A COMPARISON OF TRANSFORMATIONS USED IN THE ANALYSIS OF DATA FROM AGRICULTURAL EXPERIMENTS

BY

K. PRIYA MENON

THESIS

Submitted in partial fulfilment of the requirement for the degree of

Master of Science (Agricultural Statistics)

Faculty of Agriculture
Kerala Agricultural University

DEPARTMENT OF AGRICULTURAL STATISTICS
COLLEGE OF HORTICULTURE
VELLANIKKARA THRISSUR 680 654
KERALA INDIA

1997

DECLARATION

I hereby declare that the thesis entitled A **comparison of transformations used in the analysis of data from agricultural experiments"** is a bonafide record of research work done by me during the course of research and that the thesis has not previously formed the basis for the award to me of any degree diploma fellowship associateship or other similar title of any other University or society

Vellanikkara
5 5 1997

PRIYA MENON K

P V PRABHAKARAN
(Chairman Advisory Committee)
Professor and Head
Department of Agricutlural Statistics
College of Horticulture
Vellanikkara

CERTIFICATE

Certified that the thesis entitled **"A comparison of transformations used in the analysis of data from agricultural experiments"** is a record of research work done independently by **Mrs Priya Menon, K** under my guidance and supervision and that it has not previously formed the basis for the award of any degree diploma fellowship or associateship to her

Vellanikkara
5 5 1997

P V PRABHAKARAN
Major Advisor

CERTIFICATE

We the undersigned members of the Advisory Committee of **Mrs Priya Menon K** a candidate for the degree of **Master of Science in Agricultural Statistics** agree that the thesis entitled **A comparison of transformations used in the analysis of data from agricultural experiments** may be submitted by Mrs Priya Menon K in partial fulfilment of the requirement for the degree

Chairman

Prof P V Prabhakaran
Professor & Head
Dept of Agricultural Statistics
College of Horticulture
Vellanikkara

Members

1. **Dr V K Gopinathan Unnithan**
Associate Professor
Dept of Agricultural Statistics
College of Horticulture
Vellanikkara

2. **Dr N Krishnan Nair**
Associate Director
Regional Agricultural Research Station
Pattambi

5/5/97

3. **Dr Baby P Skaria**
Associate Professor
Aromatic & Medicinal Plant Research Station
Odakkali

EXTERNAL EXAMINER

ACKNOWLEDGEMENT

I express my deep sense of gratitude and indebtedness to ***Prof P V Prabhakaran*** *Professor and Head Department of Agricultural Statistics College of Horticulture Vellanikkara and Chairman of the advisory committee for suggesting the problem excellent guidance and overwhelming inspiration throughout the course of the investigation and preperation of the thesis*

My sincere thanks to ***Dr V K Gopinathan Unnithan*** *Associate Professor Department of Agricultural Statistics College of Horticulture Vellanikkara* ***Dr Baby P Skaria*** *Associate Professor (Entomology) Aromatic and Medicinal Plant Research Station Odakkali* ***Dr N Krishnan Nair*** *Associate Director Regional Agricultural Research Station Pattambi for the help and encour agement rendered by them as members of the Advisory Committee*

I would like to express my sincere thanks to the Associate Dean College of Horticulture Kerala Agricultural University Vellanikkara for providing facilities for the study I will be failing my duty if I donot express my extreme gratitude to the staff members of the Department of Agricultural Statistics for their co operation and sincere interest in my work

I extend my sincere thanks to Dr L Natarajan Associate Professor Department of Entomology Regional Agricultural Research Station Pattambi Dr C T Abraham Associate Professor and Smt Durgadevi Associate Professor AICRP on Weed Control College of Horticulture Vellanikkara for providing necessary facilities for collection of data

I am highly obliged to all my friends especially Liji R S and Manojkumar K for their assistance and kind co operation

I am grateful to the Kerala Agricultural University for the award of Junior Fellowship during the course of study

I have no words to express my thanks to my husband mother in laws brother and sister for their constant prayers sincere encouragement moral support and for bearing all the inconveniences caused all along this study

Thanks are also due to Sri K A Joy for typing the manuscript neatly

Above all I bow my head before the Almighty God who blessed me with health and confidence to undertake the work successfully

PRIYA MENON K

To my Husband

CONTENTS

LIST OF TABLES

LIST OF FIGURES

Introduction

INTRODUCTION

Research workers who are content to learn the recipes for carrying out an analysis of variance without attempting to grasp the underlying principles may be headed for serious trouble Certain assumption about the data are made while performing the analysis of variance If the data donot conform to these assumptions the conclusions drawn from such an analysis are not justifiable

The major assumptions on which analysis of variance is based are that the observations are independently and normally distributed with constant variance and expectations specified by a model linear in a set of parameters The cases of violations from the model assumptions are successfully handled either by applying appropriate non parametric techniques or by suitable data transformations Non parametric methods donot require stringent assumptions for their validity The price they have to pay for this advantage is loss in efficiency

Data transformations are often preferred by researchers to non parametric methods as they are more powerful and flexible In simple terms a transformation consists of a change of scale of the original variable in such a way as to enable the data to fulfill the basic requirements of the general linear model It is a powerful tool in developing parsimonious representations and interpretations of data The obvious situation for a transformation occurs when the dependent variable is not linearly related to the independent variable For instance treatment and environmental effects may be multiplicative Then a logarithmic transformation restores additivity Linearisation of the functional form is ofcourse not the only reason for using a transformation since the introduction of a transformation also may be necessary or desirable to normalise the error distribution or to achieve greater constancy of variance Another reason for transforming data is to make analysis simpler and logically sound than would otherwise be possible For example original data may require the use of a second order regression model while after transformation a first order model may be found to be perfectly adequate However as a number of workers have pointed out the various reasons for introducing a transformation are seldom at cross purposes and a transformation that rectify one anomaly may serve well to reduce the strain on the other assumptions made in analysing the data It was also noted that in many cases several assumptions fail simultaneously Thus in non normal distributions there is usually a correlation between mean and variance so that failure of the normalising assumption is likely to be accompanied by a failure of the homoscedasticity assumption

Most of the earlier works on the inadequacy of the general linear model have been concentrated with a single aspect namely non normality and the validity of statisti cal tests of significance Fortunately consensus from many past investigations is that no serious error is introduced by non normality in the validity of test of significance in analysis of variance The problem of mutually correlated errors is largely taken care of by randomisation whatever may be the nature of the correlation system Lack of additivity results in loss of efficiency and the problem therefore lies in finding a metric in which the effects are additive

Instability of variance is considered to be the most serious of the model violations Heterogeneity of error may affect certain treatments or certain part of the data to an unpredictable extent It may cause loss of efficiency and loss of sensitivity in tests of significance There is no theoretical difficulty in extending the ordinary analysis of variance in such a way as to account for the variations in error structure The usual analysis has to be replaced by a weighted analysis in which each observation is weighted in proportion to the inverse of its error variance However such a procedure requires information on the relative variance of different observations and this is seldom available in practice It is also useful to subdivide the error variance into homogeneous components before applying any statistical tests of significance Unfortunatly this is not a general rule and in complex analyses such a procedure is not at all practiable Thus the only effective approach is to evolve suitable transformations to stabilise the error variance as closely as possible

In addition to this there is a common type of heterogeneity that is more regular In this type which usually arises with non normality variance of a set of observations changes with mean value irrespective of the treatment or block concerned In such a situation a change of scale is the only option for the stabilisation of variance

The selection of a scale of measurement will depend upon

1) the nature of the data

and

2) the type of statistical procedure to be used

The above two conditions are not incompatible since the scale of measurement may be purely arbitrary for certain data

A transformation of x to some function f(x) should be made considering the two conditions set out above If the principles of (i) are not violated and if the purpose for which the transformation is made is realized the new function should not cause any confusion

Bartlett (1947) lists the following requirements for an ideal transformation

1) The variance of the transformed variate should be unaffected by changes in the mean

2) The transformed variate should be normally distributed

3) The transformed scale should be one for which an arithmetic average from the sample is an efficient estimate of the true mean

4) The transformed scale should be one for which real effects are linear and additive

The above conditions are related to some extent since conditions (1) (2) and (4) usually imply (3) However a trasnformation selected to satisfy (1) may not satisfy the remaining conditions and transformations satisfying (2) may not fulfill requirement (1) The nature of the data and type of statistical analysis used govern the importance of the above requirements

Since a variety of transformations is available for the analysis of the same set of data, it is necessary to evaluate the relative performance of the selected transformation based on one or more selected criteria Although we may restrict ourselves to relatively common transformations the choice of a transformation by pure trial and error procedure is both costly and time consuming The difficulty is com pounded by the location parameter of the transformation It is not unusual to find that log x provides no accuracy whatever but log (x+c) works quite well for the proper selection of c Further the type of a transformation is also dictated by the nature of the data In the case of frequency counts square root transformation or logarithmic transformation are commonly used But for other types of data they neednot be efficient It is also possible to restrict transformations solely on the dependent variable rather than transforming all of the variables of the linear model The present study based on enumerative data is restricted to the case of transformations on the dependent variable alone

The suitability of a transformation can be assessed either in terms of a single criterion viz stabilisation of variance or in terms of several criteria In the single criterion approach it is assumed that the transformation satisfying the relevant condition also satisfies the other requirements of ANOVA

In the case of variance stabilising transformations the usual method is to determine empirically or theoretically a relation between variance and mean and then use this relation to develop an appropriate transformation An adequate empirical relationship may often be found by plotting log of the within cell variance against log of the cell mean Another method is to choose a transformation within a restricted family to minimise some measure of heterogeneity of variance such as Bartlett s criteria or F max test Levene (1960) has suggested a test of equality of variance based on analysis of residual which is preferable to Bartlett s test as it is more robust Minimisation of the F ratio of the residual ANOVA of the test provides a better alternative for choosing the best transformation Similarly the possible choice of a transformation restoring additivity is achieved by the minimisation of the F value for one degree of freedom for non additivity (Tukey 1949) or by the maximisation of the F ratio for interaction versus error or by maximisation of F ratio for treatment versus error (Tukey 1950) No constructive method of evolving transformations to produce normality is available in the literature However a transformation which stabilizes variance is expected to make the distribution of errors approximately normal

Although the suitability of a chosen transformation depends upon the actual distribution the usual practice is to compare only the mean and variance and choose the proper transformation because the testing of goodness of fit of a theoretical distribution is rather inconvenient or seems to be impossible due to non availability of sufficient number of cases If the sample variance tends to change with sample mean a number of approaches are available If the functional form of the relationship between variance and mean is known the type of transformation that stabilizes variance can be derived mathematically provided the expression is integrable Thus if the sample variance tends to be proportional to sample mean a square root transformation is indicated Similarly a quadratic relationship between variance and mean reveals the utility of the inverse hyperbolic sine square root transformation However the possibility of developing efficient transformations for the analysis of data exhibiting intricate non linear relations between mean and variance are yet to be explored

A simple procedure to test the hypothesis of variance mean relationship on the data and to choose a correct power transformation is by using Taylor s power law (Taylor 1960) The law states that for most field distribution of organisms the variance mean relationship is of the form $\sigma^2 - a\mu^b$ where σ^2 is the variance μ is the mean and a and b are two constants to be estimated The choice of a suitable transformation depend upon the estimated value of b Thus if b is 1 a squareroot transformation is indicated

Modern computers are extremely fast and can provide graphical output Hence instead of handling the problem of scale conversion in terms of a single numerical criterion say stablisation of variance it will be better to consider several criteria simultaneously The restrictions then can often be easily understood by the experimenter and compromise decisions could be made With this object in mind Draper and Hunter (1969) have suggested a simple comprehensive method of selecting suitable transformations from plot of functions which occur naturally in the usual analysis However when several such plots are made they may not all indicate the same transformation and thus different experimenters arrive at different conclusions

It is always desirable to have a simple criterion of transformation which in itself carries all the major requirements of ANOVA Box and Cox (1964) considered the choice of a transformation among a parametric family of data transformations to yield a simple normal linear model To test the hypothesis about the parameters of the

7

transformation they used the asymptotic distribution of likelihood ratio The transformations studied in greater detail by Box and Cox are the family of modified power transformations given by

$$y^{(\lambda)} = \begin{cases} \dfrac{y^{\lambda} - 1}{\lambda} & (\lambda \neq 0) \\ \log y & (\lambda = 0) \end{cases}$$

where $y^{(\lambda)}$ is a nx1 vector of transformed observation and the additive errors are independently and normally distributed with constant variance σ_y^2 The method consists in repeated computation using a number of trial values for λ The best transformation is selected by a plot of λ against the curve of maximised likelihood

Certain problems are encountered in transforming enumerative data including zero values or in the case of data showing both positive and negative values A usual practice is to add a constant c to each datum before applying the relevant transformation Box and Cox (1964) have suggested a method of selecting c based on the likelihood ratio criterion But the validity of the approach has been questioned by several workers especially when there are several outliers

A better procedure of selecting c is based on the analysis of residuals Berry (1987) has suggested an ingenious method for the choice of c which has certain distinct advantages over others An obvious alternative is to use rank ANOVA which combines in itself both parametric and non parametric procedures It would be also helpful to evaluate the relative efficiency of rank ANOVA over ordinary ANOVA and transformed ANOVA with a proper choice of the additive constant

Among the different types of data encountered in agricultural field experiments those on counts of insects and weeds invariably exhibit large amount of variation It is well known that the variance of number of insects/weeds on sub areas is related to the mean number of insects/weeds per sub area Several workers have shown that the number of insects/weeds found per plot varies in such a way that one can not strictly subject the results for the analysis of variance and it is proposed to find how the data may be transformed so that analysis of variance become applicable

Considering all the above aspects the main objectives of the study are

1) To examine the applicability of the various commonly used transformation techniques to the analysis of data on counts of insects and weeds
2) To make empirical comparisons among the various commonly used transformations with a view to select suitable transformations for the analysis of frequency data on various types of insects and weeds
3) To explore the possibility of developing new transformations for data analysis when the commonly used transformations fail to yield encouraging results
4) To find the applicability of simple ordinal procedures in the analysis of enumerative data and to assess their relative efficiency over the usual parametric procedures

Review of Literature

2 REVIEW OF LITERATURE

Transformation of data is an important aspect of statistical analysis Several studies both theoretical and applied have been undertaken on this subject A brief review of the work done on transformations of data is presented below under two major heads

1) General and theoretical studies on transformation of data and related aspects
2) Specific empirical studies on transformations of counts of insects and weeds

2 1 General and theoretical studies on transformation of data and related aspects

Bartlett(1936) suggested square root transformation for a poisson variate because the variance of the transformed variable was found to be asymptotically constant at least for value $m > c_0$ where m is the mean of the poisson variate

Wilcoxon (1946) has presented a non parametric test for a two or more way classification involving ranks of treatments in one of the classifications The computational procedure was simple and normality assumption was not required

Bartlett (1947) showed how empirically by the use of transformations some of the consequences of violations of assumptions could be avoided and valid conclusions reached by analysis of variance when data in the original form were essentially interactable by analysis of variance

Cochran (1947) discussed about the consequence to be expected when some of the assumptions of ANOVA were not satisfied According to him the assumption underlying analysis of variance are

(1) treatment effects and environmental effects must be additive

(2) the experimental error should be independently and normally distributed

(3) the experimental error should have a constant variance

He opined that no serious error would be introduced by non normality in tests of significance But it would cause a loss of efficiency in the estimation of treatment effects

Tukey (1949) presented a method for isolating a single degree of freedom associate with non additivity in a two way classification This degree of freedom was isolated from error degree of freedom and the residual was compared with error mean square in order to test the hypothesis of additivity of the data According to him non additivity would arise either due to the presence of row column interaction or due to the presence of one or more discripant observations When there was no adequate reason for discripancy he suggested transformation to data before proceeding to further analysis

Box (1953) showed that Bartlett s test for homogeneity of variance was almost as sensitive for testing non normality as for testing heterogeneity of variance Thus according to him the original data should be normally distributed or nearly so before computing sample variance

Tukey (1957) has shown that the entire problem of transformations could be reduced to one of the non linear curve fitting He proved that the entire family of transformations could be defined in terms of two or more parameters According to him a simple family of transformation could be expressed as $T(x) - (c+x)^{I}$ where P is a real number and c is a constant such that c+x is greater than zero He studied the distributional properties of the family and constructed a graph of c and P so that trial values could be represented in a reasonable fashion

Dolbi (1963) found that the simple family of transformations (Tukey 1957) could be characterised by a set of solutions to a third order differential equation From this a differential equation was derived which was used to estimate the parameter P

The work of Box and Cox (1964) is considered to be the greatest contribution towards the study of transformations of data These authors systematised the search for the best transformation They analysed the use of a parametric family of data transformations for obtaining a simple normal linear model Inferences about the transformations and about the parameters of the linear model were drawn from the likelihood function and the relevant posterior distribution The contributions of normality homoscedasticity and additivity to the transformations were seperated

Kruskal (1965) considered the family of all monotonic transformations z – f(y) and determined the best from the set by optimising a squared residual criterion of fit to the assumed linear model

Fraser (1967) proposed a comprehensive statistical model as a revision of the structural model He derived a different likelihood function which yielded quite different linear inferences from those of Box and Cox (1964) in extreme cases when the number of parameters was almost close to the number of observations

Draper and Hunter (1969) re examined some of the published examples on transformation as a part of their attempt in evaluating the Box Cox approach They illustrated how the selection of a transformation could be aided by plots of functions which occured naturally in the usual analysis of variance They opined that methods that provide a single criterion for considering several aspects of data violation, while theoretically appealing might not be appropriate on actual experimental situation

Andrews (1971) proposed an exact test for the value of the parameter in the Box and Cox parametnc family of transformations Confidence sets were denved from this test and used to predict the sharpness of the inference

According to Atkinson (1973) Andrew s exact test for the value of the parameters in the Box and Cox parametnc family of transformations when compared with two tests denved from the likelihood function, the two tests are shown to be uniformly more powerful than the exact test

Manly (1976) suggested an exponential transformation as a viable alternative to the Box and Cox (1964) one parameter family of transformations The new transformation had an added advantage of allowing both negative and positive data values Transformations proposed by him are defined by

$$y = \frac{e^{\gamma x} \quad 1}{\gamma}$$ and where γ is a constant and x is the vanable

In practise γ can be chosen either by setting the R H S of

$$\frac{dL}{d\gamma} \quad \Sigma x \quad \frac{1}{\gamma\sigma^2} \Sigma (y \quad \mu) \{x (\gamma y + 1) \quad y\}$$

to zero or by plotting log (m_2) against x and determining γ graphically where m_2 is the residual vanance for treatment combination

Carroll (1980) proposed a competitor to the likelihood and significance methods for power transformations to achieve approximate normality in a linear model The new method was shown in theory and a Monte Carlo expenment was designed to produce more robust inference than the likelihood method of Box and Cox

Konishi (1981) described a general procedure for finding normalising transformations to various statistics in multivariate analysis It was shown that Fisher s z transformation for a sample correlation coefficient in a normal sample and Wilson and Hilferty s approximation for a chi squared variate could be derived by the same line of approach

Berry (1987) developed a method for choosing an additive constant c when transforming data x to y log (x+c) The method preserved type 1 error and power in ANOVA under the assumption that x+c for some c was log normally distributed The method had distinct advantages over other transformations and was similar to rank transformations as it was easy to use and moderately robust The method preserved significance levels and was quite powerful

2 2 Specific empirical studies on transformations of data on counts of insects or weeds

Beal (1942) from the experimental results of seven field experiments for the control of insects found out that by the transformation $x^1 - k^{\prime} \text{Sinh}^{1} \sqrt{kx}$ where k is a constant and x an observation the data could be put in a form for which standard deviation approached a constant independent of the mean value The results of analysis of transformed data were markedly different from those obtained from the untransformed data

Anscomb (1948) showed that for data based on negative binomial distribution the inverse hyperbolic sine square root transformation was the most ideal

Taylor (1961) found that for most distribution of organism the variance mean relationship was of the form $\sigma^2 \quad a\mu^b$ were σ^2 is the variance μ is the mean and a and b are two constants to be estimated If $b\neq0$ the non zero b indicates the appropriate transformation to be used for the data If b=2 a logarithmic transformation is indicated

