

Abstract

Discriminant analysis has been used to compare patients with the following diagnoses: psychotic depression (PD), bipolar disorder (BD) and unipolar depression (UD). The data used for statistical analyses consist of EEG recordings from seizures induced by ECT-treatment from 28 depressed patients of whom five had PD, seven had BD and 16 had UD. EEG recordings from two channels were considered, Fp1 and Fp2. The obtained EEG-series were split into two parts representing different phases. The fractal dimension was calculated for these series using Higuchi's fractal dimension algorithm. This yielded four variables for each patient representing fractal dimension of two phases in Fp1 and Fp2. A linear discriminant function was constructed. No separation between the UD and the BD was observed, only between PD and the other two groups. Classification using resubstitution, with equal priors, yielded a total misclassification rate of 36%. Correct classification rate for PD patients was 5/5, with two overlaps from UD and one from BD. When leave-one-out validation was used, error rate rose to 62% and correct classification rate for BD became 4/5.

Contents

1. Introduction	5
2. Survey of the Field	7
3. Material and Methods.....	9
3.1 Patients.....	9
3.2. Data Transformation: Higuchi's Fractal Dimension	9
3.5 Multivariate Normality	11
3.6 Tests for Multivariate Normality	12
3.6.1 Multivariate Skewness	12
3.6.2 Multivariate Kurtosis	12
3.6.3 Graphical Test.....	12
3.3 Discriminant Analysis	13
3.3.1 Discriminant Function with several Groups	14
3.4 Classification of Observations.....	16
3.4.1 Equal Population Covariance Matrices.....	16
3.4.2 Unequal Population Covariance Matrices.....	17
3.7 Statistical Analysis of EEG data.....	18
4. Results	20
5. Discussion	25
References	27
Appendix 1. Box Plots of Individual Variables by Diagnose	29
Appendix 2. Univariate and Multivariate Significance Tests	30
Appendix 3. Discriminant Analysis	31
Appendix 4. Test of Multivariate Normality with SAS	32

1. Introduction

Electroconvulsive therapy (ECT) is a psychiatric treatment in which the brain is subjected to electrical current which causes seizures in anesthetized patients. It is used as a treatment for severely depressed patients and the subgroups of these, including unipolar depression, bipolar depression and psychotic depression.

Electroencephalographic examinations during those seizures produce very characteristic electroencephalograms (EEGs). They do show large biological variation from different subjects and they also present non-stationary patterns. Algorithms, partly based on EEG during seizure that are implemented in ECT equipment have been used to monitor responses from different types of treatment and to predict diagnoses.

ECT is used on pharmacological treatment resistant depression such as those mentioned above. Patients with psychotic depression can experience delusions that are beliefs that are unsubstantiated. Such delusions may be paranoia or delusion of guilt. Some patients believe that people around them are paying special attention to them, that people want to persecute them and some even have hallucinations. They are also characterized by low mood and other depressive symptoms. Unipolar patients experience episodes of depression that can vary in frequency and length. Bipolar patients may shift between low mood and elevated mood which can manifest in mania or hypomania.

There is no commonly accepted statistical approach for analyzing EEG-data. Classical statistical approaches where measurement errors are regarded as generators of uncertainty are not valid when considering data from EEG-time series (Wahlund et al., 2009). Up to now, no model has been proposed where residuals around a fitted model are to be considered random. It could be due to uncovered dependency structures, but it is not clear how these dependencies should be estimated in non-stationary series. 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.

There are a few works on sub-grouping depressive disorders by analyzing EEG from ECT seizures. In the work by Wahlund et al. (2009) smoothed density functions of time series obtained by recording EEG during ECT-induced seizures have been related to different subgroups of depressed patients using discriminant analysis.

Recently, in a similar research setting, an approach was proposed where one looked at fractal dimension (complexity) of the curve as a representation of EEG data from ECT seizures to identify differences in responses between parts of the brain, differences in phases of the seizures and to see if subgroups of depressed patients were different with respect to fractal dimension (Wahlund et al., 2010).

Estimation of fractal dimension of similar biological signals has been used in other works. Gómez et al. (2008) used data from magnetoencephalogram (MEG) to identify complexity of those signals in Alzheimer's patients and healthy elderly. They have found significantly lower complexity of MEG rhythms in Alzheimer's patients compared to the control subjects.

Georgiev et al. (2008) have investigated how fractal dimension of EEGs changes as a response to different mental activities in healthy volunteers.

In this work, discriminant analysis is used to analyze EEG data that was used in the work by Wahlund et al. (2010). The goal is to elucidate differences between subgroups of depressed patients, using fractal dimension of EEG-recordings from two parts of the brain. The results provide important information concerning the potential of fractal dimension to “predict” the correct diagnose of a patient. Since there is a turnover in subgroups of depressed patients where unipolar depressed become bipolar (Angst, 1996), identification of changes in diagnosis at an early stage might yield an appropriate complimentary treatment.

The next chapter is a survey of the field. In chapter three, the data material and methods used in the statistical analysis will be presented, including discriminant analysis, estimation of fractal dimension and tests for multivariate normality. In chapter four, results will be revised. In chapter five we will discuss the obtained results.

2. Survey of the Field

The acquisition of data has been the primary focus when studying mental disorders. Advanced equipment has provided researchers with a lot of data that has been very hard to understand and make use of. A number of mathematical models have been proposed to transform the data and make it more usable for statistical procedures. Different frequencies in the brain have been found to be associated with different states and activities. For instance, EEG from a person during excitement will show rapid and irregular waves while slow waves are present during sleep (Encyclopædia Britannica, 2010).

The EEG-series that we are interested in have specific properties that make them unsuitable for the common statistical approaches. They are non-stationary and they are predictable over short time periods but unpredictable for longer time periods (Wahlund et al., 2010). EEG-series cannot be considered random nor non-random and no suitable model has been found that fits EEG-series (Culic et al., 2009).

When studying patients with different types of depression Fourier analysis was used in combination with principal component analysis in the work by Wahlund et al. (2009). Significant differences were found in the frequency domain between psychotic depressed and bipolar depressed patients. Psychotic depressed patients had a higher occurrence of the lower frequencies while the bipolar depressed had almost all activity in the higher frequency range and the unipolar depressed showed a heterogeneous range of frequencies. Fourier EEG was recorded during seizures induced by ECT. Furthermore, principal components were used for discriminant analysis to separate between the three different subgroups of depressed patients: bipolar, unipolar and psychotic. Almost perfect separation between the different subgroups was observed. It was also established that the differences between the groups were reduced after successive treatments. No clear answer for the reasons for it was given. To have been able to separate between the subgroups in a study with small and unequal sample sizes is still to be considered a success. Although, some limitations were noted. For example, using only two electrodes when recording EEGs may not be sufficient.

