



MULTIVARIATE ANALYSIS OF EEG DATA: SOME ASPECTS OF DIAGNOSTIC OF MANOVA MODEL

Md Rokonzaman¹

*Associate Professor Department of Statistics University of Chittagong
Chittagong-4331 Bangladesh
Email: rokonuzzaman@cu.ac.bd*



ABSTRACT

Article History

Received: 7 September 2017

Revised: 21 August 2018

Accepted: 24 August 2018

Published: 3 September 2018

Keywords

Electroencephalogram
Fractal dimension
MANOVA.

The main focus of this study is the multivariate analysis of Electroencephalogram data which included multivariate analysis of variance. The multivariate model diagnostics comprise checking number of assumptions of MANOVA model such as multivariate normality, homogeneity of covariance matrices. In this paper, the model $\mathbf{X}=\mathbf{BC}+\mathbf{E}$ is used and estimate the different parameters. Also by using a general form, $H_0: \mathbf{GBF}=\mathbf{0}$ to test the different types of null hypothesis. Here G and F is known matrices and obtained from the hypothesis. This study gives mathematical ideas from multivariate statistical analysis to find a solution or a good approximation of a complex scientific problem.

Contribution/Originality: This study gives mathematical ideas from multivariate statistical analysis to find a solution or a good approximation of a complex scientific problem.

1. INTRODUCTION

Electroconvulsive therapy (ECT) formerly known as electroshock, is a psychiatric treatment used to treat severe depressed patients. People with symptoms such as delusions, hallucinations, or suicidal thoughts or when other treatments such as psychotherapy and antidepressant medicines have not worked, are exposed to ECT. The composite activity or measure of brain waves is called an electroencephalogram (EEG). Nowadays, there is no common statistical approach to analyze the EEG data. Some analysis of EEG signals using dynamic models have been conducted in neurophysiology (e.g. [1-3]). Recent research efforts have been based on various numerical algorithms that estimate individual dynamics and statistical information characteristics have been considered in more detail [4, 5]. The dynamical approach was subsequently applied to a variety of physiological time series, including EEG recordings [6-10]. Various mathematical techniques developed for the studies of complex, nonlinear and chaotic systems have also been used to explore the dynamics of human brain physiology [8, 11, 12].

The fractal dimension (FD) is used to quantify the complexity and self-similarity of a signal [13]. FD analysis of EEG data has been applied in the biomedical research, such as the routine detection of dementia [14, 15] EEG analysis of sleeping newborns [16] and fractal spectral analysis of pre-epileptic seizures [17]. The use of FD in the analysis of EEG data was suggested by Popivanov, et al. [18]. Wahlund, et al. [19] proposed an approach where one looked at FD of the curve as the representation of EEG data obtained from ECT seizures to identify differences in responses between parts of the brain. In some other studies, estimation of FD of similar biological signal has been used. For the data used from magneto encephalogram (MEG), Gómez, et al. [20] estimated the complexity of those

signals in Alzheimer's patients and healthy elderly. They found significantly lower complexity of MEG in Alzheimer's patients compared to the control subjects.

Classical approaches where measurement errors are regarded as generators of uncertainty are not valid when considering EEG time series data [21]. Due to strong individual responses to ECT one possible approach is to use mixed models where individual effects are taken into account via random effects. These individual effects can then be discarded or incorporated in the analysis. The multivariate analysis of variance (MANOVA) is a type of multivariate statistical technique to analyze data that involves more than one response at a time. One of the aims of this work is to apply multivariate statistical techniques for the analysis of EEG signals recorded from depressed patients. Complexity of EEG signal is assessed by using FD according to Wahlund, et al. [19].

In the next section, background characteristics of EEG data including the assessing complexity of EEG signal will be discussed. In Section 3, multivariate approach to the statistical analysis of EEG data will be presented. Issue of model checking as assessment of the multivariate normality, and equality of covariance matrices also discussed here. In Section 4, the obtained results will be presented. Finally, the last section comprises the discussion.

2. BACKGROUND

ECT is used for people whose depression is severe or life threatening or who cannot take antidepressant medication. ECT is then often recommended for use as a treatment for severe depression that has not responded to other treatment, and is also used in the treatment of mania and catatonia. Before ECT, a patient is given anesthesia, and medicines to loosen up his muscles. Then electrical current is briefly sent to the brain through electrodes placed on the temples or elsewhere on the head, depending on the condition and type of ECT. Usually ECT treatments are done 2 to 3 times a week for 2 to 3 weeks and continuing for several months to a year by reducing time frequency, to reduce the risk of relapse. ECT is usually given in combination with medicine, psychotherapy, family therapy, and behavioral therapy.

EEG is the recording of electrical activity along the scalp and measures the voltage fluctuating result from ionic current flows within the neurons of the brain. In clinical contexts, EEG refers to the recording of the brain's spontaneous electrical activity over a short period of time, usually 20–40 minutes, as recorded from multiple electrodes placed on the scalp. The EEG is used in the evaluation of brain disorders. Most commonly it is used to show the type and location of the activity in the brain during a seizure. The EEG is typically described in terms of rhythmic activity and transients. The rhythmic activity is divided into bands by frequency. To some degree, these frequency bands are a matter of nomenclature, but these designations arose because rhythmic activity within a certain frequency range was noted to have a certain distribution over the scalp or a certain biological significance. Most of the cerebral signal observed in the scalp EEG falls in the range of 1–20 Hz. The activity below or above this range is likely to be artifactual, under standard clinical recording techniques. However, brain wave EEG has a complex pattern of frequencies. Experimentally, it has been established that there are several frequently bands that's associated with particular processes in the brain : delta waves (<4 Hz), thetawaves (4-8 Hz), alpha waves (8-13 Hz), beta waves (13-36 Hz) and gamma waves (> 36 Hz).

In the next section, analysis and assessing the complexity of EEG signals will be discussed.

2.1. Analysis and Assessing Complexity of EEG Signals

Multivariate analysis of variance is one of the techniques for analyzing multivariate data. Now-a-days, multivariate statistical analysis is a very popular technique applied in such research fields as education, biology, medicine and social sciences. If there have categorical independent variables with two or more levels and a set of dependent variables and one can interested in determining if the categorical independent variables affect the dependent variables. MANOVA can be used to address of the preceding problem. One reason for conducting a MANOVA is to determinate how variates can be combined to maximally discriminating between district subgroups

of the responses. Many research articles and books have been devoted to MANOVA in the last twenty five years. Weinfurt [22] and Huberty and Petoskey [23] introduced complex and conceptually powerful statistical procedure for MANOVA in their books.