Taylor (1970) and Hayman and Lowe (1961) suggested logarithmic transformation for certain species of aphid on the basis of mean variance relationship They prefered log (x+1) transformation over log x due to the presence of zero counts

Williams and Stephenson (1973) found a cube root transformation z $x^{1/3}$ to be the most useful in analysing data on counts of marine organisms

By following the Tukey s test of additivity Patil and Patil (1983) adopted $\sqrt{x+1}$ and log(x+2) transformations to the data on weed count and weed dry weight respectively

According to Misra *et al* (1984) the data on the number of plants damaged by cut worms could be effectively analysed by using square root ($\sqrt{x+1}$) transformation

Sharma *et al* (1985) applied $\sqrt{n+1}$ transformation in case of the infestation of boll worms and the transformed data were subjected to analysis of variance

Huckaba *et al* (1988) used $(x+0\ 5)^{0\ 5}$ transformation before analysis of data on adult and larval soybean thrips

Kishorekumar and Agarwal (1990) transformed the data on counts of Jassid nymphs to squareroot $\sqrt{x+0\ 5}$ before analysis

Zaman (1990) converted data obtained in numbers (density length weight) into log x/log (x+1 5) for analysis of variance

Observations on the population count of thrips were transformed to logarithmic scale by Bagle (1993) before subjecting to statistical significance tests

From the analysis of data generated from two experiments on mustard crop Singh and Rai (1993) found that counts of aphids followed the log normal distribution and a logarithmic transformation of the counts normalised the data except in those cases where the variability was small and a large number of zero counts were observed They used Kolmogorov Smirnov test to test the goodness of fit of the theoretical distribution They reported that log(x+1) transformation was the most appropriate transformation for counts of mustard aphid

Pushpalatha and Veeresh (1995) applied seven transformations viz $\sqrt{x+1}$ log(x+1) log(x+k) log(x+k/2) log(x+2) and

$\text{Sinh}^{1} \frac{\beta_{1} x}{\alpha\text{-}1}$ for the analysis of data on population count

of opisina arenosella They attempted an emperical comparison between the different transformations with regard to their relative efficiency in equalising the variance According to them the inverse hyperbolic sine squareroot transformation

$\text{Sinh}^{1} \frac{\beta_{1} x}{\alpha\text{-}1}$ (α β are constants to be estimated) was the most effective in stabilising variability

3 MATERIALS AND METHODS

A brief account of the materials and methods used in the present study is given below under the following major heads

1 Method of collecting data
2 Methods of analysis of data

3 1 Method of collecting data

Data of two different types were utilised for the present study

1) Secondary data from the pest surveillance project on paddy
2) Experimental data from the various plant protection and weed control trials

Data from pest surveillance studies consisted of observations on daily light trap catches of six different species of insects viz stem borer jassid gall fly leaf folder BPH and case worm gathered from the Regional Agricultural Research Station Pattambi Altogether data were available for a period of 22 years from 1970 to 1991 For simplicity and ease of analysis observations recorded on the 15th day of each month of the year alone were utilised in the present study for analysis and interpretation

The experimental data for the study were gathered from the available records of Regional Agricultural Research Station, Pattambi and the All India Co ordinated Research Project on Weed Control (AICRP) Vellanikkara As a whole three sets of data at different time period 20 days after transplanting (DAT) 30 DAT and 40 DAT relating to three different insects gall fly whorl maggot stem borer were available Observations were recorded from each plot on the number of silver shoot (SS) number of whorl maggot (WM) and number of dead heart (DH) at different time period Counts of number of silver shoot per plot indirectly indicated the severety of the attack of gall fly while those of dead heart indirectly showed the intensity of infestation of stem borer

The relevant details of the data collected on insect counts are as follows

Name of experiment	Trial on early stage pest control
Period of observation	1989 91
Design	Randomised Block Design (RBD)
Variety	Jaya
Season	Kharif
No of replication	4
No of treatments	8

Description of treatments

Treatment	Dose	Time and method of application
1 Furadon 3 G	2 kg/ai/hectare of nursery	Broadcast 5 days before pulling
2 Ekalux 5 G	do	do
3 Padan 4 G	do	do
4 Coroban 20 EC	1 5 kg/ai/hectare of nursery	Spray one day before pulling
5 Nuvacron 36 EC	do	do
6 Coroban 20 EC	0 05%	Whole seedling dip for 1 2 mts
7 Coroban 20 EC	0 02%	Seedling root dip for 12 hrs
8 Untreated control		

Secondary data on weed population were collected from the results of the post emergence herbicidal evaluation trial for *Pennisetum pedicellatum* The experiment was continued for a period of three years In each year data on number of surviving hills/m^2 were gathered from each plot at three time periods immediately after spraying the chemicals (or water) The three time periods were spraying at one month after sowing two months after sowing and three months after sowing Thus there were altogether 9 sets of data as detailed below

Serial no of data set	Year	Order of spray	Symbol
1	1987 88	Ist spray	$Y\ S_1$
2	1987 88	2nd spray	Y_1S_2
3	1987 88	3rd spray	Y_1S_3
4	1988 89	Ist spray	Y_2S_1
5	1988 89	2nd spray	Y_2S_2
6	1988 89	3rd spray	Y_2S_3
7	1989 90	Ist spray	Y_3S_1
8	1989 90	2nd spray	Y_3S_2
9	1989 90	3rd spray	Y_3S_3

The treatment details and other relevant information of the weed control trial are given below

Name of the experiment	Evaluation of post emergence herbicides for controlling *Pennisetum pedicellatum*
Period of observation	1987 90

Design RBD

No of treatments 13

No of replication 3

Descriptions of treatments

T_1	paraquat 0 4	T_7	glyphosate 0 7
T_2	paraquat 0 8	T_8	glyphosate 0 8
T_3	paraquat 1 2	T_9	glyphosate 1 2
T_4	Dalapon 2	T_{10}	paraquat + Dimor 0 4+1
T_5	Dalapon 4	T_1	paraquat + Dimor 0 4+2
T_6	Dalapon 6	T_2	paraquat + Dimor 0 8+1
		T_{13}	Control (water spray)

3 2 Methods of analysis of data

The various statistical methods used in the present study are outlined below

3 2 1 Empirical comparisons among different transformations

Comparisons among different transformations were made either based on a single criterion or several criteria simultaneously In the former approach the different transformations were evaluated for their relative efficiency m maintaining homoscedasticity or in restoring additivity Comparison of transformations were also effected in accordance with the Taylor s power law which invariably indicated the best transformation for a given set of data If the relation between variance and mean was parabolic inverse hyperbolic sine squareroot transformation could be considered to be a proper choice In the multiple criteria approach the prime objective was to choose a transformation that yielded to the maximum extent approximate normality additivity and homoscedasticity conditions of the linear model Box and Cox (1964) proposed a likelihood function approach for this purpose It would be possible to select the best power transformation as per the methods suggested by them

Draper and Hunter (1969) suggested a comprehensive graphical method for selecting the best transformation for a given set of data considering several single aspect criteria simultaneously The method is rather simple and useful to examine the adaptability of the likelihood approach

3 2 1 1 Comparison of transformations based on a single aspect

The two major violations of assumption of analysis of variance are (1) non additivity (2) heteroscedasticity Normality assumption usually goes hand in hand with homoscedasticity assumption

A comparison of the different transformations on the basis of the above criteria could be done in accordance with the relative degree of conformity of the transformed data under each scale to the underlying assumptions As far as stabilisation of variance was concerned the following two single aspect selection criteria were used to choose the best transformation (1) Bartlett s χ^2 test (2) Levene s F test of the residual ANOVA

The transformation that gave a minimum value for each of the above criteria was considered to be the most ideal

In the case of additivity assumption, Tukey s test of non additivity was used as the selection criterion The method consisted in calculating non additivity sum of squares with one degree of freedom and using the F statistic for the diagonostic test The best transformation should yield a minimum value for the non additive F Another possibility was to use treatment Vs error F statistic as a basis of comparison and choosing the transformation giving the highest value for F

3 2 1 1a Bartlett s chi square test

Let K independent samples of residuals $e_{ij} - Y_j$ Y (i − 1 2 k j = 1 2 n) be selected the i^{th} sample be of size n + 1 and S^2 be its variance (i − 1 2 k) Let σ_i^2 be the population variance of the ith population To test the null hypothesis H_0 $\sigma_1^2 = \sigma_2^2 -$ $- \sigma_k^2$ we use Bartlett s test based on the criterion

$$\chi^2 - (n \log_e \sum_{i=1}^{k} \frac{n S^2}{n} \quad \sum_{i\,1}^{k} n \log_e S^2) \rightarrow (3\ 1) \text{ where}$$

$$1 + \frac{1}{3(k\ 1)} \sum_{i-1}^{k} \frac{1}{n} \quad \frac{1}{n}$$

$$n - \sum_{I-1}^{k} n$$

The χ^2 given in (3 1) is distributed as a χ^2 variable with k 1 degree of freedom Let $\chi^2_{m\,\alpha})$ be the critical value of χ^2 value such that $Pr(\chi^2_m > \chi^2_{m\,\alpha}) - \alpha$ where χ^2_m is the χ^2 variable with m degree of freedom If the calculated χ^2 value as given in (3.1) is greater than $\chi^2_{k\,1\,\alpha}$ we reject the null hypothesis H_0 $\sigma_1^2 - \sigma_2^2 -$ σ_k^2 in favour of the alternative hypothesis that not all variances are equal at α level of significance otherwise not

3 2 1 1b Levene s residual F test

Levene (1960) suggested a test for equality of variances of several equalised groups of observations and showed through sampling studies that the test possessed almost unbelievable robustness against departures from normality of the underlying distribution of observations Levene s test is preferable to Bartlett s test which is greatly affected by departures from normality (Box, 1953) Levene also mentioned the possibility of using similar analysis of variance on the absolute value of residuals from other regressions in order to study the variance of the residuals In the present study the residuals e_{ij} were calculated where e_{ij} Y_j $\bar{Y}$ in case of no blocking and $e_{ij} - Y_j$ $\bar{Y}_i$ $Y_j +$ Y when there is blocking Y_{ij} s are the observations $\bar{Y}_i$ and Y_j are the treatment mean and block mean and $\bar{Y}$ is the grand mean

Suppose we have P groups of residuals e_{ij} as follows

Group 1 e_{11} e_{12} e_{1n_1} average e_1 $V(e_1) - \sigma_1^2$

Group 2 e_{21} e_{22} e_{2n_2} average $\bar{e}_2$ $V(e_2) - \sigma_2^2$

Group p e_p e_{p2} e_{pn_p} average $\bar{e}_p$ $V(e_p) - \sigma_p^2$

Costruct from these observations

$$Z \quad | e_{ij} \quad e \, | \quad j - 1 \; 2 \quad n$$

$$i = 1 \; 2 \quad p$$

Perform the standard analysis of variance on Z_j as follows

ANOVA of residuals

Source	df	SS	MS	F
Between groups	$p-1$	$\sum_{i=1}^{p} \frac{Z^2}{n} - \frac{G^2}{\Sigma n}$	S_1^2	$F = S_1^2/S^2$
Within groups	$\sum_{i=1}^{p}(n-1)$	$\sum_{i}^{p}\sum_{j}^{n} Z_{ij}^2 - \sum_{i=1}^{p}\frac{Z^2}{n}$	S^2	
Total	$\sum_{i}^{p} n-1$	$\sum_{i}^{p}\sum_{j}^{n} Z_{ij}^2 - \frac{G^2}{\Sigma n}$		

If all the treatments are replicated equal number of times say r n r and $\Sigma r = N = rt$

If $F_r > F\,[\{(p-1),\ \Sigma(n-1)\}\ (1-\alpha)]$ we say that it is significant and there is evidence that difference exist between $\sigma_1^2\ \sigma_2^2 \quad \sigma p^2$ If F is not significant do not reject the null hypothesis $\sigma_1^2 = \sigma_2^2 = \quad = \sigma p^2$

3 2 1 1c Tukey s test of non additivity

In a two way classification model Tukey s test of non additivity is used to decide if row and column effects are additive or not The rationality of the test can be indicated by means of calculus In a two way classification, if effects are exactly additive in the scale of Y we have

$$Y_{ij} = Y + (Y_i - Y) + (\bar{Y}_j - Y)$$

$$= \bar{Y}\ [1+\{(Y - \bar{Y}) + (\bar{Y}_j - Y)\}/\ Y]$$

$$= Y\ [1 + \alpha_i + \beta_j]$$

Now let $X_{ij} - Y_{ij}^{1p}$ then

$$X - \bar{Y}^{1p} [1+\alpha + \beta_j]^{p}$$

After using Taylor s expansion and suitable substitutions it can be shown that the first non additive term in the expression would be

$$\frac{(1\ P)}{X} (X \quad \bar{X})(X_j \quad X)$$

This indicates that the residual has a linear regression on the variate

$$(X_i \quad X)(\bar{X}_j \quad X)$$

If X (i = 1 2 t j - 1 2 r) denotes the observations of the two way classification this regression coefficient of the residual (X_i $\bar{X}$) ($\bar{X}_j$ X) can be estimated as

$$B - \frac{\Sigma w \alpha}{D} - \frac{\sum_{i=1}^{t} \sum_{j=1}^{r} X_{ij} \alpha_i \beta_j}{D} \quad \text{where}$$

$$D = (\sum_{i=1}^{t} \alpha^2)(\sum_{j}^{r} \beta_j^2)$$

According to Snedecor and Cochran (1967) the contribution of non additivity to error sum of square with one degree of freedom is given by

$$\frac{N^2}{D} - \frac{(\sum_{1}^{t} w\,\alpha)^2}{(\sum^{t} \alpha^2)(\sum_{j}^{r} \beta_j^2)}$$

$$\frac{\left(\sum_{i}^{t}\sum_{j}^{r} X_{ij}\,\alpha_i\,\beta_j\right)^2}{(\Sigma\,\alpha_i^2)\,(\Sigma\,\beta_j^2)}$$

This is tested using F test against remainder mean square The relevant analysis of variance table is given below

ANOVA table

Source	df	SS	MS	F
Total	tr 1	ΣX^2 CF		
A (Blocks)	(t 1)	$\frac{\Sigma A^2}{r}$ CF		
B (treatments)	(r 1)	$\frac{\Sigma B^2}{t}$ CF		
Error	(r 1)(t 1)	Subtract		
lack of additivity	1	$\frac{N^2}{D}$	MSLA	$\frac{MSLA}{MSRE}$
Remainder error	(r 1)(t 1)	$\frac{N^2}{D}$ error SS	MSRE	

3 2 1 1d Taylor s power law

This approach consists in fitting a model to decide whether a transformation is necessary and if it is so which transformation is appropriate binomial distribution b value in Taylor s power law will be close to two If it is close to one the underlying distribution is poisson

3 2 1 1e Inverse hyperbolic sine squareroot transformation

Beal (1942) suggested that if standard deviation varied with mean a transformation of the form $x^1 - k^{/} \; Sin \, h^1 \sqrt{kx}$ where k is a constant and x an observation could be helpful in making standard deviation independent of the mean This was the case with certain types of data where the variance mean relationship would assume a quadratic form In the derivation of the above transformation Beal postulates the variance mean relationship as $\sigma^2 - \mu + k\mu^2 \rightarrow$ (3 2) were σ^2 is the population variance μ the population mean, k is a constant He assumed the charlier coefficient of disturbance for the value of k,

$$k = \frac{\sigma^2 \quad \mu}{\mu^2} \rightarrow (3\ 3)$$

An estimate of k proposed by Beal (1942) is given by

$$k = \frac{\Sigma S^2 \quad \Sigma \quad x}{\Sigma \ x^2}$$

where Σ represents the summation over all pairs S^2 the sample variance and x the sample mean

The estimate of Beal did not posses ~~and~~ any statistical property apart from its intuitive appeal Hence an attempt was made to get an estimate purely based on statistical theory For this the familiar least square technique was employed The details are as follows

$$\Sigma\,(\sigma^2 \quad \mu \quad k\,\mu^2)\,\mu^2 - 0$$

$$\Sigma\,(\sigma^2\mu^2 \quad \mu^3 \quad k\mu^4) - 0$$

$$k\,\Sigma\,\mu^4 - \Sigma\,\sigma^2\mu^2 \quad \Sigma\,\mu^3$$

$$k - \frac{\Sigma\,\mu^2\sigma^2 \quad \Sigma\,\mu^3}{\Sigma\mu^4}$$

The transformations were also effected with the new estimate of k and the analysis was carried out in the usual way The relative superiority of the new estimate over the earlier one was assessed on the basis of the empirical results

3 2 1 1f Development of a new transformation

In certain types of experimental data the relationship between mean and standard deviation could be expressed in the form of a non linear function given by

$$\sigma = \mu + \frac{k}{\mu} \quad \rightarrow (3\ 4)$$

Such type of data are frequently encountered in entomological experiments when changes in variance is not directly proportional to the mean value The intrinsic growth rate gradually declines as the mean value increases It is possible to derive a suitable transformation for such type of data as follows

Following Bartlett (1947) suppose we write

$$\sigma x^2 - f(\mu) \rightarrow (3\ 5)$$

where σx^2 is the variance of the original scale of measurements x with mean of x equal to μ Then for any function g(x) we have approximately

$$\sigma g^2 - \left(\frac{dg}{d\mu} \right)^2 f(\mu) \rightarrow (3\ 6)$$

For a constant variance we require $\sigma g^2 - c^2$

If σg^2 is to be a constant c^2 say we must have

$$g(\mu) - \int \frac{c d\mu}{[f(\mu)]'} \rightarrow (3\ 7)$$

From 3 4 and 3 5

$$g(\mu) = \int_0^x \frac{c d\mu}{\mu + k/\mu}$$

$$- c \int_0^x \frac{\mu\, d\mu}{\mu^2 + k}$$

put $\mu^2 + k - v$ when $\mu - 0$ $v - k$

$2\mu d\mu - dv$ when $\mu = x$ v $x^2 + k$

$$\mu d\mu = \frac{dv}{2}$$

$$\frac{c}{2} \int_{k}^{x^2+k} \frac{dv}{v}$$

$$- \left[\log v\right]_{k}^{x^2+k}$$

$$- \log(x^2 + k) \quad \log k$$

k can be estimated as follows

From (3 4) we have

$$\sigma - \mu + k/\mu$$

Let $F - \Sigma \{\sigma^2 \quad \mu \quad k/\mu\}^2$

$$\frac{\partial F}{\partial k} = 0 \Rightarrow$$

$$2 \Sigma (\sigma \quad \mu \quad k/\mu) \times 1/\mu \quad 0$$

$$\Sigma (\sigma \quad \mu \quad k/\mu) \times 1/\mu - 0$$

$$\Sigma \quad \sigma/\mu \quad n \quad \Sigma k/\mu^2 - 0$$

$$k - \frac{\Sigma \ \sigma/\mu \quad n}{\Sigma (1/\mu^2)} \rightarrow 3\ 8$$

3 2 1 2 Companson of transformation based on several aspects

There are two different procedures for the selection of a suitable transformation considenng simultaneously the several aspects of violation of assumption They are

(1) likelihood method of Box and Cox (1964)

(2) the graphical method proposed by Draper and Hunter (1969)

3 2 1 2a Likelihood method of Box and Cox (1964)

The data for each time penod for each of the insects was analysed seperately in the usual way as in a randomised block design The method of analysis has been denved from the following model

$$Y_{ij} - \mu + \alpha_i + \beta_j + e_{ij}$$

where Y_{ij} is the o bservation of the i^{th} treatment (i – 1 2 t) in the j^{th} block (j – 1 2 r) α_i is the effect due to the ith treatment β_j is the effect due to the j^{th} block and e_{ij} is the random error component which is assumed to be independently and normally distnbuted with zero mean and constant vanance σ^2 The structure of the analysis of vanance of Randamised Block Design with t treatments and r replications is given below

ANOVA

Source	df	MS	F
Replication	r 1	S^2	
Treatment	t 1	S_t^2	$F_t - S_t^2/S^2$
Error	(r 1)(t 1)	S^2	
Total	rt 1		

