Original Research Paper

Transformation Box-Cox for Stabilisation of Diversity in Group Random Design

Yenny Sylvia and Djunaidy Santoso

School of Computer Science, Bina Nusantara University, Jakarta, Indonesia

Abstract: In the field of research, there are a variety of statistical methods used. Hypothesis testing is one test that is often used to obtain the conclusions of the research conducted. One hypothesis testing that is often used is the testing of hypotheses through ANOVA (Box *et al.*, 1978). However, in determining the validity of a hypothesis is obtained, there are assumptions that must be met, namely the assumption of homogeneity range between treatments (Box *et al.*, 1978; Box and Cox, 1964; Gomez and Gomez, 1984; Guerrero, 1993; Mattjik and Made, 2006). Examination of a wide homogeneity assumption has been made in research involving plants, which use a randomized block design. Wide homogeneity assumption was tested using Bartlett test, but from the tests that have been done show no results were less consistent (Box *et al.*, 1978; Box and Cox, 1964; Gomez and Gomez, 1984; Guerrero, 1993; Mattjik and Made, 2006). Transformation Box-Cox used to transform data that is not homogeneous (Box *et al.*, 1978; Box and Cox, 1964). Calculation of the variance range and test computer-based transformation has been compiled using C # programming language.

Keywords: ANOVA, Assumption of Homogeneity of Variance, Test Bartlett, Transformation Box-Cox, Randomized Block Design (RAK)

Introduction

In general Research, will use statistical methods to analyze the data obtained. From the research data, will usually take a hypothesis and then will be continued to test this hypothesis. One method of this hypothesis testing can be done by Analysis of Variance (ANOVA). However, in determining the validity from the calculation of ANOVA, there are several assumptions that must be considered is the influence of additives, freedom of error/galat, homogeneity of variances and normal distribution error. So it is better if the existing data tested first according to the assumptions set because if the data is tested failed to meet one or more of the assumptions that have been mentioned, then this will affect the sensitivity of the results of the F test in the ANOVA. Therefore, any deviation of one or more of these assumptions must be corrected before continuing again into calculating ANOVA test (Box *et al.*, 1978; Box and Cox, 1964; Gomez and Gomez, 1984; Guerrero, 1993; Mattjik and Made, 2006).

Problem Formulation

Issues that will be examined in this research is formulated as follows, (a) Bartlett test will be used to test the homogeneity variance range of the experimental data research plants, (b) Box-Cox transformation method is used to transformed existing research data that has properties of a wide heterogeneity among treatments were tested in experimental plant, wich will become a new data. Furthermore, the data that has been transformed to be used in the calculation of the F test in ANOVA, (c) To facilitate and speed up the calculation, it will be made an application program for the calculation of the transformation Box-Cox research data (Box *et al.*, 1978; Box and Cox, 1964; Gomez and Gomez, 1984; Guerrero, 1993; Mattjik and Made, 2006).

Scope

In this research, the problem to be discussed is about the variance range of stability testing in a Randomized

block design (RAK) and the use of transformations Box-Cox. Here is the scope of this research study, (a) The research data used in this study were taken from the Center for Research and Development of Agricultural Biotechnology and Genetic Resources, (b) The data will be used is the result of an experiment in a randomized block design in plants, (c) Research to be done is to test to examine whether there is homogeneity in the variance range of experimental data to be analyzed, (d) Statistical calculation method to be used is the method of calculation of a randomized block design ANOVA, Bartlett test, Box-Cox transformation method and Least Significant Different test (Box *et al*., 1978; Box and Cox, 1964; Gomez and Gomez, 1984; Guerrero, 1993; Mattjik and Made, 2006).

Objectives and Benefits

The purpose of this research is (a) To detect the presence or absence of homogeneity in the data range to be tested, so that the homogeneity of the test results obtained, it can be seen that the next step should be done that is necessary to transform the data or not, (b) Using statistical sciences in examining research data that has been there, in order to get better results and proper, (c) Creating a program that can help speed up the calculation of data analysis results. The benefits of this research are (a) For the researcher: (1) Can obtain information about the research data which has a wide homogeneity properties, (2) Can obtain results more accurate data analysis, (3) Can take a more informed decision, (4) Information systems developed to simplify and accelerate the researcher in the process of calculating the test data, (b) For authors and other readers: (1) Deepen the knowledge of a wide variance of homogeneity test and the Box-Cox transformation method, (2) Directly apply statistical sciences that have been studied, (3) Apply directly sciences programming that have been studied.

Literature Review

Research and Development Center for Biotechnology and Genetic Resources (BB-BIOGEN)

Vision of BB-Biogen is a leading center (center of excellence) in research and development of agricultura biotechnology and genetic resources that can support sustainable food security and agribusiness is highly competitive. BB-Biogen program developed and formulated based on the key issues facing conventional solution which is difficult or impossible. To achieve this mission specifically BB-Biogen is (1) Increase the quantity of Human Resources (HR) qualified in the field

of agricultural biotechnology and genetic resources, (2) Manage and utilize agricultural genetic resources to support research in the field of biotechnology and plant breeding, (3) Develop a strong research program in the genetic improvement of plants and microbial properties, as well as components of agricultural cultivation technology with biotechnology approaches to technology and product BB-Biogen highly competitive, (4) Contribute to the development of national agricultural development with the development and dissemination of appropriate technologies to improve the competitiveness of Indonesian agricultural products in the national and global markets (Gomez and Gomez, 1984; Mattjik and Made, 2006).

