IRAQI Academic Scientific Journals

University of Kirkuk Journal For Administrative and Economic Science (2024) 14 (4): 259-267

UKJAES

University of Kirkuk Journal For Administrative and Economic Science

ISSN 2222-2995
University of Kirkuk Journal For Administrative and Economic Science

Abdel Amir Muthana Ibrahim & Ahmed Ahmed Dheyab. Estimating the Missing Value of Sugars Data in a Restricted One-Way Analysis of Variance Model. *University of Kirkuk Journal For Administrative and Economic Science* (2024) 14 (4):259-267.

Estimating the Missing Value of Sugars Data in a Restricted One-Way Analysis of Variance Model

Muthana Ibrahim Abdel Amir [1], Ahmed Dheyab Ahmed [2]

[(1,2)] *University of Baghdad, College of Administration and Economics, Department of Statistics, Baghdad, Iraq*

mothana.abd2201m@coadec.uobaghdad.edu.iq [1]
ahmedstatistic@coadec.uobaghdad.edu.iq [2]

Abstract: The issue of missing values has attracted much attention from many researchers, sometimes an observation may be dropped in a particular experiment or study, and thus one or more observations may be missing in the data. In this paper, an estimate of the missing value for sugars data was obtained in a restricted one-way analysis of variance model, assuming that the error follows a long-tailed symmetric distribution. The problem of missing values was highlighted, and discuss different methods of estimating them, through methods specific to the restricted one-way analysis of the variance model that seeks to make the sum of squared errors as small as possible And compare the data in terms of its distribution Long-Tailed Symmetric Distribution (LTS) or normal distribution(N) through use Akaike information criterion (AIC) , These methods include the arithmetic mean method (M), the maximum likelihood method (MLE), and the modified maximum likelihood method (MMLE), and an attempt is made to compare these methods used in estimating the missing value of real agricultural data through the effect of adding the concentration of three types of sugars to the length of the multiplying branches. (cm) in the rootstock of tissue-rich apples (MM106). Important results have been reached regarding the research, the most important of which is that it has been proven that the modified maximum likelihood method is more efficient than the rest of the estimation methods used.

Keywords: Missing Values, Restricted One - Way Analysis of Variance Model, Long-Tailed Symmetric Distribution, Modified Maximum Likelihood Method, Akaike information criterion.

تقدير القيمة المفقودة لبيانات السكريات في أنموذج تحليل التباين احادي الاتجاه المقيد

الباحث: مثنى إبراهيم عبد الامير [١]، أ.د. احمد ذياب احمد [٢]

[(١،٢)] *جامعة بغداد، كلية الادارة والاقتصاد، قسم الاحصاء، بغداد، العراق*

المستخلص: لقد جذبت مشكلة القيم المفقودة الكثير من الاهتمام من قبل العديد من الباحثين، فمن الممكن في بعض الأحيان قد يتم إسقاط مشاهدة في تجربة أو دراسة معينة، وبالتالي قد تكون مشاهدة واحدة أو أكثر مفقودة من البيانات، في هذا البحث، تم الحصول على تقدير للقيمة المفقودة لبيانات السكريات في أنموذج تحليل التباين احادي

الاتجاه المفيد بأفتراض ان الخطأ يتبع توزيع طويل الذيل المتماثل (Long-Tailed Symmetric Distribution). وتم تسليط الضوء على مشكلة القيم ألمفقودة ومناقشة الطرائق المختلفة لتقديرها، من خلال أساليب خاصة بأنموذج تحليل التباين احادي الاتجاه المفيد الذي يسعى إلى جعل مجموع مربعات الأخطاء أقل ما يمكن ومقارنة البيانات من حيث توزيعها طويل الذيل المتماثل(LTS) او التوزيع الطبيعي N)) من خلال استعمال معياراكايكي للمعلومات (AIC), ومن هذه الطرائق طريقة المتوسط الحسابي (M) و طريقة الامكان الاعظم (MLE) وطريقة الامكان الاعظم المعدلة (MMLE) ومحاولة المقارنة بين هذه الطرائق المستعملة في تقدير القيمة المفقودة لبيانات زراعية حقيقية وذلك من خلال تأثير اضافة تركيز ثلاثة انواع من السكريات في طول الافرع المتضاعفة (سم) في اصل التفاح MM106)) المكثر نسيجيا وقد تم التوصل الى نتائج مهمة خاصة بالبحث اهمها انه ثبت ان طريقة الامكان الاعظم المعدلة اكثر كفاءة من باقي طرائق التقدير المستعملة.

