Available online www.unicrossjournals.com

UNICROSS JOURNAL OF SCIENCE AND TECHNOLOGY, UJOST

RESEARCH ARTICLE VOL. 2(2) SEPTEMBER, 2023 ISSN:2814-2233

Date Accepted 30th September, 2023 Pages 61 - 69

AN APPRAISAL OF $\chi 2$ AND RAO'S *F* APPROXIMATIONS FOR WILKS' LAMDA (Λ) ON A MULTIVARIATE TEST IN CRUTECH FARMS

Okim Inyang Ikpan[1], John Effiong Usen[2], Emmanuel Eta Omini[3] and Bright Ozar Obeten[4]
Department of Statistics
University of Cross River State, Calabar, Nigeria.
[1]ikpanlouis@gmail.com
[2]usenjohn@unicross.edu.ng
[3]emmanuelomini74@gmail.com
[4]brightobeten86@yahoo.com
Corresponding author: ikpanlouis@gmail.com

Abstract

This research was designed to appraise the two approximate methods of Wilks' likelihood ratio in studying the effect of different types of soils on the yield of groundnuts in a growing season. Three different soil types based on the plant requirements were selected as variables of interest. The population of the study consisted of yields (in kilograms) of three (out of several) varieties of groundnuts from three different soil groups of ten farms each. The data for the work was collected as secondary data obtained from the Faculty of Agriculture and Forestry in Obubra Campus of the University of Cross River State. The data were then subjected to the two approximate methods of multivariate likelihood ratio(MLR) analysis. Adequate literature on the use of the likelihood ratio (LR) or Wilk's lamda (Λ) was consulted. The data was partly analyzed manually and partly with the use of EXCEL and R-Core Team, version 49. The result of the multivariate analysis showed that the approximate tests methods, based on $\chi^2 -$and *F*-distributions of functions of the test statistics for the Wilk's lamdaa t $\alpha = 0.05$, both rejected the multivariate null hypothesis of equality of centroids. Therefore, the three different soil groups differed overall on the set of three varieties of groundnuts, certifying that the two approximation methods (Bartlett's $\chi^2 -$and Rao's *F*) are good approximations of Wilks' lamda, in a sample of size 10. Any of the two methods could be used to approximate Wilk's lamda when the sample size is lower than 30.

Keywords: Multivariate likelihood ratio (MLR), Chi-square, F-distribution, Soils, Yield.

1.0 Introduction

Multivariate statistical techniques are used in a variety of fields, including research in the social sciences (e.g., education, psychology, and sociology), natural sciences, agriculture, and medical fields. Their use has become more commonplace due largely to the increasingly complex nature of research designs and related research questions. Multivariate analysis of variance (MANOVA) is designed to test the significance of group differences (Lindman, 1992); including different soil types. The multivariate analysis of variance provides the techniques for a single test of a combined null hypothesis.

The dependent variable in the design is a vector variable corresponding to different attributes (responses) being investigated for each subject. The attributes or responses, in this case, are the yield of the different varieties of groundnuts with the subject being the different types of soils supporting the growth of the crop.

Unfortunately, the combined null hypothesis cannot always be tested with a simple F ratio. Instead, more complicated procedures, leading to statistics for which good tables may not be readily available, have to be used. The likelihood ratio procedure, commonly called Wilks's lambda, is based on a very general method for obtaining parameter estimates and doing statistical tests. Likelihood ratio tests have three desirable properties: they are usually fairly easy to derive, they tend to be very powerful, and when no exact distribution can be found for a likelihood ratio statistic, approximations are readily available (Lindman, 1992).

CRUTECH farms cultivate crop varieties that differ in environmental requirements such as soil types, among others. Among these crops are Groundnuts. The soils considered to favor the cultivation of groundnuts are Silt, Sandy soils, and Loamy soils. Multivariate Likelihood Ratio test will therefore be used to study the effect of the different soils on the crops' growth.

Numerous works abound to obtain exact or asymptotic forms of the central and non-central distributions of the test statistic so that exact or approximate significance levels by means of a statistic may be obtained. Although tables of those statistics are available, attempts have been made to obtain good approximate tests based on $\chi^2 -$ and F-distributions of functions of the test statistic desired. Researchers have it that both approximations have small differences in terms of sample sizes and accuracy. Raycov and Marcoulides, (2008) hold that for moderate to large sample sizes, Bartlett's $\chi 2$ is a good approximation while for smaller sample sizes, Rao's F is a better approximation.