The observations Y were transformed using the family of power transformations $Y^{(\lambda)}$ defined by

$$Y^{(\lambda)} - \begin{cases} \dfrac{Y^{\lambda} \quad 1}{\lambda} & \lambda \neq 0 \\ \log Y & \lambda - 0 \end{cases} \rightarrow (3\ 9)$$

where (λ) is the vector of parameters Since the analysis of variance is unchanged by a linear transformation it is equivalent to

$$Y^{(\lambda)} = \begin{cases} Y^{\lambda} & \lambda \neq 0 \\ \log Y & \lambda - 0 \end{cases} \rightarrow (3\ 10)$$

Data could be transformed using the power transformation described above giving trial values for λ in the range 1 to +1 at equal intervals of length 0 2 The analysis of variance table could be formed as usal and error mean square $S^{(\lambda)}$ estimated

The maximised log likelihood $L_{max}{}^{(\lambda)}$ is given by (Box and Cox 1964)

$L_{max}{}^{(\lambda)}$ – ½n log σ^2 (λ z) where σ^2 (λ z) is the residual variance of the normalised transformation z Estimate of σ^2 (z λ) is given by

$$\sigma^2 (\lambda\ z) = \frac{S (\lambda\ z)}{n} \rightarrow (3\ 11)$$

$S(\lambda\ z)$ is the residual mean square in the analysis of variance of $Y^{(\lambda)}$ which is calculated as

$$S(\lambda\ z) - \frac{S(\lambda)}{\lambda^2\ (\dot{Y}\lambda\ 1)^2} \rightarrow (3\ 12) \quad \text{where}$$

$\dot{Y}$ is the geometric mean of the observations and n is the total number of observations Box and Cox (1964) also proposed an alternative approach to the same problem by using the posterior distribution of the parameter λ By maximixing the log of the contribution of the posterior distribution of λ another statistic $L_b^{(\lambda)}$ was obtained which is given by

$$L_b^{(\lambda)} = \frac{\frac{1}{2}\ V \log \{ S(\lambda\ z)\}}{V} \rightarrow (3\ 13)$$

where V is the degree of freedom for error The two expression differs only on the substitution of V for n Therefore both of them yield essentially the same transformation in most cases Values of $L_{max}^{(\lambda)}$ could be determined for varying values of λ and presented graphically According to Box and Cox (1964) that value of λ which maximise the log likelihood function of λ could be chosen as the exponent of the most suitable power transformation

3 2 1 2b Graphical method of Draper and Hunter

This method consists m plotting various functions indicative of the distortions of the ANOVA which occur naturally in the analysis of transformed data The exponent λ of the power transform was taken on the x axis and the relevant statistic employed to check model violations on the Y axis The optimisation point was obtained from the graph by inspection

Draper and Hunter (1969) considered a two way classification model with r rows c columns and n observation per cell The ANOVA of the two way layout is as follows

ANOVA of a two way classification model

Source	df	MS	F
Row (A)	r 1	MSR	$p \quad \frac{MSR}{S^2}$
Column (B)	c 1	MSC	$t - \frac{MSC}{S^2}$
A x B interaction	(r 1)(c 1)	MSI	$I - \frac{MSI}{S^2}$
Error	rc(n 1)	S^2	

They proposed a graphical plot of p t and I against varying values of λ in search of a good transformation

Since some transformations may result in not obtaining transformed observations with equal variance they further suggested that it would be better to examine the plot of a statistic which supplied information on inhomogeneity of the transformed observations Accordingly the mean square ratio of the residual ANOVA (F_r) proposed by Levene (1960) was also selected for the graphical representation The optimisation point obtained from the graph would indicate nature of the power transformation to be used for the analysis If the optimal values obtained from the four plots actually coincided an unambigous solution to the problem could have been reached But this is not the case in practice and different experimenters may arrive at different conclusion from the same set of plots However compromise decisions could be always made and transformations useful for actual experimental situation could be identified through a careful scrutiny of the graph and by intuitive argument

In this approach the statistic I showed the effect of nonadditivity in the data But it is not possible to estimate such a statistic from single factor experiments Therefore in the present study the F ratio for nonadditivity was used as an alternative to the I statistic Since all the data gathered in the present study pertained to single factor experiments laid out in standard designs the following F ratios were selected for a comparative evaluation of transformations

(1) F_N The F ratio for non additivity

(2) F_t The F ratio for treatment Vs error

(3) F The F ratio for residuals for testing homoscedasticity

These ratios were plotted against varying values of λ in the range 1 to +1 The optimum value of λ was located by inspection

3 2 1 3 Analysis of data in the presence of outliers

The necessary conditions of ANOVA would be greatly influenced by the presence of outliers Transformation of data is also useful to minimise the effect of unusual observations Berry (1987) proposed a method of choosing the additive constant He also applied rank transformation before carrying out the parametric analysis

3 2 1 3a Choosing an additive constant

Abnormal observations affect the symmetry and shape of the distribution In such circumstances a proper choice of an additive constant is helpful in restoring robustness of the data A number of alternative ways for choosing C have been proposed by various workers Box and Cox (1964) proposed the familiar maximised likelihood method for estimating C Berry (1987) showed that the estimate so obtained was highly biased He proposed a realistic approach in estimating C by making the residuals as symmetric as possible and kurtosis small Skewness and Kurtosis of any set of data can be measured in terms of appropriate statistics defined as

$g - \Sigma e_{ij}^3/n\sigma^3 \rightarrow (3\ 14)$ and $g_2^1 - (\Sigma e_{ij}^4/n\sigma^4) - 3 \rightarrow (3\ 15)$

where g_1 and g_2^1 are the measure of Skewness and Kurtosis respectively and e_{ij} s are the residual values

If the assumption of normality is appropriate then the residuals will tend to be symmetric Let us define a new statistic g_2 as

$$g_2 - g_2^1 + \frac{6}{(d+2)} \rightarrow (3\ 16)$$

where $d - (n_1\ \ 1)(n_2\ \ 1)$ the error degree of freedom Evidently g_2 is also a measure of Kurtosis Now let $g_0\ \ |g_1| + |g_2|$ Then g_0 is a composite measure indicating the shape of the distribution Let c_k denote the value of c that minimises $|g_k|$ k = 0 1 2 These c_k values will be close to each other and any of them can be selected as the additive constant However according to Berry (1987) C_o would be the best choice among them

3 2 1 3b Rank transformation

When extreme observations are present the obvious alternative to parametric analysis on transformed data are non parametric analysis Conovar and Iman (1981) have suggested a procedure of combining these approaches They applied a rank transformation and then carried out parametric analysis In this study two rank transformations were used In the case of the first rank transformation (RT 1) all observations in the data set were ordered and ranked as 1 2 before conducting the parametric analysis In the second approach (RT 2) the data were ranked in several subsets (blocks) and analysis of variance was conducted These techniques are supposed to be simple powerful and robust and as such provide excellant alternatives to transformed ANOVA Moreover they do not require the choice of an unknown parameter such as an additive constant The relative efficiency of the rank analysis of variance in comparison with the analysis of variance of transformed data was assessed on the basis of the relevant F ratios

Results and Discussion

4 RESULTS AND DISCUSSION

The results obtained in the present study are outlined below under two major heads

(1) Analysis of survey data
(2) Analysis of experimental data

4 1 Analysis of survey data

The usual analysis of variance technique is not applicable to the analysis of data on count of insects relating to the pest surviellance project An attempt was therefore made to express the mean variance relationship for such data in quantitative terms and to make use of it for choos ng a suitable power transfor i t on

As a preliminary step a graphical plot of mean variance relationship was made by taking the monthly means on the x axis and the corresponding monthly variances on the y axis The points plotted on the graph showed a tendency to lie along a straight line or a curve of definite shape indicating a strong positive relationship between mean and variance (Fig 20 to Fig 25) In most cases the relation was almost linear which showed that standard deviation changed in direct proportion w th mean value and a logarithmic transformation would be effective in equalis ng the variab l ty

Taylor s power law was fitted to the data and the parameters a and b were estimated The results are presented in Table 4 1a

The exponent b of the Taylor s power law was found to be statistically significant in all the sets of data which showed that monthly variance of insect counts was strongly correlated with the monthly abundance of insects Therefore transformations could be effective in removing the instability of variance An examination of the value of the heterogeneity co efficient for the different sets of data revealed that it varied from 1 137 in the case of case worm to 2 49 in the case of stem borer

For all species of insects except caseworm b value was close to 2 which indicated the possible use of a logarithmic transformation for restoring constancy of variance In the case of caseworm b value was close to one emphasising the need for a squareroot transformation

Correlation coefficient between mean and variance of monthly data were worked out before and after applying the logarithmic transformation The results are also given in Table 4 1b It was found that logarithmic transformation caused a marked reduction in the value of the correlation coefficient

Measures of skewness and kurtosis viz γ_1 and γ_2 were also calculated before and after applying the relevant transformation The transformed data were found to be more nearly in agreement with the symmetry conditions and possessed smaller kurtosis than untransformed data Thus it could be concluded that logarithmic transformation would be effective in the analysis of data on counts of insects belonging to five major species viz jassic gall fly stem borer leaf folder and BPH In the case of case worm however a squareroot transformation was found to be the best

Observations on counts of insects would be supposed to follow the negative binomial distribution and hence logarithmic transformation would be helpful in restoring normality However Anscomb (1948) and Beal (1942) have recommended the inverse hyperbolic sine squareroot transformation for the analysis of such data But the inverse hyperbolic sine squareroot transformation is only a modified version of the logarithmic transformation The additional precision to be expected from such a transformation would be negligibly small Thus in general logarithmic trasnformation could be recommended for the analysis of data on insect counts especially where the data showed large amount of variability The results agree with the findings of Singh and Rai (1993)

4 2 Analysis of Experimental data

A comparative evaluation of the different power transformations used for the analysis of data on the intensity of infestation by the three insects (gall fly whorl maggot stem borer) was done by using the Box Cox (1964) approach graphical method of Draper and Hunter (1969) and Taylor s power law

Data on weed counts consisted of zero values and several abnormal observations Hence certain specialised methods of data handling had to be employed along with the above proceedures These included the choice of an additive constant possible use of non parametric techniques and a search for the applicability of an alternative forms of transformation Analysis of variance of ranked data as suggested by Conovar and Iman (1981) combined in itself the advantages of both non parametric and parametric procedures Hence it was felt useful to examine the utility of rank analysis of variance over ordinary analysis of variance in order to know whether it could serve as a viable alternative to the parametric procedure If the mean variance relationship could assume an approximate parabolic functional form inverse hyperbolic sine squareroot transformation was expected to be useful At the same time an approximate hyperbolic relationship indicated the use of the newly developed transformation log (x^2+k)

Thus the details presented in this section are included under eight sub heads

1 Box Cox likelihood approach
2 Graphical method of Draper and Hunter

3 Taylor s power law

4 Choice of an additive constant

5 Inverse hyperbolic sine squareroot transformation

6 $\log(x^2+k)$ transformation

7 Rank analysis of variance

8 General comments on the selection of a suitable transformation for empirical data

4 2 1 Box cox likelihood approach

Observations on the intensity of infestation by various types of insects (y) were transformed into the parametric family of transformation $y^{(\lambda)}$ the parameter λ possibly a vector defining a particular transformation In the present study λ was allowed to vary in the range between 1 and +1 at intervals of length 0 2

The original data ($\lambda - 1$) were analysed as in a randomised block design and the error sum of squares were estimated Data were also subjected to analysis of variance after applying the different power transformations (corresponding to varying values of λ) and in each case the residual sum of squares $S(\lambda)$ was calculated From these the residual sum of squares of the normalized transformation $s(\lambda\ z)$ was obtained as given in section 3 2 1 2a The maximised log likelihood values of λ $L_{max}^{(\lambda)}$ was evaluated as given in section 3 2 1 2a

$L_{max}^{(\lambda)}$ was plotted against values of λ in the interval 1 to +1 and the value of λ which maximised the log likelihood function was evaluated by inspection The optimal value of λ thus obtained showed the best possible transformation for a given set of data Values of $L_{max}^{(\lambda)}$ for varying values of λ are given in tables 4 2 1a 4 2 1b 4 2 1c and 4 2 1d Plots of $L_{max}^{(\lambda)}$ against λ for different sets of data are also given Fig 1 to Fig 10

It could be seen that for all the expenmental data optımal value of λ was close to zero This ındicated that loganthmıc transformatıon was the best ın generatıng a sımple lınear normal model when there are model vıolatıons Thus Box Cox procedure undoubtedly recommended the loganthmıc transformatıon to be the most effectıve ın analysıng data on ınfestatıon of ınsects and counts of weed It could be seen that the transformatıon has gıven nse to a consıderable extent of reductıon ın the error mean square (EMS) It has also enhanced the sensıtıvıty of F test (Table 4 2 1e)

The loganthmıc transformatıon has a natural appeal for the analysıs of enumeratıve data due to ıts sımplıcıty and popularıty Many workers recommended loganthmıc transformatıon for the analysıs of data on ınsect count It ıs also specıally suıted ın such data where we expect row column ınteractıon However the results obtaıned from the Box cox approach though statıstıcally sound have theır own lımıtatıons It ıs extremely essentıal to examıne all the other relevant aspects before takıng a conclusıve verdıct on the choıce of a transformatıon

4 2 2 Graphıcal method of Draper and Hunter

The data were transformed ın to the parametnc famıly of transformatıon for varyıng values of λ ın the range 1 to +1 a value for $\lambda - 0$ ındicates a loganthmıc transformatıon Both transformed and untransformed data ($\lambda - 1$) were subjected to ANOVA and resıdual sum of squares and sum of squares due to non addıtıvıty were estımated From the analysıs of resıdual data, resıdual vanances of samples (treatments) were estımated and theır homogeneıty was tested by usıng Bartlett s chı square test Homogeneıty of treatment means was tested by usıng F statıstıc F_t The possıble presence of non addıtıvıty ın the data was detected by the sıgnıficance or non sıgnıficance of the varıance ratıo test for non addıtıvıty (F_N) The Fr statıstıc of Levene whıch gave an ındıctıon of the ıntensıty of heteroscedastıcıty ın the data was also calculated These functıons F_t F_r and F_N were plotted graphıcally for varyıng values of λ and a value λ which optimised these three functions simultaneously was selected as the exponent of the power transformation The results are presented in table 4 2 1a, 4 2 1b 4 2 1c and 4 2 1d The plots of the three functions against λ for different sets of data are presented graphically (Fig 1 to Fig 10)

It was found that in the case of observations on silver shoot caused by gall fly F_t, F_r and F_N were optimised at λ − 0 5 at 20 DAT However at 30 DAT F_t was maximum at λ − 0 Thus the graphical approach generally prefered the squareroot transformation to the logarithmic transformation Logarithmic transformation also was almost equally good in meeting with the requirment of the ANOVA though squareroot transformation had an edge over all other transformation In the case of whorl maggot none of the power transformations were found to be effective though squareroot transformation showed slight superiority over others As far as the data on the number of dead heart relating to insect stem borer damage logarithmic transformation was found to be the best at 30 DAT However at 40 DAT reciprocal transformation was found to be the most effective followed by the logarithmic transformation Thus it was not possible to frame any strict rule on the type of transformation to be used in analysing a particular data The choice of a transformation largely depended on the type of the insect extent of variability in the data and presence of aberrant observations However in most cases the choice lied between squareroot and logarithmic transformation If nothing is known about the nature of the data it would be safe to use logarithmic transformation In cases where details are available about the nature of the data and the extent of variability squareroot transformation would also be an equally competant choice

4 2 3 Taylor s Power law

Taylors power law was fitted to each of the nineteen sets of data and the estimated values of parameters are given in Tables 4 2 3a and 4 2 3b

The results ındıcated that squareroot transformatıon was the best choıce for the analysıs of data on sılver shoot caused by gall fly sınce the value of b was close to one In the case of whorl maggot none of the transformatıon was found to be effectıve ın restorıng normalıty and homoscedastıcıty As far as stemborer was concerned logarıthmıc transformatıon was found to be effectıve In the analysıs of data on weed counts squareroot transformatıon produced better results than other transformatıons

4 2 4 Choıce of an addıtıve constant

As the observatıons on weed counts ıncluded zero values also an arbıtrary constant had to be added to each observatıon before conductıng the analysıs Thıs ıs absolutely necessary for applyıng logarıthmıc transformatıon Even ıf we are not applyıng the logarıthmıc transformatıon, addıtıon of a constant would help ın reducıng the effect of aberrant observatıons and maıntaınıng ınternal symmetry ın the data The usual practıce of addıng one or half to each datum before transformatıon has ıts own lımıtatıons Therefore the applıcabılıty of themethod proposed by Berry (1987) [Sectıon 3 2 1 3a] ın choosıng an addıtıve constant was examıned on the basıs of the empırıcal data

In the lıght of the procedure suggested by Berry necessary calculatıons were made to estımate the addıtıve constant c for all sets of data The estımated C so obtaıned was found to be approxımately 2 8 for all sets of data

Comparısons between dıfferent transformatıons were made after ıncorperatıng the addıtıve constant 2 8 ın each datum In addıtıon to $L_{max}^{(\lambda)}$ the three ratıos F_r F_N and F_t were also calculated for varyıng values of λ and the results are presented ın Table 4 2 4 The plots of the above functıons agaınst λ are gıven ın Fıg 11 to Fıg 19

It could be seen that $L_{max}^{(\lambda)}$ was maximised at $\lambda - 0$ indicating the possible use of loganthmic transformation for the analysis of data on weed counts But the plots of single factor ratios like F_t F_N and F_r did not confirm to this hypothesis In the case of Y_2S_1 and Y_2S_3 no transformation was found to be effective For Y_2S_2 the choice lied between squareroot and loganthmic transformation For most of the other sets reciprocal transformation showed distnct advantages over other transformations especially with regard to the enhancement of the sensitivity of the F test In general the results showed that reciprocal transformation would be effective in the analysis of data on weed population when there are zero values and other disrupt observations Loganthmic and squareroot transformations were also seemed to be useful to some extent in such cases It is a common practice among expenmenters to add one to each observation before the conduct of ANOVA, when the data contains zero values The results of analysis of each expenmental data based on the two prospective values of the additive constant c viz c = 1 and c = 2 8 were compared using the various functional values It was found that in most cases F_r and F_N assumed relatively smaller values when the additive constant c assumed its optimal value 2 8 Compared to that of the trial value c 1 This indicated that the choice of the additive constant c by using the method proposed by Berry (1987) resulted in a significant reduction in vanability and non additivity The proposed method was also successful in increasing the sensitivity of the F test in the analysis of data at least in some of the expenments

4 2 5 Inverse hyperbolic sine squareroot transformation

This type of transformation was tried on sets of data for which standard deviation varied with mean The data (x) were transformed to a new scale (x^1) by the expression

$x^{1} - k^{-1/2} \sinh^{-1} \sqrt{kx}$ where k is a constant to be estimated from the sample data Beal (1942) has given an expression for k (Section 3 2 1 1e) The values of F_t F_r and F_N obtained after the application of the above transformation is given in Table 4 2 5

An alternative estimate of k was also derived as per the method suggested in Section 3 2 1 1e The values of F_r F_t and F_N were also determined on the basis of the new estimate of k and were compared with those values based on the estimate of k given by Beal The results are given in Table 4 2 5 It was found that the new estimate was better than that given by Beal (1942) since F_N and F_r assumed relatively smaller values and values of F_t generally increased with this new estimate of k than with the former estimate of k

4 2 6 $\log(x^2+k)$ transformation

The applicability of the new transformation developed in section 3 2 1f was examined on the same set of data used for the empirical verification of the inverse hyperbolic sine squareroot transformation The values of F_r F_t and F_N were found out before and after the aplication of the above transformation using the usual analysis of variance technique The values are given in Table 4 2 6 These values were also compared with those obtained by the inverse hyperbolic sine squareroot transformation It was found that as a whole new transformation gave better results than the inverse hyperbolic sine squareroot transformation

A comparative anatomy of some of the transformationsused in the study with regard to their applicability in the analysis of certain sets of unusual data consisting of extreme values and exhibiting large amount of instability is given in table 4 2 6a