Design of Experiments

Definition

Definition of experimental design according to (Mattjik and Made, 2006) is a test or series of tests both use statistics and inferential statistics description which aims to transform the input variables into an output which is the response of the experiment.

Basic Principles of Experimental

In an experimental design, the data were analyzed statistically valid or invalid if said data obtained from an experiment that meets the three basic principles (Mattjik and Made, 2006), namely: (1) Replay, (2) Randomization, (3) Environmental control (Mattjik and Made, 2006).

Classification of Design Experiments

Broadly speaking, the experimental design can be classified as follows (Mattjik and Made, 2006): (1) The design of the treatment, (2) Draft Environmental, (3) Draft Measurement (Mattjik and Made, 2006).

Randomized Block Design (RAK) for Complete Group

According to Pratisto (2004), a randomized block design is a form of design that is widely used in various experiments, agricultural sciences, industry and others. This design is characterized by the presence of a number of groups in which each group is given the same treatments. Through groupings, expected treatment errors can be reduced. Thus it can be said that the purpose of grouping is to reduce the diversity of experimental units in each group, or in other words we seek homogeneity within a group. In RAK to note is that the grouping is not a new variable as well as treatment. Formation of groups based on specific criteria. Linear model of a randomized block

design in general bewritten as follows (Mattjik and Made, 2006; Pratisto, 2004):

$$Y_{ij} = \mu + \tau_i + \beta_j + \varepsilon_{ij}$$

Where:

Y_{ij} = Observation on the treatment of the i-th and j-th group
μ = The average general
τ_i = Effect of treatment i-th
β_j = Effect of the j-th group
ε_{ij} = Random effect of the i-th treatment and the j-th group

Hypotheses that can be tested from a randomized block design is the effect of treatment and the effect of grouping. Form of the hypothesis can be written as follows:

Effect of Treatment

(Treatment had no effect on the observed responses) at least one i where.

Effect of Gouping

(The group has no effect on the observed responses) at least one j where. To determine the effect of treatment was tested in the experiment, can be tested using Analysis of Variance (ANOVA) (Box *et al.*, 1978; Mattjik and Made, 2006; Pratisto, 2004).

ANOVA table structure is given as follows (Table 1).

Description of Table 1 Analysis of Variance (ANOVA):

- Diversity resources for Treatment depend on the degree of diversity of sources of free(db) is t-1
- Diversity resources for Treatment depend onSum of Squares(JK) is JKP
- Diversity resources for Treatment dependon Central Square(KT) is KTP
- Diversity resources for Treatment depend onF-Count is KTP/KTG
- Diversity resources for Group depend on the degree of diversity of sources of free(db) is R-1
- Diversity resources for Group depend onSum of Squares (JK) is JKK
- Diversity resources for Group dependon Central Square(KT) is KTK
- Diversity resources for Group depend onF-Count isKTK/KTG
- Diversity resources for Error/galat depend on the degree of diversity of sources of free(db) is (t-1) (r-1)
- Diversity resources for Error/galat depend onSum of Squares(JK) is JKG

- Diversity resources for Error/galat dependon Central Square(KT) is KTG
- Diversity resources for Total depend on the degree of diversity of sources of free(db) ist r-1
- Diversity resources for Total depend onSum of Squares(JK) is JKT

Calculation steps are outlined as follows:

$$FK = Factor\ Correction\ FK = \frac{y_{..}^2}{tr} = \frac{(\sum\sum y_{ij})^2}{tr}$$

JKP = Sum of Squares of Treatment :

$$JKP = \sum_{i=1}^{t}\sum_{j=1}^{r}(\overline{Y}_{i.} - \overline{Y}_{..})^2 = \sum \frac{Y_i^2}{r} - FK$$

JKK = Sum of Squares Group :

$$JKK = \sum_{i=1}^{t}\sum_{j=1}^{r}(\overline{Y}_{j} - \overline{Y}_{..})^2 = \sum \frac{y_j^2}{t} - FK$$

JKG = Sumof Squares Error :

$$JKG = \sum_{i=1}^{t}\sum_{j=1}^{r}(Y_{ij} - \overline{Y}_{i.} - \overline{Y}_j + \overline{Y}_{..})^2 = JKT - JKP - JKK$$

JKT = Sumof Squares Total :

$$JKT = \sum_{i=1}^{t}\sum_{j=1}^{r}(Y_{ij} - \overline{Y}_{..})^2 = \sum\sum Y_{ij}^2 - FK$$

KTP = Central squares of treatment
KTP = JKP/db treatment

CEC = Middle quadrat group
CEC = JKK/db group

KTG = Central Squares error/galat
KTG = JKG/db error/galat

Where:
T = number oftreatments
R = number ofreplications

Table 1. Analysis of Variance (ANOVA) (Box *et al.*, 1978; Mattjik and Made, 2006; Pratisto, 2004)

Diversity resources	The degree of diversity of sources of free (db)	Sum of squares (JK)	Central square (KT)	F-count
Treatment	t-1	JKP	KTP	KTP/KTG
Group	R-1	JKK	KTK	KTK/KTG
Error/gallat	(t -1) (r-1)	JKG	KTG	
Total	tr-1	JKT		

Testing of Hypothesis

Fcount = KTP/KTG follow the F distribution with degrees of freedom for t-1 numerator and denominator degrees of freedom equal to (t-1) (r-1). If the value Fcount is greater, than the null hypothesis is rejected and the opposite applies.