الكلمات المفتاحية: القيم المفقودة، انموذج تحليل التباين احادي الاتجاه المفيد، توزيع طويل الذيل المتماثل، طريقة الامكان الاعظم المعدلة، معيار اكايكي للمعلومات.

Corresponding Author: E-mail: mothana.abd2201m@coadec.uobaghdad.edu.iq

Introduction

Increasing the frequency of treatment in the experiment leads to an increase in the accuracy of the experiment, which results in an increase in the degree freedom of experimental error and thus a reduction in the mean squares error (MSE) , so the loss of some observations leads to a decrease in the number of treatment appearances and the appropriate procedure is to estimate the values of the missing observations .

For different reasons, some observations are missing , For example the researcher may fail to record some data, or crops may be destroyed in some plots of land, or a patient may not take a treatment, or during an experiment one or more animals may die.

The problem of missing values is a common problem in almost all types of research and studies, and deleting missing values naturally affects the analysis method, so before analyzing the data, the missing values are estimated.

In this paper, the missing value is estimated using the estimation methods used in the theoretical side and assuming that the random error is long-tailed symmetric distribution, and it is possible that the random error distribution is a logistic distribution as in the source (Ahmed, Abdulwahhab & Abdulah, 2019). An exponential distribution as in the source (Ahmed, Abdulwahhab & Abdulah, 2020), but in this research we will assume that the random error is distributed (LTS) and where the Akaike information criterion(AIC) is used to know how the data is distributed, and there are two types of linear models for one-way analysis of variance, which is the analysis model Restricted one-way variance or fixed and random variance. In this paper, we will focus on the restricted one-way variance analysis model.

Many researchers have addressed the problem of estimating the lost value in different ways and methods , In (1933) ,(Yates) presented two formulas for estimating more than one missing value. An iterative procedure is suggested , In (1958) ,Show (Wilkinson) When some observations are missing from the data otherwise It is consistent with the planned experimental design and linear model fitting The principle of least squares is applied to the data very simply , Use estimated values for missing observations in this way The most effective use is the residual symmetry of the experiment. In (1972), (Rubin) used the non-iterative method to estimate missing values in a randomized complete block design (RCBD) and a Latin square design (LSD) using the least square method by making the sum of the error squares Equal to zero. In (1995) , researcher (Reynolds) derived formulas to estimate missing observations for analysis of variance by minimizing the sum of the squares of the residuals. These formulas apply to cases in which there is one missing observation. If there are two missing observations, the recommended procedure is to estimate one and use the appropriate formula to estimate the other, then use the formula again to estimate the first, and continue iteratively until the estimates converge to arrive at the two missing observations. The purpose is to have unambiguous formulas for the two missing observations and thus compute them directly without the need for an iterative procedure. In (2007), the researchers (Dabdoub and Hadba) used four methods to estimate missing observations in a completely randomized block design, namely (Haseman and Gaylar, Rubin, Harry, and Yates). The comparison between these methods was made according to the following statistical criteria, which is the average The mean square error (MSE) and the mean square of the treatments (MSt) are the values of the significant difference between the treatments. It has been shown that the most efficient methods are the Haseman and Gaylar method, followed by the Harry method, and then the Yates method. The Rubin method failed to keep up with the other three methods. In (2016), the researcher (Al-Qazzaz) presented a proposed method for estimating the missing values of observations of the explanatory variables of the nonparametric multiple regression model and comparing it with the method of substituting the mean, as the idea of this method was based on how to employ the causal relationship between the variables and obtain an efficient estimate of the value. missing, through the use of kernel estimation, represented by the Nadary-Watson estimator, as well as the least squares method for legitimate crossing (LSCV) in estimating the introductory parameter, and using the simulation method in comparing the two methods, where it was concluded that the proposed method is highly efficient in estimation and matches the practical reality, especially in the case of increasing the percentage Loss also in case of data dispersion. In (2017),(Al-Mashhadani and Kamel) indicated the use of the analysis of covariance method (coons) according to two methods to estimate the missing value of the response variable or the variable adopted for the split block design, and the possibility of estimating the missing values in the experiment, as The loss of the value or value of the results of the experimental plots for the studied characteristic represented by the response variable affects the results of the analysis for the response variable. Therefore, these values must be estimated to obtain sound analysis results. In (2018) , (Ayden) proposed estimators for the missing value in the one-way analysis of variance (ANOVA) when the error distribution is long-tail symmetrical, and three methods were used to estimate the missing value, namely the least squares method LS, the method of maximum likelihood ML and the method of the modified maximum likelihood MML, compared the efficiency of the estimators through the Monte Carlo simulation study, and the results showed that the ML estimator is the most efficient.