This research, carried out under MANOVA, aimed to investigate whether the different soil types that give rise to this dependent multivariate response variable (yield of Groundnuts) could be said to have the same mean vector under the two approximation methods. The research aimed to find out a better of the two-approximation methods of Wilk's Lamda for the different soil groups to give rise to a higher yield of Groundnuts, based on a sample of size 10.

2.0 Materials/Method

Groundnut farming in Nigeria is one of the most lucrative businesses in the country because of the demand for its highly nutritious seed and most importantly the edible oil derived from the seed. Groundnut is a good source of cheap protein both for animals and human beings. Groundnut tends to be well in arid or semi-arid regions. There are various varieties of groundnut in Nigeria based on their yields and how they adapt to climate conditions. Groundnut thrives well in well-drained, sandy loam soil. Soils with a pH of 6.5-7.0 with high organic matter are ideal for groundnut to thrive. Groundnut requires an optimal temperature of 27^0C – 30^0C and 24^0C – 27^0C for good germination, vegetation growth, and reproductive growth. An optimum annual rainfall of 450mm – 1250mm is required for good growth and yield (Maxipharo, 2020).

Groundnuts, depending on the type, are affected by different environmental conditions including soils. Soil provides structural support to plants' (groundnut) growth by providing: anchorage, oxygen, water, temperature modification, and nutrients (Stack, 2016). They vary greatly in their chemical and physical properties. Processes such as leaching, weathering, and microbial activity combine to make a whole range of different soil types. Each type has particular strengths and weaknesses for agricultural production. Good soil structure contributes to soil and plant health allowing water and air movement into and through the soil profile. The type of soil in your garden plays a huge role in determining how well plants grow. Different plants are adapted to different types of soil and growing them in the wrong types of soil negatively impacts growth. Understanding the different properties of soil and how they affect your plants helps one to select the best plants for one's garden (Fenil, 2021).

The researchers selected three of the different types of soils (Silt, Sandy, and Loamy Soils) as the variables of interest, which was done based on the plant requirements, for the collection of data; certifying the work as experimental design. The population of the study consisted of yields (in kilograms) of Groundnuts (that is, Peanuts (*Arachis Hypogea*), Bambara Groundnuts (*Vigna subterranea*), Hausa Groundnuts (*Macrotyloma Geocarpum*)) from three different soil groups of ten farms each. The data obtained focused on the entire population of the yield of Groundnuts in CRUTECH Farms. With the nature of the information gathered, no specific technique of sampling was employed. The data for the research was collected as secondary data. It was obtained from the Faculty of Agriculture and Forestry, Obubra Campus of the Cross River State University; where the effect of soil types was studied on the yield of the groundnuts in a growing season.

2.1 Mathematical model

Suppose that we have $N = \sum_{i=1}^{k} N_i$ observations classified into k groups. The j-th observation in the i-th group is a $p \times 1$ column vector $\mathbf{x}_{ij}$ which is assumed to be expressed as follows:

$$\mathbf{x}_{ij} = \boldsymbol{\mu} + \boldsymbol{\alpha}_i + \mathbf{e}_{ij} \qquad (1)$$

where $i = 1, 2, ..., p; \; j = 1, 2, ..., N_i$, $\boldsymbol{\mu}$ and $\boldsymbol{\alpha}_i$, are $p \times 1$ constant vectors such that $\sum_{i=1}^{p} N_i \boldsymbol{\alpha}_i = 0$, and the $p \times 1$ random error vector: $\mathbf{e}'_{ij} = \left(\mathbf{e}_{1ij}, \mathbf{e}_{2ij}, ..., \mathbf{e}_{pij}\right)$ is assumed to be distributed according to a $p-$variate distribution which may be characterized by as many moments as desired. Specifically, the first two moments of eta are given by