Amplitude, frequency, and degree of synchronization constitute three basic parameters of background EEG activity. The first two parameters are easy to quantify, but recently synchronization only has become a subject of computerized EEG analysis. The opposite of synchronization is complexity of EEG signal. Signal complexity [24] can be analyzed either in frequency domain, or in time domain, or in the system's phase space. Analysis in the frequency domain requires fourier or wavelet transform of the signal, while analysis in the phase space requires embedding of the data in a multidimensional space. Complexity in time domain may be characterized by fractal dimension, D_f . It characterizes complexity of the curve representing the signal on a plane. The fractal dimension D_f has values between 1 and 2. D_f calculated in time domain should not be confused with fractal dimension of the attractor in the system's phase space. An object in Euclidean space, like a coastline of an island, may also be characterized by a fractal dimension. In a rough sense, it measures 'how many points' belong to a given set. A plane is 'larger' than a line, while Sierpinski triangle sits somewhere in between these two sets. On the other hand, all three of these sets have the same number of points in the sense that each set is uncountable. Somehow, though, fractal dimension captures the notion of 'how large a set is' while its fractional part captures the notion of 'how complex a set is'. In 1988, Higuchi proposed an efficient algorithm for measuring the fractal dimension (FD) of discrete time sequences [25]. Higuchi's algorithm calculates the FD directly from time series. Higuchi's fractal dimension (HFD) has already been used to analyze the complexity of brain recording [13, 26] and other some biological signals [27].

Let the curve of fractal dimension D_f is denoted by $L(k)$, representing amplitude of the signal under consideration in a plane as a function of time. If the length of this curve is scales as

$$L(k) \approx K^{-D_f}, \quad 1 < D_f < 2,$$

the curve is said to have fractal dimension D_f .

If $D_f = 1$ then it is a simple line and if $D_f = 2$ then the curve nearly fills out the whole plane. In general, using D_f a huge number of data points would be summarized in an interpretable way. Higher values of D_f correspond to the presence of higher frequencies in the signal's Fourier spectrum.

Let us define $x(1), x(2), \dots, x(N)$ be the values of a finite set of time series observations, which are taken in a regular interval, where $i = 1, 2, \dots, N$ and N is the number of points in the time series. Here x are the successive EEG values.

For a ranges of k values, $k = 1, 2, \dots, k_{max}$, construct k new times series x_m^K ,

$$x_m^K = \{x(m), x(m+k), \dots, x(m+ik), \dots, x\left(m + \text{int}\left(\frac{N-m}{k}\right)k\right)\}, \quad m=1, 2, \dots, k,$$

where k and m are intergers and $\text{int}(\bullet)$ is the integer part, k indicates the discrete time interval between points, whereas $m=1, 2, \dots, k$, represents the initial time value. Now the length $L_m(k)$ of each new time series, i.e. the length of the curve x_m^K , can be defined as follows;

$$L_m(k) = \frac{1}{k} \left[\sum_{i=1}^{N^*} |x(m+ik) - x(m+(i-1)k)| \frac{(N-1)}{N^* \times k} \right],$$

where $\frac{(N-1)}{N^* \times k}$ is a normalization factor, and $N^* = \text{int}\left(\frac{N-m}{k}\right)$.

Thus, the length of the curve for the time interval k is defined as the average of the average of the k values $L_m(k)$, $m= 1,2,\dots,k$,

$$L(k) = \frac{1}{k} \sum_{m=1}^k L_m(k),$$

$$\ln L(k) = \text{constant} - D_f \ln(k) + \text{error}(k),$$

where $k = k_{\min}, \dots, k_{\max}$.

The value of k_{\min} and k_{\max} have to be chosen in advance. Finally, when $L(k)$ is plotted against $(1/k)$ on a double logarithmic scale, with $k = 1, 2, \dots, k_{\max}$, the data should exhibit a straight line, with a slope D_f equal to the FD of EEG. Thus HFD is defined as the least square estimator of a slope of the line that fits the $\{\ln [L(k)], \ln (1/k)\}$.

In the next section, we will discuss about the multivariate tests that are common used in MANOVA. Also multivariate tests based on contrasts will discuss and multivariate modeling of fractal dimension with estimation of parameters.

3. MULTIVARIATE ANALYSES

Let us consider the following MANOVA model:

$$\mathbf{X} = \mathbf{BC} + \mathbf{E},$$

where \mathbf{X} : $p \times n$, \mathbf{B} : $p \times k$, \mathbf{C} : $k \times n$ and $\mathbf{E} \sim N_{p,n}(0, \Sigma, \mathbf{I})$. Matrices \mathbf{B} and Σ are unknown parameters. \mathbf{X} is the data matrix; \mathbf{C} is known between individual design matrix and the columns of \mathbf{E} are independent and the dispersion of each column equal to Σ .

To test the following null hypothesis about equality of the mean responses with the assumption; the covariance matrices of the different groups of responses are assumed to be equal, Wilks' lambda, Lawley-Hotelling's trace, Pillai's Trace and Roy's largest root tests [28] can be used to test the equality of the mean responses. These tests, however, may become seriously biased when the assumption of equality of covariance matrices is violated.

Now we will discuss test statistics for contrasts.

3.1. Multivariate Significance Test for the Contrasts

A contrast is a linear combination of two or more group means. We have a contrast among the population mean vectors $\mu_1, \mu_2, \dots, \mu_k$ given by

$$\mathbf{C}_i = c_{i1}\mu_1 + c_{i2}\mu_2 + \dots + c_{ik}\mu_k,$$

where μ_k is the mean vector for the k th group and \mathbf{C}_i is the i th contrast vector.

This contrast is estimated by replacing the population mean vectors by the sample vectors:

$$\hat{\mathbf{C}}_i = c_{i1}\mathbf{\mu}_1 + c_{i2}\mathbf{\mu}_2 + \dots + c_{ik}\mathbf{\mu}_k.$$

The test statistic for the above contrast with the null hypothesis $H_0: C_i=0$ versus $H_a: C_i \neq 0$ is given by

$$T^2 = \left(\sum_{j=1}^k \frac{C_{ij}^2}{n_j} \right)^{-1} C_i' S_w^{-1} C_i,$$

where $S_w = f^{-1} \sum_{i=1}^k \sum_{j=1}^n (X_{ij} - \bar{X}_i)(X_{ij} - \bar{X}_i)'$ is the sample pooled within-group covariance matrix. The test

statistic T^2 can be transformed into an F statistic with p and $(df_e - p + 1)$ degrees of freedom,

$$F = \left(\frac{df_e - p + 1}{df_e \times p} \right) T^2,$$

where df_e is the error degrees of freedom.

