PLANNING AND APPLICATION OF STATISTICAL TECHNIQUES IN AGRICULTURAL EXPERIMENTATION

B. RAVINDRA REDDY*, P. SUMATHI AND G. MOHAN NAIDU

Department of Statistics and Maths, S.V. Agricultural College, Tirupati-517 502

ABSTRACT

Any researcher may follow these phases in his experimentation – Planning, Design, Data collection, Analysis and Dissemination. Statistical methodologies can be used to conduct better scientific experiments if they are incorporated into entire scientific process right from inception of the problem to experimental design, data analysis and interpretation. After collection of data various tools techniques viz., Descriptive statistics, Correlation, Regression, ANOVA, ANCOVA, Experimental designs and Technology adaptation experiment analysis over sites etc., may be used for valid results and conclusions.

INTRODUCTION

Scientists in the Agricultural and biological fields who are involved in research constantly face problems associated with planning, designing and conducting and finally analysis of experiments. Basic familiarity and understanding of the statistical methods that deal with issue of concern would be helpful in many ways. Researcher who collect data and then look for a statistical technique that would provide valid results will find that there may not be solutions to the problem and that the problem could have been avoided first by a properly designed experiment. Obviously it is important to keep in mind that we cannot draw valid conclusions from poorly planned experiments. Second, the time and cost involved in many experiments are enormous and poorly designed experiment increases such costs in time and resources. For example, an agronomist who carries out fertilizer experiment knows the time limitation of the experiment. He knows that when seeds are to be planted and harvested. The experimental plot must include all components of a complete design. Otherwise what is omitted from the experiment will incur additional time and expenditure and could be minimized by a properly planned experiment that will produce valid results as efficiently as possible. Good experimental designs are products of the technical knowledge of one's field, an understanding of statistical techniques and skill in designing experiments. In this paper, the importance of proper planning, designing the agricultural experiments and the important statistical tools and techniques used in analysing agricultural research data have been emphasized scientifically.

BASIC STATISTICAL TECHNIQUES

Descriptive Statistics: The 'Descriptive Statistics' analysis tool in any statistical software, it generates a report of univariate statistics for data in the input range, which includes information about central tendency (Mean, Median, Mode etc.,) and variability (Variance, Standard deviation, Range, Skewness, Kurtosis etc.) of the data (Gupta and Kapoor, 1970). The preliminary conclusions or behaviour of the data can be drawn from these techniques.

Correlation: The Correlation analysis measures the relationship between two data sets that are scaled to be independent of the unit of measurement. It can be used to determine whether two ranges of data move together i.e., The change or variation in one variable due to change in other variable(s) may follow a linear or non-linear (curvilinear) relationship or any other form. The problem lies in measuring the change in one variable due to the change in other variables and vice-versa (Sahu, 2007). Examples: The relationship between variables like Rainfall and Yield, Number of chemical sprayings and pest incidences, Income & Expenditure, Prices and Demand etc., is called correlation.

Regression: The 'Regression' analysis performs linear regression analysis by using the 'least squares'

*Corresponding author, E-mail: balamrr72@gmail.com

method to fit a line through a set of observations. You can analyse how a single dependent variable is affected by the values of one or more independent variables. When all independent variables are assumed to affect the dependent variable in a linear fashion and independently of one another, the procedure is called multiple linear regression analysis.

In regression analysis independent variable is also known as regressor or predictor or explanatory variable while dependent variable is also known as regressed or explained variable. Similarly, in regression analysis one of the two variables is identified as cause and the other is effect. Hence, regression models are also known as *cause and effect models* (Sarma *et al.,* 2013). For example, one can analyze how grain yield of barley is affected by factors like ears per plane, ear length (in cms), 100 grain weight (in gms) and number of grains per ear. Further in regression analysis, the value of the dependent variable for a given value of the independent variable can be estimated.