Fcount = KTK/KTG follow the F distribution with degrees of freedom for r-1 numerator and denominator degrees of freedom equal to (t-1) (r-1). If the value is greater than the null hypothesis is rejected and the opposite applies (Box *et al.*, 1978; Mattjik and Made, 2006; Pratisto, 2004).

Coefficient of Diversity

Value of coefficient of variance showed that the level of accuracy compared with the treatment and a good index of the state of the experiment. It is expressed as a percentage of the average experimental error/galat, so that the greater the diversity coefficient, the lower the reliability of the experiment. Calculation of the value of diversity coefficient (KK) can be done after the value of variety/error experiments obtained. Steps of calculation coefficient values diversity in RAK described as follows:

$$KK = \left(Square\ (KTG) * 100\% \right) / general\ everage$$

$$General\ everage := \sum\sum Y_{ij}\ /\ \text{tr}$$

Where:
Y_{ij} = the value of the i-th observation of treatment and replicates/group to-j

Least Significant Difference Test (LSD/BNT Test)

Least Significant Difference test (LSD) is a method used to test for differences between treatments were tested. LSD made if H0 conclusions on the ANOVA F test was rejected. LSD decision criteria are, If the difference between the two treatments on average larger than the LSD/BNT value of both treatments are significantly different at significance level alpha, If the difference between two average treatment less than or equal to the value of the BNT, the two treatments were not significantly different at significance level alpha. Steps of Least Significant Difference test calculations are:

$$LSD\ /\ NT = t_{a/2;\ db\ error\ /\ galat} Square\left(2KTG\ /\ r \right)$$

where, t_α distribution with degrees of freedom equal to (t) (r-1) (Box *et al.*, 1978; Box and Cox, 1964; Gomez and Gomez, 1984; Guerrero, 1993; Mattjik and Made, 2006; Pratisto, 2004).

Variety Analysis (ANOVA)

Definition

According to Walpole *et al.* (1995), analysis of variance is a method to describe the total diversity of data into components that measure various sources of variability. Analysis of variance was used to test some middle value simultaneously.

Assumptions Analysis of Variance

ANOVA calculation results will be valid if one or more mathematical assumptions about the data are met. Assumptions-The assumptions are (1) Effect of additives, (2) Freedom error, (3) Homogeneity range, (4) Normal spreads.

Homogenity Range

Wide heterogeneity can be classified into two types (Gomez and Gomez, 1984), namely: (1) Where the change is a variety of functions with the average relationship, (2) where there is no relationship between the range and the averaging function.

Variety Test Homogeneity (Bartlett Test)

Bartlett test is a test that can be used to test the homogeneity range.

Hypotheses to be tested are:

$$H_0 : \sigma_1^2 = \sigma_2^2 = ... = \sigma_k^2$$
$$H_1 : \sigma_1^2 \neq \sigma_2^2 \neq ... \neq \sigma_k^2$$

Calculation of the Bartlett test using chi-squared distribution approach with k-1 degrees of freedom.

Calculation steps are as follows:

$$s_p^2 = \frac{\sum_{i=1}^{k} s_i^2}{k}$$

$$\chi^2 = \frac{(2,3026)(f)(k\log s_p^2 - \sum_{i=1}^{k} \log s_i^2)}{1 + [(k+1)/3kf]}$$

Where:
Si^2 = Range of ith treatment
K = number of treatments
F = number of replications-1

If, then accepted and mean homogeneity range fulfilled. Otherwise it is rejected.

Transformation Box-Cox (Box et al., 1978; Box and Cox, 1964)

Data transformation is a common technique used to solve the measurement problem where a wide range of heterogeneity and Average have a relationship function. With the transformation, the original data is converted into a new scale and make a new group which is expected to eventually be fulfilled so that the stability range can approach the validity of the testing process. One of transformation that is used is the transformation of Box-Cox.

The general form of the transformation Box-Cox is:

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

$Y(\lambda)$ is a transformation of the Box-Cox with the y at specific variable lambda (λ). In the transformation of Box-Cox, heavily dependent on the value of lambda (λ) are used.

Therefore the value of lambda will be sought and will be raised on the original data to obtain new data that has been transformed.

Lambda value calculation can be done with two approaches, by using range and standard deviation.

With a variety of approaches (Guerrero, 1993):

$$\lambda = \tfrac{1}{2}(2 - b)$$

To find the value of a and b used formula (Walpole *et al.*, 1995):

$$b = \frac{n\sum_{i=1}^{n} x_i y_i - (\sum_{i=1}^{n} x_i)(\sum_{i=1}^{n} y_i)}{n\sum_{i=1}^{n} x_i^2 - (\sum_{i=1}^{n} x_i)^2}$$

$$a = \bar{y} - b\bar{x}$$

Where:
X_i = LOG evereges
Y_i = Log variance

Approach with a standard deviation (Box *et al.*, 1978) $\lambda = 1-b$.