3.2. Multivariate Modeling of Fractal Dimension

In the model $X = BC + E$, our interest is to test the hypothesis $GB=0$, where G is known and formulated according to null hypothesis. However, we shall extend the model (3) and consider

$$X = BC + E, \quad GBF=0, \quad (5)$$

where F is known. According to Kollo and Rosen [29] this model is equivalent to an extended Growth Curve model, i.e.

$$X = A_1 B_1 C_1 + A_2 B_2 C_2 + E,$$

where $E \sim N(0, \Sigma, I)$ and $A_1 = (G')^o$, $C_1 = C$, $A_2 = (G')$ and $C_2 = F'C$, $(G')^o$ is any matrix spanning the orthogonal complement to the space generated by column of (G') . It is important to observe that the column space of C_2' is included in the column space of C_1' . According to Kollo and Rosen [29] the estimator of B_1 , B_2 and Σ of this model obtained as well as moments and density approximations for each of them.

In this work we are interested in model (5) and the following hypotheses

$$H_0: GBF=0 \text{ against } H_a: GBF \neq 0$$

The test statistic is

$$U = \frac{|V|}{|V+W|} = \frac{1}{|I+V^{-1}W|}, \quad (6)$$

where

$$V = G(A'S^{-1}A)^{-1}G', \quad (7)$$

$$W = (GBF)(F'RF)^{-1}(F'\hat{B}'G'), \quad (8)$$

$$\hat{B} = (A'S^{-1}A)^{-1}AS^{-1}XC'(CC')^{-1}, \quad (9)$$

$$R = (CC')^{-1} + (CC')^{-1}CX'(S^{-1} - S^{-1}A(A'S^{-1}A)^{-1}A'S^{-1})XC'(CC')^{-1},$$

$$S = X(I - C'(CC')^{-1}C)X', \quad (10)$$

To test the $H_0: \mathbf{GB}=\mathbf{0}$, i.e. $\mathbf{F}=\mathbf{I}$, the identity matrix, Srivastava [30] suggested to use the likelihood ratio test statistic. The null hypothesis H_0 is rejected if

$$U = \frac{|\mathbf{SSE}|}{|\mathbf{SSE} + \mathbf{SSTR}|} \leq c,$$

where c is chosen such that the error of first type is α , and

$$\mathbf{SSE} = \mathbf{V} = \sum_j^K \sum_i^{n_j} (\mathbf{x}_{ij} - \bar{\mathbf{x}}_j)(\mathbf{x}_{ij} - \bar{\mathbf{x}}_j)'$$

$$\mathbf{SSTR} = \mathbf{W} = \sum_j^K \mathbf{n}_j (\bar{\mathbf{x}}_j - \bar{\mathbf{x}}_{..})(\bar{\mathbf{x}}_j - \bar{\mathbf{x}}_{..})'$$

where $f = n - k$, $i = 1, 2, 3, \dots, n_j$ and $j = 1, 2, 3, \dots, k$.

The following asymptotic expansion gives a good approximation for any constant $z > 0$,

$$P[-(f - \frac{1}{2}(p-m+1)) \ln U > z] = P[\chi_{pm}^2 > z] + \frac{1}{f^2} \gamma [P(\chi_{pm+4}^2 \geq z) - P(\chi_{pm}^2 \geq z)] + O(f^{-4})$$

where $m = k - 1$, $\gamma = \frac{pm(p^2 + m - 5)}{48}$, χ_{pm}^2 denotes the chi-square random variate with DF pm , and $O(f^{-4})$ represents terms that goes to zero.

For large f ,

$$P[-(f - \frac{1}{2}(p-m+1)) \ln U > z] \cong P(\chi_{pm}^2 > z)$$

Therefore the $H_0: \mathbf{GB}=\mathbf{0}$ is rejected if $-(f - \frac{1}{2}(p-m+1)) \ln U \geq \chi_{pm}^2$.

Next, we present the hypotheses which will be of interest when analyzing EEG data.

Let μ_{ijk} is the mean D_f of k th phases with j th electrode locations and i th diagnosis groups. Here $i = 1, 2, 3$; $j = 1, 2$; and $k = 1, 2$. Also define the following matrices that will use in different types of null hypothesis:

$$\mathbf{G}_1 = \begin{pmatrix} 1 & 0 & -1 & 0 \\ 0 & 1 & 0 & -1 \end{pmatrix},$$

$$\mathbf{G}_2 = \begin{pmatrix} 1 & -1 & 0 & 0 \\ 0 & 0 & 1 & -1 \end{pmatrix},$$

$$\mathbf{F} = \begin{pmatrix} 1 & 0 \\ -1 & 1 \\ 0 & -1 \end{pmatrix},$$

and the parameter \mathbf{B} in (4) is defined as,

$$\mathbf{B} = \begin{pmatrix} \mu_{111} & \mu_{211} & \mu_{311} \\ \mu_{112} & \mu_{212} & \mu_{312} \\ \mu_{121} & \mu_{221} & \mu_{321} \\ \mu_{122} & \mu_{222} & \mu_{322} \end{pmatrix}.$$

The following hypotheses are of interest:

- (1) H_{01} : There is no difference between the two electrodes locations (Fp1 and Fp2), which is equivalent to $\mathbf{G}_1\mathbf{B} = \mathbf{0}$, where \mathbf{G}_1 is defined earlier.

If $H_0: \mathbf{G}_1\mathbf{B} = \mathbf{0}$ then

$$-(f - \frac{1}{2}(r - t + 1)) \ln U \sim \chi_{rt}^2 \quad (11)$$

asym.

where $r = \text{rank of } \mathbf{G}_1$ and $t = \text{rank of } \mathbf{F}$. U and f is defined in page 14.

- (2) The second hypothesis of interest is

H_{02} : There is no difference between the two phases (Phase I-III and Phase IV), i.e. which is equivalent to $H_{02}: \mathbf{G}_2\mathbf{B} = \mathbf{0}$, where \mathbf{G}_2 is defined above.

To test $\mathbf{G}_2\mathbf{B} = \mathbf{0}$, the test statistic in (11) will be used.

- (3) The third hypothesis of interest is

H_{03} : There is no difference between the diagnostic groups, i.e. $H_0: \mathbf{BF} = \mathbf{0}$, where \mathbf{F} is defined in previous page.

To test $\mathbf{BF} = \mathbf{0}$, the test statistic in (11) will be used.

Now, one can extend the null hypothesis from $H_0: \mathbf{GB} = \mathbf{0}$ to $H_0: \mathbf{GBF} = \mathbf{0}$.