To better understand EEG, the combination of computational analysis, mathematical modeling and multivariate statistical methods has been used. In Wahlund et al. (2010) a version of fractal dimension was used to analyze EEG-data. Fractal dimension calculated for a time series is a number between 1 and 2 which can be regarded as a measure of complexity of the curve of the time series where a greater value means higher complexity. Nowadays this is a valuable tool in assessing changes in EEG-signals from either the same person in different trials or from one trial, looking at different states, for instance awake and different sleep stages (Klonowski, 2007). More details on the fractal dimension are given in chapter 3.

In Wahlund et al. (2010), data from EEG recordings during ECT-induced seizures were partitioned into four different phases. They were not partitioned analytically, but rather picked after visual examination of the EEG. Patients then had 8 time series associated with them, four from Fp1 and four from Fp2; and the fractal dimension was estimated for these series. No statistical difference has been found between the first three phases, so they were averaged

together. Sufficient evidence has been found that there is difference in fractal dimension from two electrode placements Fp1 and Fp2. Statistically significant differences in fractal dimension between phase 1-3 and phase 4 of the seizure have also been discovered. The difference in fractal dimension between subgroups of depressed patients has been found to be non significant (at $\alpha = 0.05$).

3. Material and Methods

3.1 Patients

The sample consists of 29 depressed patients where seven have bipolar depression, six have psychotic depression and the rest 16 have unipolar depression. None of the patients had received ECT during the last three months. The patients were all right handed. The patients received a bi-directional pulse. EEG-signals were recorded from two channels, Fp1 and Fp2, see Figure 3.1. The continuous signals were digitized at a sampling rate of 128 hz.

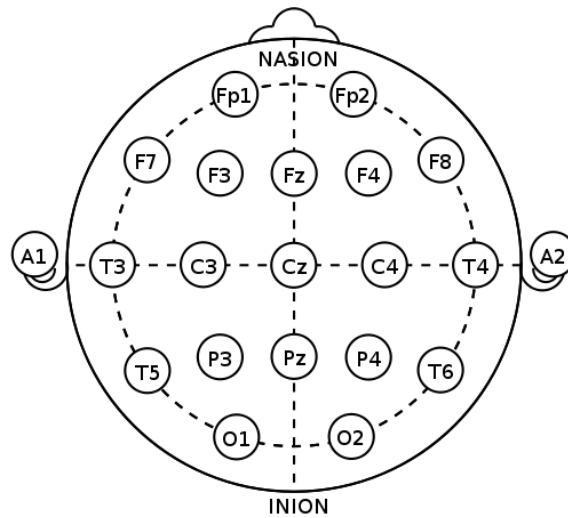


Figure 3.1. Map of locations of the EEG electrodes/channels on the scalp according to the 10-20 system. In our material, EEGs were recorded from Fp1 and Fp2, which are the standard locations when treating patients with ECT. Electrode and channel represent the same concept and will be used interchangeable in this work.

3.2. Data Transformation: Higuchi's Fractal Dimension

Fractal Dimension (FD) can be used to assess the complexity and self-similarity of a signal. Highly irregular or complex curves representing the signal will show higher FD. Higuchi proposed an efficient algorithm for measuring FD in which FD is calculated directly from the time series (Higuchi, 1988). It has gained popularity in applications involving biomedical data such as EEG measurements due its efficiency. It has been shown to be more accurate than other methods.

In this work, the data from the EEG time series has been processed with Higuchi's FD, D_f , which measures local complexity of the curve representing the EEG signal. The, mean D_f has been calculated for each of the identified phases of the seizures from two electrodes. The

three first phases were averaged together and the fourth was kept. This gives four variables, for each patient.

Let $L(k)$ represent the amplitude of the signal as a function of time. The curve is said to have a fractal dimension of D_f if the length of this curve scales as:

$$L(k) \approx k^{-D_f} . \quad (3.1)$$

Higuchi's fractal dimension of the curve will measure the filling of the plane and can be regarded as a measure of complexity. A line is of dimension one and a plane is of dimension two. Hence, D_f is always between 1 and 2. Fractal dimension may be used to summarize a large number of data points or as few as $t = 100$ observations. Higher values of $D_f(t)$ correspond to higher frequencies in the Fourier spectrum of the signal. From the time sequence $x(1), x(2), \dots, x(N)$, k new time series have been constructed:

$$x_m^k = \left\{ x(m), x(m+k), \dots, x\left(m + \left\lfloor \frac{N-m}{k} \right\rfloor k\right) \right\},$$

where $m = 1, \dots, k$. The notation $\lfloor \bullet \rfloor$ stands for the integer function, and k is the time between consecutive observations. For each of the above time series, the length of the curve is estimated by calculating

$$\hat{L}_m(k) = \frac{(1/k) \sum_{i=1}^{\lfloor (N-m)/k \rfloor} (x(m+ik) - x(m+(i+1)k))(N-1)}{[(N-m)/k]k}$$

and then the average is calculated

$$\hat{L}(k) = \frac{1}{k} \sum_{m=1}^k \hat{L}_m(k)$$

From (3.1) we have a log-log relation which can be used to estimate the unknown D_f :

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

$k = k_{min}, \dots, k_{max}$. The values of k_{min} and k_{max} were chosen by the medical researchers. If the error term can be assumed to be the same, with a mean of 0 and with a symmetric distribution for all k , then the least squares approach can be used to estimate D_f .

3.5 Multivariate Normality

The multivariate normal distribution is a generalization of the univariate normal distribution.

Definition 1. A random vector $\mathbf{X} = (X_1, \dots, X_p)'$ is said to have a p-variate normal distribution with mean vector $\boldsymbol{\mu} = (\mu_1, \dots, \mu_p)'$ and the covariance matrix $\boldsymbol{\Sigma} = \begin{pmatrix} \sigma_{11} & \cdots & \sigma_{1p} \\ \vdots & \ddots & \vdots \\ \sigma_{p1} & \cdots & \sigma_{pp} \end{pmatrix}$ if its probability density function has the following form:

$$f_{\mathbf{X}}(\mathbf{x}) = \frac{1}{(2\pi)^{p/2} |\boldsymbol{\Sigma}|^{1/2}} e^{-\frac{1}{2}(\mathbf{x}-\boldsymbol{\mu})'\boldsymbol{\Sigma}^{-1}(\mathbf{x}-\boldsymbol{\mu})},$$

where $-\infty < x_i < \infty, i = 1, \dots, p$, and $|\boldsymbol{\Sigma}| > 0$.

If $p = 1$, the above density reduces to the univariate normal probability density function. The multivariate normal distribution possesses some characteristics that are necessary but not sufficient to indicate multivariate normality.