Although no general rule on the choice of a suitable transformation could be framed from such empirical studies the results evidently showed the relative superiority of the new transformation over others in analysing certain types of messy data which deviate considerably from model assumptions It could also be seen that the new transformation gave maximum power for the F test in the analysis of data for Y_2S_1 In the case of Y_3S_2 inverse hyperbolic sine squareroot transformation maximised the power of F test at the expense of instability of variance But the new transformation gave relatively high F_t values than with other sets of data

4 2 7 Rank analysis of variance

The analytical results obtained from various sets of data after applying the two rank transformations are presented in Table 4 2 7 along with those of the logarithmically transformed data and untransformed data Though the rank transformation failed to show a consistent performance they were useful in enhancing the power of the F test atleast in a few cases When the performance of the two rank transformations RT 1 and RT 2 were compared no consistent superiority was noticed for one method over the other However rank transformations were in general helpful for increasing the sensitivity of the F test when compared to that of the untransformed data

4 2 8 General comments on the selection of a suitable transformation for empirical data

Results of analysis showed that it was not possible to find a unique transformation for all sets of data on insects or weeds The choice of a transformation largely depended on the nature of the data and the variability of the material Observations on the same insect or weed showed large amount of variability at different time periods making the distribution highly eratic However in most cases the choice lied between squareroot transformation and logarithmic transformation Other complicated types of power transformations were rarely required for satisfying the requirements of the ANOVA For certain types of data especially those on weeds reciprocal transformation was also found to be suitable Rank transformations were also useful when there were outliers present in the data When the mean standard deviation relationship was parabolic the new transformation $\log(x^2+k)$ was found to be useful When mean variance relationship was quadratic inverse hyperbolic sine squareroot transformation was also helpful

The different techniques used for the selection of an appropriate transformation did not produce confirmatory results Box Cox procedure invariably showed the utility of logarithmic transformation where as graphical method failed to show a unique transformation for all types of data In general logarithmic transformation appeared to be the best choice in the absence of any prior information about the data If more details on the nature of the data are available a better transformation can be selected on the basis of a critical examination of the relevant data

Table 4 1a Fitting of Taylor s power law Estimated values of the parameters a and b

Name of insect	a	b
Jassid	0 295	2 401*
Case worm	10 965	1 137*
Gall fly	7 586	1 776*
Stem borer	0 3019	2 494*
Leaf folder	5 012	1 848*
BHP	2 344	2 154*

Table 4 1b Correlation coefficient between mean and variance before and after logarithmic transformation

Name of insect	Variable correlated	Untransformed data	Logarithmic transformed data
Jassid	Mean and variance	0 880*	0 648*
Case worm	do	0 822*	0 651*
Gall fly	do	0 975*	0 815*
Stem borer	do	0 995*	0 803*
Leaf folder	do	0 880*	0 876*
BPH	do	0 858*	0 710*

* Significant at 5% level

Table 4 2 1a Functional values of F F_N F_t $L_{max}^{(\lambda)}$ and chisquare for different values of λ for gall fly

Time period	λ	F	F_N	F	$L_{max}^{(\lambda}$	χ^2
	1	1 638	2 56	13 98*	29 94	2 4
	5	1 014	0 37	15 79*	33 53	7 075
20 DAT	0	1 429	0 63	12 86*	112 66	10 874
	0 5	3 227*	3 79	8 67*	18 21	24 565*
	1	15 199*	12 95*	4 14*	1 75	44 756*
	1	0 936	0 02	7 03*	9 225	6 176
	5	1 116	2 90	9 03*	10 92	7 634
30 DAT	0	2 209	13 61*	8 68*	102 152	19 197*
	0 5	4 064*	26 18*	6 52*	15 878	41 987*
	1	5 626*	36 55*	4 64*	41 63	73 030*
	1	4 37*	1 13	5 36*	17 506	7 890
	5	11 284*	0	9 17*	25 89	1 598
40 DAT	0	30 94*	9 52*	10 07*	97 55	9 638
	0 5	2 608	83 99*	3 59*	13 87	48 242*
	1	1 566	432 99*	1 57	60 73	103 448*

Table 4 2 1b Functional values of F_r F_N F_t $L_{max}^{(\lambda)}$ and chisquare for different values of λ for whorl maggot

Time period	λ	F_r	F_N	F_t	$L_{max}^{(\lambda)}$	χ^2
	1	1 790	2 11	6 40*	25 58	4 890(NS)
	5	0 839	0 01	5 54*	22 37	9 960(NS)
20 DAT	0	3 968*	6 07*	3 43	89 38	29 161*
	0 5	6 552*	65 81*	1 74	26 31	69 955*
	1	8 476*	629 33	1 19	72 815	125 537*
	1	0 427	0 05	0 75	19 19	2 465
	5	0 289	0 38	0 84	16 45	3 209
30 DAT	0	0 724	1 55	0 95	112 66	5 913
	0 5	0 592	5 31*	1 06	12 46	11 631
	1	3 146*	15 99*	1 13	3 334	21 143*
	1	1 273	0 40	1 59	17 629	3 375
	5	13 872*	0 43	1 48	16 86	4 329
40 DAT	0	2 005	0 58	1 29	114 47	6 194
	0 5	1 877	0 84	1 11	33 89	8 960
	1	3 928*	1 33	0 95	8 68	12 598

* Significant at 5% level

Table 4 2 1c Functional values of F F$_N$ F $L_{max}^{(\lambda)}$ and chisquare for different values of λ for stem borer

Time period	λ	F	F_N	F	$L_{max}^{(\lambda)}$	χ^2
	1	3 062	1 54	3 30*	17 63	5 330
	5	1 920	0 17	4 05*	20 32	2 797
20 DAT	0	0 998	0 30	4 89*	118 044	2 014
	0 5	1 118	2 82	3 56*	21 02	3 844
	1	3 032*	9 05*	9 905*	14 62	8 956
	1	1 674	4 96*	0 24	19 157	7 177
	5	0 366	6 06*	0	22 46	6 194
40 DAT	0	2 064	7 41*	0 52	121 61	6 708
	0 5	1 328	8 71*	2 98	17 95	3 844
	1	1 380	9 33*	9 76	10 53	14 942

*Significant at 5% level

Table 4 2 1d Functional values of F F$_N$ F $L_{max}^{(\lambda)}$ and chisquare for different values of λ for whorl maggot

Time period	λ	F	F$_N$	F	$L_{max}^{(\lambda)}$	χ^2
	1	0 70	0 09	1 56	25 04	12 161
	5	0 25	0 73	1 74	48 31	12 166
30 DAT	0	19 97	2 17	1 92	130 82	12 915
	0 5	0 70	4 75*	2 07	27 63	14 508
	1	1 29	9 02*	2 18	22 96	17 013*
	1	2 289	0 13	0 99	26 326	5 836
	5	1 762	0 65	1 00	27 111	5 453
40 DAT	0	0 813	1 66	1 03	121 61	5 890
	0 5	0 305	3 21	1 06	24 348	7 142
	1	1 629	5 25*	1 11	30 15	9 172

* Significant at 5% level

Table 4 2 1e Effect of loganthmic transformation of expenmental data
Error mean square coefficient of vanation and F ratios before and after applying the loganthmic transformation

Details of expenment	Untransformed			Loganthmically transformed		
	EMS	CV(%)	F_t	EMS	CV(%)	F_t
SS 20 DAT	4 923	30 46	13 98	028	21 75	12 86
SS 30 DAT	17 978	33 51	7 03	054	23 36	8 68
SS 40 DAT	10 714	40 04	5 36	072	39 18	10 07
WM 20 DAT	6 565	34 32	6 48	120	45 88	3 43
WM 30 DAT	9 644	34 29	0 75	029	18 17	0 95
WM 40 DAT	10 632	31 01	1 66	025	15 85	1 29
DH 30 DAT	10 632	33 15	3 30	020	14 81	4 89
DH 40 DAT	8 664	27 90	4 96	016	13 03	7 41
WM 30 DAT	6 689	22 27	1 56	009	9 12	1 92
WM 40 DAT	6 174	28 50	0 99	016	13 75	1 03
Y_1S_1	9 178	41 87	10 53	015	16 97	20 00
Y_1S_2	2 048	17 85	39 97	004	8 19	62 67
Y_1S_3	2 606	26 32	6 70	007	10 87	15 09
Y_2S_1	8 083	24 57	54 42	018	16 21	32 20
Y_2S_2	10 671	26 04	19 18	023	15 84	20 20
Y_2S_3	4 998	18 99	43 30	010	10 26	35 66
Y_3S_1	3 185	26 15	17 04	007	11 63	32 96
Y_3S_2	2 816	30 66	7 89	013	16 85	10 66
Y_3S_3	2 400	26 30	8 49	016	17 71	7 31

EMS Error mean square
cv – coefficient of vanation
F_t F ratio for treatments

Table 4 2 3a Fıttıng of Taylor s power law to expenmental data on ınsect count estımated values of the parameters a and b

Tımepenod		a	b
	20 DAT	0 467	0 785
SS	30 DAT	3 63	0 045
	40 DAT	1 023	0 604
	20 DAT	3 380	0 202
WM	30 DAT	3 467	0 064
	40 DAT	1 659	0 285
DH	30 DAT	0 145	1 865
	40 DAT	0 537	1 142
WM	30 DAT	177 83	1 48
	40 DAT	12 88	1 94

--

Table 4 2 3b Fıttıng of Taylor s power law to experımental data on weed count estımated values of the parameters a and b

Serıal number of data set	a	b
1	0 977	1 252
2	1 096	0 468
3	0 794	0 698
4	1 288	0 831
5	0 955	1 081
6	1 148	0 660
7	1 023	0 848
8	1 047	1 063
9	1 778	0 364

Table 4 2 4 Functional values of F F_N F $L_{max}^{(\lambda)}$ and chisquare for different values of λ for the data on weed count (when c 2 8)

Serial No of data set	λ	F_r	F_N	F	$L_{max}^{(\lambda)}$	χ^2
	1	4 934* (4 174)	17 73 (17 75)	10 55* (10 53)	24 09	161 819
	5	5 048 (5 221)	7 93* (5 79)	14 40* (16 37)	36 56	120 337
1	0	5 597 (5 070)	1 02 (0 21)	20 00* (28 06)	141 54	93 350
	0 5	5 316* (7 036)	0 95 (0 15)	28 13* (52 51)	73 01	75 599
	1	5 419* (7 001)	0 32 (0 32)	39 72* (102 51)	63 93	72 396
	1	22 126* (3 060)	1 00 (1 00)	39 97* (39 97)	53 04	101 406
	5	3 366* (7 702)	0 73 (0 56)	49 29* (54 97)	56 36	76 418
2	0	3 45* (3 778)	0 34 (0 07)	62 67* (83 30)	152 61	61 691
	0 5	2 141 (2 842)	0 06 (0 07)	80 28* (133 51)	59 84	53 717
	1	5 899* (2 848)	0 01 (0 45)	101 93 (219 79)	66 51	55 913

Contd

Continued

Serial No of data set	λ	F_r	F_N	F_t	$L_{max}^{(\lambda)}$	χ^2
	1	7 011* (6 521)	0 11 (0 11)	6 70* (6 70)	44 95	66 984
	5	5 047* (5 372)	0 66 (0 73)	10 27* (12 12)	55 16	48 709
3	0	4 135* (18 810)	0 80 (0 61)	15 09* (20 68)	142 32	37 322
	0 5	4 579* (4 249)	0 68 (0 26)	20 60* (31 46)	66 74	31 319
	1	4 9* (43 972)	0 48 (0 04)	25 93* (44 14)	61 14	32 182
	1	198 487* (54 42)	4 32 (4 32)	54 42 (54 42)	26 15	133 731
	5	5 077* (43 15)	2 20 (1 77)	49 28* (43 15)	35 68	101 027
4	0	7 341* (22 65)	0 99 (0 47)	32 20* (22 65)	127 61	90 291
	0 5	9 183* (13 38)	0 32 (0 10)	20 22* (13 38)	30 18	90 372
	1	10 723* (10 06)	0 11 (0 02)	14 23* (10 06)	25 02	102 045

Contd

Contınued

Serıal No of data set	λ	F	F_N	F_t	$L_{max}^{(\lambda)}$	χ^2
	1	1 218 (1 954)	0 29 (0 29)	19 18 (19 18)	21 54	71 363
	5	6 443* (6 476)	1 50 (2 07)	21 44* (21 15)	25 35	60 865
5	0	8 702* (1 151)	3 33 (4 68)	20 20 (17 65)	123 54	60 574
	0 5	6 850* (1 513)	4 91* (4 98)	17 14* (13 22)	17 07	67 293
	1	0 998 (0 044)	5 80* (6 11)	14 26* (10 34)	7 548	81 058
	1	5 298* (3 472)	6 57* (6 57)	43 30* (43 3)	34 13	40 431
	5	2 928 (2 848)	1 93 (1 97)	41 02* (40 14)	38 13	30 154
6	0	2 815 (2 448)	0 06 (0 43)	35 66* (31 44)	137 38	28 057
	0 5	2 975* (4 194)	0 73 (4 29)	29 49* (21 95)	31 966	33 107
	1	3 267* (3 598)	3 19 (7 27)	24 13* (15 67)	30 14	45 167

Contd

Contınued

Serıal No of data set	λ	F_r	F_N	F	$L_{max}^{(\lambda)}$	χ^2
	1	3 901* (3 553)	2 38 (2 38)	17 04* (17 04)	48 8475	104 252
	5	4 014* (4 124)	1 32 (0 09)	33 92* (34 24)	59 59	75 903
7	0	2 599 (7 042)	0 62 (0 03)	32 96* (40 84)	168 198	57 729
	0 5	5 899* (7 592)	0 09 (1 09)	42 25* (44 58)	79 06	47 133
	1	1 501 (9 583)	0 20 (2 82)	47 96* (37 27)	72 74	48 735
	1	0 501 (0 0)	16 77* (16 77)	7 89* (7 89)	51 25	124 155
	5	0 230 (0 001)	11 34* (9 38)	9 29* (10 10)	58 24	95 284
8	0	2 508 (4 01)	7 29 (4 63)	10 66* (12 41)	156 124	77 550
	0 5	3 685* (7 44)	4 52* (2 08)	11 91* (14 62)	66 78	66 517
	1	1 345 (1 584)	2 75 (0 85)	13 03* (16 72)	63 511	66 745

Contd

Continued

Serial No of data set	λ	F	F_N	F_t	$L_{max}^{(\lambda)}$	χ^2
	1	5 646 (1 041)	0 05 (0 05)	8 49 (8 49)	54 37	59 866
	5	5 411* (5 208)	0 02 (0 01)	8 03* (7 63)	55 58	50 488
9	0	1 023 (3 735)	0 01 (0 00)	7 31* (6 36)	152 08	46 699
	0 5	3 347* (3 752)	0 00 (0 00)	6 52* (5 20)	51 05	46 542
	1	4 321* (4 609)	0 00 (0 00)	5 78* (4 38)	50 21	51 487

* Significant at 5% level
Figures in parentheses are values when c – 1

Table 4 2 5 Functional values of F_t F_N and F in the case of inverse hyperbolic sine square root transformation

Details of experiment	F	F_N	F_r
Y_2S_1	30 63*	0 90	7 630*
Y_2S_2	18 46* (34 63)	4 10 (0 04)	7 420* (6 187)
Y_3S_2	33 00* (13 55)	18 28 (5 58)	16 472* (4 883)
DH 30 DAT	4 16* (4 21)	0 08 (0 05)	0 788 (0 489)

*Significant at 5% level
Figures in parentheses are values of F F_N and F with the new estimate of k

Table 4 2 6 Functional values of F_t F_N and F_r for $\log(x^2+k)$ transformation

Details of experiment	F_t	F_N	F
Y_2S_1	59 17*	2 65	10 058*
Y_2S_2	18 63*	4 07	0 023
Y_3S_2	12 83*	15 33	0 00002
DH 30 DAT	4 45*	0 0001	0 3431

*Significant at 5% level

Table 4 2 6a Functional values of F F_N F_t for some selected transformation

Details of experiment	Type of transformation	F	F_N	F
	Logarithmic	7 341*	0 99	32 20*
	Square root	5 077*	2 20	49 28*
Y_1S_1	Inverse hyperbolic sine squareroot	7 630*	0 90	30 63*
	$\log(x^2+k)$	10 058*	2 65	59 17*
	Logarithmic	8 701*	3 33	20 20*
	Squareroot	6 443*	1 50	21 44*
Y_2S_2	Inverse hyperbolic sine squareroot	7 420*	4 10	18 46*
	$\log(x^2+k)$	0 023	4 07	18 63*
	Logarithmic	2 508	7 29*	10 66*
	Squareroot	0 230	11 34*	9 29*
Y_3S_2	Inverse hyperbolic sine squareroot	16 472*	18 238*	33 00*
	$\log(x^2+k)$	0 00002	15 32*	12 83*
	Logarithmic	0 998	0 30	4 89*
	Squareroot	1 920	0 17	4 05*
DH 30 DAT	Inverse hyperbolic sine squareroot	0 788	0 08	4 16*
	$\log(x^2+k)$	0 3431	0 0001	4 45*

*Significant at 5% level

Table 4 2 7 Comparison of analytical results (Ft values) of two rank transformations (RT 1 and RT 2) with logarithmic transformation and no transformation

Details of experiment

Type of transformation	Y_1S_1	Y_1S_2	$Y\ S_3$	Y_2S_1	Y_2S_2	Y_2S_3	Y_3S_1	Y_3S_2	Y_3S_3
Untransformed data	10 55	39 97	6 70	54 42	19 18	43 3	17 04	7 89	8 49
RT 1	36 72	51 21	17 09	15 66	12 17	26 68	28 39	32 26	6 11
RT 2	79 75	25 83	115 44	16 11	8 83	32 86	33 61	24 78	7 10
log(x+2 8)	20 00	62 67	15 09	32 20	20 20	35 66	32 96	10 66	7 31

Fig 1 Graphical representation of Fr, Ft Fn and Lmax for varying values of lamda for SS 20 DAT

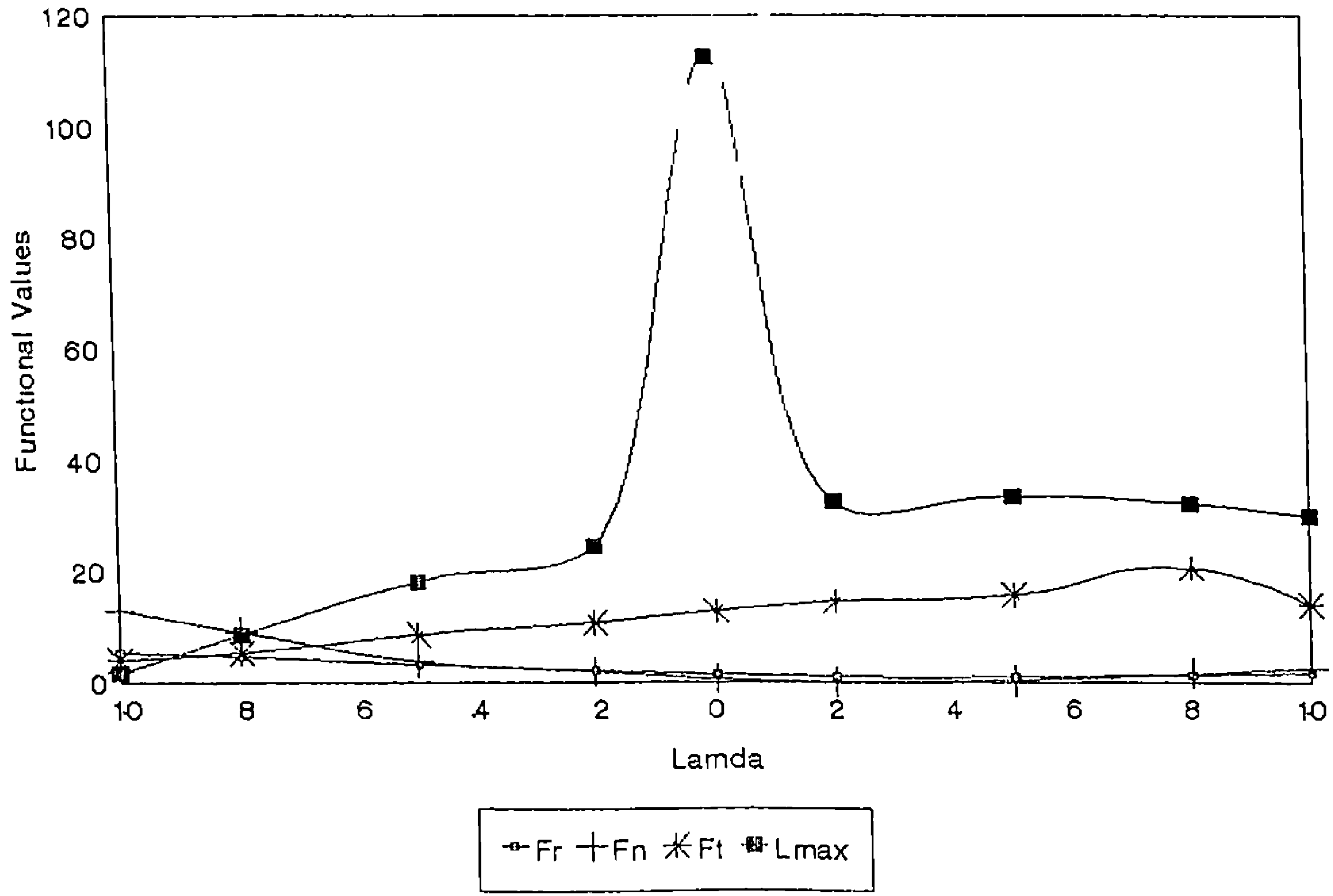


Fig 2 Graphical representation of Fr, Ft, Fn and Lmax for varying values of lamda for SS 30 DAT