1. Objective of the Research

This research aims to compare several methods for estimating the missing value of the restricted one-way analysis of variance model, assuming that the error is distributed with long-tailed symmetric distribution, These methods include the arithmetic mean method (M), the maximum likelihood method (MLE), and the modified maximum likelihood method (MMLE) using the mean squared error (MSE) criterion, and to choose the best estimation method from among the estimation methods used in this research and apply it to real data.

2. Material and Methods

A. Analysis of variance

Analysis of variance is the mathematical method according to which the total sum of squares of data about a phenomenon is divided into its different sources. The results of the analysis are put in a table known as the analysis of variance table. As (Taha and Al-Qadi,1994) mention, the emergence of this type of analysis is due to the scientist Fisher. (R.A.Fisher) who is considered the first to establish the rules of analysis of variance in 1923), as the discovery of this method led to broad and rapid progress in the field of statistical science and experimental design.
As for (Al-Khalili, 1988) , analysis of variance is defined as a smart way to test the difference in the means of two or more groups at once through contrast. (Adas, 1997) also states that analysis of variance “provides us with a convenient statistical method for comparing many means with each other at the same time.” .

Whereas the variance is the rate of difference or difference in the items of any phenomenon (Al-Otaibi and Jad Al-Rub, 2016) , and also the variance is the square of the standard deviation, or it is the average of the squares of the deviations of the items from the arithmetic mean, and we cannot obtain the standard of dispersion or difference from the same deviations because their sum equals zero for any distribution From the distributions) ,and variance is also the average sum of the squares of the deviations of the observation values of the population from their arithmetic mean, according to the following relationship.

$$\sigma^2 = \sum_{i=1}^{N} (x_i - M)^2 / N \qquad (1)$$

B. One- Way analysis of variance

There are different forms of analysis of variance, and these forms depend on the number of independent and dependent variables. One Way Analysis of Variance is one of these forms, as it is concerned with detecting differences or discrepancies between a number of groups in one dependent variable and each of these groups. each group is called treatment. There is also a two-way analysis of variance, and in this research we will focus on a one-way analysis of variance, which is similar to a completely randomized design in which the data is arranged according to treatments

One Way Analysis of variance is also a statistical form concerned with detecting differences in a phenomenon for many groups in one dependent variable. It is mentioned (Al-Shamrani and Taha, 2000) that analysis of variance for one way helps the researcher test the differences between a number of groups in one dependent variable.

C. Restricted one- way analysis of variance model

It is to estimate the effect of treatment and compare the effect of treatments and thus compare the averages of treatments, where the null hypothesis and the alternative hypothesis for this model are as follows:

$H_0: M_1 = M_2 = \cdots \dots \dots \dots \dots = M_a$

H_1: at Least two mean are not equael

and that the formula of the statistical model for the restricted one way analysis of variance model is as follows:

$$x_{ij} = M_i + e_{ij} \qquad (2)$$

$i = 1,2,3,\dots\dots\dots,a \quad ; \quad j = 1,2,3,\dots\dots\dots,n$

$\mu_i = M + T_i$

Whereas

M: is the general average which is equal to $\bar{x}_{..}$

$$M = \frac{\sum_{i=1}^{a} \mu_i}{a}$$

Treatment Effect (i) T_i :

D. Estimation methods

For the purpose of estimating the missing value of the restricted one way analysis of variance model ,we will use three estimation methods : the mean method , the maximum likelihood method , and the modified maximum likelihood method .