To find the value of a and b used the formula:

$$b = \frac{n\sum_{i=1}^{n} x_i y_i - (\sum_{i=1}^{n} x_i)(\sum_{i=1}^{n} y_i)}{n\sum_{i=1}^{n} x_i^2 - (\sum_{i=1}^{n} x_i)^2}$$

$$a = \bar{y} - b\bar{x}$$

Where:
X_i = Log average
Y_i = Log standard deviation

Software Engineering (Pressman, 2010)

Software engineering is defined by the Fritz Bauer (Pressman, 2010) is the activity of applying and using the principles of good engineering in order to produce software that is economical, reliable and work efficiently. In software engineering, there are three main elements (Pressman, 2010), namely: (a) Process (b) Methods (c) Tools. In software engineering there are various kinds of models. Models are often commonly used is Waterfall model. This model is called the waterfall model for each running process must first wait for the completion of the process previously and run sequentially. In general, the stages in the waterfall model can be seen in the following (Fig. 1).

According to Pressman (2010), there are 6 stages in the waterfall model are (1) System/Information Engineering and Modeling, (2) Software Requirements Analysis, (3) Design, (4) Coding, (5) Testing and (6) Maintenance.

Human-Computer Interaction (Shneiderman and Plaisant, 2010)

Five criteria that must be met by a system that is user friendly (Shneiderman and Plaisant, 2010) are (1) Learning time is not long.

(Time to learn), (2) The rapid pace of information presentation. (Speed of performance), (3) Use of low error rate. (Rate of errors by users), (4) After a period of time beyond memorization. (Retention over time), (5) Personal satisfaction of users who used it. (Subjective satisfaction). In designing an interface, there is a rule that has been known as the Eight Golden Rules of Interface Design (Shneiderman and Plaisant, 2010), namely: (1) Trying to be consistent. (Strive for consistency), (2) Allows users to use shortcuts. (Enable frequent users to use shortcuts), (3) Provide informative feedback. (Offer informative feedback), (4) Designing the dialogue to produce a final state. (Design dialogs to yield closure), (5) Provide error prevention and simple error handling. (Offer error prevention and simple error handling), (6) Allow easy reversal of actions.
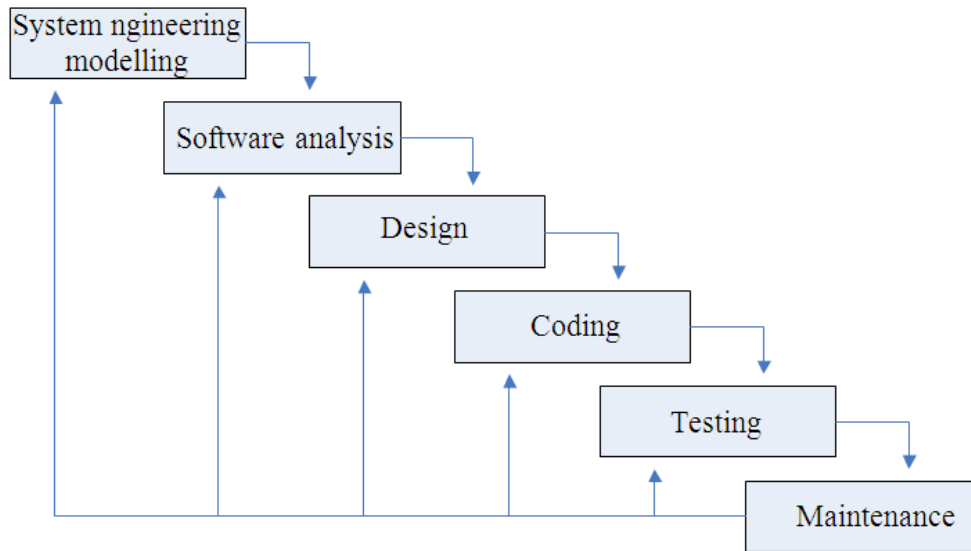
Fig. 1. Waterfall model (Pressman, 2010)

(Permit easy reversal of actions), (7) Provides internal control. (Support internal locus of control), (8) Reducing the burden of memorization. (Reduce short-term memory load).

Database

According Farthansyah, 2004, the database is one component of the system data base. The database consists of three things: The well-organized collection of data, relationships between data and objective. There are many options to organize the data and there are a lot of considerations in forming relationships among the data, but in the end the important thing is the main objective that we should always remember that the speed and ease of interacting with managed data/processed.

Microsoft Excel

Microsoft Excel is a powerful spreadsheet program that works under windows operating system. There are many conveniences that obtained during use microsoft excel as working with lists of data, calculate numbers, create reports, charts and graphs. Microsoft Excel can also be used to process statistical data. Input data from excel can also be connected and read the application that created the program.

State Transition Diagram

State Transition Diagram (STD) is used to describe the time dependence of the properties contained within the system. Components-the main component of the

STD are (1) State, (2) Transition, (3) Condition and Action:

Flowchart

Is a flow chart diagram showing the sequence and relationships between processes and their instructions. This diagram is denoted by the symbol, so each symbol illustrates a particular process. Symbols used such as rectangular, rhombus and oval are used to express the operation, while the relationship between the process described by a hyphen. But an outline flowchart consists of three main parts, namely: Input, Process, Output. Flowchart compiled with symbols. This symbol is used as a tool to describe the process in the program.