- (4) The fourth hypothesis of interest is

H_{04} : There is no interaction effect between electrodes locations and the diagnostic groups, i.e. $H_0: \mathbf{G}_1\mathbf{BF} = \mathbf{0}$, where \mathbf{G}_1 and \mathbf{F} are defined in previous page.

To test $\mathbf{G}_1\mathbf{BF} = \mathbf{0}$, the test statistic in (11) will be used.

- (5) The fifth hypothesis of interest is

H_{04} : There is no interaction effect between phases and the diagnostic groups.

i.e. $H_0: \mathbf{G}_2\mathbf{BF} = \mathbf{0}$, where \mathbf{G}_2 and \mathbf{F} are defined in previous page.

To test $\mathbf{G}_2\mathbf{BF} = \mathbf{0}$, the test statistic in (11) will be used.

In next section, we will focus on model validation.

3.3. Checking the Equality of Variance-Covariance Matrices

Another important assumption of MANOVA is that the covariance matrices across all groups are equal. It implies that there exists no substantive difference between the variances of different groups of dependent variables. In a simple regression model, it is defined as the homogeneity of residual variances.

Let X_{ij} be the j th set of response variables in the i th cell; n_i is number of cases in the i th cell; total sample size, $n = n_1 + n_2 + \dots + n_k$. Hence

$$\text{Means, } \bar{\mathbf{X}}_i = \sum_{j=1}^{n_i} \frac{\mathbf{X}_{ij}}{n_i},$$

Cell covariance matrix; $\hat{\Sigma}_i = S_i = \frac{1}{n_i - 1} \sum_{j=1}^{n_i} (\mathbf{x}_{ij} - \bar{\mathbf{x}}_i)(\mathbf{x}_{ij} - \bar{\mathbf{x}}_i)'$,

Pooled covariance matrix; $\hat{\Sigma} = S = \frac{1}{n - k} \sum_{i=1}^k (n_i - 1)S_i$,

Box's M test statistic is used to test the null hypothesis about the homogeneity of covariance matrices. In the next section, about the data are discussed.

4. DATA

The data used in this example are synthetic data, i.e. responses from patients are created so that they resemble responses from real study patients. The sample consists of 28 patients with severally depressed patients and with three different diagnoses. Five patients have psychotic depression (PD), seven have bipolar disorder (BD) and sixteen patients have unipolar depression (UD). Two channels of EEG of left and right prefrontal-to-ipsilateral coded Fp1 and Fp2, respectively, were recorded. The Fp1 and Fp2 were assessed according to the international 10-20 system (see Figure1), an internationally recognized method to describe and apply the location of scalp electrodes in the context of an EEG test. This system is based on the relationship between the location of an electrode and the underlying area of cerebral cortex. The "10" and "20" refer to the actual distances between two electrodes are either 10 or 20 percent of the total front-back or right-left distance of the scalp.

The digitalized data from the digital port of MECTA equipment were simultaneously recorded and transferred to computer hard disc drive. The continuous EEG signals were digitalized at a sampling frequency of 128 Hz. None of the patients had received ECT during the last three months. Every patient was treated 6-12 times with recording of EEG and the average of the readings was used in the analysis.

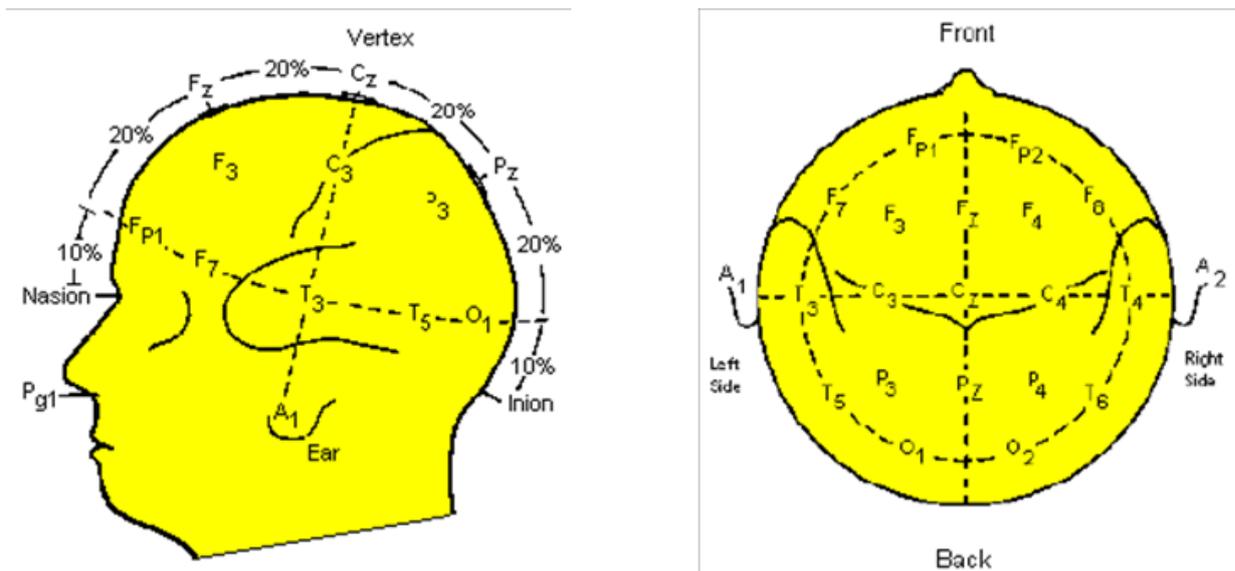


Figure-1. Map of locations of the EEG electrodes/channels

Source: Map of locations of the EEG electrodes/channels on the scalp according to International 10-20 System. EEGs were recorded from Fp1 and Fp2, which are the standard locations for treating patients with ECT.

Now we will discuss the obtained results.

5. STATISTICAL ANALYSIS

5.1. Descriptive Statistics

Table 5.1 contains the mean values of D_f of two phases Phase I-III and Phase IV for the two different electrodes/channels from different diagnoses groups of depressed patients. The result shows that the mean values of D_f for phase I-III are lower than those for phase IV. Lower SE values implies lower variability of D_f are observed for the both channels Fp1 and Fp2 for Phase I-III.

Table-5.1. Average of fractal dimension D_f (SE) value for different diagnoses groups of depressed patients; from Phase I-III and Phase IV at different electrode locations.