DIAGNOSTIC ANALYSIS OF REGRESSION

The R^2 value is the commonly used measure of the adequacy of the linear model. We can test for its statistical significance with the help ANOVA and understand the p-value. Now the question arises for the researcher. If the regression model is used to predict the values of Y at the know values of X, will these estimates match with the actual Y values or will there be any difference? If so what would be the average error? The Software's (Excel/ SPSS) resolves these issues as a part of the regression analysis by selecting the *Residuals and residual plots, Line fit plot, and the Normal probability plots (Q-Q plots).*

Residuals and residual plots: The difference between the actual and the predicted value of Y form a given model is called the residual. If the model is a good fit to the data, we can expect these residuals to be close to zero. However they do differ and software shows these differences in the form of residual output. Software also provides the plot of these standardized residuals.

Line fit plot: The plot of the actual and predicted values obtained from the regression model is shown in line fit graph. This is useful to judge the adequacy of the model. If many of the predicted values differ from the actual values we may consider the model is not good representative of the relationship.

Normal probability plots (Q-Q plots): The Normal probability plots are used to test the normality of a data which is one of the assumptions in regression analysis. It is a plot of the observed data against the normal quartiles. If the data has truly come from a normal population, we can expect a linear fit between normal quartiles and observed values. This method works well when the number of values is moderately large (Kachigan Sam Kash, 1986).

Covariance: The term 'Covariance' is a measure of the relationship between two ranges of data. It can be used to determine whether two ranges of data move together, i.e., whether large values of one set are associated with large values of the other (positive covariance), whether small values of one set are associated with large values of the other(negative covariance), or whether values in both sets are unrelated(covariance near zero).

ANOVA: The 'Analysis of Variance (ANOVA)' is a powerful statistical tool for tests of significance. ANOVA is the 'separation of variance ascribable to one group of causes from the variance ascribable to other group'. By this technique the total variation in the sample data is expressed as the sum of its non-negative components where each of these components is a measure of the variation due to some specific independent source or factor or cause. The ANOVA consists of the estimation of the amount of variation due to each of the independent factors (causes) separately and then comparing these estimates due to assignable factors (causes) with estimate due to chance factor (causes), the later being known as experimental error or simply error.

ANCOVA: The 'Analysis of Covariance (ANCOVA)' simultaneously examines the variances and covariance's among selected variables such that the treatment effect on the character of primary interest is more accurately characterized than by use of analysis of variance only. Analysis of covariance requires measurement of the character of primary interest plus the measurement of one or more variable known as covariates. It also requires that the functional relationship of the covariate with the character of primary interest is known beforehand. The ANOVA and ANCOVA tools will be used in all Basic and Factorial designs.

EXPERIMENTAL DESIGNS

Any scientific investigation involves formulation of certain assertions (or hypotheses) whose validity is examined through the data generated from an experiment

conducted for the purpose. Thus experimentation becomes an indispensable part of every scientific endeavour and designing an experiment is an integrated component of every research programme. Three basic techniques (principles) fundamental to designing an experiment are replication (repetition of the treatment under investigation), local control (homogeneous blocking), and randomization (random allocation of the treatments to experimental units). Whereas the first two help to increase precision in the experiment, the last one is used to decrease bias (Nigam and Gupta, 1979).

Based on these principles, the widely used experimental designs are; Completely Randomized Design (CRD), Randomized Complete Block Design (RCBD), Spilt Plot Design (SPD) and Factorial Designs with RCBD.

Completely Randomized Design (CRD):

Designs are usually characterized by the nature of grouping of experimental units and the procedure of random allocation of treatments to the experimental units. In a completely randomized design the units are taken in a single group. As far as possible the units forming the group are homogeneous. This is design in which only randomization and replication are used. There is no use of local control here. The source of variation in this design is due to treatment and error.

Randomized Complete Block Design (RCBD):

When the experiments require a large number of experimental units, the experimental units may not be homogeneous, and in such situations CRD cannot be recommended. When the experimental units are heterogeneous, a part of the variability can be accounted for grouping the experimental units in such a way that experimental units within each group are as homogeneous as possible. In RCBD, the treatments are then allotted randomly to the experimental units with each group (or block). The principle of first forming homogeneous groups of the experimental units and then allotting at random each treatment once in each group is known as local control. This results in an increase in precision of estimates of the treatment contrasts due to the fact that error variance that is a function of comparisons within blocks is smaller because of homogeneous blocks. If the number of experimental units within each group is same as the number of treatments and if every treatment appears precisely once in each group then such an arrangement is called a ***randomized complete block design.*** The source of variation in this design is due to blocks (replications), treatments and error.