Materials and Methods

Analysis of Issues

ANOVA test results obtained from the experimental design will be valid or invalid value if one of the assumptions that have been set. One such assumption is the homogeneity range and testing of these assumptions is a test that is used by the author. Of the assumptions of the test results obtained conclude whether or not the data is homogeneous. If it is not homogeneous, then there is a method to homogenize the data. One commonly used method is the method of transformation. Transformation is a writer who used the transformation method Box-Cox. The variables that are used to support these

simulations is the value of the treatment and test scores of the experiment. By entering the data-data that has been available, it will obtain a homogeneous assay results, the value of lambda (otherwise homogeneous) which will be used to get the value of the new data, ANOVA and LSD values.

Data Collection

The data used are secondary data in the form of crop research data obtained from the Center for Research and development of agricultural Biotechnology and Genetic resources (BB-BIOGEN). Based on the research data will be retrieved value of the treatment and test scores.

Place and Time Research

Research conducted at the Center for Research and Development of Agricultural Biotechnology and Genetic Resources (BB-BIOGEN) in September-November 2008.

Data Collection Method

The data and information required in this study were collected with the following steps (a) Research field, where the data and information gathered is obtained by direct observation in an institution that became the object of this research study. The research was done by (1) Observation, which conducted direct observation of activities in the field related to the problem under study, (2) Interview, where questions related to the data that is needed in this research study submitted to the company.

Data Analysis Techniques

Analysis techniques that will be used to analyze the data is to use a test of homogeneity of variance (Bartlett test) to test the homogeneity of the data and the transformation of the Box-Cox to get the value of lambda functions in getting new data. Stepsto process the data with the test and transformation Bartlett Box-Cox can be seen (1) Collect the data, (2) Visualisation variety of relationships and the average of the data in the form of scatter plots, (3) The data obtained were then tested homogeneity using Bartlett test manifold. Bartlett test statistic used in the test of chi-squared with degrees of freedom k-1. If the H0 is accepted, which means that the data is tested is homogeneous, (4) If the data is homogeneous, followed by calculation of ANOVA and coefficient of variance, (5) If the conclusion of F count in ANOVA rejected, followed by Least Significant Difference/BNT (LSD) test, (6) If data is not homogeneous, then continued

with the transformation Box-Cox to find the value of lambda which serves to transform the old data into the new data, (7) Once the data is transformed, then proceed with the calculation of ANOVA, coefficient of variance and LSD test (if the calculated F value in ANOVA rejected) (Box *et al.*, 1978; Box and Cox, 1964; Gomez and Gomez, 1984; Guerrero, 1993; Mattjik and Made, 2006; Pratisto, 2004).

Screen Design (Pressman, 2010; Shneiderman and Plaisant, 2010)

The design of the main screen for calculation program consists of a screen with two tabs (Main and Box-Cox), this is so that users can easily use it. Function to be placed in the calculation of the main screen. There are 2 buttons ie browse and count. Serves browse button to locate and open the excel file desired by the user. Count function button to display the results of calculation and testing data that includes display scatter plots, ANOVA and all the conclusions from the data calculations (Fig. 2).

The display screen on the main tab and the main screen is the same. When a user opens the main screen, meaning users are in the main tab screen. Here the user will press the browse button to open the excel data is desired. After showing the data, if the user has not pressed the button count, so that users will see is just excelnya data display. To do as well as see the results of the calculation process, the user must press the calculate button. Calculation results are shown is a scatter plot of the data, the value of chi-square, the value of chi-squared table, conclusion which states homogeneous or heterogeneous, the coefficient of variability, the value of Least Significant Difference (LSD), the transformation of Box-Cox stating yes or no need for transformation, the value of lambda, ANOVA and conclusions. LSD/BNT will exit if the value in the ANOVA rejected the conclusion that F count>F. Lambda values and calculation results tab Box-Cox would come out if the results in the transformation of the Box-Cox is yes. Tab Box-Cox will work if the results of the transformation of the Box-Cox on the main tab stated. Display the data that is displayed on this tab screen is value-the value of the data that has been transformed and their averages, a scatter plot of data transformation, the data value of chi-squared transformation (χ^2), the value of chi-squared table ((χ^2 table), conclusion which states homogeneous transformation or heterogeneous, the coefficient of variability of data transformation, the value of Least Significant Difference (LSD) of data transformation, transformation of data and conclusions ANOVA.
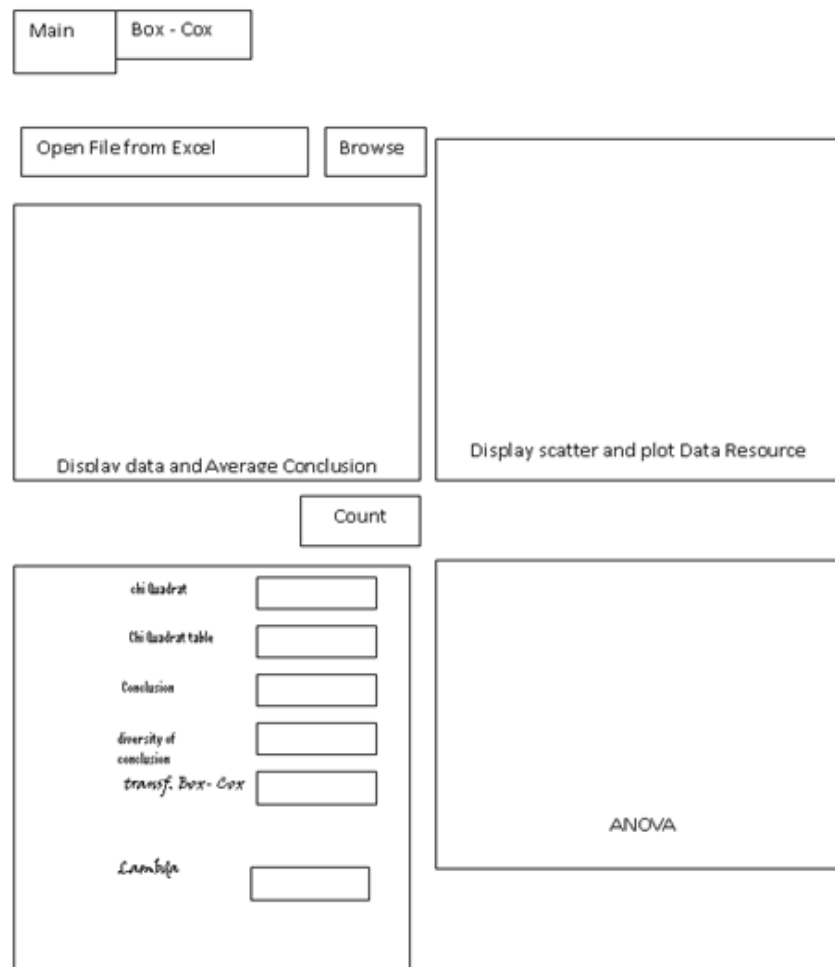
Fig. 2. The design of the Main Screen of Graphical User Interface of the Research with the goal userfrendly (Pressman, 2010; Shneiderman and Plaisant, 2010)