Diagnosis	N	Fp1		Fp2	
		Phase I-III	Phase-IV	Phase I-III	Phase-IV
Psychotic depression	5	1.210 (0.018)	1.538 (0.049)	1.194 (0.022)	1.712 (0.064)
Bipolar disorder	7	1.266 (0.025)	1.513 (0.053)	1.261 (0.023)	1.553 (0.069)
Unipolar depression	16	1.250 (0.015)	1.584 (0.036)	1.251 (0.015)	1.630 (0.044)

5.2. Multivariate Analysis of Variance

Multivariate tests discussed in Section 3.1, are conducted to determine if the mean vectors for three different groups (PD, BD and UD) of the four dependent variables are significantly different or not (an overall or global test for equality of mean vectors). The results presented in Table 5.2 (significant that indicates at the 5% level of significance) indicate that mean vectors for three groups are statistically different.

Table-5.2. Different test statistics for the null hypothesis; H_0 : there is no significance vector mean difference between the different diagnosis groups of the fractal dimension D_f .

Name of the statistic's	Statistic	p-value
Wilks' Lamda	0.4918	0.0342
Pillai's Trace	0.5404	0.0499
Hotelling-Lawley Trace	0.9677	0.0236
Roy's Greatest Root	0.8945	0.0041

The univariate test for individual dependent variables with the null hypothesis about equality of the means of the different groups, are not significant at 5% level (see Table 5.3).

Table-5.3. Univariate analysis of variance for individual response variables with H_0 : there have no mean differences between the different diagnoses of the individual fractal dimension D_f .

Variables	Between Group MSS	Between Group MSS	F Statistics (p-values)	Bartlett's Chi-Square
Fp1_Phase I-III	0.0061	0.0032	1.88(0.173)	1.299 (0.522)
Fp1_Phase IV	0.0126	0.0192	0.66(0.526)	0.702(0.704)
Fp2_Phase I-III	0.0090	0.0034	2.68(0.089)	0.985(0.611)
Fp2_Phase IV	0.0484	0.0290	1.67(0.209)	0.333(0.847)

5.3. Analysis of Contrasts

The overall multivariate test showed significant (see the Section 3.2), which means that considerable differences between different groups exist. We proceed with studying pair wise difference and some predefined contrasts. We set the multivariate test of PD versus the average of BD and UD, then BD versus the average of PD and UD, and UD versus the average of PD and BD. To find where the differences are, some contrasts shown in Table 5.4 are studied. PD shows statistical difference with the average of BD and UD, BD is also different from the average of PD and UD but UD is not significantly differ from the average of PD and BP. Further the pair wise comparison shows that in every pair the average are statistically different from one diagnosis group to another except (PD, UP) pair.

Table-5.4. Test statistics for different types of contrasts testing equality of group means of the fractal dimension D_f .

Contrasts	Wilk's Λ	Pillai's Trace	Hotelling-Lawley Trace	Roy's Greatest Root	F Statistic	p-value
<i>Contrasts with three groups of diagnoses</i>						
$2(\text{PD})-\text{BD}-\text{UP}=0$	0.6169	0.3831	0.6211	0.6211	3.42	0.0257
$2(\text{BD})-\text{PD}-\text{UP}=0$	0.5410	0.4590	0.8484	0.8484	4.67	0.0070
$2(\text{UD})-\text{PD}-\text{BP}=0$	0.8511	0.1489	0.1750	0.1750	0.96	0.4476
<i>Pair wise contrasts</i>						
$\text{PD}-\text{BD}=0$	0.5331	0.4669	0.8757	0.8757	4.82	0.0061
$\text{PD}-\text{UP}=0$	0.8613	0.1387	0.1611	0.1611	0.89	0.4886
$\text{BD}-\text{UP}=0$	0.6305	0.3695	0.5859	0.5859	3.22	0.0317

PD, BD and UP stand for the mean vector of fractal dimension for psychotic depression, bipolar disorder and unipolar depression respectively.

5.4. Multivariate Modeling of Fractal Dimension

For the modeling of fractal dimension D_f will use the general multivariate model defined in Section 3.2, i.e.

$$\mathbf{X} = \mathbf{B}\mathbf{C} + \mathbf{E},$$

where \mathbf{X} : 4×28 , \mathbf{B} : 4×3 , \mathbf{C} : 3×28 and $\mathbf{E} \sim N_{4,28}(\mathbf{0}, \mathbf{\Sigma}, I)$. Matrices \mathbf{B} and $\mathbf{\Sigma}$ are unknown parameters. In this model, each column of \mathbf{X} represents one individual in Phase I-III and Phase IV for FP1 and FP2, respectively.

Here, the data matrix \mathbf{X} is the following

$$\mathbf{X}' = \begin{pmatrix} 1.22 & 1.45 & 1.13 & 1.55 \\ 1.19 & 1.40 & 1.19 & 1.62 \\ \dots & \dots & \dots & \dots \\ 1.24 & 1.49 & 1.23 & 1.57 \\ 1.28 & 1.51 & 1.29 & 1.78 \end{pmatrix},$$

The columns here represent 4 repeated measurement on patients, i.e. Fp1_phase I-III, Fp1_phase IV, Fp2_phase I-III, Fp2_phase IV.

The between-individual design matrix \mathbf{C} corresponds to the 3 diagnoses groups is

$$\mathbf{C} = \begin{pmatrix} \mathbf{1}'_5 & 0 & 0 \\ 0 & \mathbf{1}'_7 & 0 \\ 0 & 0 & \mathbf{1}'_{16} \end{pmatrix}$$

where $\mathbf{1}'_5, \mathbf{1}'_7, \mathbf{1}'_{16}$ denote row vectors of size 5, 7 and 16 respectively having all element equal to one.

Now, we would like to test the following hypothesis i.e. equality of two multivariate means,

H_{01} : There is no difference in the average D_f between the two electrodes locations Fp1 and Fp2, is equivalent to $\mathbf{G}_1\mathbf{B} = 0$,

where \mathbf{G}_1 and \mathbf{B} is defined in Section 3.2.

Under the null hypothesis the test statistic (described in the Section 3.2)

$$-(f - \frac{1}{2}(r - t + 1)) \ln U,$$

approximately follows chi-squared distribution with rt DF as f is large. Here $f = n$ - number of column in $\mathbf{G}'_1 = 28 - 2 = 26$, $r = \rho(\mathbf{G}_1) = 2$ and $t = \rho(\mathbf{F}) = 2$.