Theorem 1 (Johnson and Wichern, 2007): If a random vector $\mathbf{X}: p \times 1$ is distributed as $N_p(\boldsymbol{\mu}, \boldsymbol{\Sigma})$, then any linear combination $\mathbf{a}'\mathbf{X} = a_1X_1 + a_2X_2 + \dots + a_pX_p$ is distributed as $N(\mathbf{a}'\boldsymbol{\mu}, \mathbf{a}'\boldsymbol{\Sigma}\mathbf{a})$.

Also, if $\mathbf{a}'\mathbf{X}$ is distributed as $N(\mathbf{a}'\boldsymbol{\mu}, \mathbf{a}'\boldsymbol{\Sigma}\mathbf{a})$ for every $\mathbf{a} \neq \mathbf{0}$, then \mathbf{X} must be $N(\boldsymbol{\mu}, \boldsymbol{\Sigma})$.

An immediate corollary to Theorem 1 implies that each component X_i in \mathbf{X} must have a univariate normal distribution. Therefore, examination of histograms and tests for normality for each individual random variable, $X_i, i = 1, \dots, p$, give an indication whether the data departs from multivariate normality, but they cannot be used to confirm multivariate normality.

Theorem 2. All subvectors of a random vector $\mathbf{X} \sim N_p(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ are normally distributed. If we, respectively, partition $\mathbf{X} = (\mathbf{X}'_1: \mathbf{X}'_2)'$, its mean vector $\boldsymbol{\mu} = (\boldsymbol{\mu}'_1: \boldsymbol{\mu}'_2)'$, and its covariance matrix $\boldsymbol{\Sigma} = \begin{pmatrix} \boldsymbol{\Sigma}_{11} & \boldsymbol{\Sigma}_{12} \\ \boldsymbol{\Sigma}_{21} & \boldsymbol{\Sigma}_{22} \end{pmatrix}$ then $\mathbf{X}_1: q \times 1 \sim N_q(\boldsymbol{\mu}_1, \boldsymbol{\Sigma}_{11})$ and $\mathbf{X}_2 \sim N_{p-q}(\boldsymbol{\mu}_2, \boldsymbol{\Sigma}_{22})$.

For more details concerning the properties of the multivariate normal distribution, see Johnson and Wichern (2007).

Since every q-subvector of a random vector $\mathbf{X} \sim N_p(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ follows a q-variate normal distribution one can for example plot each pair of random variables in \mathbf{X} to get a visual indication of a bivariate normality. In case of a bivariate normal distribution we will get a circular or elliptical scatter plot with greater concentration around the mean with very few or zero outliers.

In the next section we present some possibilities to test multivariate normality.

3.6 Tests for Multivariate Normality

Skewness and kurtosis are two characteristics of the shape of a distribution. Skewness is a measure of asymmetry in a distribution. Zero value implies a perfectly symmetric distribution, a negative skewness means that there is a heavy tail to the left of the distribution and a positive skewness indicates that there is a heavy tail to the right. The normal distribution is symmetric.

Kurtosis is a measure to what degree the variance is a result of infrequent extreme deviations from the mean. The normal distribution has a kurtosis of 3. Higher values of kurtosis mean that more of the variation is due to distant extremes, and lower values would indicate the opposite.

For a random sample $\mathbf{X} = (\mathbf{X}_1, \dots, \mathbf{X}_n)$ from a p-variate normal distribution $N_p(\boldsymbol{\mu}, \boldsymbol{\Sigma})$, Mardia (1970, 1974) defined the p-variate skewness and kurtosis statistics.

3.6.1 Multivariate Skewness

The p-variate skewness statistic is given by:

$$b_{1,p} = \frac{1}{n^2} \sum_{i=1}^n \sum_{j=1}^n ((\mathbf{X}_i - \bar{\mathbf{X}})' \mathbf{S}^{-1} (\mathbf{X}_j - \bar{\mathbf{X}}))^3,$$

where n is the total sample size, $\bar{\mathbf{X}}$ and \mathbf{S} are the sample mean and the covariance matrix, respectively. Mardia (1970) showed that $(n/6)b_{1,p}$ is approximately distributed as a χ^2 random variable with $p(p+1)(p+2)/2$ degrees of freedom.

3.6.2 Multivariate Kurtosis

The p-variate kurtosis statistic is given by

$$b_{2,p} = \frac{1}{n} \sum_{i=1}^n ((\mathbf{X}_i - \bar{\mathbf{X}})' \mathbf{S}^{-1} (\mathbf{X}_i - \bar{\mathbf{X}}))^2,$$

which follows a normal distribution with mean $p(p+2)$ and a variance of $8p(p+2)/n$ under the null hypothesis of multivariate normality.

These statistics can be used to test multivariate normality.

3.6.3 Graphical Test

One can also use a graphical test to check multivariate normality. It is based on the squared Mahalanobis Distances (MD^2), see Sharma (1996). Johnson and Wichern (2007) showed that the squared Mahalanobis distance of each observation from the sample centroid behaves like a chi-square random variable if the sample is sufficiently large, e.g. 25 or larger, and it is drawn from a multivariate normal population.

In order to perform a graphical test the following steps should be done:

1. Order the $MD_i^2, i = 1, 2, \dots, n$, from smallest to largest: $MD_1^2 < MD_2^2 < \dots < MD_n^2$
2. Compute the $(j - 0.5)/n$ percentile, for each MD_j^2 , where j is the observation number.
3. Obtain the χ^2 values for the percentiles in step 2 from the χ^2 distribution with p degrees of freedom, where p is the number of variables.
4. The χ^2 and MD^2 are then plotted against each other.

The plot should exhibit a linear pattern under H_0 of multivariate normality and any deviation from it would indicate non-normality.

3.3 Discriminant Analysis

Discriminant analysis (DA) is a method that is used to find out in which way a set of variables can be combined to best describe differences between certain groups in a population. One of the goals of DA is to describe the group separation, in which linear functions of the original variables are used to elucidate the differences between the predefined groups. This knowledge can then be used to serve another purpose: to assign new observations to the predefined groups.

Consider a random vector $\mathbf{X} = (X_1, \dots, X_p)' \sim N_p(\boldsymbol{\mu}, \boldsymbol{\Sigma})$, where $\boldsymbol{\Sigma}$ is the known covariance matrix of \mathbf{X} and $\boldsymbol{\mu}$ is the mean vector of \mathbf{X} , measured on the elements belonging to $k \geq 2$ groups. In DA we are interested in finding a linear combination:

$$y = \mathbf{a}'\mathbf{X} = a_1X_1 + \dots + a_pX_p \quad (3.2)$$

such that it can provide the best separation between the k different groups. That is, the discriminant scores computed with (3.2) from subjects belonging to one group should be maximally dissimilar from the scores of subjects in other groups. The linear combination in (3.2) is called Fisher's linear discriminant function, after Fisher (1936).