BNT values will exit if the value in the ANOVA rejected the conclusion that the value F count>F.

Steps for application Program with algorithm as follow:

(1) Start
(2) Browsing file Exel
(3) If calculation ? then goto (5)
(4) Else Goto Start
(5) Calculate Scatter Plot
(6) Calculate χ^2, χ^2 table → if conclution ? no goto (8) homogen
(7) yes Heterogen : (a) Calculate Lambda
   (b) data transformation goto (6)
(8) Homogen : (x) Calculate correction factor
(y) calculate JKP, JKK, JKG, JKT
(z) calculate KTP, KTK, KTG, KK

(w) Calculate F, F table
If conclusion? yes: Accepted goto end
Reject: Test BNT(LSD/BNT)
end: end
   Steps for module Calculation of:
   Module scatter plot:

(1) input x,y
(2) plot scatter
(3) return

Module calculation χ^2 and table of χ^2:
(1) Input 2.3026,f,k,log S_p^2, log S_i^2,1,3,α:

$$(2)\ \chi^2 = \frac{(2,3026)(f)(klogs_p^2 - \sum_{i=1}^{k} logs_i^2)}{1+[(k+1)/3kf]}$$

(3) χ^2 table = χ^2 (α ; x-1)
(4) Print χ^2, χ^2 table
(5) Return

Module Calculate correction factor:

(1) Input Y_{ij}, t, r
(2) I=0
(3) J=0
(4) $\sum\limits_{i=0}^{m}\sum\limits_{j=0}^{m} Y_{ij} = A_{ij}$
(5) If i<m then i = i+1 goto (3)
(6) If j<m then j = j+1 goto (4)
(7) Calculate A_{ij}
(8) Calculate A_{ij}^2/t r
(9) Print Results
(10) Return

Module Calculate JKP, JKK, JKT, JKG:

(1) Input t,r, FK,Yi, Yj, Yij
(2) Process JKP = Σ Yi^2/r-FK
     Process JKK = ΣYj^2/t-FK
     Process JKT = $\Sigma\Sigma$Yij^2/-FK
     Process JKG = JKT-JKP-JKK
(3) Print JKP, JKK, JKT, JKG
(4) Return

Module Calculate KTP, KTK, KTG, KK:

(1) Input JKP, JKK, JKG, treatment db,
     Group db, galat db, t, r, Yij
(2) Process KTP = JKP/treatment db
     KTK = JKK/group db
     KTG = JKG/galat db
     KK = (SQRT(KTG)/$\Sigma\Sigma$Yij/t r)*100%
(3) Print KTP, KTK, KTG, KK
(4) Return

Module Calculate F, table F:

(1) Input KTP,KTK,KTG,t,r, α
(2) Process treatment F = KTP/KTG
     group F = KTK/KTG
     treatment table F = Fα,t-1;
     (t-1)(r-1)
     group table F = Fα,r-1,
(t-1)(r-1)
(3) Print F, table F
(4) Return

Module Calculate Lambda:

(1) Input n, Xi, Yi
(2) Process:

$$b = \frac{n\sum\limits_{i=1}^{n}x_i y_i - (\sum\limits_{i=1}^{n}x_i)(\sum\limits_{i=1}^{n}y_i)}{n\sum\limits_{i=1}^{n}x_i^2 - (\sum\limits_{i=1}^{n}x_i)^2}$$

Lambda = ½(2-b)
(3) Print Lambda
(4) Return

Module Calculate data transformation:

(1) Input Y, λ
(2) Process Y$^\lambda$
(3) Print data transformation
(4) Return

Module Testing BNT/LSD:

(1) Input α, db, galat, KTG, r
(2) ProcessBNT = T$_{\alpha/2}$,dbgalat*SQRT(2KTG/r)
(3) Print BNT
(4) Return

Design State Transition Diagram (STD): (Fig. 3).