We obtain the following results:

The estimated total sum squared matrix described in (10),

$$\mathbf{S} = \begin{pmatrix} 0.0835714 & 0.0154857 & 0.0736429 & 0.0115857 \\ 0.0154857 & 0.4772166 & 0.0239677 & 0.4519629 \\ 0.0736429 & 0.0239677 & 0.0870995 & 0.0689314 \\ 0.0115857 & 0.4519629 & 0.0689314 & 0.7476229 \end{pmatrix}$$

where

$$\mathbf{S} = \mathbf{n} \cdot \hat{\Sigma}$$

and the estimated \mathbf{B} describe in (9) is,

$$\hat{\mathbf{B}} = \begin{pmatrix} 1.210 & 1.265714 & 1.250000 \\ 1.538 & 1.512857 & 1.584375 \\ 1.194 & 1.261429 & 1.250625 \\ 1.712 & 1.552857 & 1.630000 \end{pmatrix}$$

Further, \mathbf{V} in (7) is,

$$\mathbf{V} = \begin{pmatrix} 0.5298166 & 0.4900523 \\ 0.4900523 & 0.6968595 \end{pmatrix},$$

and \mathbf{W} in (8) is,

$$\mathbf{W} = \begin{pmatrix} 0.038726 & 0.051062 \\ 0.051062 & 0.150012 \end{pmatrix}.$$

So $|\mathbf{V}| = 0.129056$, and $|\mathbf{V} + \mathbf{W}| = 0.188678$, giving $U = 0.684004$.

Now the observed value of the test statistic is $-(f - \frac{1}{2}(r - t + 1)) \ln U = 9.685$, leading to the p-value 0.046. Thus the null hypothesis is rejected at 5% level of significance which implies that there is a significant difference in the average D_f between the two points Fp1 and Fp2.

The second hypothesis to test is

H_{02} : There is no difference the average D_f between the two phases (Phase I-III and Phase IV), i.e. which is equivalent to

$$H_{02}: \mathbf{G}_2 \mathbf{B} = 0,$$

where \mathbf{G}_2 and \mathbf{B} is defined in Section 3.2.

Under the null hypothesis the test statistic

$$-(f - \frac{1}{2}(r - t + 1)) \ln U,$$

approximately follows chi-squared distribution with rt df as f is large.

$$\text{So } |\mathbf{V}| = 0.005117 \quad \text{and } |\mathbf{V} + \mathbf{W}| = 0.007873, \text{ and } U = 0.649965.$$

where

$$\mathbf{V} = \begin{pmatrix} 0.023385 & 0.048864 \\ 0.048864 & 0.150012 \end{pmatrix},$$

$$\mathbf{W} = \begin{pmatrix} 0.001058 & -0.00795 \\ -0.00795 & 0.069658 \end{pmatrix}.$$

The observed value of test statistic is $-(f - \frac{1}{2}(r - t + 1)) \ln U = 10.986$, leading p-value 0.027. Here $r = \rho(G_2) = 2$ and $t = \rho(F) = 2$. We reject the null hypothesis at 5% level of significance which implies that there is a significant difference in mean response of the average D_j between the two phases Phase I-III and Phase IV.

Now the third null hypothesis is;

H_{03} : There is no difference between the three diagnostic groups, i.e. $\mathbf{BF} = \mathbf{0}$, where \mathbf{B} and \mathbf{F} is defined in Section 3.3.

The following test statistic is used

$$-(f - \frac{1}{2}(r - t + 1)) \ln U .$$

Under H_0 it approximately follows chi-squared distribution with rt DF as f is large.

We have $|\mathbf{V}| = 0.155$ and $|\mathbf{V} + \mathbf{W}| = 0.698$, giving $U = 0.222$. here

$$\mathbf{V} = \begin{pmatrix} 0.1326523 & -0.0260233 \\ -0.0260233 & 1.1727706 \end{pmatrix},$$

$$\mathbf{W} = \begin{pmatrix} 0.1568740 & -0.0009620 \\ -0.0009620 & 1.2400094 \end{pmatrix}.$$

The observed value of the test statistic $-(f - \frac{1}{2}(r - t + 1)) \ln U = 38.38741$ leading p value < 0.001 , where $r = \rho(G_2) = 2$ and $t = \rho(F) = 2$. We reject the null hypothesis at 5% level of significance which implies that there is a significant difference in mean response of D_j between the three diagnoses.

Next null hypothesis is;

H_{04} : There is no interaction effect between electrodes locations and the diagnostic groups, i.e. $\mathbf{G}_1 \mathbf{BF} = \mathbf{0}$, where \mathbf{G}_1 , \mathbf{B} and \mathbf{F} is defined in Section 3.3.

The following test statistic is used

$$-(f - \frac{1}{2}(r - t + 1)) \ln U .$$

Under H_0 it approximately follows chi-squared distribution with rt DF as f is large.

We have $|\mathbf{V}| = 0.00512$ and $|\mathbf{V} + \mathbf{W}| = 0.007873$, giving $U = 0.649962$. here

$$\mathbf{V} = \begin{pmatrix} 0.023385 & 0.048864 \\ 0.048864 & 0.320914 \end{pmatrix},$$

$$\mathbf{W} = \begin{pmatrix} 0.0010577 & -0.0079495 \\ -0.0079495 & 0.0696577 \end{pmatrix}.$$

The observed value of the test statistic $-(f - \frac{1}{2}(r - t + 1)) \ln U = 10.986$ leading p value = 0.0267, where $r = \rho(G_2) = 2$ and $t = \rho(F) = 2$. We reject the null hypothesis at 5% level of significance which implies that there is interaction effect between electrodes locations and the diagnostic groups.

Next null hypothesis of our interest is;

H_{05} : There is no interaction effect between phases (Phase I-III and Phase IV) and the diagnostic groups, i.e. $\mathbf{G}_2 \mathbf{BF} = \mathbf{0}$, where \mathbf{G}_2 , \mathbf{B} and \mathbf{F} is defined in Section 3.3.

The following test statistic is used

$$-\left(f \frac{1}{2} (r - t + 1)\right) \ln U .$$

Under H_0 it approximately follows chi-squared distribution with rt DF as f is large.

We have $|\mathbf{V}| = 0.12906$ and $|\mathbf{V} + \mathbf{W}| = 0.188678$, giving $U = 0.684004$.

here

$$\mathbf{V} = \begin{pmatrix} 0.529817 & 0.490052 \\ 0.490052 & 0.696859 \end{pmatrix},$$

$$\mathbf{W} = \begin{pmatrix} 0.0387263 & 0.051062 \\ -0.051062 & 0.150012 \end{pmatrix},$$

The observed value of the test statistic $-\left(f \frac{1}{2} (r - t + 1)\right) \ln U = 9.684$ leading p value = 0.04608, where $r = \rho(\mathbf{G}_2) = 2$ and $t = \rho(\mathbf{F}) = 2$. We reject the null hypothesis at 5% level of significance which implies that there is interaction effect between phases and the diagnostic groups.

5.5. Assessing Equality of Group Covariance Matrices

The value of Box's M statistic to test the equality of these three group covariance matrices is 17.82 with p -value greater than 0.10, indicates that there exists no substantive difference between the variances of different response groups.