- The mean method (M) [21]

This method is used to estimate the missing value in the data series and compensate for it with the average value of the studied variable .

- Maximum likelihood method (MLE)

The mathematical formula for deriving the maximum likelihood method for estimating model parameters, assuming that the error is distributed with Long-Tail Symmetric Distribution (LTS) , is as follows:

$$f(y)=\frac{1}{\sigma\sqrt{q}B(\frac{1}{2},P-\frac{1}{2})}\left(1+\frac{(y-\mu)^2}{q\sigma^2}\right)^{-p} \qquad (3) \qquad -\infty<y<\infty$$

$E(y-\mu)=0$, $Var(y-\mu)=\sigma^2$

$q = 2p-3$

$$t=\sqrt{\frac{v}{q}}\cdot\frac{(y-\mu)}{\sigma}$$

Whereas

(Shape Parameter) : P

(Scale Parameter) : σ

(Location parameter) : μ

(Beta function) : B

Degrees of freedom of the Student – (t) distribution : v

V = 2p-1

$$f(y)=a\frac{1}{\sigma}\left(1+\frac{(y-\mu)^2}{q\sigma^2}\right)^{-p} \qquad (4)$$

$$a=\frac{1}{\sqrt{q}B(\frac{1}{2},\ P-\frac{1}{2})}$$

$$L=\prod_{i=1}^{n} f(y)$$

$$=\frac{(x_{ij}-\mu_i)}{\sigma}S_{ij}$$

Suppose that the x_{kl} is observation lth in the kth treatment is missing, the likelihood function can be analyzed as a product of two components The first component is the observed values (x_{ij} ; $i\neq k$; $j\neq l$) and the second component is the missing value (x_{kl})

We take the natural logarithm of the function (likelihood)

$$lnL=n\ln\alpha-n\ln\sigma-p\sum_{i=1}^{n} ln\left(1+\frac{(y-\mu)^2}{q\sigma^2}\right) \qquad (5)$$

$$\sum_{j=1}^{n}\frac{S_{ij}}{(1+\frac{S_{ij}^2}{q})}=0 \qquad (6) \qquad =\frac{2p}{q\sigma}\frac{\delta lnL}{\delta\mu_i}$$

$$\frac{\delta lnL}{\delta\mu_k}=\frac{2p}{q\sigma}\sum_{j=1}^{n}\frac{S_{kj}}{(1+\frac{(S_{kj})^2}{q})}+\frac{2p}{q\sigma}\frac{S_{kl}}{(1+\frac{(S_{kl})^2}{q})}=0 \qquad (7)$$

$$\frac{S_{kl}}{(1+\frac{(S_{kl})^2}{q})}=0 \qquad (8) \qquad =\frac{2p}{q\sigma}\frac{\delta lnL}{\delta m}$$

$$\frac{\delta lnL}{\delta\sigma}=-\frac{n}{\sigma}+\frac{2p}{q\sigma}\sum_{\substack{i=1\\ i\neq k}}^{a}\sum_{j=1}^{n}S_{ij}\frac{S_{ij}}{(1+\frac{(S_{ij})^2}{q})}+\frac{2p}{q\sigma}\sum_{\substack{j=1\\ j\neq l}}^{n}S_{kj}\frac{S_{kj}}{(1+\frac{(S_{kj})^2}{q})}+\frac{2p}{q\sigma}S_{kl}\frac{S_{kl}}{(1+\frac{(S_{kl})^2}{q})}=0 \qquad (9)$$

$$g(S_{ij})=\frac{S_{ij}}{(1+\frac{S_{ij}^2}{q})}$$

We will use one of the numerical methods, which is Newton Raphson's method. To obtain an estimate of the model parameters due to the difficulty of solving these equations because they are non-linear equations.