Design Specifications

Software Specifications (Pressman, 2010; Shneiderman and Plaisant, 2010)

Specification of the software used at the time of the application program designing are (1) Operating System Microsoft Windows XP Professional, (2) Programming language Microsoft Visual Studio 2005, using Visual C#2005, (3) Microsoft Excel 2003 and Microsoft Access 2003, (4) Infragistics.

Hardware Specifications

Hardware specifications are used when designing the application program are (1) Intel Pentium 4 processor 1.80 GHz, (2) Memory 512 Mb, (3) Hard disk capacity of 40 GB.

Results for this Research

Description of Research Data

Research data have been obtained and used in this study includes data from four experiments are all plant

data. The fourth trial data are (1) Experiments the effect of lime on the corn cob on the number and weight of seeds sterile corn, Amount of sterile Corn Cob, In Table 2, K0 treatment is the control treatment. Table 2, looks range from the data are divided into 2 groups where lime treatment K0, K1 and K2 has a great range of values, while the lime treatment K3, K4 and K5 has a small

range of values. Lime treatment K0 has the largest range of variety and value, while the lime treatment K5 has the smallest value of the variety and range (Fig. 4).

From Table 3, it can be seen that the average number of cobs sterile because K0 lime treatment had average-the highest average, while the average treatment on the K5 has-the lowest average (Fig. 2).

1. View the data transformation and its average
2. View scatter plot
3. Show the value of Chi Square
4. Show table of the value of Chi Square
5. View conclusion
6. Show the Coefficients of variance
7. Show BNT or LSD
8. Show ANOVA

Show result of calculation of Box Cox

click box-cox
Main

Click browse
Show open dialog

Click calculation

Show result of calculation

1. View the data transformation and its average
2. View scatter plot
3. Show the value of Chi Square
4. Show table of the value of Chi Square
5. View conclusion
6. Show the Coefficients of variance
7. Show BNT or LSD
8. Show result of box-cox transformation
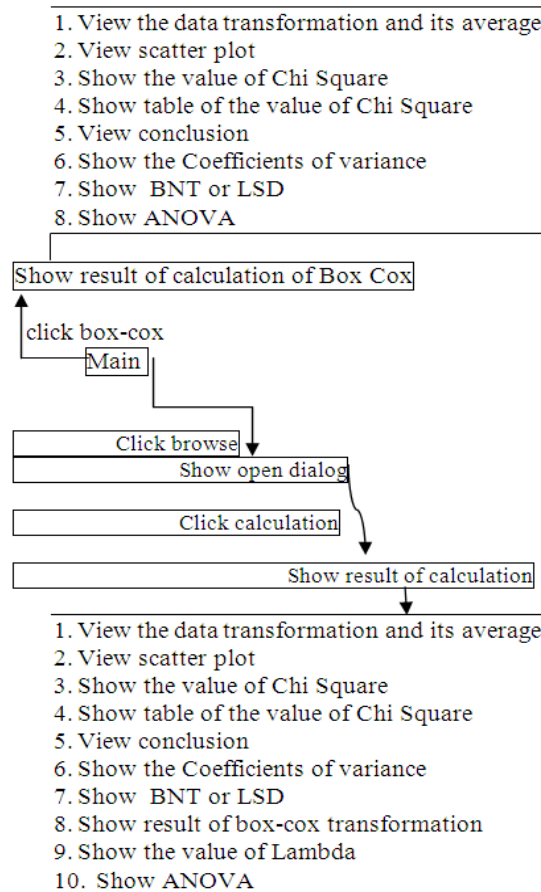9. Show the value of Lambda
10. Show ANOVA

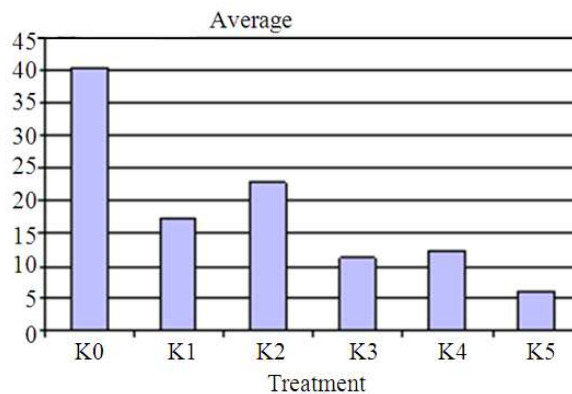Fig. 3. STD program application (Pressman, 2010; Shneiderman and Plaisant, 2010)

Fig. 4. Is the Effect of lime against average number of sterile corn cob, that the K0 = 40 the greatest effect of lime and the K5 the small effect of the lime