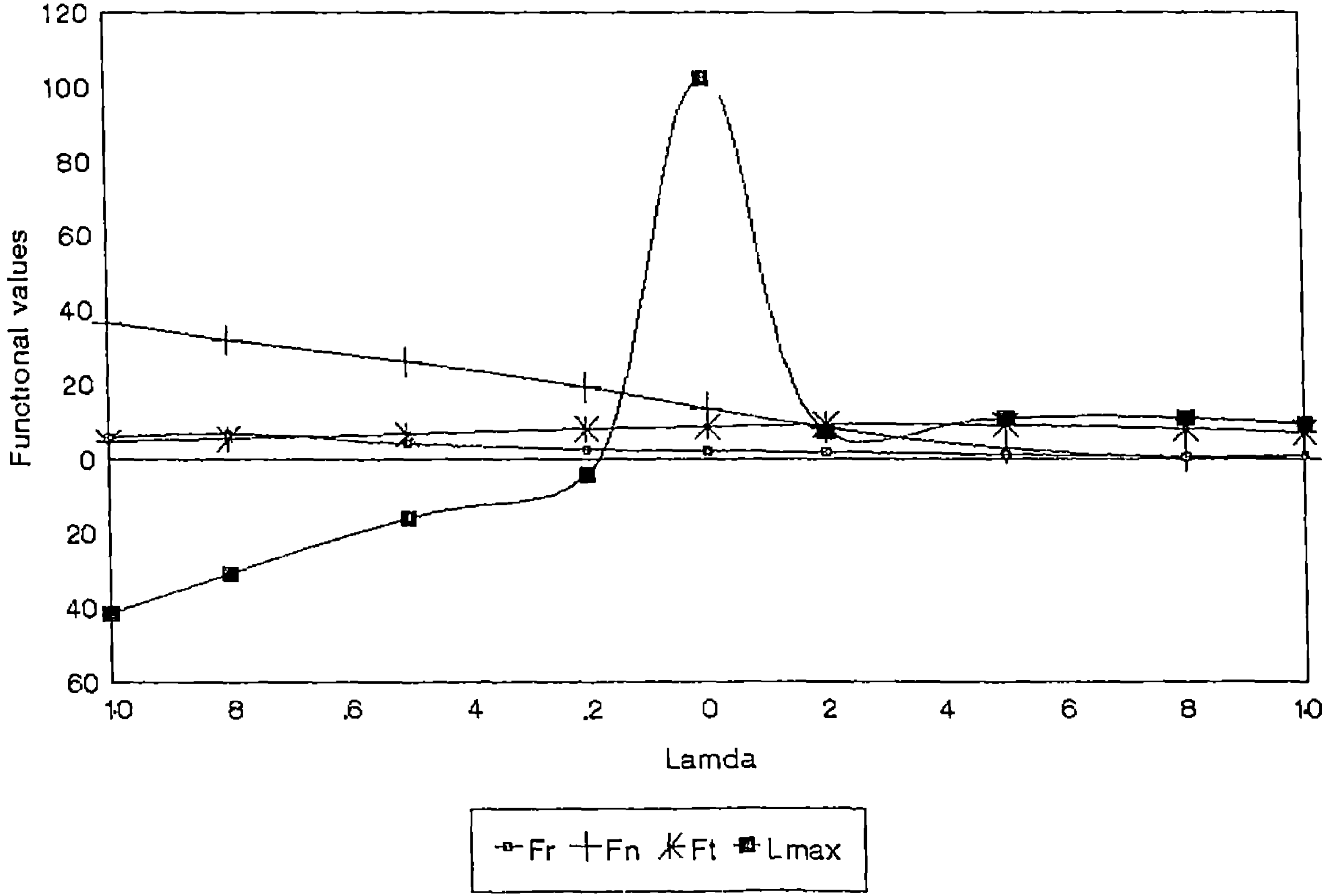


Fig 3 Graphical representation of Fr, Ft, Fn and Lmax for varying values of lamda for SS 40 DAT

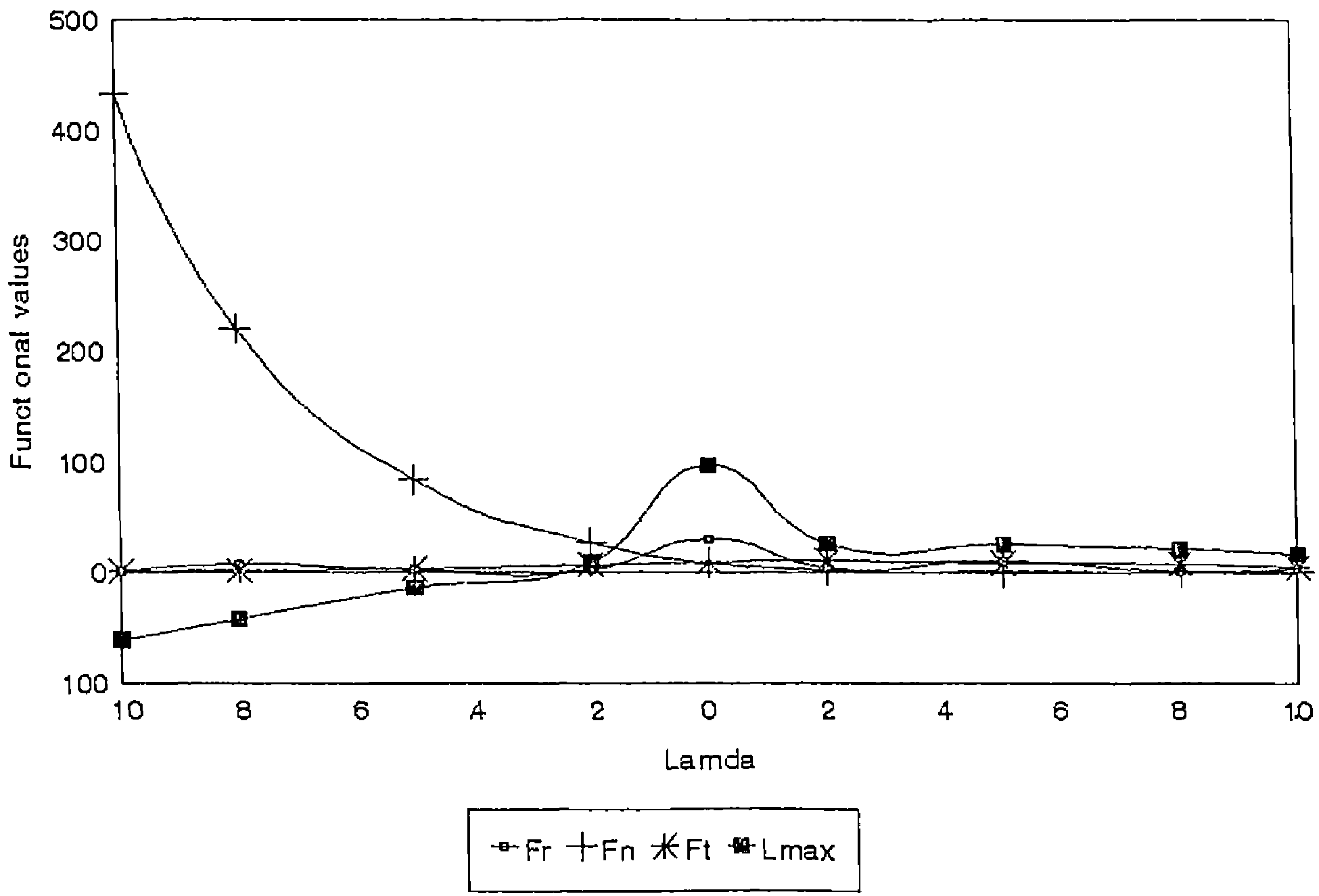


Fig 4 Graphical representation fo Fr, Ft Fn and Lmax for varying values f lamda for WM 20 DAT

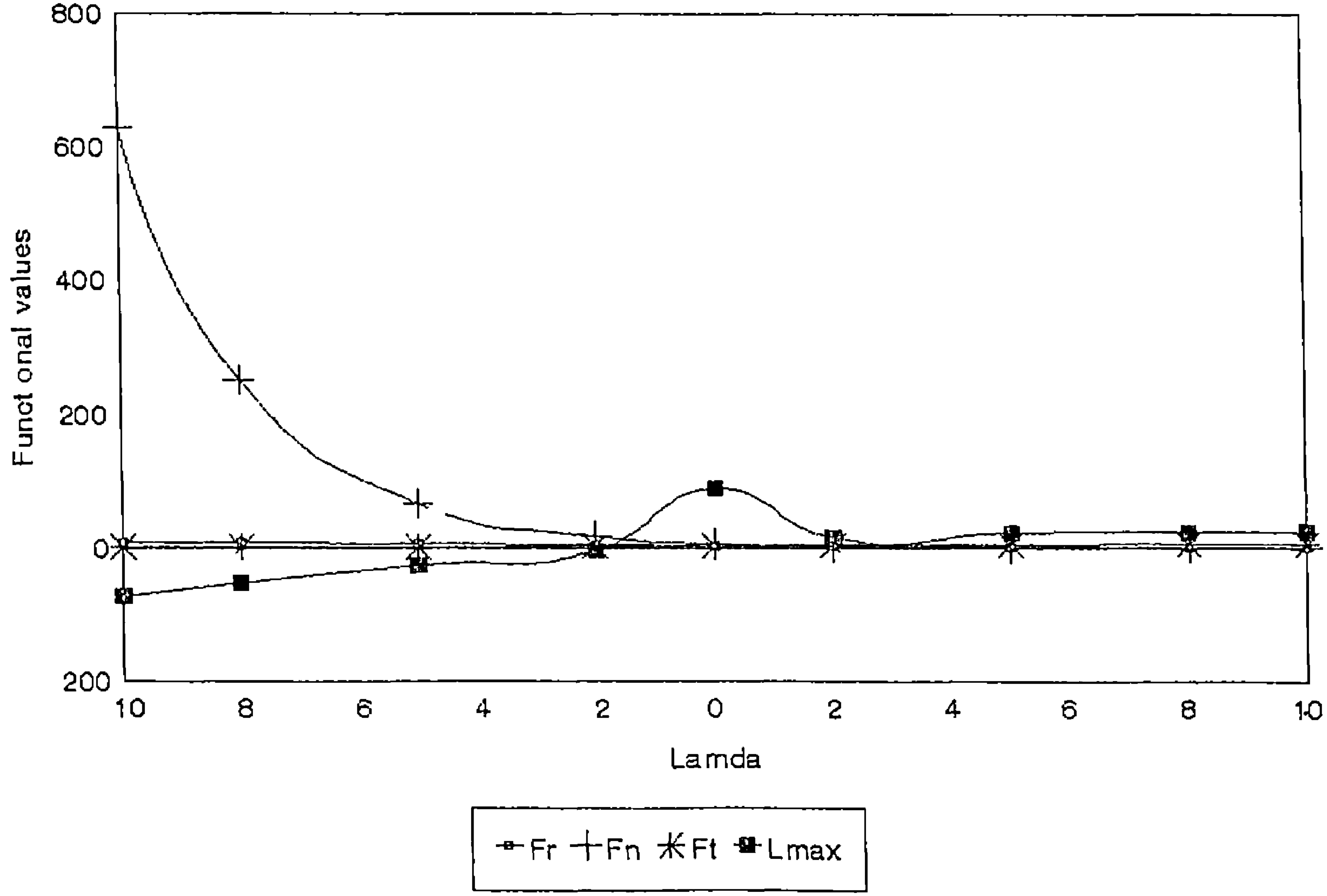


Fig 5 Graphical representation of Fr Ft,Fn and Lmax for varying values of lamda for WM 30 DAT

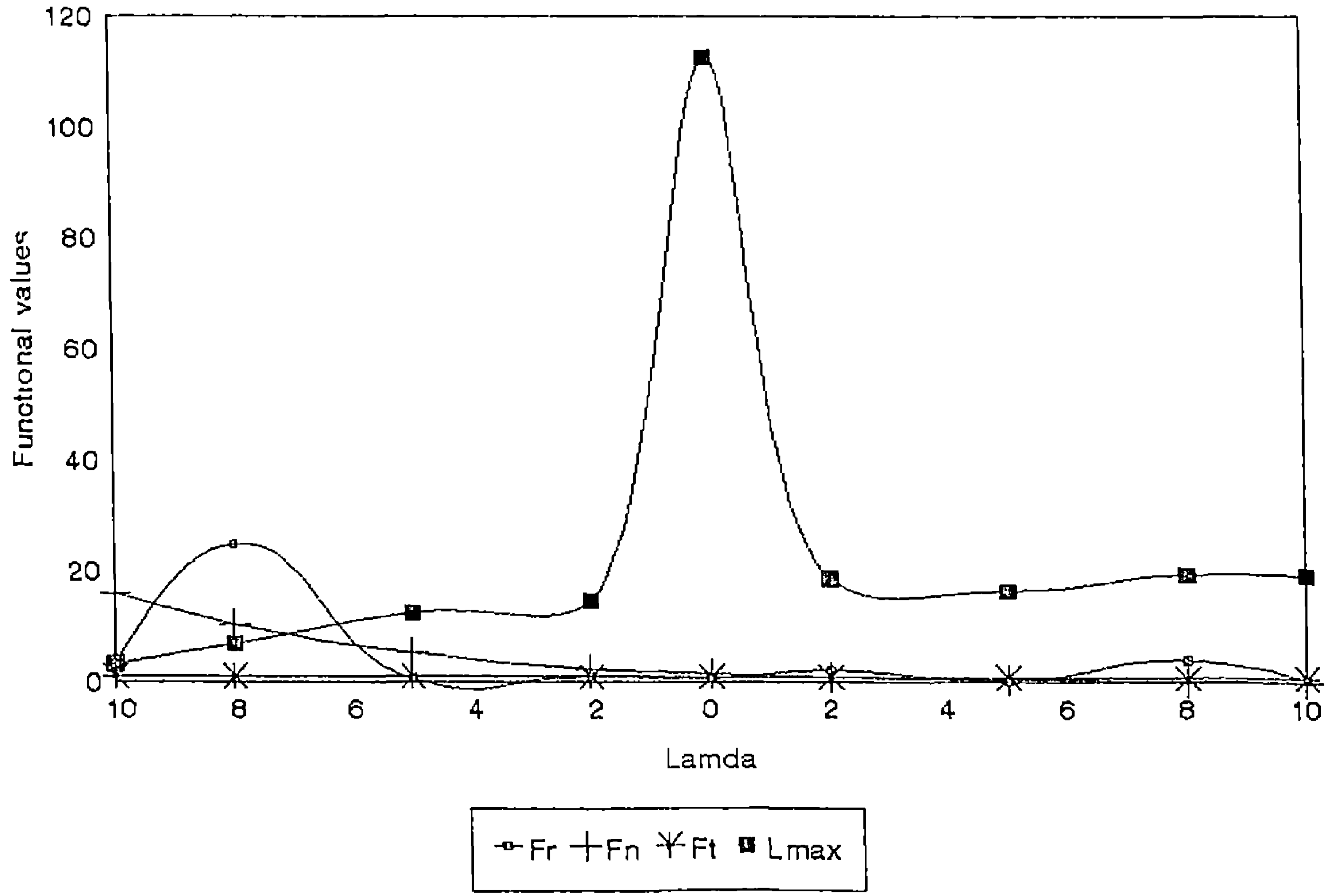


Fig 6 Graphical representation of Fr Ft Fn and Lmax for varying values of lamda for WM 40 DAT

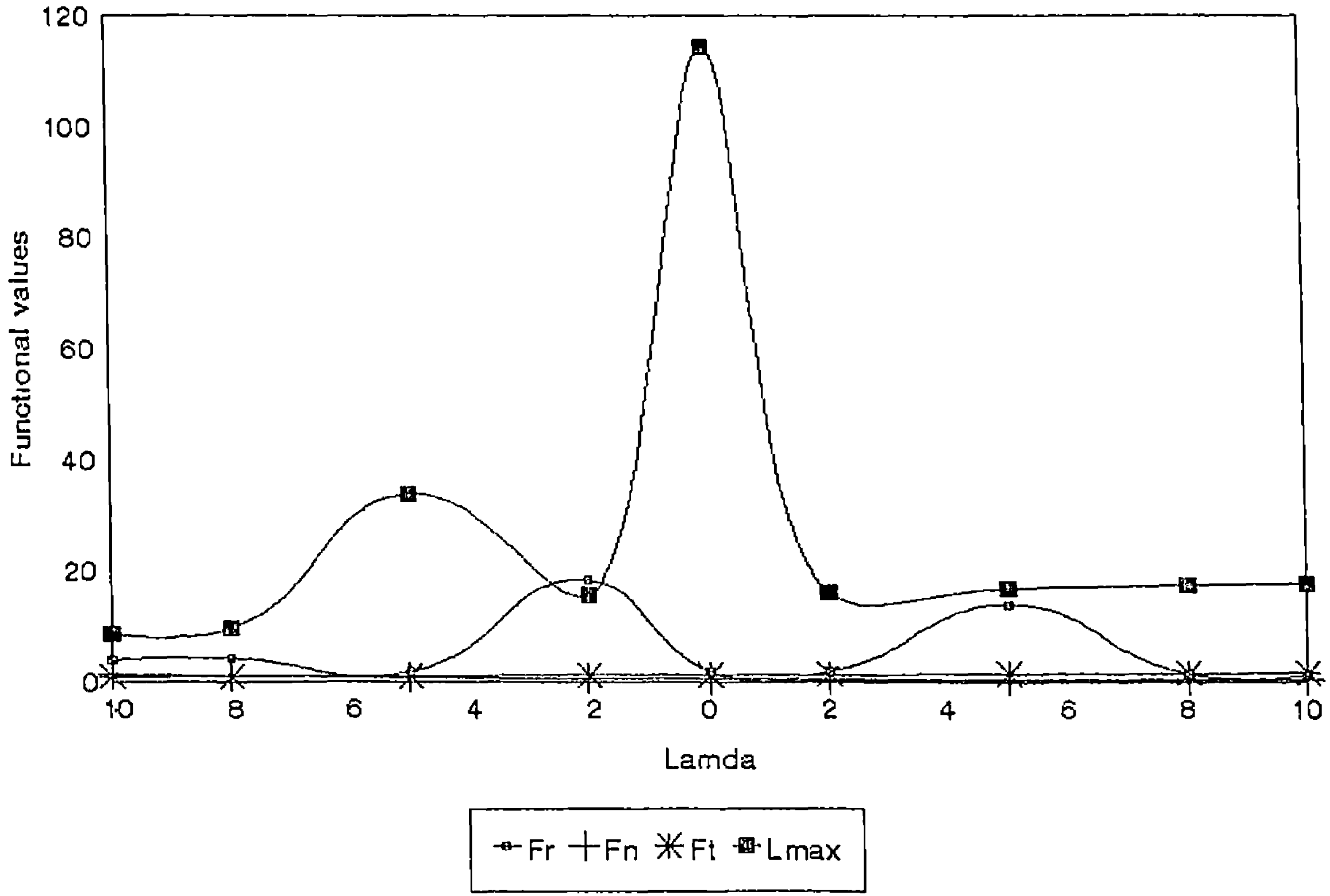


Fig 7 Graphical representation of Fr Ft, Fn and Lmax for varying values of lamda for DH 30