$$E\left(\mathbf{e}_{ij}\right) = 0 \qquad (2)$$

$$E\left(\mathbf{e}_{ij}\mathbf{e}'_{ij}\right) = \Sigma_t \qquad (3)$$

where $\Sigma_t = \sigma_{ij}^{(t)}$ is a $p \times p$ positive definite, symmetric matrix called the variance-covariance matrix of $\mathbf{e}_{ij}$ whose elements are assumed to be finite. The model (1) for the one-way MANOVA can be rewritten in terms of grand centroid, treatment effects, and error as

$$\begin{pmatrix} x_{ij1} \\ x_{ij2} \\ \vdots \\ x_{ijp} \end{pmatrix} = \begin{pmatrix} \mu_1 \\ \mu_2 \\ \vdots \\ \mu_p \end{pmatrix} + \begin{pmatrix} \alpha_{i1} \\ \alpha_{i2} \\ \vdots \\ \alpha_{ip} \end{pmatrix} + \begin{pmatrix} e_{ij1} \\ e_{ij2} \\ \vdots \\ e_{ijp} \end{pmatrix} \qquad (4)$$

The multivariate model assumes that

(i) The dependent (response) variable is vector valued and is distributed as multivariate normal, with the same dispersion matrix (homoscedastic) for all the k groups ie. $N(0, \Sigma)$, where $\Sigma = \Sigma_1 = \Sigma_2 = \cdots \Sigma_k$

(ii) The observations within each sample must be randomly sampled and must be independent of each other.

2.2 Sum of squares

The MANOVA is based on an assumption of a linear model of the form

$$x_{ij} = \bar{x}_{..} + (\bar{x}_{i.} - \bar{x}_{..}) + (x_{ij} - \bar{x}_{i.}) \qquad (5)$$

where, x_{ij} is the dependent vector variable for the jth subject in the ith sample for $i = 1, 2, \ldots, k$ and k is the number of populations under study. $\bar{x}_{..}$ is the MANOVA grand centroid; that is, the vector of total sample means and $\bar{x}_{i.}$ is the centroid for the ith sample.

Deducting this grand centroid vector from both sides of Equation (5), the resultant equation is

$$x_{ij} - \bar{x}_{..} = (\bar{x}_{i.} - \bar{x}_{..}) + (x_{ij} - \bar{x}_{i.}) \qquad (6)$$

From (5), $(\bar{x}_{i.} - \bar{x}_{..})$ represents the hypothesis we are testing (no differences in locations of the means of the groups); $(x_{ij} - \bar{x}_{i.})$ represents the error or residual effect (i.e., the deviations of the responses from the centroids of the samples).

The total sum of squares can be obtained using vector multiplication as follows:

$$\sum_{i=1}^{k}\sum_{j=1}^{N_i}(x_{ij} - \bar{x}_{..})(x_{ij} - \bar{x}_{..})' = \sum_{i=1}^{k}\sum_{j=1}^{N_i}(\bar{x}_{i.} - \bar{x}_{..})(\bar{x}_{i.} - \bar{x}_{..})' + \sum_{i=1}^{k}\sum_{j=1}^{N_i}(x_{ij} - \bar{x}_{i.})(x_{ij} - \bar{x}_{i.})' \qquad (7)$$

From (6), the total sum of squares, the between group sum of squares and the within group variation are respectively given as

$$T = \sum_{i=1}^{k}\sum_{j=1}^{N_i}(x_{ij} - \bar{x}_{..})(x_{ij} - \bar{x}_{..})' \qquad (8)$$

$$B = \sum_{i=1}^{k} N_i (\bar{x}_{i.} - \bar{x}_{..})(\bar{x}_{i.} - \bar{x}_{..})' \qquad (9)$$

and

$$W=\sum_{i=1}^{k}\sum_{j=1}^{N_i}\left(x_{ij}-\bar{x}_{i.}\right)\left(x_{ij}-\bar{x}_{i.}\right)' \qquad (10)$$

Therefore, (9) can be written as

$$T=B+W \qquad (11)$$

The estimators of the common population dispersion matrix, Δ, if the null hypothesis of equality of group centroids holds, are obtained as

$$\Delta=\left[\frac{1}{k-1}\right]B \qquad (12)$$

$$\Delta=\left[\frac{1}{N-1}\right]W \qquad (13)$$

Where $(k-1)$, $(N-k)$ are their respective degrees of freedom, and $N=\sum_{i=1}^{k}N_i$.