Now we will summarize the results of this study.

6. DISCUSSION

In this work, multivariate analysis of EEG data comprises multivariate analysis of variance and discriminant analysis. We also assessed local complexity of EEG signals by fractal dimension. Usually in ECT, data for EEG is obtained from a small sample. The main goal of the master thesis was to work on diagnostic of the multivariate model that we used to study the complexity of EEG. In particular, the performance of a newly proposed method for identifying outliers based on the eigenvalues of the sample covariance matrix was investigated here, and compared to Mahalanobis distance as based method.

The descriptive statistics showed that the mean values of fractal dimension D_f for phase I-III are always lower than that for phase IV, also with lower SE values. That implies the low variability of D_f with low mean are observed for the both channels Fp1 and Fp2 for Phase I-III. At the 5% level of significance, mean vectors for the three group diagnoses are statistically different.

There is a significant difference in mean response of D_f between the electrodes locations Fp1 and Fp2 as well as significant difference between the phases Phase I-III and Phase IV. Furthermore, it has been shown that there is a significant difference in mean response of D_f between the three diagnoses. There have been detected significant interaction effects between electrode locations and diagnoses. Significant interaction effects were found between phases and diagnoses group.

Using graphical method, the response variables also do not deviate from multivariate normality. Insignificant Mardias statistic based on skewness and kurtosis showed that there is no significant deviation from the multivariate normal distribution. Chi-square (χ^2) test statistic based on skewness and kurtosis shows all the response variables are normal individually except Fp1_Phase IV. The value of Box's M statistic to test the equality of these three group

covariance matrices is not significant, indicates that there exists no substantive difference between the variances of different response groups.

The outlying observations were identified based on Mahalanobis distance and MED method. By MD method, there are identified two multivariate outlying observations whereas using MED method four multivariate outlying observations were found. In general, one could see that MED works better compare to MD method as we can identify the outlying observations at first time calculation whereas in MD method needs fourth time calculation. This result confirms [31] results. According to Gao, et al. [31] MED method is a quite powerful tool for identifying multivariate outlying observations. This thesis also gives mathematical ideas from multivariate statistical analysis to find a quick solution or a good approximation of a complex scientific problem. Further simulation studies are needed to investigate the performance of MED method in case of small samples with covariance matrices having different structures, for example with specific pattern or heteroscedastic covariance matrices.

Funding: This study received no specific financial support.

Competing Interests: The author declares that there are no conflicts of interests regarding the publication of this paper.

Contributors/Acknowledgement: I would like to express the deepest appreciation to Dr Tatjana von Rosen, Associate Professor, Stockholm University, Sweden who always gives me a great support in regard to this work.

REFERENCES

- [1] W. J. Freeman, *Societies of brains*. Hillsdale, NJ: Erlbaum, 1995.
- [2] L. V. M. Quyen, J. Martinerie, M. Baulac, and F. Varela, "Anticipating epileptic seizures in real time by a non-linear analysis of similarity between EEG recordings," *Neuroreport*, vol. 10, pp. 2149-2155, 1999. [View at Google Scholar](#) | [View at Publisher](#)
- [3] A. J. MacLennan, P. R. Carney, W. J. Zhu, A. H. Chaves, J. Garcia, J. R. Grimes, K. J. Anderson, S. N. Roper, and N. Lee, "An essential role for the H218/AGR16/Edg-5/LPB2 sphingosine 1-phosphate receptor in neuronal excitability," *European Journal of Neuroscience*, vol. 14, pp. 203-209, 2001. [View at Google Scholar](#) | [View at Publisher](#)
- [4] K. Kaneko, *Complex system: Chaos and beyond: A constructive approach with applications in life sciences*. Berlin: Springer, 2001.
- [5] L. Diambra, J. B. de Figueiredo, and C. P. Malta, "Epileptic activity recognition in EEG recording," *Physica A: Statistical Mechanics and its Applications*, vol. 273, pp. 495-505, 1999. [View at Google Scholar](#) | [View at Publisher](#)
- [6] H. Haken, *Principles of brain functioning: A synergetic approach to brain activity, behaviour and cognition*. Berlin, Heidelberg, New York: Springer, 1996.
- [7] C. E. Elger and K. Lehnertz, "Seizure prediction by non-linear time series analysis of brain electrical activity," *European Journal of Neuroscience*, vol. 10, pp. 786-789, 1998. [View at Google Scholar](#) | [View at Publisher](#)
- [8] T. Ozaki, P. V. Sosa, and V. Haggan-Ozaki, "Reconstructing the nonlinear dynamics of epilepsy data using nonlinear time series analysis," *Journal of Signal Processing*, vol. 3, pp. 153-162, 1999. [View at Google Scholar](#)
- [9] T. Frank, A. Daffertshofer, C. Peper, P. Beek, and H. Haken, "Towards a comprehensive theory of brain activity: Coupled oscillator systems under external forces," *Physica D: Nonlinear Phenomena*, vol. 144, pp. 62-86, 2000. [View at Google Scholar](#)
- [10] R. G. Andrzejak, G. Widman, K. Lehnertz, C. Rieke, P. David, and C. Elger, "The epileptic process as nonlinear deterministic dynamics in a stochastic environment: An evaluation on mesial temporal lobe epilepsy," *Epilepsy Research*, vol. 44, pp. 129-140, 2001. [View at Google Scholar](#) | [View at Publisher](#)
- [11] K. Lehnertz, J. Arnold, P. Grassberger, and C. E. Elger, In: *Lehnertz, K., Arnold, J., Grassberger, P., Elger, C.E. (Eds.). Chaos in Brain?* Singapore: World Scienti Mc, 2000.