The maximum number of discriminant functions that can be constructed is $\min(p, k - 1)$. In the case of two groups only one discriminant function can be constructed. With three groups, it may be the case that one linear combination of X_1, \dots, X_p is good at separating between the first group on the one hand and the second and third group on the other hand. Furthermore, if the second and the third groups are indistinguishable by the scores of the first discriminant function, a second discriminant function may then differentiate between the second and third group.

How these functions are obtained and how their significance is assessed will be presented in the next section.

3.3.1 Discriminant Function with several Groups

When evaluating group differences between several groups in the univariate case with respect to some variable y the following F statistic is used:

$$F = \frac{SS_b/(k-1)}{SS_w/(n-k)} \quad (3.3)$$

where $SS_b = \sum_{i=1}^k (\bar{Y}_i - \bar{Y})^2$ measures heterogeneity of groups, \bar{Y}_i is the mean of the i th group, $i = 1, \dots, k$, and \bar{Y} is the grand mean; $SS_w = \sum_{i=1}^k \sum_{j=1}^n (Y_{ij} - \bar{Y}_i)^2$ measures the variability of observations within their respective group, n is the total number of observations and k is the number of groups. Under the null hypothesis of equality of group means, i.e. $H_0: \mu_1 = \dots = \mu_k$, the F -statistic in (3.3) follows the F distribution with $k-1$ and $n-k$ degrees of freedom.

Let y be defined as in (3.2). The ratio between SS_b and SS_w can now be written as:

$$\lambda = \frac{SS_b}{SS_w} = \frac{\mathbf{a}'\mathbf{B}\mathbf{a}}{\mathbf{a}'\mathbf{W}\mathbf{a}} \quad (3.4)$$

where \mathbf{a} is a $p \times 1$ vector of weights, $\mathbf{B} = \sum_{i=1}^k n_i (\bar{\mathbf{X}}_i - \bar{\mathbf{X}})(\bar{\mathbf{X}}_i - \bar{\mathbf{X}})'$ is a between group **SSCP** (Sum of Squares and Cross Product) matrix of the original variables and \mathbf{W} is the sum of the within group **SSCP** matrices, i.e. $\mathbf{W} = \sum_{i=1}^k \sum_{j=1}^n (\mathbf{X}_{ij} - \bar{\mathbf{X}}_i)(\mathbf{X}_{ij} - \bar{\mathbf{X}}_i)'$. The ratio in (3.4) is called the discriminant criterion and it measures how well the groups are separated with respect to the linear combination y given in (3.2). A higher value of λ implies better separation between groups. Moreover, higher values of the numerator in (3.4) indicate a greater distance between groups, and lower values of the denominator mean homogeneity within groups.

Thus, the goal is to find the vector \mathbf{a} containing the weights of the original variables X_1, \dots, X_p that maximizes (3.4).

Proposition 1 (Raykov and Marcoulides, 2008, pp. 354): The discriminant criterion λ in (3.4) is maximized when the vector of weights \mathbf{a} in (3.4) is taken as an eigenvector corresponding to the largest eigenvalue of the matrix $\mathbf{A} = \mathbf{W}^{-1}\mathbf{B}$.

The maximal number of linearly independent rows or columns of matrix \mathbf{A} is $r = \min(p, k-1)$, where r stands for the rank of the matrix \mathbf{A} , p is the dimension of vector \mathbf{X} and k is the number of groups, i.e. there are r nonzero eigenvalues $\lambda_1 \geq \dots \geq \lambda_r$ with the corresponding eigenvectors $\mathbf{a}_1, \dots, \mathbf{a}_r$. Using these eigenvectors one can construct r discriminant functions that are uncorrelated:

$$\mathbf{a}'_1 \mathbf{X}, \dots, \mathbf{a}'_r \mathbf{X}. \quad (3.5)$$

The eigenvectors in (3.5) are standardized so that $\mathbf{a}'_i \mathbf{a}_i = 1, i = 1, \dots, r$.

The obtained discriminant functions $\mathbf{a}_i' \mathbf{X}, i = 1, \dots, r$, that correspond to different eigenvalues λ_i of matrix \mathbf{A} are not of equal importance. In order to decide how many discriminant functions to include in the analysis, we can evaluate the so called statistical and practical significance of these functions. This will be explained in the following sections.

Statistical Significance of the Discriminant Functions

Let $\mathbf{X} = (X_1, \dots, X_p)' \sim N_p(\boldsymbol{\mu}, \boldsymbol{\Sigma})$, where $\boldsymbol{\Sigma}$ is known. The null hypothesis of equality of population mean vectors of the k groups, that is

$$H_0: \boldsymbol{\mu}_1 = \boldsymbol{\mu}_2 = \dots = \boldsymbol{\mu}_k, \quad (3.6)$$

can be tested using Wilk's Lambda:

$$\Lambda = \frac{|\mathbf{W}|}{|\mathbf{T}|},$$

where \mathbf{T} is the total-sample **SSCP** matrix, \mathbf{W} is the within group **SSCP** matrix, $|\bullet|$ stands for the determinant of a matrix, and $\mathbf{T} = \mathbf{W} + \mathbf{B}$. The test statistic Λ can be rewritten in inverse form (Raykov and Marcoulides, 2008):

$$\begin{aligned} 1/\Lambda &= |\mathbf{T}|/|\mathbf{W}| = |\mathbf{W}^{-1}\mathbf{T}| = |\mathbf{W}^{-1}(\mathbf{W} + \mathbf{B})| = |\mathbf{I}_p + \mathbf{W}^{-1}\mathbf{B}| \\ &= (1 + \lambda_1) \dots (1 + \lambda_r), \end{aligned} \quad (3.7)$$

where $\lambda_1, \dots, \lambda_r$ are the eigenvalues of $\mathbf{W}^{-1}\mathbf{B}$. According to (3.7), testing the H_0 specified in (3.6) is equivalent to testing whether all eigenvalues of the matrix $\mathbf{W}^{-1}\mathbf{B}$ are zero. Rejection means that at least one eigenvalue is significantly different from zero which implies that we have at least one significant discriminant function. Thus, finding the linear combination that maximizes the discriminant criterion given in (3.4), which will be $\mathbf{a}_1' \mathbf{X}$ for λ_1 , and testing the univariate null hypothesis using the F statistic in (3.3) on this new variable is the same as testing the multivariate null hypothesis given in (3.6). If we want to assess the significance of the remaining discriminant functions we simply compute the value of the F-statistic for the linear combination corresponding to the next largest discriminant criterion and continue until we have tested all r discriminant functions or until we encounter an insignificant function, in which case we choose not to include that or any remaining function in our analysis.