2.3 Test of hypothesis in MANOVA

The null hypothesis in MANOVA states that the population mean vectors are equal (note that bold font indicates that the variables are vector, not scalar):

$$H_0:\boldsymbol{\alpha}_1=\boldsymbol{\alpha}_2=\cdots=\boldsymbol{\alpha}_k=0 \qquad (14)$$

Wilks' likelihood ratio (Wilks' Lambda), is one of a number of tests that could be used for the multivariate analysis of variance. Wilks Criterion due to S. S. Wilks (1906–1964), is the oldest and perhaps the most widely used criterion. This statistic (Wilks, 1932), which is often called the Wilks lambda criterion usually denoted by Λ, may be defined as the ratio of two determinants, the within groups and the total sum of squares,

$$\Lambda=|W|/|B+W|=\prod_{i=1}^{p}1/(1+c_i) \qquad (15)$$

where $\boldsymbol{W}$**,** is the pooled estimate of within variability on the set of variables, the multivariate error term. The range on Λ is 0 to 1.0 for a perfectly fitting model under H_0 in which there is no multivariate effect. Unlike the F-ratio in univariate ANOVA, smaller values of Λ lead to a rejection of H_0 and an inference of the statistical alternative H_1.

Source	Degree of freedom	Sum of squares and crossproducts
Treaments (B)	$k-1$	$B=\sum_{i=1}^{k}\sum_{j=1}^{N_i}\left(x_{i.}-\bar{x}_{..}\right)\left(x_{i.}-\bar{x}_{..}\right)'$
Within Treatments (W)	$N-k$	$W=\sum_{i=1}^{k}\sum_{j=1}^{N_i}\left(x_{ij}-\bar{x}_{i.}\right)\left(x_{ij}-\bar{x}_{i.}\right)'$
Total (T)	$N-1$	

There have been numerous works to obtain exact or asymptotic forms of the central and non-central distributions of the test statistic so that exact or approximate significance levels and powers of the tests for $H_0(p,k)$ by means of the statistic may be obtained. Although tables of those statistics are available, attempts have been made to obtain good approximate tests based on $\chi^2-$and *F*-

distributions of functions of the test statistics described.

(1) Bartlett's χ^2:

The χ^2 approximation obtained by Bartlett (1947) depends on Wilks' Lambda, and is of the form

$$\chi_f^2 = -\left[v_e + v_k - 0.5\left(p + v_k + 1\right)\right]\ln\left\{\sum_{i=1}^{s}\frac{1}{1+\lambda_i}\right\} \quad (16)$$

where v_e is the error degree of freedom, v_k the hypothesis degree of freedom, and p is the number of dependent variables. It is worthy of note that $v_e + v_k = N - 1$ and $v_k + 1 = k$ where N is total sample size and k is the number of groups. Thus (16) can also be written as

$$\chi^2 = -\left[\left(N-1\right) - 0.5\left(p+k\right)\right]\ln\Lambda \quad (17)$$

where $\Lambda = \sum_{i=1}^{s} 1/\left(1+\lambda_i\right)$. The degree of freedom of the χ^2 statistic is $p(k-1)$.

(2) Rao's *F*.

The approximation obtained by Rao (1952) also depends on Wilks' Lambda and has the form

$$F_c = \left(\frac{1-\Lambda^{1/s}}{\Lambda^{1/s}}\right)\frac{ds - 0.5\left(pv_k - 2\right)}{pv_k} \quad (18)$$

where

$$d = \left[\left(N-1\right) - 0.5\left(p+k\right)\right] \quad (19)$$

$$s = \sqrt{\frac{p^2\left(k-1\right)^2 - 4}{p^2 + \left(k-1\right)^2 - 5}} \quad (20)$$

For a better, though more complicated approximation, let

$$g = \begin{cases} 1 \text{ if } p^2 + v_1^2 = 5 \\ \left[\left(p^2 v_1^2 - 4\right)/\left(p^2 + v_1^2 - 5\right)\right]^{1/2}, \text{ otherwise} \end{cases}$$

$h = gm^* - \left(pv_1 - 2\right)/2$

$m^* = v_2 - \left(v_1 - p - 1\right)/2$

$p = \#$ of dependent variables

$W = \Lambda^{(1/g)}$

Then $\left[\left(1 - W\right)/W\right]\left[h/\left(pv_1\right)\right]$ has approximately an F distribution with numerator degrees of freedom $\left(pv_1\right)$ and denominator degrees of freedom $= h$.