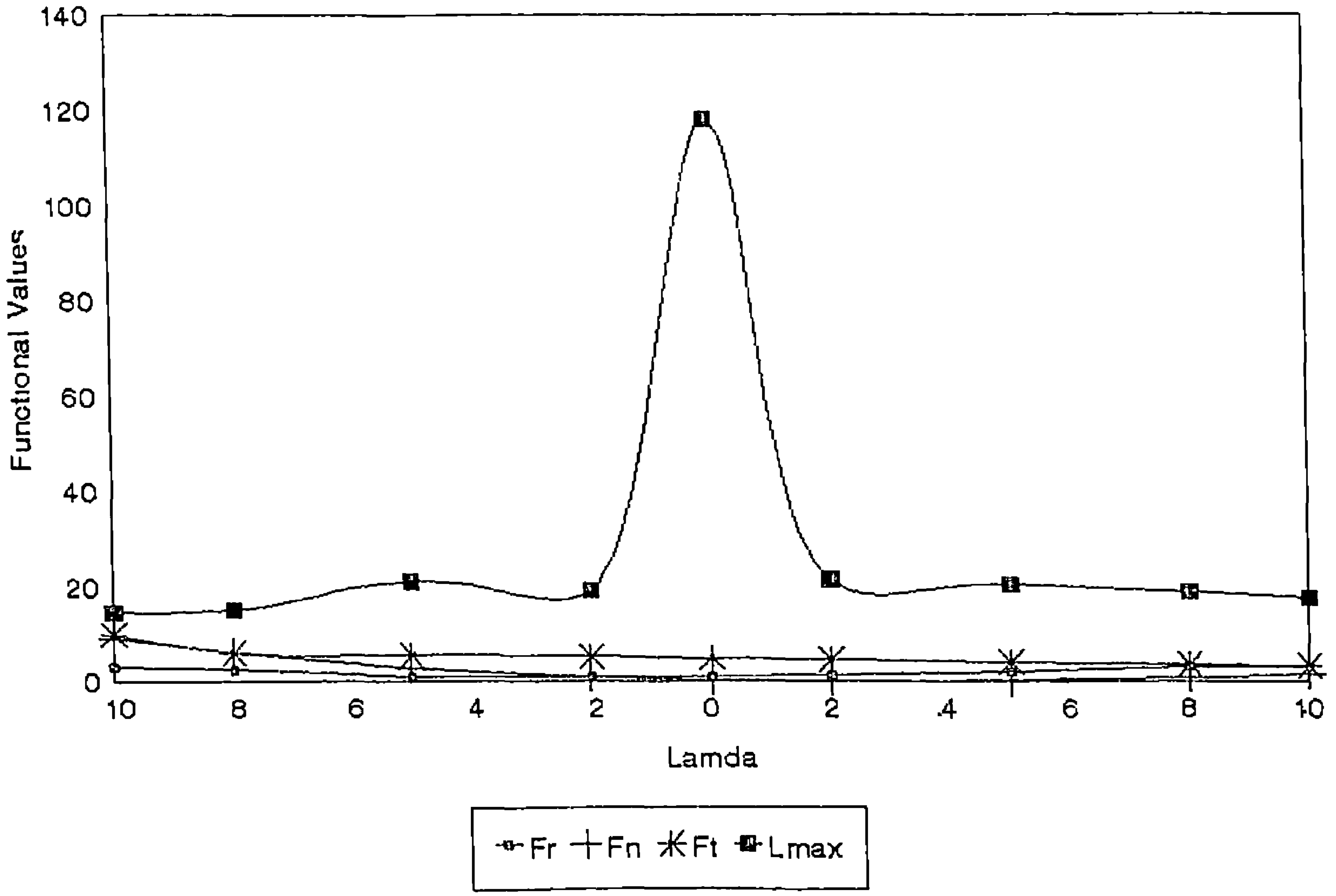


Fig 8 Graphical representation of Fr Ft Fn and Lmax for varying values of lamda for DH 40 DAT

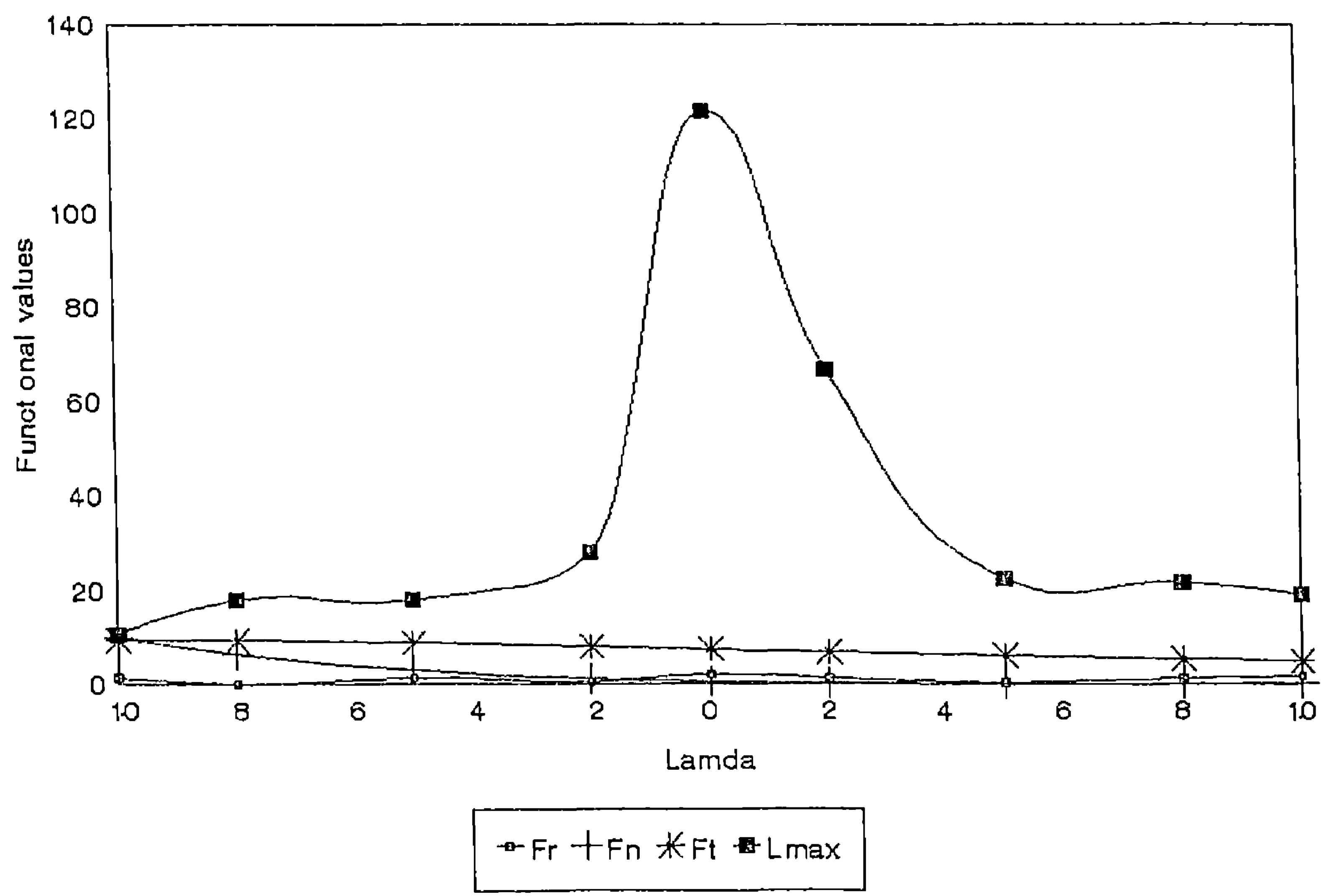


Fig 9 Graphical representation of Fr Ft Fn and Lmax for varying values of lamda for WM 30 DAT

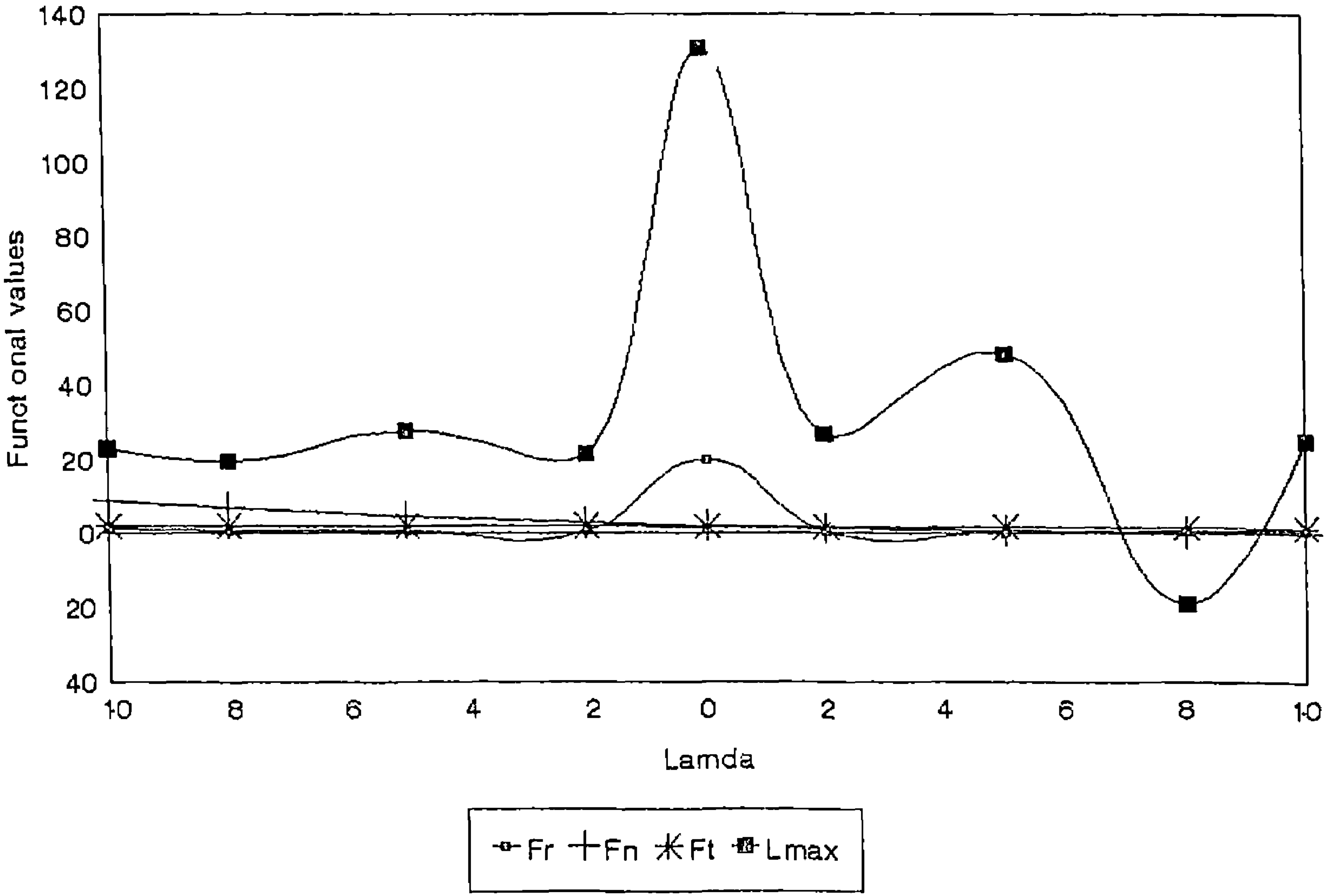


Fig 10 Graphical representation of Fr Ft Fn and Lmax
for varying values of lamda for WM 40 DAT

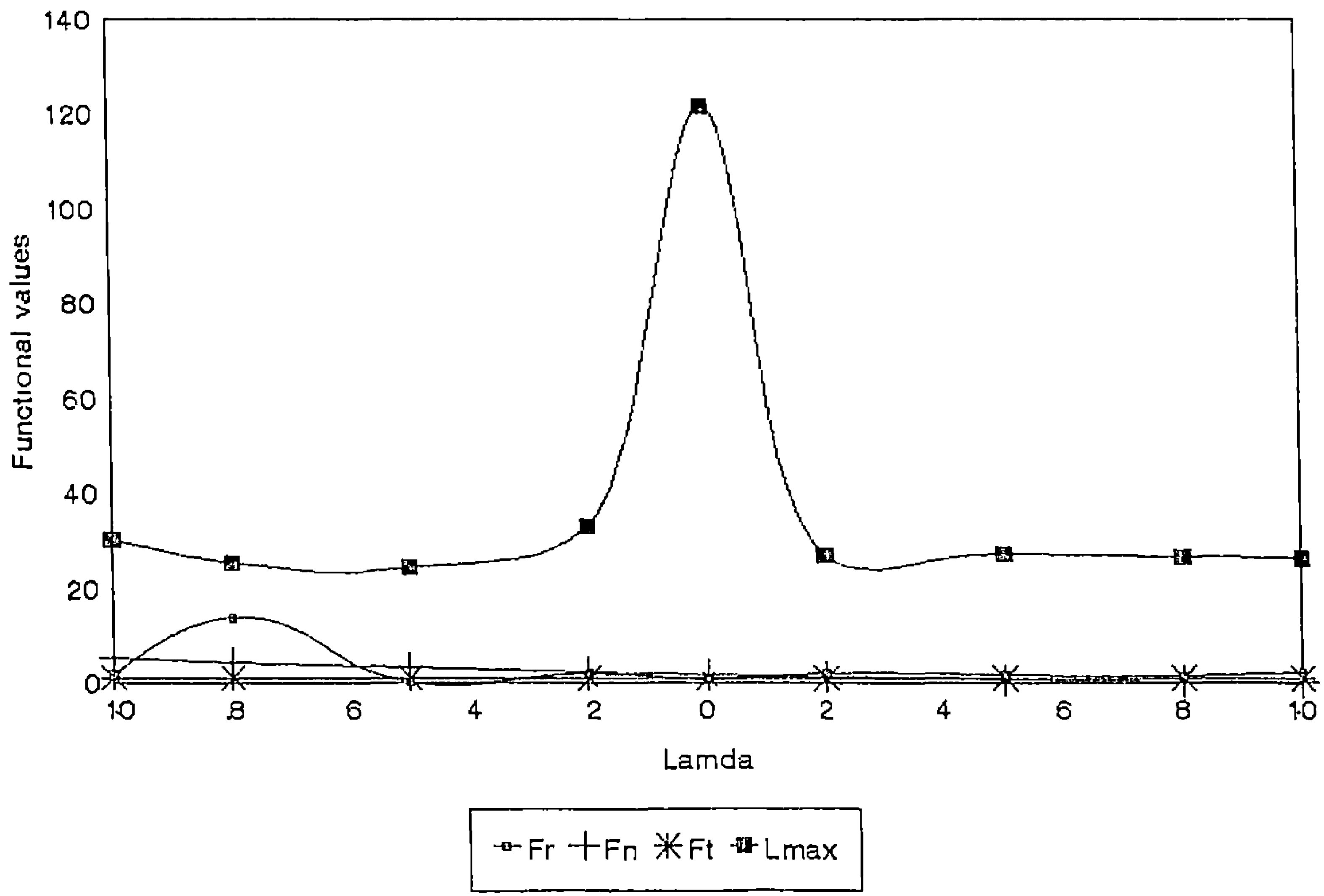


Fig 11 Graphical representation of Fr Ft Fn and Lmax for varying values of lamda for data set 1

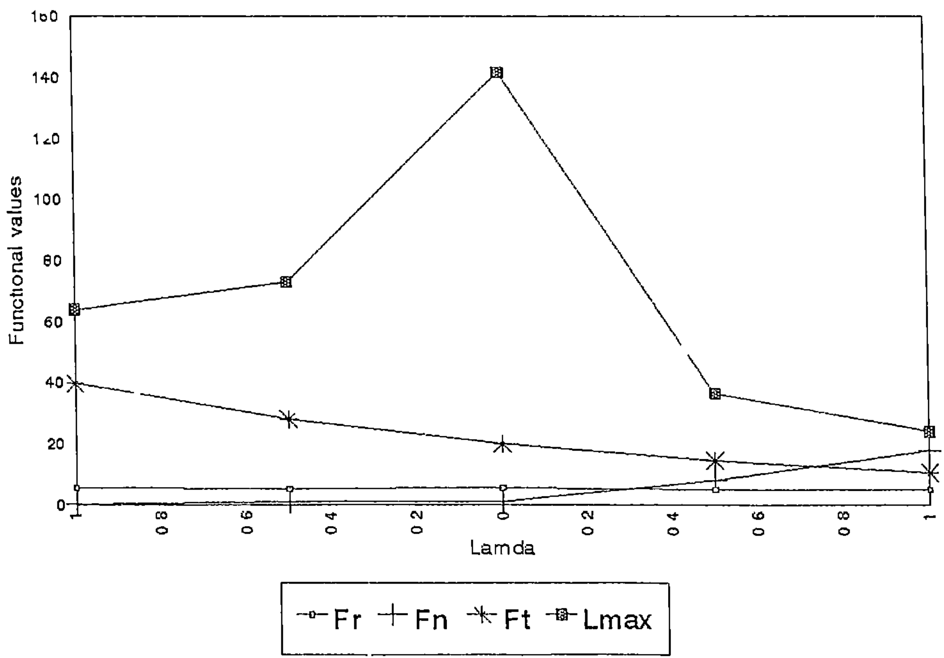
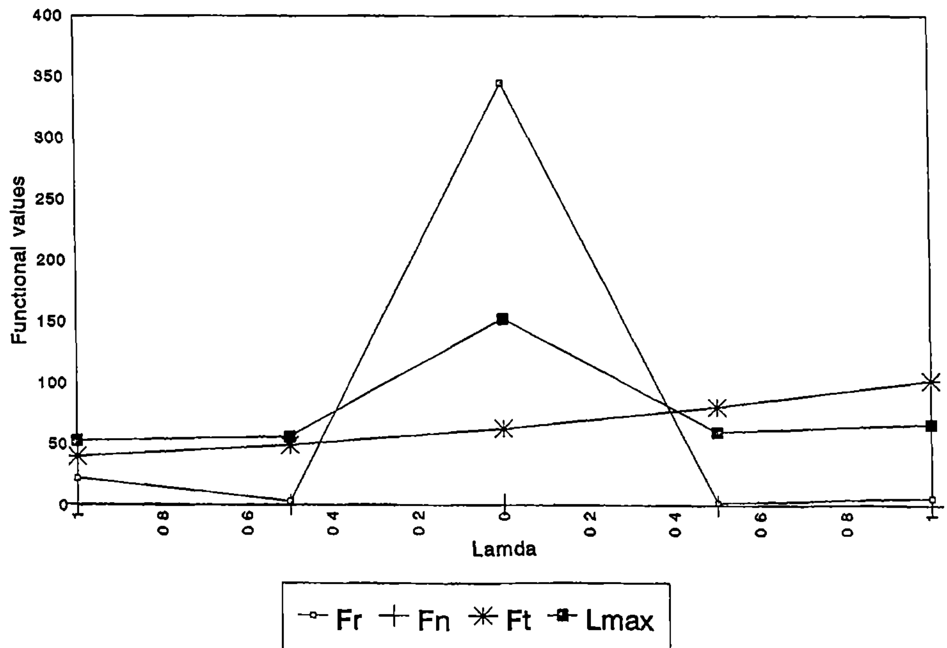