Modified maximum likelihood method (MMLE)

We apply the modified maximum likelihood method stages by placing these formulas by naming the order of the variables in ascending order, since the sums are fixed in the sense $\sum S\mathrm{i} = \sum S(\mathrm{i})$ And as it comes.

$$S_{(1)} \le S_{(2)} \le S_{(3)} \le S_{(n)} \qquad (10)$$

$$\frac{\delta \ln L}{\delta \mu_k} = \frac{2p}{q\sigma} \sum_{\substack{j=1 \\ j\neq l}}^{n} g(S_{kj}) + \frac{2p}{q\sigma}\, g(S_{kl}) = 0 \qquad (11)$$

$$g(S_{kl}) = 0 \qquad (12) \quad = \frac{2p\, \delta \ln L}{q\sigma\, \delta m}$$

From formula (11) and (12) we get

$$\frac{\delta \ln L}{\delta \mu_k} = \frac{2p}{q\sigma} \sum_{\substack{j=1 \\ j\neq l}}^{n} g(S_{k(j)}) = 0 \qquad (13)$$

Whereas

$$; \qquad g(S_{k(j)}) = \frac{S_{k(j)}}{(1+\frac{S_{k(j)}{}^2}{q})} \qquad S_{k(j)} = \frac{(x_{k(j)} - \mu_k)}{\sigma}$$

Using the first two terms of the Taylor series to find estimates of both a_j and b_j about the average values of the statistical ranks, i.e $E\, S_{k(j)} = t_{(j)}$

$$g(S_{k(j)}) = a_j + b_j\, S_{k(j)} \qquad (14)$$

$$g(S_{k(j)}) = g(t_{(j)}) + (S_{k(j)} - t_{(j)})\, g'(t_{(j)}) \qquad (15)$$

$$a_j = \frac{(\frac{2}{q}) t_{(j)}{}^3}{(1+\frac{t_{(j)}{}^2}{q})^2} \quad , \quad b_j = \frac{1-(\frac{1}{q}) t_{(j)}{}^2}{(1+\frac{t_{(j)}{}^2}{q})^2}$$

By performing some integrations, we obtain an estimate of the value of (t) .

$$F(t_{(j)}) = \frac{1}{\sigma\sqrt{q}\, B\left(\frac{1}{2},\ P-\frac{1}{2}\right)} \int_{-\infty}^{t(j)} (1+\frac{(S)^2}{q})^{-P}\, dS \qquad (16)$$

$$\frac{j}{n}; \ (j = 1,2,\cdots,n;\ j \neq l).$$

We will use the formula t = j / n to have the (MMLE) and from an formula (13) and(14) we get

$$\frac{\delta \ln L}{\delta \mu} = \frac{2p}{q\sigma} \sum_{\substack{j=1 \\ j\neq l}}^{n} (a_j + b_j\, S_{k(j)}) = 0 \qquad (17)$$

Whereas

$$S_{k(j)} = \frac{(x_{k(j)} - m)}{\sigma}$$

$$\frac{\delta \ln L}{\delta \mu} = \frac{2p}{q\sigma} \sum_{\substack{j=1 \\ j\neq l}}^{n} \left(a_j + b_j\, \frac{x_{k(j)} - m}{\sigma}\right) = 0 \qquad (18)$$

The value of a_j is zero because it is not related to μ , so the equation is as follows

$$\frac{\delta lnL}{\delta\mu} = \frac{2p}{q\sigma}\sum_{\substack{j=1\\ j\neq l}}^{n}\left(b_j \frac{x_{k(j)} - m}{\sigma}\right) = 0 \qquad (19)$$

$$\hat{m} = \frac{\sum_{\substack{j=1\\ j\neq l}}^{n} b_j x_{k(j)}}{\sum_{\substack{j=1\\ j\neq l}}^{n} b_j} \qquad (20)$$

$$\hat{m} = \frac{1}{w}\sum_{\substack{j=1\\ j\neq l}}^{n} b_j x_{k(j)} \qquad (21)$$

Whereas

$$w = \sum_{\substack{j=1\\ j\neq l}}^{n} b_j$$

3. Results and Discussion

Real data application

An agricultural experiment dating back to the year (2017) was applied to determine the effect of adding three types of sugars on the length of the multiplying branches (cm) of the tissue-rich apple rootstock (MM106), at the University of Baghdad / College of Agriculture / Department of Horticulture and Landscape Engineering / Plant Tissue Culture .Type of sugars are a factor and each type of sugar is a treatment It will be treatments (without addition , sucrose, sorbitol , glucose) and has been repeated eight times, and for the statistical model for the restricted one-way analysis of variance according to the formula (2) mentioned in the side Theoretical .
i = 1,2,3,4 ; j = 1,2,3,4,5,6,7,8

Table (1): Akayki Information criterion for Long Symmetric Distribution (LTS) and Normal Distribution (N) .

t	p	AIC (LTS)	AIC (N)
4	3	434.1	561.1

Where AIC = - 2 Log (L) + 2 Φ
L : The value of the maximum likelihood function of the distribution
Φ : Number of distribution parameters
After using the Akaike information criterion, it was found that the distribution of the data is long tail symmetric (LTS) compared to the normal distribution, through The value calculated for the Akaike information criterion for the (LTS) distribution was less than the value calculated for the Akaike information criterion for the normal distribution.