Practical Significance of the Discriminant Functions

Another measure of significance of a discriminant function is the practical significance. If the sample size is very large, even small differences between the groups may yield significant discriminant functions. While it is true that they possess statistical significance, the differences may be so small that practitioners might find it practically negligible.

Practical significance of the j th discriminant function can be assessed in the following way (Sharma, 1996):

$$\delta = \frac{\lambda_j}{\sum_{i=1}^r \lambda_i} \times 100. \quad (3.8)$$

The quantity in (3.8) represents the amount of difference between groups, in percent, that is accounted for by the discriminant function corresponding to eigenvalue $\lambda_j, j = 1, \dots, r$.

The practical significance can also be assessed by the squared canonical correlation coefficient (CR^2) which can be defined as follows:

$$CR^2 = \frac{SS_b}{SS_t}.$$

The statistic CR^2 measures the proportion of the total sum of squares of the discriminant scores that is due to group differences.

Another goal of DA is the classification of new observations which is explained in the following section.

3.4 Classification of Observations

Depending on the nature of the groups in the sample, e.g. whether they are from a normal or nonnormal population or whether the covariance matrices of the groups are equal or unequal we end up with different optimal classification rules. An optimal classification rule minimizes the probability of misclassification. Two factors can be incorporated in the classification rules. One of them is the prior probability of a group that is the probability that a randomly selected observation belongs to that group. The other one is the cost of misclassification which is a weight that specifies relative cost that is associated with misclassification of a member of that group. In this work we are only interested in finding out whether it is possible to correctly classify a patient to a diagnose group based on the fractal dimension of EEG-data. We will thus assume equal misclassification costs and find a classification rule that minimizes overall misclassification rate. Prior probabilities will be assumed to be equal for all groups.

3.4.1 Equal Population Covariance Matrices

Let $\mathbf{X} = (X_1, \dots, X_p)' \sim N_p(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ and $\boldsymbol{\mu} = (\mu_1, \dots, \mu_p)'$. To obtain an optimal classification rule we need to know the density function of \mathbf{X} for each group. Let $f_{\mathbf{X}}(\mathbf{x}|G_i)$ be the density function of \mathbf{X} in the i th group.

The optimal classification rule satisfies the following inequality:

$$p_i f_{\mathbf{X}}(\mathbf{x}|G_i) \geq p_j f_{\mathbf{X}}(\mathbf{x}|G_j), \text{ for all } j = 1, 2, \dots, k, \quad (3.9)$$

where p_i is the prior probability that a randomly chosen observation belongs to the i th group.

When the population covariance matrices are equal for the k groups, that is:

$$\boldsymbol{\Sigma}_1 = \boldsymbol{\Sigma}_2 = \dots = \boldsymbol{\Sigma}_k = \boldsymbol{\Sigma},$$

and only $\boldsymbol{\mu}_i, i = 1, \dots, k$, may differ between the groups, we obtain:

$$p_i f_{\mathbf{X}}(\mathbf{x}, G_i) = p_i \frac{1}{(2\pi)^{\frac{p}{2}} |\boldsymbol{\Sigma}|^{\frac{1}{2}}} e^{-\frac{1}{2}(\mathbf{x} - \boldsymbol{\mu}_i)' \boldsymbol{\Sigma}^{-1}(\mathbf{x} - \boldsymbol{\mu}_i)}. \quad (3.10)$$

Taking the logarithm of (3.10) we get:

$$\ln(p_i f_{\mathbf{X}}(\mathbf{x}, G_i)) = \ln p_i + \frac{p}{2} \ln 2\pi - \frac{1}{2} \ln |\boldsymbol{\Sigma}| - \frac{1}{2}(\mathbf{x} - \boldsymbol{\mu}_i)' \boldsymbol{\Sigma}^{-1}(\mathbf{x} - \boldsymbol{\mu}_i).$$

If we delete the terms that are equal for all groups and plug in the sample estimates \mathbf{S} and $\bar{\mathbf{x}}_i$ of the respective population parameters $\boldsymbol{\Sigma}$ and $\boldsymbol{\mu}_i$ we get:

$$L_i(\mathbf{x}) = \ln p_i + \bar{\mathbf{x}}_i' \mathbf{S}^{-1} \mathbf{x} - \frac{1}{2} \bar{\mathbf{x}}_i' \mathbf{S}^{-1} \bar{\mathbf{x}}_i, \quad (3.11)$$

An observation \mathbf{x} will be classified to the group for which the value of $L_i(\mathbf{x})$ is the largest, $i = 1, \dots, k$. The function (3.11) is called a linear classification function. It is important that the sampled groups have the same population covariance matrix, otherwise the observations will tend to be classified into groups with larger variances more often than they should be. As will be shown in the next section, unequal covariance matrices will yield a quadratic classification function.

3.4.2 Unequal Population Covariance Matrices

If we instead draw samples from a multivariate normal population $\mathbf{X} \sim N_p(\boldsymbol{\mu}, \boldsymbol{\Sigma})$, where the k groups have unequal covariance matrices $\boldsymbol{\Sigma}_i$, we get

$$\ln(p_i f_{\mathbf{X}}(\mathbf{x}, G_i)) = \ln p_i + \frac{p}{2} \ln 2\pi - \frac{1}{2} \ln |\boldsymbol{\Sigma}_i| - \frac{1}{2}(\mathbf{x} - \boldsymbol{\mu}_i)' \boldsymbol{\Sigma}_i^{-1}(\mathbf{x} - \boldsymbol{\mu}_i).$$

Replacing population parameters with the corresponding sample estimates and after some calculations, we obtain the following (Rencher, 1998):

$$Q_i(\mathbf{x}) = \ln p_i - \frac{1}{2} \ln |\mathbf{S}_i| - \frac{1}{2} \bar{\mathbf{x}}_i' \mathbf{S}_i^{-1} \bar{\mathbf{x}}_i + \bar{\mathbf{x}}_i' \mathbf{S}_i^{-1} \mathbf{x} - \frac{1}{2} \mathbf{x}' \mathbf{S}_i^{-1} \mathbf{x}. \quad (3.12)$$

Function (3.12) is called a quadratic classification function. An observation \mathbf{x} is classified into the group for which $Q_i(\mathbf{x})$ gives the highest value, $i = 1, \dots, k$.

Although this quadratic classification function is the optimal with respect to misclassification rate when dealing with unequal group covariance matrices, it has been shown to be unstable when dealing with small samples and small groups because the sample covariance matrices, $\mathbf{S}_i, i = 1, \dots, k$, can be highly heterogeneous between samples. However, if the groups' covariance matrices $\boldsymbol{\Sigma}_i$ are known, then the quadratic classification functions will perform

best (Lachenbruch, 1975). It should be noted that quadratic classification functions are more sensitive to nonnormality than the linear classification functions (Michaelis, 1973; Huberty and Curry, 1978).