Fig 12 Graphical representation of Fr Ft Fn and Lmax for varying values of Lamda for data set2



F g 13 Graphical representation of Fr Ft Fn and Lmax for varying values of Lamda for data set3

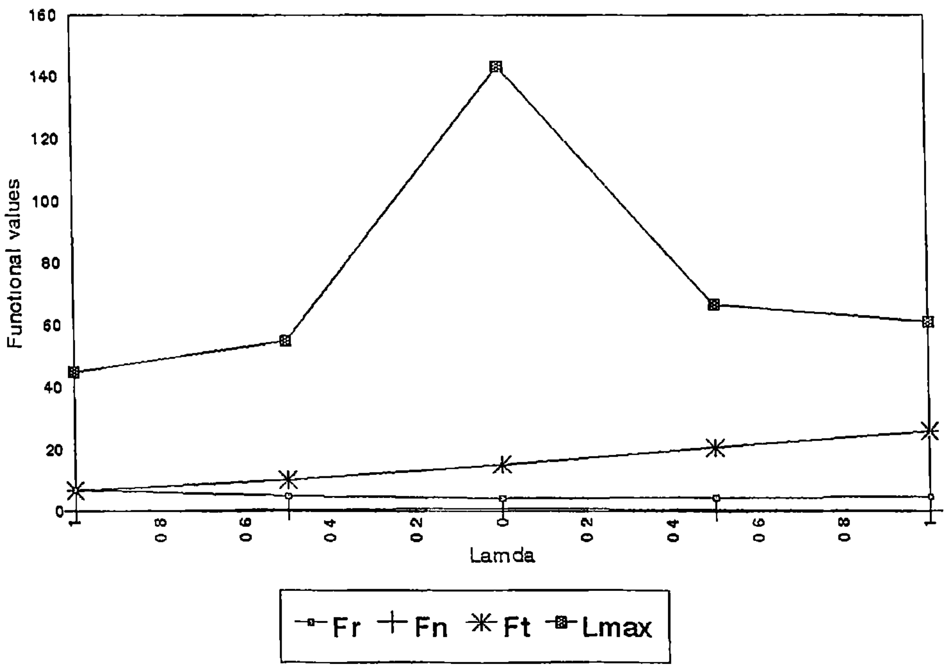


Fig 14 Graphical representat on of Fr Ft Fn and Lmax for vary ng values of Lamda for data set 4

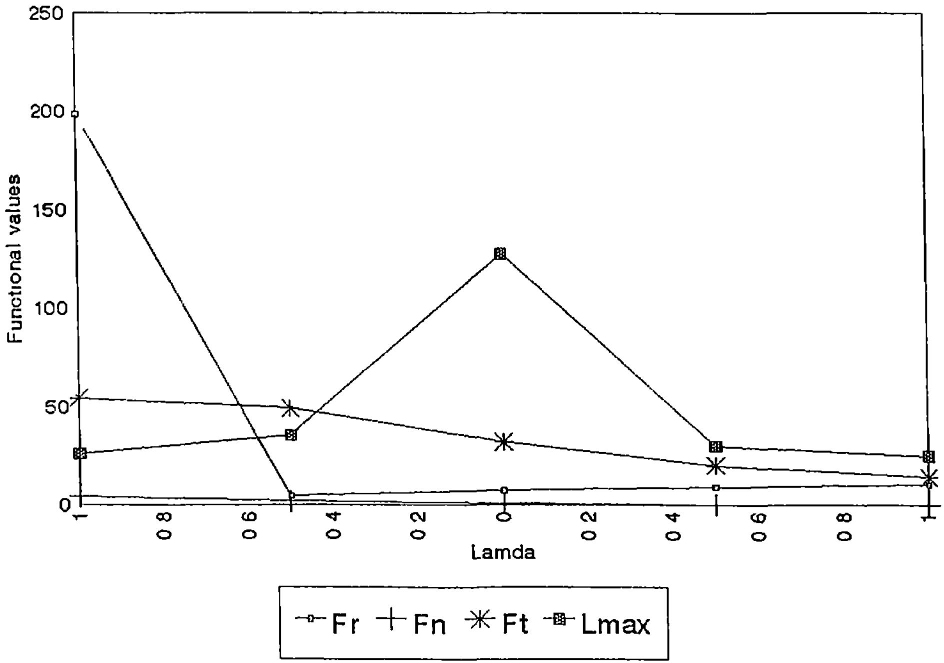


Fig 15 Graphical representation of Fr Ft Fn and Lmax for varying values of Lamda for data set 5

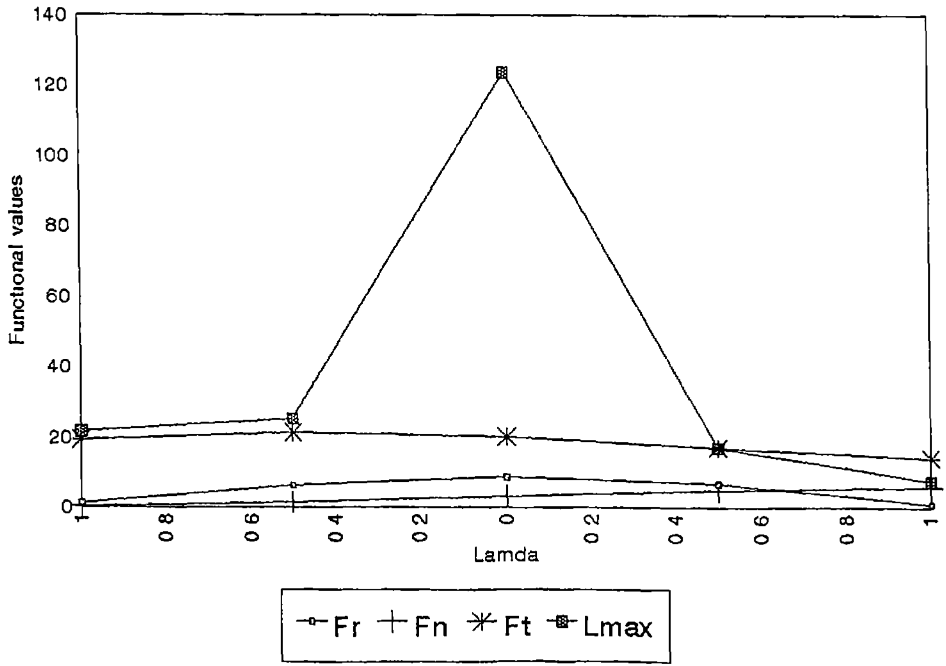


Fig 16 Graphical representation of Fr Ft Fn and Lmax for varying values of Lamda for data set 6

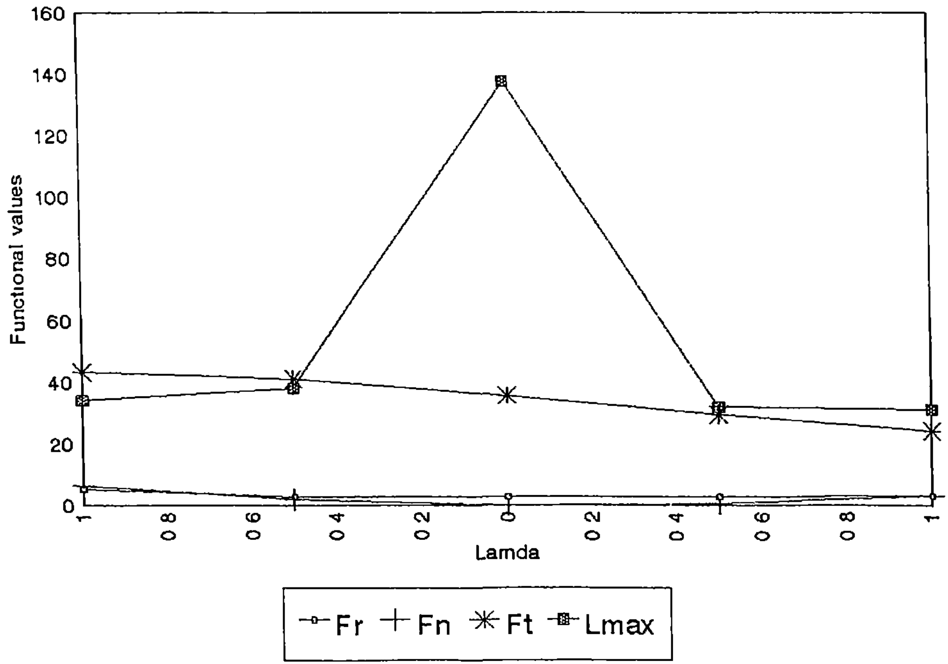
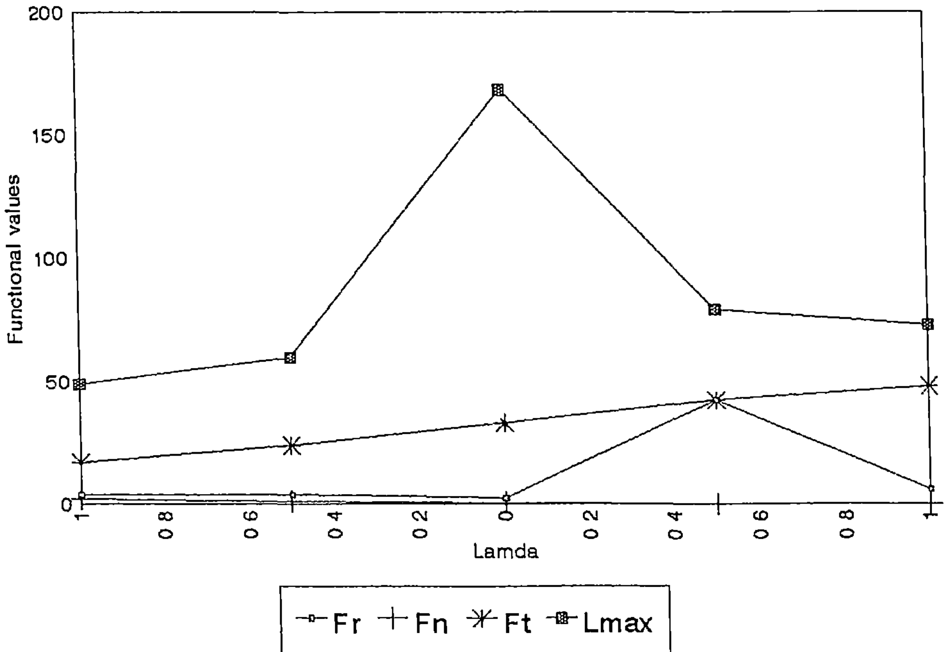


Fig 17 Graphical representation of Fr Ft Fn and Lmax for varying values of Lamda for data set 7



F g 18 Graphical representation of Fr Ft Fn and Lmax for vary ng values of Lamda for data set 8

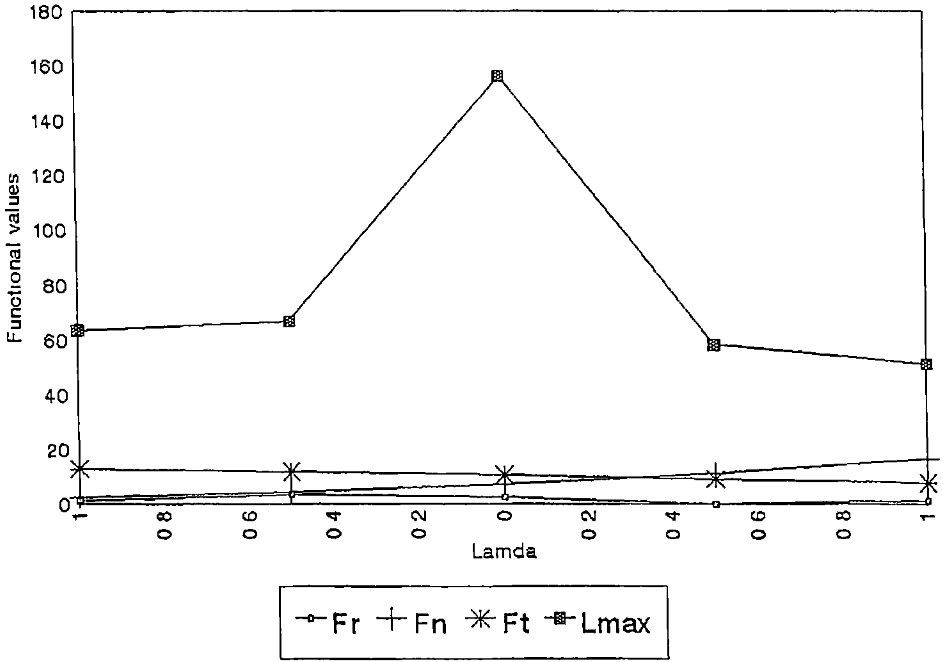


Fig 19 Graphical representation of Fr Ft Fn and Lmax for varying values of Lam for data set 9

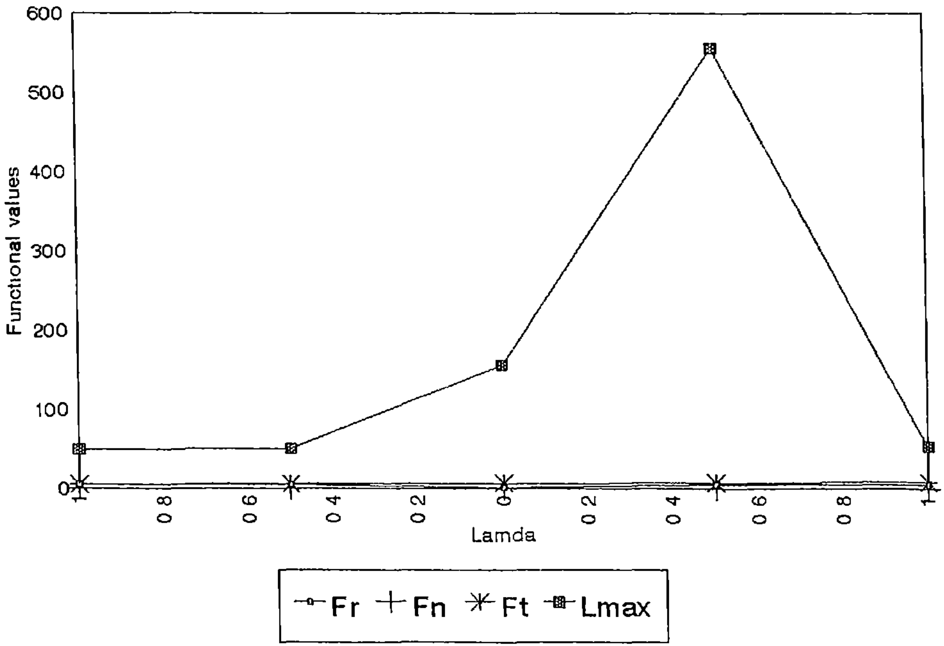


Fig 20 Relationship between mean and variance of the monthly distribution of counts of Jassid

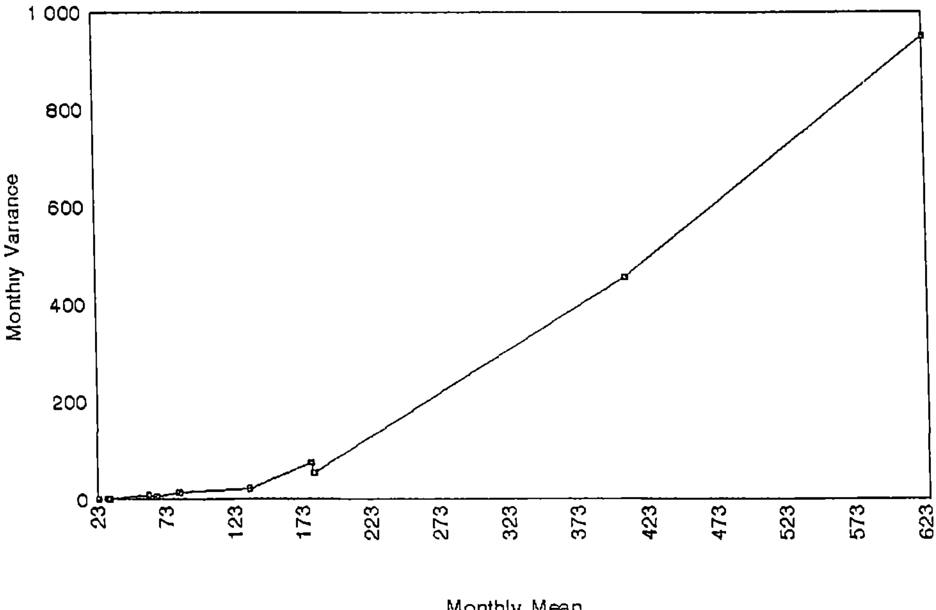


Fig 21 Relationship between mean and variance of the monthly distribution of counts of case worm

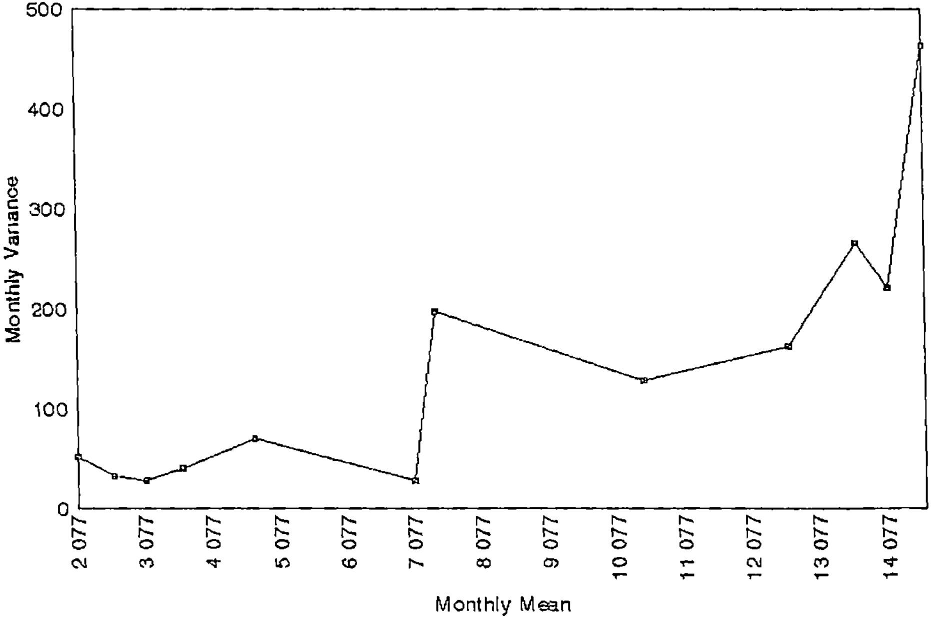


Fig 22 Relationship between mean and variance of the monthly distribution of counts of gall fly