Spilt Plot Design:

In conducting experiments, sometimes some factors have to be applied in larger experimental units while some other factors can be applied in comparatively smaller experimental units. Further some experimental materials may be rare while the other experimental materials may be available in large quantity or when the levels of one (or more) treatments factors are easy to change, while the alteration of levels of other treatment factor are costly, or time-consuming. One more point may be that although two or more different factors are to be tested in the experiment, one factor may require to be tested with higher precision than the others. In all such situation, a design called the split plot design is adopted.

A split plot design is a design is with at least one blocking factor where the experimental units within each block are assigned to the treatment factor levels as usual and in addition, the blocks are assigned at random to the levels of a further treatment factor. The designs have a nested blocking structure. In a block design, the experimental units are nested within the blocks, and a separate random assignment of units to treatments is made within each block. In a spilt plot design, the experimental units are called spilt-plots (or sub-plots), and are nested within whole plots (or main plots).

In split plot design, plot size and precision of measurement of effects are not the same for both factors, the assigning a particular factor to either the main plot or the sub-plot is extremely important.

The source of variation in this design is due to Replication, Main plot treatment (A), Main plot error (E1), Sub-plot treatment (B), Interaction (A x B), Sub-plot error (E2).

Factorial Designs:

The factorial experiment designs are particularly useful in experimental situations which require the examination of the effects of varying two or more factors. In a complete exploration of such situation it is not sufficient to vary one factor at a time, but that all combinations of the different factor levels must be examine in order to elucidate the effect of each factor and the possible ways in which each factor may be modified by the variation of the others. In the analysis of the

experimental results, the effect of each factor can be determine with the same accuracy as if only one factor had been varied at a time and the interaction effects between the factors can also be evaluated. For example consider three fertilizers (factors) N, P, K with two dosages (levels) forms a 2^3 –factorial experiment. The factorial experiments general conducted in RCBD layout.

The source of variation in the above example is due to blocks(or replicates), Main effects of N, P and K, Interaction effects of N × P, N × K, P × K, Interaction effect of N × P × K and Error.

Technology Adaptation Experiment analysis over sites (OR) Groups of Experiments:

Technology adaptation experiments are designed to estimate the range of adoptability of new production technologies, where adaptability of a technology at a given site is defined in terms of its superiority over other technologies compared simultaneously at that site. The primary objective of such a trial is to recommend one or more new practices that are an improvement upon or can be substituted for, the currently used farmers' practices (Gomez and Gomez, 1984). Thus, a technology adaptation experiment has three primary features.

- The primary objective is the identification of the range of adaptability of a technology. A particular technology is said to be adapted to a particular site if it is among the top performers at that site. Furthermore, its range of adaptability includes areas represented by the test sites in which the technology has shown superior performance.

- The primary basis for selecting test sites is representation of a geographical area. The specific sites for the technology adaptation experiments are purposely selected to represent the geographical area, or a range of environments, in which the range of adaptability of technology is to be identified. Such areas are not selected at random. In most cases, these test sites are research stations in different geographical area. However, when such research stations are not available, farmers' fields are sometimes used as test sites.

- The treatments consist mainly of promising technologies. Only those technologies that have shown excellent promise in at least one environment (e.g. selection from a preliminary evaluation

experiment) are tested. In addition, at least one of the treatments tested in usually a control, which represents either a no-technology treatment (such as no fertilizer application or no insect control) or a currently used technology(such as local variety).

Two common examples of technology adaptation experiments are crop variety trials or a series of fertilizer trials at different research station in a region or country. For the variety trials, a few promising newly developed varieties of a given crop are tested at several test sites and for several crop results of such trials are used as the primary basis for identifying the best varieties as well as the range of adaptability of each of these varieties. For fertilizer trails, on the other hand several fertilizer rates may be tested at different test sites and for several crop seasons – in order to identify groups of sites having similar fertilizer responses.