Table (2): Missing value estimation for all estimation methods

Missing value	**Estimation methods**		
m	M	MLE_N	MMLE
4.21	4.4843	4.5465	4.1592

Assuming that the value of the second observation of sucrose was missing, the best way to estimate the missing value for this experiment was the modified maximum likelihood method (MMLE), then comes the mean method (M), followed by the maximum likelihood method by Newton-Raphson (MLE_N).

4. Conclusions

A. The modified maximum likelihood method (MMLE) was found to be more efficient to estimate the missing value of the restricted one-way analysis of variance model compared to the rest of the estimation methods used based on the statistical criterion Mean Squares of Error (MSE).
B. The estimators of the maximum likelihood method (MLE_N) are less efficient in estimating the missing value of the restricted one-way analysis of variance model.
C. The data follow the long-tail symmetric distribution (LTS) because the calculated value of (AIC) for (LTS) distribution is lower compared to (AIC) for normal distribution.
D. The value of the second observation of Sucrose is (4.21), which is close to the assumed lost value.

References

1- Abu Hatab, Fouad and Sadiq, Amal (2010) “Research Methods and Methods of Statistical Analysis in the Psychological, Educational, and Social Sciences,” Cairo, Anglo-Egyptian Library, 165 Mohamed Farid Street.
2- Adas, Abdul Rahman (1997). “Principles of Statistics in Education and Psychology,” Part Two. Amman: Dar Al-Fikr.
3- Ahmed, Ahmed Dhiyab & Lateef, Z. K., (2021). "Use of modified maximum likelihood method to estimate parameters of the multiple linear regression model". Int. J. Agricult. Stat. Sci, 17(Supplement 1), 1255-1263.
4- Ahmed, Ahmed Dhiyab., Abdulwahhab, B. I., & Abdulah, E. K. (2019). " Comparison of Estimate Methods of Multiple Linear Regression Model with Auto-Correlated Errors when The Error distributed with General Logistic". Journal of Engineering and Applied Sciences, 14(19), 7072-7076
5- Ahmed, Ahmed Dhiyab., Abdulwahhab, B. I., & Abdulah, E. K. (2020). " A comparison among Different Methods for Estimating Regression Parameters with Autocorrelation Problem under Exponentially Distributed Error" . Baghdad Science Journal, 17(3 (Suppl.)), 0980-0980.. DOI: http://dx.doi.org/10.21123/bsj.2020.17.3(Suppl.).0980
6- Al-Aqabi, Farah Abdel Karim, Al-Amiri, Lamia Khalifa Jawad(2017) “The effect of the type and concentration of sugars on the micropropagation of apple rootstock (MM106) ex vivo” Zagazig Journal of Agricultural Research 44 (3).
7- Al-Imam, Muhammad Muhammad Al-Tahir (1994) "Design and Analysis of Experiments", King Saud University, Riyadh, Mars Publishing House
8- Al-Khalili, Khalil Youssef (1988). “Statistics for the researcher in education and human sciences.” Amman: Dar Al-Fikr.
9- *Al-Mashhadani, Kamal Alwan Khalaf and Kamel, Saimaa Firas (2017) “Estimating the missing value using the analysis of covariance method to design split sectors” Journal of Economic and Administrative Sciences (College of Administration and Economics - University of Baghdad) Volume (23) Issue (100). DOI: https://doi.org/10.33095/jeas.v23i100.225*
10- Al-Mashhadani, Kamal Alwan Khalaf,(2010) “Design and Analysis of Experiments - Using Computers,” Al-Jazeera Printing and Publishing Office, Iraq, Baghdad.
11- Al-Otaibi, Ashraf Ahmed Awad, Jad Al-Rub, Hisham Fathi. (2016) “An evaluation study of the validity of using the analysis of variance method in master’s and doctoral theses in the College of Education at Umm Al-Qura University (during the period 1421 AH - 1430 AH). (Arab Studies in Education and Psychology. 73 (73), 427-491. DOI: 10.12816/0030325 .
12- Al-Qazzaz, Qutaiba Nabil Nayef (2016) “A proposed method for estimating missingness in a nonparametric multiple regression model” Journal of Economic and Administrative Sciences (College of Administration and Economics - University of Baghdad) Volume 22), Issue (89), pages 396-406
13- Al-Shamrani, Muhammad Musa Muhammad, Taha, Rabei Saeed (2000) “Problems of using one-way analysis of variance and dimensional comparisons and methods for treating them” Master’s thesis, College of Education, Umm Al-Qura University in Mecca, Department of Psychology, specialty in statistics and research.
14- Al-Sulaiman, Muthanna Subhi,(1995) “Using linear and nonlinear models in estimation,” Master’s thesis, College of Administration and Economics, University of Baghdad.
15- Aydın, D., & Şenoğlu, B. (2018)." Estimating the missing value in one-way anova under long-tailed symmetric error distributions". Sigma Journal of Engineering and Natural Sciences, 36(2), 523-538.
16- Choudhary, D. and Mehta, A. (2010) Fruit Crops. Oxford Book Company. Jaipur, India