In order to estimate how accurately a predictive model will perform, resubstitution and crossvalidation techniques can be used.

Resubstitution is the procedure where the same observations are used to find the classification functions and to classify observations. The classification results and misclassification rate (proportion of misclassified observations), using resubstitution, will be positively biased compared to the procedure where external data is used for classification (i.e. the misclassification rate will be lower).

One method that has been proposed (Lachenbruch,1967) to correct this bias is called the leave-one-out crossvalidation method. In order to classify and estimate misclassification rates each observation is excluded from the analysis prior to classification. Thus, each observation that is classified will act as validation data. Leave-one-out method is usually very expensive from a computational point of view.

In the next section we briefly describe what types of analyses have been performed in this work using EEG data.

3.7 Statistical Analysis of EEG data

The following variables were used for the analysis.

X1_Fp1: The estimated fractal dimension of the EEG-signal during phases 1 to 3, recorded from Fp1 channel.

X1_Fp2: The estimated fractal dimension of the EEG-signal during phases 1 to 3, recorded from Fp2 channel.

X2_Fp1: The estimated fractal dimension of the EEG-signal during phase 4, recorded from Fp1 channel.

X2_Fp2: The estimated fractal dimension of the EEG-signal during phase 4, recorded from Fp2 channel.

The SAS System for Windows (Version 9.2, SAS Institute Inc., Cary, NC, USA) has been used to perform all the statistical analyses.

We described the four variables in the study group using graphical analysis with PROC BOXPLOT. Box-and-whisker plots were used to get an indication of difference between the groups of patients with respect to the variables of interest.

Descriptive statistics (means and standard deviations) were calculated with the help of PROC MEANS for the four variables based on the whole sample and by groups. Pearson's coefficient of correlation has been calculated using PROC CORR.

Univariate and multivariate tests were performed with PROC ANOVA and PROC GLM to test whether there were significant differences between the groups of depressed patients with respect to the individual variables and all the variables considered jointly.

A SAS macro, %MULTINORM, was used to test whether the assumption of multivariate normality holds (SAS, 2010).

For discriminant analysis PROC CANDISC and PROC DISCRIM were used. Linear discriminant functions were obtained. The scores of the obtained discriminant functions were calculated, and plotted using PROC GPLOT.

Classification functions were constructed for the different groups of patients. The observations were classified into groups and misclassification rates were calculated using the resubstitution method and equal priors. Crossvalidation (leave-one-out method) was also used to get an indication of how the classification works for validation data.

Due to missing values, one PD observation was excluded from the multivariate analysis

4. Results

The mean values of D_f for phase 1-3 and phase 4 recorded from Fp1 and Fp2 are presented in Table 4.1. The results show that the mean values of D_f for phase 1-3 in Fp1 and Fp2 are lower for all subgroups compared to the values of D_f for phase 4 in Fp1 and Fp2, respectively. Lower spread of D_f was observed for all subgroups in both channels Fp1 and Fp2 for phase 4. Pronounced differences between the means of D_f for phase 4 in Fp1 and Fp2 were observed in the PD group.

Table 4.1. Comparison of the fractal dimension for subgroups of depressed patients: bipolar disorder (BD), psychotic depression (PD), unipolar depression (UD), from Fp1 and Fp2

	X1 (phase 1-3)		
	BD	PD	UD
Fp1	1.26±0.07	1.23±0.05	1.25±0.06
Fp2	1.26±0.06	1.19±0.05	1.25±0.06
	X2 (phase 4)		
	BD	PD	UD
Fp1	1.51±0.14	1.47±0.19	1.58±0.14
Fp2	1.55±0.18	1.71±0.14	1.63±0.18

The lowest observed value of D_f , 1.15, was found for phase 4 in Fp2 of a BD patient.

Strong correlations have been observed between D_f for phase 4 comparing Fp1 and Fp2, in all subgroups. Strong correlations were also observed for phase 1-3 from Fp1 and Fp2 in BD and UD groups. The correlation coefficients calculated by diagnosis (BD, PD and UD) are presented in Table 4.2-4.4, where (*) indicates statistically significant coefficients at $\alpha = 0.01$:

Table 4.2. Pearson correlation coefficients between the D_f obtained from two channels Fp1 and Fp2 for phase 1-3 and phase 4 in BD group

	X1_fp1	X2_fp1	X1_fp2	X2_fp2
X1_fp1	1			
X2_fp1	0.116	1		
X1_fp2	0.847*	0.416	1	
X2_fp2	-0.048	0.753*	0.425	1

Table 4.3. Pearson correlation coefficients between the D_f obtained from two channels Fp1 and Fp2 for phase 1-3 and phase 4 in PD group

	X1_fp1	X2_fp1	X1_fp2	X2_fp2
X1_fp1	1			
X2_fp1	-0.572	1		
X1_fp2	0.486	0.589	1	
X2_fp2	-0.177	0.906*	0.629	1

Table 4.4. Pearson correlation coefficients between the D_f obtained from two channels Fp1 and Fp2 for phase 1-3 and phase 4 in UD group

	X1_fp1	X2_fp1	X1_fp2	X2_fp2
X1_fp1	1			
X2_fp1	0.060	1		
X1_fp2	0.932*	-0.080	1	
X2_fp2	0.125	0.734*	0.142	1

Table 4.5. Pearson correlation coefficients between the D_f obtained from two channels Fp1 and Fp2 for phase 1-3 and phase 4 in total sample

	X1_fp1	X2_fp1	X1_fp2	X2_fp2
X1_fp1	1			
X2_fp1	0.067	1		
X1_fp2	0.875*	0.116	1	
X2_fp2	-0.050	0.727*	0.134	1

None of the univariate normality tests provided sufficient evidence that the assumption of normality has been violated. The multivariate tests also did not provide sufficient evidence of deviation from multivariate normality.