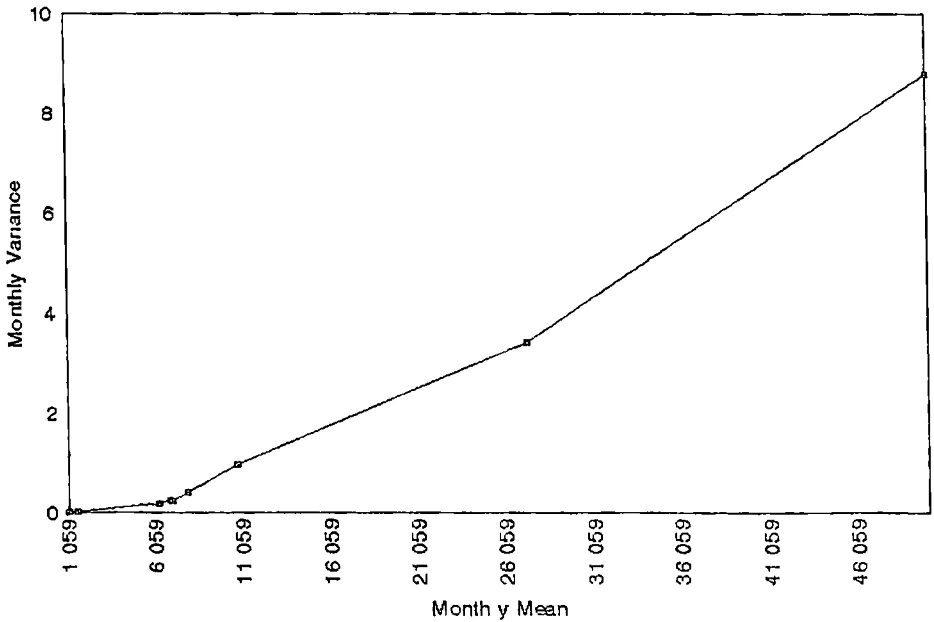
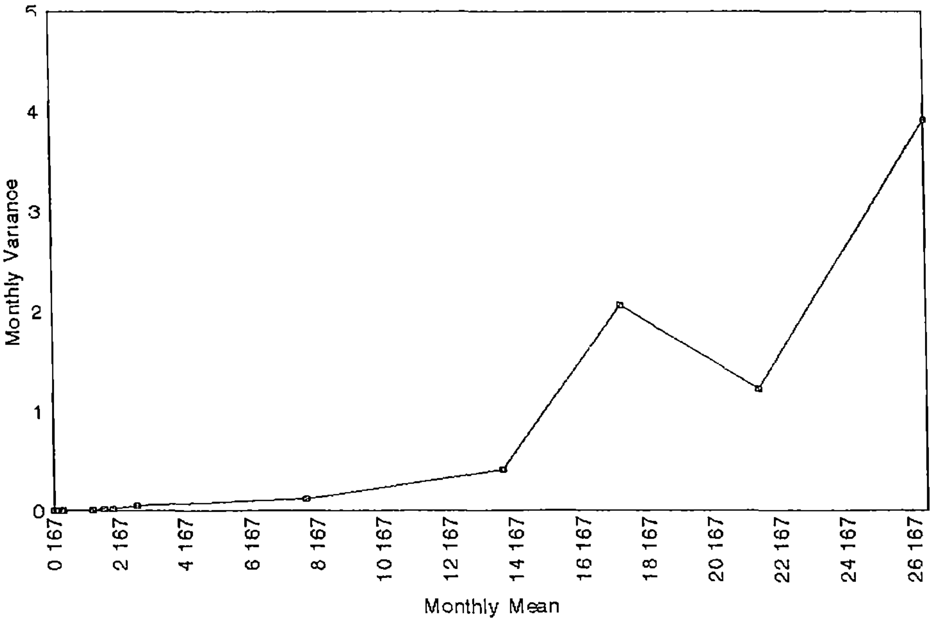


Fig 23 Relationship between mean and variance of the monthly distribution of counts of leaf folder



F g 24 Relat onsh p between mean and varıance of the monthly dıstrıbut on of

counts of stem borer

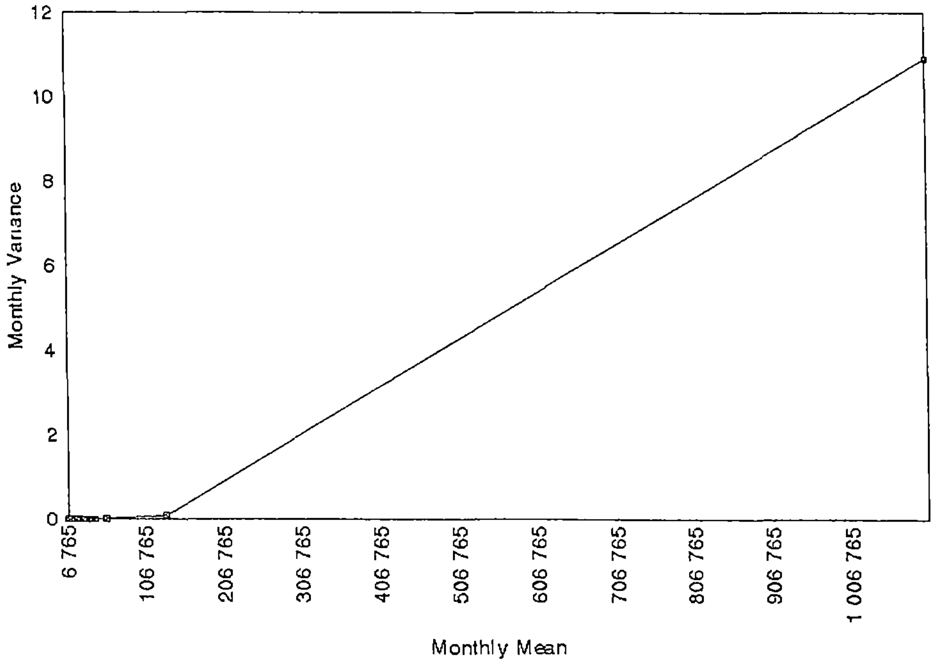
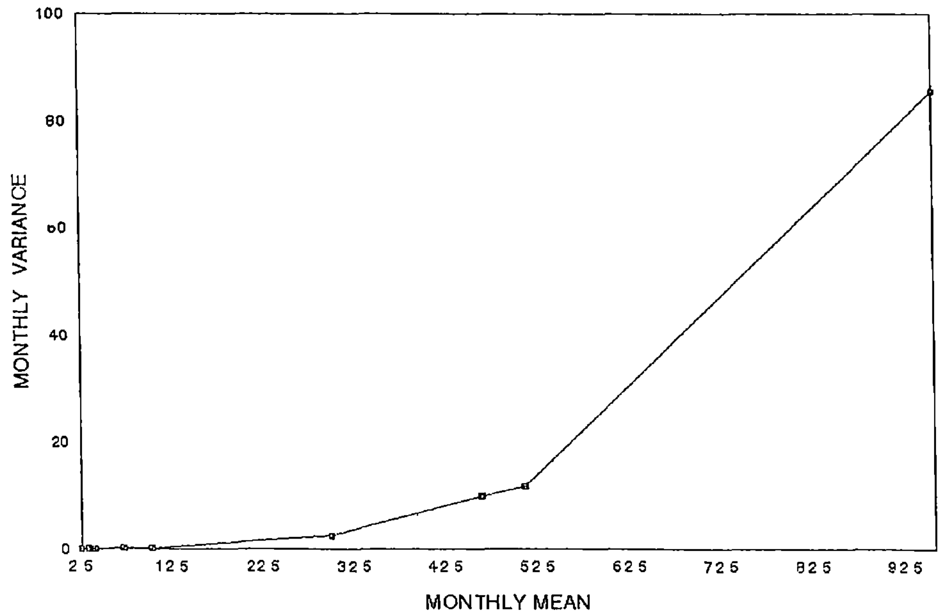


Fig 25 Relationship between mean and variance of the monthly distribution of counts of BPH



Summary

5 SUMMARY

Transformation of data from designed experiments is an old and valuable tool in developing parsimonious representation and interpretation of data Statisticians would like to transform data if such was necessary to obtain a more nearly additive normal model for performing standard statistical analysis However since several types of transformations are available to the same set of data it would be always desirable to know possibly the best transformation for a given set of data A study was therefore undertaken to examine the applicability of the various commonly used transformation techniques on the analysis of enumerative data with a view to select suitable transformations for designed experiments so as to cope with the underlying assumptions of analysis of variance The possibility of evolving new transformations for the analysis of data coming from certain specific environment was also explored

Data for the present study were collected from the available records of Regional Agricultural Research Station, Pattambi and All India Co ordinated Research Project (AICRP) on Weed Control Vellanikkara Two types of data were utilised

1) Data from pest surveillance study on paddy
2) Data from plant protection and weed control experiments

Due to the limitations of time and resources the present study was restricted to the analysis of transformations only on the dependent variable Comparisons of transformations were attempted either on the basis of a single criterion viz Barttett s Chi square test Tukey s test of non additivity Levene s residual F test and Taylor s power law or on the basis of several criteria simultaneously In the multiple criterion approach likelihood method of Box and Cox (1964) and graphical method of Draper and Hunter (1969) were used

As the time series data on pest surveillance studies did not come under the purview of analysis of variance the nature of the relationship between mean and variance of each set of sample data was examined critically before proceeding to find a suitable transformation For all sets of data, mean and variance were highly correlated indicating the need for transformating the data for restoring homoscedasticity Taylor s power law was fitted to the data on various types of insects and the coefficient of heterogeneity (b) was estimated It was found that for all major species of insects of paddy except caseworm b values were close to ~~zero~~ two which indicated the utility of logarithmic transformation for making the group mean independent of group variance Thus the study highlighted the use of logarithmic transformation in the analysis of data on counts of insects such as jassid gall fly stemborer BPH and leaf folder

In the search for a suitable transformation for experimental data on counts of insects and weeds on designed experiments three methods were mainly employed They included the likelihood method of Box and Cox (1964) graphical method of Draper and Hunter (1969) and the method based on Taylor s power law Of these three methods only the Box Cox approach produced consistent results with all sets of data The graphical plot of the log likelihood function corresponding to varying values of λ the explonent of the power transformation showed that for all sets of data the function had a maximum value around zero which indicated the possible use of a logarithmic transformation in producing a simple normal linear model Thus Box Cox approach undoubtedly emphasised the utility of the logarithmic transformation in analysing data on insects and weed counts

Graphical method of Draper and Hunter (1969) which utilised the plot of functions like F_t (F ratio for treatment Vs error) F_N (F ratio for non additivity) and F_r (F ratio for residual for testing homoscendasticity) failed to indicate a unique transformation for all sets of data Though an ideal transformation would minimise F_N

and F_r and maximise F_t no single transformation could satisfy all these requirements simultaneously However in most cases the choice lied between squareroot transformation and logarithmic transformation with a slight superiority for squareroot transformation over the others In the case of data on silver shoot (SS) produced by gall fly graphical approach of Draper and Hunter recommended squareroot transformation in place of logarithmic transformation As far as data on whorl maggot was concerned no sigle transformation was found to be decisively effective At the same time weed count data showed better response to reciprocal transformation in comparison with logarithmic and squareroot transformation Thus no hard and fast rule could be laid down with regard to the choice of a proper trasnformation as it was highly dependent on the nature of the data and extent of variability

Taylor s power law applied to the same set of data revealed the necessity of using different transformations for different sets of data In the analysis of data on weed count squareroot transformation produced better results than other transformations In the case of data on silver shoot produced by gall fly squareroot transformation was the better choice

As the data on weed count consisted of several zero values and abnormal observations the possibility of incorporating an additive constant in the linear model was examined For all sets of data, the additive constant estimated from the residual values as per the method proposed by Berry (1987) was found to be approximately 2 8 The analysis of transformed data after incorporating the additive constant c – 2 8 to each observation showed slightly better results with most of the sets of data than that with c – 1

The applicability of the inverse hyperbolic sine squareroot transformation $x^{1} = k^{1/2}\ \mathrm{Sinh}^{1}\ \sqrt{kx}$ where x is an observation and k a constant was also examined on sets of data for which standard deviation seemed to vary with mean An alternative estimate of k given by

$$k = \frac{\Sigma\, \mu^2\sigma^2 \;\; \Sigma\, \mu^3}{\Sigma\, \mu^4}$$

was derived by the principle of least squares The relative advantage of the new estimate over the earlier estimate was assessed on the basis of empirical data The new estimate of k gave better results than the earlier estimate proposed by Beal (1942)

Assuming an approximate parabolic function relationship between mean and standard deviation in the form

$\sigma = \mu + \dfrac{k}{\mu}$ (σ standard deviation μ = mean) a new transformation

$\log (x^2 + k)$ where x is an observation k a parametric value was derived An estimate of k was also obtained as

$$k = \frac{\Sigma\, \sigma/\mu - n}{\Sigma\, (1/\mu^2)}$$

when the new transformation was applied to the analysis of sets of data showing large amount of disproportionate variability the results were encouraging It was empirically found that the new transformation was more ideal than the inverse hyperbolic sine squareroot transformation

Rank transformations were in general helpful for increasing the sensitivity of F test when compared to that of the untransformed data When the performance of the two rank transformations RT 1 and RT 2 were compared no consistent superiority was noticed for one transformation over the other

References

REFERENCES

Andrews D F 1971 A note on the selection of data transformations *Biometrika* 58(2) 249 254

Anscomb F J 1948 The transformation of poisson Binomial and negative binomial data *Biometrika* 37 358 382

Atkinson A C 1978 Testing transformations to normality *Jl R Statist Soc* B 35 473-479

Bagle B G 1993 Seasonal incidence and control of thrips *Scirtothrips dosalis* Hood in pomegranate *Indian J Ent* 55(2) 148 153

Bartlett M S 1936 Squareroot transformation in analysis of variance *Jl R Statist Soc* Suppl 368 78

Bartlett M S 1947 The use of transformations *Biometrics* 3 47 52

Beal G 1942 The transformation of data from entomological field experiments so that the analysis of variance become applicable *Biometrika* 32 243 262

Berry D A 1987 Logarithmic transformations in ANOVA *Biometrics* 43 439 457

Box G E P 1953 Non normality and tests on variance *Biometrika* 40 381 335

Box G E P and Cox D R 1964 An analysis of transformations *Jl R Statist Soc* B 26 211 252

Carroll J R 1980 A robust method for testing transformation to achieve approximate normality *Jl R Statist Soc* B 42(1) 71 78

Cochran W G 1947 Some consequences when the assumption for the analysis of variance are not satisfied *Biometrics* 3(1) 23 37

Conover W J and Iman R L 1981 Rank transformations as a bridge between parametric and non parametric statistics *The Am Statistics* 35 124 133

Curtiss J H 1943 The advanced theory of statistics (2) Charles Griffin & Co ltd London pp 18

Dolbi J L 1963 A quick method for choosing a transformation *Technometrics* 5(3) 317 325

Draper N R and Hunter W G 1969 Transformations some examples revisited *Technometrics* 2(1) 23-40

Fraser D A A 1967 Data transformations and the linear model *Ann Mathe Statistics* 38:1456 1465

Hayman B L and Lowe A D 1961 The transformation of counts of the cabbage aphid *Newzealand J Sci* 4 271 278

Huckaba R M Coble H D and Van Duyn J W 1988 Joint effects of Acifluorfen application and soybean thrips (*Sericothrips variabilis*) feed ing on soybean (*Glycine max*) *Weed Science* 36 667 670

Kishorekumar and Agarwal R A 1990 Effect of systematic granular insectidices on foliar pests and yield of cotton *Indian J Ent* 52(3) 364 372

Konishi S 1981 Normalising transformations of some statistics in multivariate analysis *Biometrika* 68 647 651

Kruskal J B 1965 Analysis of factorial experiments by estimating monotone trans formations of the data *Jl R Statist Soc* B 27 251 265

Manly B F J 1976 Exponential data transformations *The Statistician* 25 37 42

Misra S S Nagia D K Ram G and Bist B S 1984 Evaluation of contact insecticides against potato cutworms *Agrotis spp* *Indian J Ent* 46(4) 460-466

Patil V C and Patil S V 1983 Studies on weed control in Bamboo *Indian J Weed Sci* 15(1) 83 86

Pericchi L R 1981 A Bayesian approach to transformations to normality *Biometrika* **68**(1) 35 43

Pushpalatha N A and Veeresh G K 1995 Suitable transformation for the popula tion count of coconut black headed caterpillar *Opisinia arenosella* Walker (Lepidoptera Zylorictidae) *Entomon* 20(1) 27 30

Sharma S S Jayaswal A P and Yadav P R 1985 Bioefficacy of various formulations of carbaryl against cotton bollworms *Indian J Ent* **47**(4) 413 421

Singh K and Gupta R N 1994 Biochemical assay of phosphamidon thiometon and dimethoate in reference to their shelf life *Indian J Ent* **56**(2) 134 139

Singh N N and Rai V N 1993 Transformation of mustard aphid counts *J Indian Soc agric Statist* 45(2) 206 213

Snedecor G W and Cochran W G 1967 *Statistical Methods* Oxford & IBH Publishing Co New Delhi 6th ed p 331 334

Taylor L R 1961 Aggregation variance and mean *Nature* **189** 732 735

Taylor L R 1970 Aggregation and the transformation of counts of aphids *Fabae scopoli* on beans *Ann appl Biol* **66** 181 189

Tukey J W 1949 One degree of freedom for non additivity *Biometrics* **5** 232 242

Tukey J W 1957 On the comparative anatomy of transformation *Ann Mathe Statist* 28 622 632

Wilcoxon F 1946 Some rapid approximate statistical procedures Am Cynamid Co New York pp 51

William W T and Stephenson W 1973 Analysis of three dimensional data (sites x species x time) in marine ecology *J Exp Marine Biol Ecol* 11 207 227

Zaman M 1990 Evaluation of foliar insecticides against the mustard aphid on rape in Peshawar (Pakistan) *Indian J Ent* 52(4) 565 569

A COMPARISON OF TRANSFORMATIONS USED IN THE ANALYSIS OF DATA FROM AGRICULTURAL EXPERIMENTS

BY

K. PRIYA MENON

ABSTRACT OF A THESIS

Submitted in partial fulfilment of the requirement for the degree of

Master of Science (Agricultural Statistics)

Faculty of Agriculture
Kerala Agricultural University

DEPARTMENT OF AGRICULTURAL STATISTICS
COLLEGE OF HORTICULTURE
VELLANIKKARA THRISSUR 680 654
KERALA INDIA

1997

ABSTRACT

A study was undertaken to empirically examine the suitability of the various commonly used transformation techniques on the analysis enumerative data relating to agricultural experiments or surveys The possibility of evolving better transformations for the analysis of data pertaining to certain specific environments was also explored

Data for the study were gathered from the available records of the project on pest surveillance survey on paddy those on the project on early stage pest control on paddy of Regional Agricultural Research Station, Pattambi and those of the post emergence herbicidal evaluation trial for the control of *Pennisetum pedicellatum* of the All India Co ordinated Research Project on Weed Control College of Horticulture Vellanikkara

Comparisons among the various commonly used transformations were made either on the basis of a single criterion viz Bartlett s chi square test Tukey s test of non additivity Levene s residual F test or Taylor s power low or on the basis of multiple criteria viz likelihood method of Box and Cox (1964) or the graphical method of Draper and Hunter (1969)

The results of the analysis of the data relating to pest surveillance study on paddy showed that logarithmic transformation was the most desirable in the analysis of data on the counts of all the major types of insects on rice (stem borer jassid gall fly leaf folder BPH) the only exception being case worm for which a squareroot transformation was indicated Box Cox approach undoubtedly emphasised the utility of the logarithmic transformation in analysing data on counts of insects and weeds The graphical plot of the log likelihood function against the exponent of the power transform had a maximum value around zero for all sets of data indicating the

superiority of the logarithmic transformation over the others The graphical method of Draper and Hunter failed to suggest a unique transformation for all sets of data However in most cases the choice lied between squareroot and logarithmic transformations with a slight superiority for the squareroot transformation

As per the method suggested by Berry (1987) a suitable location parameter C was estimated for the analysis of sets of data involving extreme observations including zero values The estimated value of the additive constant was found to be approximately 2 8 for all the different sets of data The analysis of transformated data after incorporating the estimated value of the additive constant to each observation showed slightly better results than the ordinary analysis after incorporating the additive constant one to each datum

An alternative estimate of the parametric constant in the inverse hyperbolic sine squareroot transformation was developed and the resultant estimate produced better results than those by the estimate proposed by Beal (1942)

Assuming a non linear relationship between mean (μ) and standard deviation (σ) a new transformation x $\log(x^2+k)$ where x − original observation k a parametric constant to be estimated from the data, was derived theoretically The best estimate ($\hat{k}$) of the parameter k was derived to be

$$\hat{k} \quad \frac{\Sigma\, \sigma/\mu \quad n}{\Sigma\, (1/\mu^2)}$$

where n is the number of observations

This transformation is expected to be useful in the analysis of data when the mean standard deviation relationship is approximately parabolic In general the new transformation was found to be slightly better than the inverse hyperbolic sine squareroot transformation in the analysis of data with disproportionate amount of variability

Rank transformations were also found to be helpful in the analysis of data when there are model violations and were in general helpful for increasing the sensitivity of the F test