Table 2. Data observations lime treatment against sterile corn cob number

Flour/powder treatment of corn	Testing I	II	III	Range	Variation	Average
K0	51	30		21	110,333	40,333
K1	19	24	9	15	58,333	17,333
K2	32	21	15	17	74,333	22,667
K3	8	12	14	6	9,333	11,333
K4	10	12	15	5	6,333	12,333
K5	6	8	4	4	4	6

Table 3. Data observations treatment for test BNT and average

Treatment herbicides	Average	Test BNT
Kontrol	213	
Manual	5,667	207,333*
Oksadiazon	23,333	189,667*
Benthonil	10	203*
2,4D amine	42	171*

Table 4. Data observations treatment of herbicide everage and test BNT

Treatment of herbicide	Herbicide average	Test BNT
Control	4,257	
Manual	1,599	2,657*
Oksadiazon	2,350	1,907*
Benthonil	1,873	2,384*
2,4D amine	2,771	1,485*

Maize seed weight With the same Experiment we have others Table that the Lime treatment K0 is the control treatment. From this table, note the data range for lime treatment K1, K2, K3, K4 and K5 has a great range of value, while the value range for the lime treatment K0 has a small range of values. Treatment K1 has the largest range of variety and value, while the K0 treatment has a value range and the smallest range. LSD/BNT count value is greater than the LSD/BNT value table. LSD/BNT Table is Experiment Influence Treatment Insecticides Against Rice Pests Number.

Discussion for this Research

From Table 3 and 4, note that for the data before and after transformations, experimental effects of herbicide treatment manual, oksadiazon, benthonil and 2,4 D amine to weight weed/grass control with herbicide treatments, has a different herbicide treatments.

The Original Data (Box et al., 1978; Box and Cox, 1964; Gomez and Gomez, 1984; Guerrero, 1993; Mattjik and Made, 2006; Pratisto, 2004)

From the calculated BNT/LSD values obtained:

$$BNT = t_{(\alpha/2; db\ galat)}\sqrt{\frac{2KTG}{r}} = 89,688$$

Table 4 BNT original data table to trial against herbicide treatment effect weight weeds/grass at rice.

Data Transformation (Box et al., 1978; Box and Cox, 1964; Gomez and Gomez, 1984; Guerrero, 1993; Mattjik and Made, 2006; Pratisto, 2004)

From the calculated BNT values obtained:

$$BNT = t_{(\alpha/2; db\ galat)}\sqrt{\frac{2KTG}{r}} = 0,583$$

Table 4 BNT data transformation to trial against herbicide treatment effect of weight weeds/grass on rice.

Conclusion

Discussion of the results of experiments that test 8 data examined using Bartlett test method, transformation Box-Cox, ANOVA and LSD test, it was concluded as follows (1) Bartlett test conducted mostly not successful enough to prove whether the range of experimental data is homogeneous or heterogeneous. This is evidenced by the 2 experimental (treatment insecticide against rice pests and the amount of lime treatment on the number of sterile corn cobs) were there, which is visualization, the data clearly seen not homogeneous, but the Bartlett test results indicate otherwise, (2) Transformation function Box-Cox went pretty well, because of the existing experiments in which the results of a variety of heterogeneous data, after the data is transformed into a homogeneous variety, (3) Programs that have been made provide analytical results quickly and accurately. The advice can be given of the results of this study are (1) To test homogeneity of variance would be better to try other methods such as test Neyman-Pearson to get better results, (2) Application programs can be developed further with the addition of other statistical calculations in order to be better.

Funding Information

Author's Contributions

All authors equally contributed in this work.

Ethics

This article is original and contains unpublished material. The corresponding author confirms that all of the other authors have read and approved the manuscript and no ethical issues involved.

References

Box, G.E.P., W.G. Hunter and J.S. Hunter, 1978. Statistics for experimenters: An Introduction to Design, Data Analysis and Model Building. 1st Edn., Wiley, New York, ISBN-10: 0471093157, pp: 653.

Box, G.E.P. and D.R. Cox, 1964. An analysis of transformations. J. Royal Stat. Society, 26: 211-252.

Gomez, K.A. and A.A. Gomez, 1984. Statistical Procedures for Agricultural Research. 1st Edn., John Wiley and Sons, New York, ISBN-10: 0471870927, pp: 680.

Guerrero, V.M., 1993. Time-series analysis supported by power transformations. J. Forecast., 12: 37-48. DOI: 10.1002/for.3980120104

Mattjik, A.A. and I.S. Made, 2006. Design of Experiments with SaS and Minitab Applications. 1st Edn., Press IPB, Bogor.

Pratisto, A., 2004. Easy ways to overcome problems of statistics and experiment.

Pressman, R.S., 2010. Software Engineering: A Practitioner's Approach. 7th Edn., McGraw-Hill, New York, ISBN-10: 0071267824, pp: 895.

Shneiderman, B. and C. Plaisant, 2010. Designing the User Interface: Strategies for Effective Human-Computer Interaction. 5th Edn., Illustrated, Pearson Education India, Addison-Wesley, ISBN-10: 0321537351, pp: 606.

Walpole, J., C. Pu, T. Autrey, A. Black, C. Consel *et al.*, 1995. Optimistic incremental specialization: Streamlining a commercial operating system. Proceedings of the fifteenth ACM symposium on Operating systems principles, Dec. 03-06, Copper Mountain, CO, New York, pp: 314-321. DOI: 10.1145/224056.224080