The graphical test didn't show the evidence of deviations from multivariate normality. The result of the graphical test is presented in Figure 4.1:

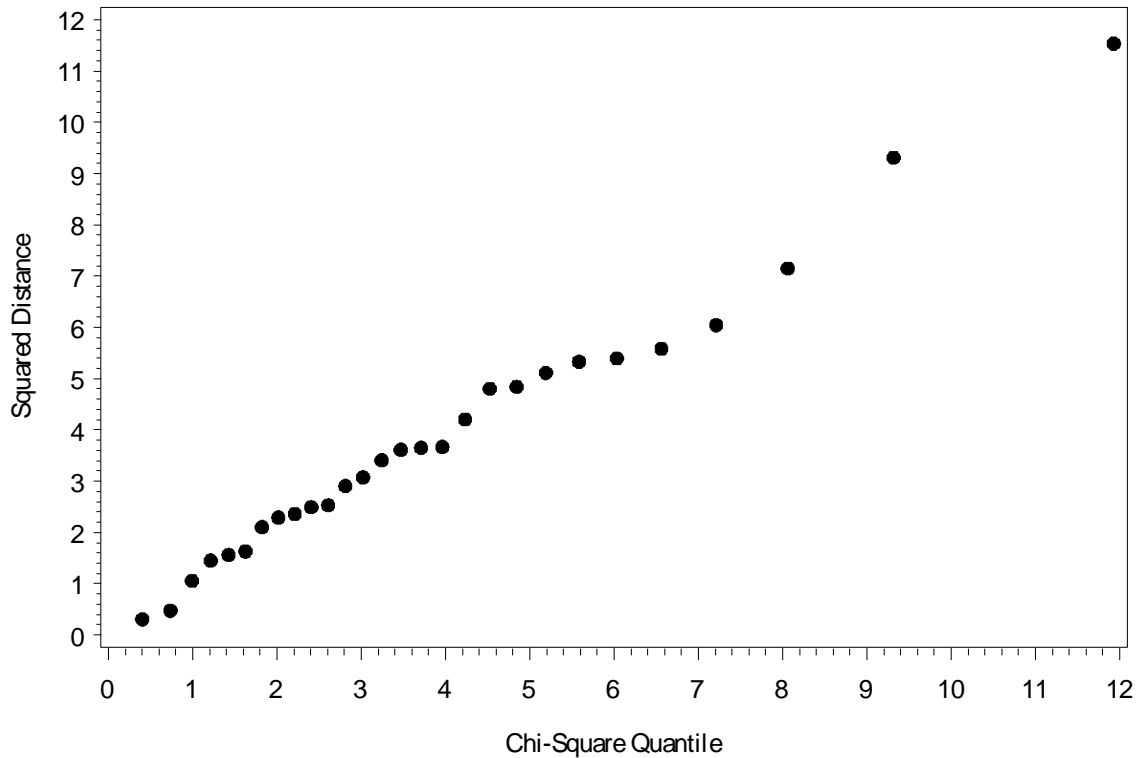


Figure 4.1. Chi-square plot for total sample

We did not observe significant differences between the groups with respect to any of the individual variables when a one-way ANOVA was performed. The results of the one-way ANOVAs are presented in Table A2.1 of Appendix 2.

We found however significant differences between the subgroups when performing a one-way multivariate ANOVA using all variables ($p=0.064$). The results are presented in Table A2.2 of Appendix 2.

Two discriminant functions have been obtained:

$$y_1 = 0.369 \times X1_{Fp1} + 0.067 \times X2_{Fp1} + 0.478 \times X1_{Fp2} - 0.297 \times X2_{Fp2} \quad (4.1)$$

and

$$y_2 = -0.382 \times X1_{Fp1} + 0.878 \times X2_{Fp1} - 0.212 \times X1_{Fp2} + 0.687 \times X2_{Fp2}. \quad (4.2)$$

Only the first discriminant function in (4.1) was statistically significant ($p=0.064$), i.e. the groups can be distinguished based on the predictor variables in (4.1), see Table A3.3 of Appendix 3. The second discriminant function in (4.2) was not statistically significant (p -value=0.662).

Practical significance of the discriminant function y_1 as measured by δ given in (3.8) was 0.92. For y_2 , $\delta = 0.08$. The CR^2 for y_1 was 0.430 and 0.065 for y_2 .

When plotting scores of the two discriminant functions for all patients (see Figure 4.2) we found that patients belonging to the PD subgroup were separated from the patients belonging to the other two, BD and UD, along the axis corresponding to the y_1 function. No separation was observed between the scores of BD and UD in any of the discriminant functions. The lower scores of the y_1 function were associated with observations belonging to PD and three of the highest scores belonged to UD. In between the extremes were observations belonging to any group. All standardized scores from PD along the y_1 function were negative. A possible outlier was observed, belonging to PD. Scores of y_2 were scattered evenly between all subgroups.

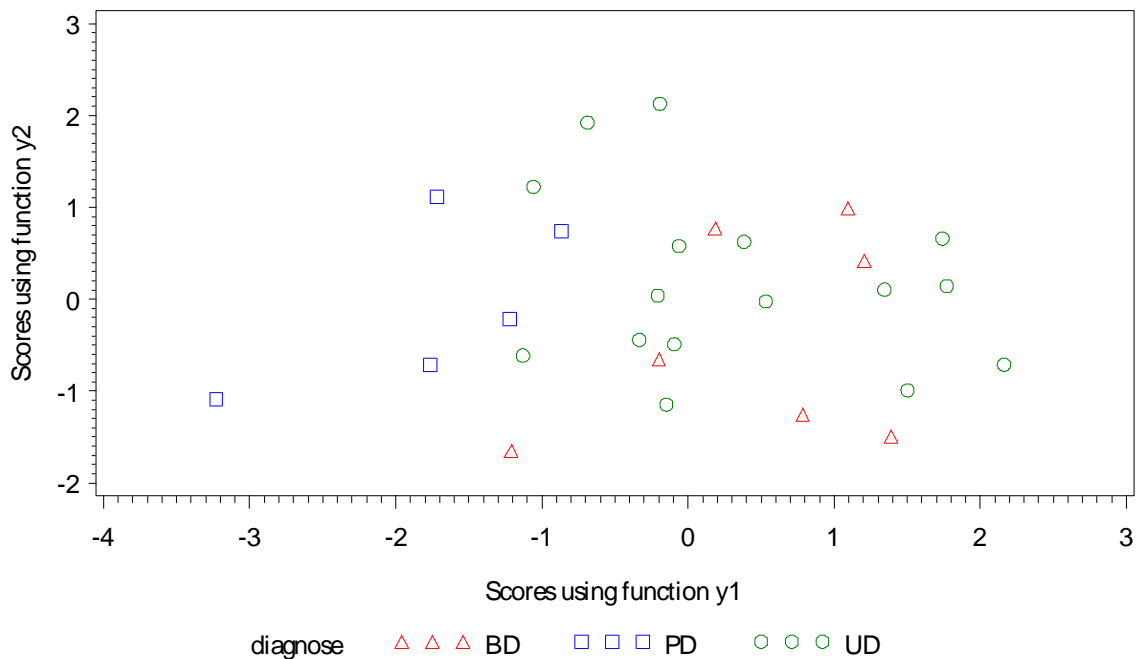


Figure 4.2. The scores of the two discriminant functions plotted by Diagnose

The hypothesis of homogeneity of within covariance matrices was not rejected (p -value=0.6). The linear classification functions were calculated:

$$C_{BD} = -291.21 + 224.97 \times X1_{Fp1} + 74.93 \times X2_{Fp1} + 160.41 \times X1_{Fp2} - 11.65 \times X2_{Fp2},$$

$$C_{PD} = -270.42 + 268.55 \times X1_{Fp1} + 58.49 \times X2_{Fp1} + 92.26 \times X1_{Fp2} + 9.22 \times X2_{Fp2},$$

$$C_{UD} = -290.59 + 219.44 \times X1_{Fp1} + 78.22 \times X2_{Fp1} + 160.60 \times X1_{Fp2} - 10.99 \times X2_{Fp2}.$$