3.0 Analysis and Results

The data presented in Table 1 below shows the yield (in Kilograms) of three varieties of groundnuts obtained from three different groups of soils in one growing season

Table 1: Yield of Groundnuts in Kg

Silt (Group 1)			Sandy Soil (Group 2)			Loamy Soil (Group 3)		
P	B	H	P	B	H	P	B	H
56	87	76	39	66	79	35	56	96
51	84	69	32	60	64	33	50	80

42	86	77	38	74	80	47	50	89
40	74	69	43	76	83	44	52	87
42	80	70	42	60	82	38	50	84
45	78	71	37	60	72	38	42	84
40	72	62	39	63	67	42	55	86
55	55	58	33	75	77	39	60	79
54	85	69	35	51	68	43	52	87
40	80	72	40	60	70	50	58	70

P stands for – Peanuts (Arachis Hypogea)

B – Bambara Groundnuts (Vigna Subterranea)

H – Hausa Groundnuts – (Macrotyloma Geocarpum)

The multivariate null hypothesis with the χ^2 approximation for Wilks'Λ is calculated for groups 1, 2, and 3, respectively as follows.

$$\boldsymbol{W}_1 = \begin{bmatrix} 818.9 & -23.5 & 429.7 \\ -23.5 & 408.5 & -44.5 \\ 429.7 & -44.5 & 296.1 \end{bmatrix},$$

$$\boldsymbol{W}_2 = \begin{bmatrix} 600.5 & 50 & 327 \\ 50 & 117.6 & 132.4 \\ 327 & 132.4 & 419.6 \end{bmatrix}$$

$$\boldsymbol{W}_3 = \begin{bmatrix} 234.5 & 57.5 & -73 \\ 57.5 & 252.9 & -108.8 \\ -73 & -108.8 & 427.6 \end{bmatrix}$$

Thus

$$\boldsymbol{W} = \sum_{i=1}^{3} \boldsymbol{W}_i = \begin{bmatrix} 1653.9 & 84 & 683.7 \\ 84 & 779 & 69.1 \\ 683.7 & 69.1 & 1143.3 \end{bmatrix}$$

The **B and T** matricesare

$$\boldsymbol{B} = \begin{bmatrix} 3281.06 & 748.267 & -1893.6 \\ 748.267 & 388.87 & -316.9 \\ -1893.6 & -316.9 & 1153.4 \end{bmatrix} \text{ and}$$

$$\boldsymbol{T} = B + W = \begin{bmatrix} 4934.96 & 832.27 & -1209.9 \\ 832.27 & 1167.87 & -247.8 \\ -1209.9 & -247.8 & 2296.7 \end{bmatrix}$$

The Wilk's lamdaΛ, is

$$\Lambda = 0.1086468$$

The Bartlet's χ^2 is

$$\chi^2 = 57.71098, \text{ with with } 3(3-1) = 6 \text{ df}$$

The Rao's F - Approximation is

$$F_c = 16.94861$$

For the both approximations, the multivariate null hypothesis is:

$$H_0 : \begin{pmatrix} \mu_{11} \\ \mu_{21} \\ \mu_{31} \end{pmatrix} = \begin{pmatrix} \mu_{12} \\ \mu_{22} \\ \mu_{22} \end{pmatrix} = \begin{pmatrix} \mu_{13} \\ \mu_{23} \\ \mu_{33} \end{pmatrix}$$

That is, the population means in the three groups on variety 1 are equal, and as well, the population means on variety 2 and variety 3 are equal.

The critical value at $\alpha = 0.05$ for the χ^2 approximation is 12.6. We reject the multivariate null hypothesis of equality of centroids and conclude that the three different soil groups differ overall on the set of three groundnut varieties.

The critical value $F_{(2,6)}$ at $\alpha = 0.05$ is 5.14. We reject the multivariate null hypothesis of equality of centroids and conclude that the three different soil groups differ overall on the set of three groundnut varieties.