17- Dempster, A. (1977). "Maximum likelihood estimation from incomplete data via the EM algorithm". Journal of the Royal Statistical Society, 39, 1-38.
18- El-Sherbiny, Zakaria (1995). “Statistics and Experimental Design in Psychological, Educational, and Social Research,” Cairo: Anglo-Egyptian Library.
19- Hadba, WKaa Ali, Dabdoub, Marwan Abdel Aziz (2007) “A comparison of some methods for estimating missing values in a completely randomized block design,” Journal of Research of the College of Basic Education, 4(3), 242-257
20- Islam, M. Q., & Tiku, M. L. (2005). "Multiple linear regression model under nonnormality. Communications in Statistics-Theory and Methods", 33(10), 2443-2467 , https://doi.org/10.1081/STA-200031519
21- Jamshidian, M., & Mata, M. (2007). "Advances in analysis of mean and covariance structure when data are incomplete. In Handbook of latent variable and related models" (pp. 21-44). North-Holland , DOI: 10.1016/S1871-0301(06)01002-X
22- Keller.G and warrack, 2003, “Statistics for management and conomics” Duxbury Press, US
23- McLachlan, G. J., & Krishnan, T. (2007)." The EM algorithm and extensions". John Wiley & Sons.
24- Montgomery, D. C. (2017)." Design and analysis of experiments". John wiley & sons.
25- Muhammad, Mahdi Younis and Ahmed, Ahmed Dhiyab. (2022) “Estimating a polynomial regression model when the error is distributed with a symmetric long-tail distribution with an application” Master’s thesis, College of Administration and Economics, University of Baghdad.
26- Rao C.R. and Toutenburg H., (1999) "Linear Models: Least Square and Alternatives", Springer, New York
27- Reynolds J.F., (1995) Estimating Missing Observations in ANOVA, International Journal of Mathematical Education in Science and Technology 26(6), 895–901. doi.org/10.1080/0020739950260610.
28- Rubin, D., 1972, "A non-iterative algorithm for least square estimation of missing values in any analysis of variance design", J. of App. Stat., 21, P 136-141.
29- Taha, Rabie Saeed and Al-Qadi, Diaa (1994). “Fundamentals of Applied Statistics in the Agricultural Field,” Cairo University, Cairo: University Book.
30- Tiku, M. L., & Akkaya, A. D. (2010)."Estimation in multifactor polynomial regression under non-normality ". Pakistan Journal of Statistics, 26(1).
31- Wilkinson, G. N. (1958). Estimation of missing values for the analysis of incomplete data. Biometrics, 14(2), 257-286. doi.org/10.2307/2527789
32- Yates, F. (1933). The analysis of replicated experiments when the field results are incomplete. Empire Journal of Experimental Agriculture, 1(2), 129-142.