- [12] L. D. Iasemidis, P. Pardalos, J. C. Sackellares, and D.-S. Shiau, "Quadratic binary programming and dynamical system approach to determine the predictability of epileptic seizures," *Journal of Combinatorial Optimization*, vol. 5, pp. 9-26, 2001. [View at Google Scholar](#)
- [13] A. Accardo, M. Affinito, M. Carrozzi, and F. Bouquet, "Use of the fractal dimension for the analysis of electroencephalographic time series," *Biological Cybernetics*, vol. 77, pp. 339-350, 1997. [View at Google Scholar](#) | [View at Publisher](#)
- [14] G. Henderson, E. Ifeachor, H. Wimalaratna, E. Allen, and N. Hudson, "Prospects for routine detection of dementia using the fractal dimension of the human electroencephalogram," *IEE Proceedings-Science, Measurement and Technology*, vol. 147, pp. 321-326, 2000. [View at Google Scholar](#) | [View at Publisher](#)
- [15] R. Saji and H. Konno, "Dynamical features of the local fractal dimension of brain waves and its applicability for diagnosis of senile dementia," *Japanese Journal of Applied Physics*, vol. 39, pp. 679-684, 2000. [View at Google Scholar](#) | [View at Publisher](#)
- [16] A. P. Accardo, M. Affinito, M. Carrozzi, S. Cisint, and F. Bouquet, "Comparison between spectral and fractal EEG analyses of sleeping newborns," in *Conference Proceedings - IEEE Engineering in Medicine and Biology Society*, 1998, pp. 1569-1571.
- [17] X. Li, J. Polygiannakis, P. Kapisiris, A. Peratzakis, K. Eftaxias, and X. Yao, "Fractal spectral analysis of pre-epileptic seizures in terms of criticality," *Journal of Neural Engineering*, vol. 2, pp. 11-16, 2005. [View at Google Scholar](#) | [View at Publisher](#)
- [18] D. Popivanov, S. Jivkova, V. Stomonyakov, and G. Nicolova, "Effect of independent component analysis on multifractality of EEG during visual-motor task," *Signal Processing*, vol. 85, pp. 2112-2123, 2005. [View at Google Scholar](#) | [View at Publisher](#)
- [19] B. Wahlund, W. Klonowski, P. Stepien, R. Stepien, T. von Rosen, and D. von Rosen, "EEG data, fractal dimension and multivariate statistics," *Journal of Computer Science and Engineering*, vol. 3, pp. 10-14, 2010. [View at Google Scholar](#)
- [20] C. Gómez, Á. Mediavilla, R. Hornero, D. Abásolo, and A. Fernández, "Use of the Higuchi's fractal dimension for the analysis of MEG recordings from Alzheimer's disease patients," *Medical Engineering & Physics*, vol. 31, pp. 306-313, 2009. [View at Google Scholar](#) | [View at Publisher](#)
- [21] B. Wahlund, P. Piazza, D. von Rosen, B. Liberg, and H. Liljenström, "Seizure (Ictal)—EEG characteristics in subgroups of depressive disorder in patients receiving electroconvulsive therapy (ECT)—a preliminary study and multivariate approach," *Computational Intelligence and Neuroscience*, vol. 2009, p. 1-8, 2009. [View at Google Scholar](#) | [View at Publisher](#)
- [22] K. P. Weinfurt, *Multivariate analysis of variance*. In L. Grimm and P. Yarnold (Eds), *Reading and understanding multivariate statistics*. Washington DC: American Psychological Association, 1995.
- [23] C. J. Huberty and M. D. Petoskey, *Multivariate analysis of variance and covariance*. In H. Tinsley and S. Brown (Eds.), *Handbook of applied multivariate statistics and mathematical modeling*. NY: Academic Press, 2000.
- [24] J. Wackermann, "Beyond mapping: Estimating complexity of multichannel EEG recordings," *Acta Neurobiologiae Experimentalis*, vol. 56, pp. 197-208, 1996. [View at Google Scholar](#)
- [25] T. Higuchi, "Approach to an irregular time series on the basis of the fractal theory," *Physica D: Nonlinear Phenomena*, vol. 31, pp. 277-283, 1988. [View at Google Scholar](#) | [View at Publisher](#)
- [26] R. Ferenets, T. Lipping, A. Anier, V. Jantti, S. Melto, and S. Hovilehto, "Comparison of entropy and complexity measures for the assessment of depth of sedation," *IEEE Transactions on Biomedical Engineering*, vol. 53, pp. 1067-1077, 2006. [View at Google Scholar](#)
- [27] N. Kannathal and S. Krishnan, "Comprehensive analysis of cardiac health using heart rate signals," *Physiological Measurement*, vol. 25, pp. 1139-1151, 2004. [View at Google Scholar](#) | [View at Publisher](#)
- [28] T. W. Anderson, *An introduction to multivariate statistical analysis*, 3rd ed. New York: Wiley, 2003.
- [29] T. Kollo and V. D. Rosen, "A unified approach to the approximation of multivariate densities," *Scandinavian Journal of Statistics*, vol. 25, pp. 93-109, 1998. [View at Google Scholar](#) | [View at Publisher](#)

Appendix-2. SAS code to test the normality of the fractal dimension D_f of EEG data.

```

proc iml ;
x = {
1.22    1.45    1.13    1.55,
1.19    1.4    1.19    1.62,
.       .       .       .
.       .       .       .
1.28    1.51    1.29    1.78} ;
/* Matrix x can be created from a SAS data set as follows:
data eeg;
infile 'eeg.dat';
input x1 x2 x3 x4;
run;
proc iml;
use eeg;
read all into x;
*/
/* Here we determine the number of data points and the dimension of the vector. The variable dfchi is the degrees of freedom for
the chi square approximation of Multivariate skewness. */
n = nrow(x) ;
p = ncol(x) ;
dfchi = p*(p+1)*(p+2)/6 ;
/* q is projection matrix. s is the maximum likelihood estimate of the variance covariance matrix. g_matrix is n by n the matrix
of g(i,j) elements. beta1hat and beta2hat are respectively the Mardia's sample skewness and kurtosis measures. statistic1 and
statistic2 are the test statistics based on skewness and kurtosis to test for normality and pvalskew and pvalkurt are
corresponding p values. */
q = i(n) - (1/n)*j(n,n.1);
s = (1/n)*x`*q*x ;
s_inv = inv(s);
g_matrix = q*x*s_inv`*q;
beta1hat = ( sum(g_matrix#g_matrix#g_matrix) )/(n*n);
beta2hat = trace( g_matrix#g_matrix )/n ;
statistic1 = n*beta1hat/6 ;
statistic2 = (beta2hat - p*(p+1) ) /sqrt(8*p*(p+2)/n) ;
pvalskew = 1 - probchi(statistic1.dfchi) ;
pvalkurt = 2*( 1 - probnorm(abs(statistic2)) );
print s ;
print s_inv ;
print 'TESTS:';
print 'Based on skewness: ' beta1hat statistic1 pvalskew ;
print 'Based on kurtosis: ' beta2hat statistic2 pvalkurt;
/* mjb is multivariate JB test statistic to test for normality and pvaljb is corresponding p values. */
mjb=n*(beta1hat/6+((beta2hat-p*(p+2))*(beta2hat-p*(p+2)))/(8*p*(p+2)));
pvaljb= 1 - probchi(mjb.dfchi) ;
print beta1hat beta2hat mjb pvaljb;

```

Views and opinions expressed in this article are the views and opinions of the author(s), International Journal of Mathematical Research shall not be responsible or answerable for any loss, damage or liability etc. caused in relation to/arising out of the use of the content.