Furthermore, from the result of the multivariate analysis, the approximate tests methods based on $\chi^2 -$and *F*-distributions of functions of the test statistics for the Wilk's lamdaat $\alpha = 0.05$, both rejected the multivariate null hypothesis of equality of centroids. Any of the two approximations methods is therefore a good approximations of Wilks' lamda, in a sample of size 10.

4.0 Conclusion

From the result of the research being carried out under MANOVA, it can be concluded that the different soil types that give rise to this dependent multivariate response variable (yield of groundnuts) have the same mean vector under the different approximation methods.

Based on the result, the three different soil groups differ overall on the set of three varieties of groundnuts showing that the two approximations methods (Bartlett's $\chi^2 -$and Rao's *F*) are good approximations of Wilks' lamda, in a sample of size 10. Any of the two methods can be used to approximate the Wilk's lamda when the sample size is 10.

5.0 Recommendation

It can be recommended that a further appraisal of the two methods be adopted in a multivariate likelihood ratio test (the Wilk's lamda Λ) for a sample size larger than 10.

References

Abady, S., Shimelis, H., Janila, P. & Mashilo, J. (2019). Groundnut (*Arachis hypogaea L.*). Improvement in sub-Saharan Africa: A review. *Soil and Plant Science, 69*(6), 528-545.

Aron, A., Aron, E. N. & Coups, E. (2006). Statistics for psychology (4th Ed.). Upper Saddle River NJ: Prentice Hall.

Awodun, M. A. (2007). Influence of sawdust ash on soil chemical properties and cowpea performance in Southeastern Nigeria.*International Journal of Soil Science, 2*(1), 78-86.

Azu, E. O. D., Osodeke, V. E., Ukpong, I. M. &Osisi, A. F. (2018). Chemistry and mineralogy of soils derived from different parent materials in Southeastern Nigeria. *International Journal of Plant and Soil Science*, 1-16.

Azu, EO, D., Osodeke, V. E., Ukpong, I. M., &Osisi, A. F. (2018). Chemistry and Mineralogy of Soils Derived from Different Parent Materials in Southeastern Nigeria. *International Journal of Plant & Soil Science,* 1-16.

Bartlett, M. S. (1947). Multivariate analysis. *Journal of the Royal Statistical Society, Series B, 9*, 176-197.

Femi, K. (2021). How does soil affect plant growth? https://gardenerideas.com.

Gupta, D. S. & Perlman, M. D. (1973). On the power of Wilks' U-test for MANOVA. *Journal of Multivariate Anal*, *3*, 220-225.

Johnson, R. A. & Wichern, D. W. (2002). Applied multivariate statistical analysis. Upper Saddle River, NJ: Prentice Hall.

Lindman, H. R. (1992). Analysis of variance in experimental design. New York: Springer-Verlag, Inc.

Lois, B. S. (2016). Soil and plant nutrition: A gardener's perspective cooperative extension 5741. Libby Hall Orono: Maine 04469-5741.

Maxipharo, M. (2010). Groundnut farming in Nigeria: Beginners guide. https://agricincome.com.

Msrivani, H. (2009). Groundnut soils. Indian Institute of Technology Kanpur. http://agropedia.iitk.ac.in.

Morrison, D. F. (1976). Multivariate statistical methods. San Francisco, CA: McGraw-Hill.

Rao, C. R. (1952). Advanced statistical methods in biometric research. New York: John Wiley & Sons.

Raykov, T. &Marcoulides, G. A. (2008). An introduction to applied multivariate analysis. New York: Taylor and Francis Group.

Stevens, J. (2001). Applied multivariate statistics for the social sciences (4th Ed.). Hillsdale, NJ: Lawrence Erlbaum Associates.

Tatsuoka, M. M. (1971). Multivariate analysis. New York: Macmillan Publishers.

Wilks, S. S. (1932). Certain generalizations in analysis of variance. *Biometrika, 24,* 471-494.

Zhao, C. X., Jia, L. H., Wang, Y. F. & Wang, M. L. (2015). Effects of different soil texture on peanut growth and development. *Communications in Soil Science and Plant Analysis, 46*(18).