felt ever increasingly. A number of theories and areas of statistics have been developed due to the demand of other discipline of science. Nowadays one can rarely find any discipline, starting from agriculture, biology, education, economics, business, management, medical, engineering, sociology, ecology, space science and even in war fares where statistics is not used. As a result, in every sphere of human civilization, statistics has become an indispensable part.

Though statistics, as a discipline has developed tremendously during the passage of time; its application was restricted at the initial stages because of the fact that many of the statistical theories require knowledge of mathematics and require extensive calculation. With the advancement in computing facilities, use of statistics started increasing. Starting from the era of manual facit to modern laptop computers through desk/ pocket/ scientific calculators and different generations of computers, the computing facilities, has increased tremendously. With the ever increasing facility the use of different statistical theories in various fields which were not feasible earlier become possible nowadays. Statistical theories are more extensively applied nowadays.

In the above two paragraphs our attempt was to understand the importance of statistics side by side the need for developing computing facility for the effective use of statistical theories towards the development of the human civilization. But this development is not unidirectional or flawless. Numerous reports are there where statistical calculations have been made with the help of the computer packages, without understanding the theories and situations where actually the specific calculation are required to be taken up. In brief, statistical analysis are taken up without knowing the logic and utilities in these cases. Statistical theories are used best by the subject matter specialists in consultation with an efficient statistician. Understanding of both the specialists towards the field of each other to a certain degree is essential for efficient use of statistical theories towards advancement of human civilization. Statistics is just like a moulded clay one can make God or Devil out of it as per the choice of the user. Misuse of statistical concept is increasing day by day. We must be cautious about the practice of garbage in and garbage out (GIGO) *i.e.* what are fed to the computer, what are our requirements, whether it is feasible to get from the information fed to the computer using statistical theories, what should be the direction to the computer and what is the output generated by the computer. In the following example we shall demonstrate how many types of statistical analysis can be taken up with a given set of data. But the question is which one is the appropriate one ?

Following yield (q/ha) information is pertaining to an experiment conducted with three types of manures (M1, M2, M3) , three types of weed control measure (W1, W2, W3) for four different varieties (V1, V2, V3, V4) of paddy. Problem is to identify the effects of manure, weed control measure and variety on yield of paddy.

SUBJECT INDEX

Agriculture and Applied Statistics is a holistic attempt to present the otherwise complex theories of statistics in a simple and lucid form so as to make it convenient and user friendly, even to a student not having enough mathematical orientation. By virtue of their vast experiences in the field, the authors have tried to reach major sections of the users of statistics in various fields in solving various problems of their respective fields. The book is written in two parts; Agriculture and Applied Statistics – I have already been received and accepted in huge way by the students, researchers, teachers and other users of various fields.

Agriculture and Applied Statistics – II has fourteen chapters. It starts with the index number and time series analysis to benefit the user in the field of business, economics and management. This volume also contains topics on design of experiments, experiments at farmers' field, combined experiments, stability, sustainability, path analysis, multivariate analysis etc. Special feature of this book is that each and every theory is supported by example from the respective real life field. Attempts have been made to demonstrate how a user can solve his/her problems using simple computer oriented programme, available in most of the computers. Emphasis has been given not only to demonstrate in solving problems of various fields but also to the art of drawing inferences from the problems. The importance of computer in statistical analysis side by side its misuse has also been presented to make aware the user about the correct and specific statistical technique. At the end of each chapter the questions and problems will help the students immensely. Utmost care has been taken to present varied range of problems in agriculture and allied fields keeping in mind the examination need of the students and also the problems faced by the users.

Dr. P.K. Sahu is a Reader in Agricultural Statistics, Bidhan Chandra Krishi Viswavidyalaya, West Bengal, India. Born in 1963, he has a bright academic career with 1st division or 1st class in all spheres of academic life, besides being awarded with National Scholarship and Research Scholarship. Basically a student of agricultural science, he has specialized in the field of Agricultural Statistics. He has served the Central University, North Eastern Hill University, Medziphama Campus; Central Silk Board, Ministry of Textile, Government of India and as member of the Under Graduate and Post Graduate Council for Faculty of Agriculture, Bidhan Chandra Krishi Viswavidyalaya.

Dr. A.K. Das is Professor, Department of Agricultural Statistics and has been teaching Agricultural Statistics at under graduate and post graduate classes at Bidhan Chandra Krishi Viswavidyalaya, West Bengal for more than twenty-five years. He has taught sample surveys to the under graduate and post graduate students in the departments statistics, Kalyani University, West Bengal and Ph.D students at Bose Institute, Kolkata as a guest lecturer. Prof. Das has published several research papers in different journals of statistics.

ISBN 978-81-272-5049-2

Rs. 275/-

KALYANI PUBLISHERS