Because technology adaptation experiments are generally at a large number of sites, the size of each trial is usually small and its design simple. Thus the two most commonly used designs are randomized complete block and split-plot design.

Technology adaptation experiments at a series of sites generally have the same set of treatments and use the same experimental design, a situation that greatly simplifies the required analysis. Data from a series of experiments at several sites are generally analysed together at the end of each crop season to examine the treatment X site interaction effect and the average effects of the treatments over homogeneous sites. These effects are the primary basis for identifying the best performers, and their range of adaptability among the different technologies tested.

Analysis procedure: For variety trial in RCBD or analysis for groups of experiments following steps are to be followed.

Step 1: Construct an outline of combined analysis of variance over years or for places or environment, based on the basic design used. For example variety adoption trial at m sites (places) involving seven rice varieties tested in RCBD based on grain yield.

Step 2: Perform usual analysis for variance for the given data. The analysis can be performing m sites separately. This may be done either in SPSS, SAS or EXCEL software.

Step 3: We have 'm' error mean squares, where m is the number of sites and we have to test the homogeneity of variances. Now we have following two situations.

a. When m = 2 , in this situation we apply F-test for testing the homogeneity of variances by taking the ratio of errors divided by the degrees of freedom for two places. If the calculated value of F is greater than tabulated F value then the null hypothesis of homogeneity of variances is rejected and the data is heterogeneous in different years, otherwise it is homogeneous.

b. When m > 2, in this situation, we apply Bartlett's Chi-square test for testing the equality of error variance over sites. If the calculated value of Chi-square is greater than tabulated value at m-1 d.f, then the null hypothesis of homogeneity of variance is rejected and the data is heterogeneous in different sites, otherwise it is homogeneous.

Step 4: If error variances are not homogeneous, performing the combined analysis of weighted least square is required. The weight being the reciprocals of the root mean square error. The weighted analysis is carried out by defining a new variable as newres = res/root mean square. The transformation is similar to Aitken's transformation. This new variable is thus homogeneous and thus combined analysis of variance can performed on this new variable. If error variances are homogeneous then there is no need to transform the data.

The source of variation in the above example is due to site (L), Replication within site, Variety (V), V × L and Pooled error.

CONCLUSIONS

The role of Statistics in almost every branch of study is of paramount importance. It is absolutely essential that the data generated is appropriate, relevant and accurate. In order to maintain high standards of research and improve the quality of research, it is of supreme substance that the planning and designing the experiments and also using of sound and modern statistical methodologies in the collection and analysis of data and then in the interpretation of results. The preliminary conclusions can be drawn by using basic statistics and the magnitude of relation between two are more related variables can be obtained through the correlation and regression analysis. In agricultural experimentation, the real effect of the imposed factors can be obtained by the good experimental design duly following the basic principles of experimental designs. Good experimental designs are products of the technical knowledge of one's field an understanding of statistical techniques and skills in designing experiments. The ANOVA provides a wide range of approach to the analysis of data from designed experiments. The range of adoptability of new technologies can be estimated through the technology adaptation experiment analysis over sites. The additional time and expenditure could be minimized by a properly planned experiment and analysing the data with appropriate statistical tools and techniques by using advanced softwares viz., SAS, SPSS etc., that will produce valid results as efficiency as possible.

REFERENCES

Gomez, K.W. and Gomez A.A.1984. *Statistical procedures for agricultural research*, John Wiley & Sons.

Gupta, S.C., and Kapoor, V.K.1970. *Fundamental mathematical statistics*, SC Publicatons, New Delhi, India.

Kachigan Sam Kash.1986. *Statistical Analysis-An Interdisciplinary Introduction to Univariate and Multivariate Methods*, Radius Press, New York.

Nigam.A.K. and Gupta. V.K. 1979. *Handbook on analysis of agricultural experiments*. IASRI, New Delhi.

Sahu, P.K. 2007. *Agriculture and Applied Statistics Vol-I*, Kalyani Publishers, New Delhi, India.

Sarma, K.L.A.P, Ravindra Reddy, B and Pllaiah, T 2013. *Biostatistics*, Daya Publishing House, Astral International Pvt.Ltd., New Delhi.