The observations were classified into the groups for which the classification function yielded the highest value. The estimated total misclassification rate, with equal priors, using resubstitution was 36%. With resubstitution we observed high error rates for observations belonging to BD (57%) and UD (50%). We observed zero classification errors for observations belonging to PD. The classification results using resubstitution are summarized in Table 4.7:

Table 4.7. Results of classification based on linear classification functions using resubstitution and equal priors

Number of Observations Classified into Diagnose					Misclassification Rate
From Diagnose	BD	PD	UD	Total	
BD	3	1	3	7	57%
PD	0	5	0	5	0
UD	6	2	8	16	50%
Total	9	8	11	28	

Estimated total misclassification rate, with equal priors, using crossvalidation was 62%. We observed very high error rates for observation belonging to BD (86%) and UD (81%). We observed lower error rates for PD (20%). Classification results using crossvalidation (leave-one-out method) are summarized in Table 4.8:

Table 4.8. Results of classification based on linear classification functions using crossvalidation and equal priors

Number of Observations Classified into Diagnose					Misclassification Rate
From Diagnose	BD	PD	UD	Total	
BD	1	1	5	7	86%
PD	0	4	1	5	20%
UD	10	3	3	16	81%
Total	11	8	9	28	

5. Discussion

The graphical analysis of D_f showed that fractal dimensions of EEG-series markedly differ between the two phases which agrees with results obtained in Wahlund et al. (2010). No sufficient evidence has been found to indicate that mean D_f of each phase at location Fp1 and Fp2 differ significantly between patients' groups at the significance level of 5%. Although the multivariate analysis of variance provided sufficient evidence to claim that there was a significant difference (at $\alpha = 0.10\%$). This is an indication that multivariate techniques are desirable tools to discover differences in complexity of EEGs during ECT-treatment in different subcategories of depression.

Strong positive correlation between D_f of each phase recorded from the two different location, Fp1 and Fp2, suggest that changes in complexities of brainwaves happen simultaneously in different parts of the brain. Moreover, the variation of complexities of EEG between Fp1 and Fp2 provides useful information that can be used for separation between the groups. However, the better separation between groups can be obtained when more than two channels are used (Gómez et al., 2008).

The graphical analysis of D_f also showed that D_f was higher in Fp1 during phase 4 for all groups. However, the difference appears to be larger in the PD group than in the other two groups, BD and UD. This can be connected to the experimental design and demands further investigation.

The discriminant analysis yielded the following discriminant function that differentiates between the groups of depressed patients

$$y_1 = 0.369 \times X1_{Fp1} + 0.067 \times X2_{Fp1} + 0.478 \times X1_{Fp2} - 0.297 \times X2_{Fp2}.$$

The interpretation of the obtained discriminant function is complicated since the data exhibited multicollinearity. But if one tries to do it anyway, the following can be noted. The PD group had lower scores compared to the UD and BD. The only variable with relatively large contribution to discrimination between groups is D_f of the phase 1-3 obtained from Fp2.

The practical significance as measured by CR^2 shows that only the first discriminant function accounts for a sufficient amount of difference among the three groups with 0.43. The other measure of practical significance that measures proportion of the total group difference that is accounted for by the discriminant function showed to be 91.52%.

When thinking about the improvement of the discrimination between groups of depressed patient, it is worth to mention that the complexity of EEGs just before ECT-initiation could be considered. This information can be used as baseline information.

It also seems that D_f is of a doubtful potential to separate between the unipolar and the bipolar patients. It has been noted previously that there is a specific relationship between bipolar and unipolar depressed patients.

It is worth mentioning, that the sample was only 29, and the three groups were of different size (six, seven and sixteen). A study with more observations could yield better results. In this work, no patient characteristics such as gender or age were taken into account although these are very important. For example, a specific diagnose can be age related. This is mainly due to limitation of the sample size, and the attempt to avoid heteroscedasticity in groups with respect to different patient characteristics. Moreover the purpose of this work was not a data analysis but the study of the potential of using the combination of mathematical (deterministic) and statistical (random) modeling.

One can speculate that the difficulties in classifying the UD and BD patients could be due to the turnover (Angst, 1996), when unipolar become bipolar. Perhaps some of the patients were incorrectly diagnosed compared to the date of therapy. Poor separation can be simply due to similarities between the two diagnoses: in both cases, patients suffer from episodes of depressive emotions while bipolar depressed also experience episodes of hypomania. It could be hypothesized that similarities in these subcategories are reflected in similar responses in signal complexity from EEGs of ECT seizures. Perhaps PD is a completely different physiological condition than BD and UD.

To estimate how accurately a predictive model will perform in practice, the resubstitution and cross validation have been used. The classification results using resubstitution were very inaccurate. However every patient with psychotic depression was classified into the correct group, indicating that psychotic depression is more different than the other categories with respect to variables used in our analysis. It appears to be much more difficult to distinguish between the BD and UD. Only 3 out of 7 BD were correctly classified and there were equally many patients that were classified as UD. One was also classified into psychotic depression. The same was true for UD patients. 6 out of 16 were classified as BD, 8 were correctly classified as UD and 2 were classified PD.

When crossvalidation was used the misclassification rate worsened considerably. PD stood out as the only category that was correctly classified more often than it was misclassified, with its 4 out of 5 correct classifications. And the observations from BD and UD were misclassified more often than they were correctly classified.

In our work, equal priors have been used. Since the groups' proportion in the sample do not reflect the proportions of the corresponding diagnoses in the population, the proportion in the sample cannot be used as an estimate of proportions in the population.

In this work the goal has been to see whether it is possible to get interesting results by applying the discriminant analysis to the fractal dimension of ECT-EEG series. No separation between the BD and UD group was observed. We were however able to see differences between the PD and the other two diagnoses. We believe that larger number of variables, both patient and EEG related, and larger sample size can improve the results.

References

- Angst, J. (1996). Comorbidity of mood disorders: a longitudinal prospective study. *The British Journal of Psychiatry*, 30, 31-37.
- Culic, M., Gjoneska, B., Hinrikus, H., Jändel, M., Klonowski, W., Liljenström, H., Pop-Jordanova, N., Psatta, D., von Rosen, D. and Wahlund, B. (2009). Signatures of Depression in Non-Stationary Biometric Time Series. *Computational Intelligence and Neuroscience*, 2009: 989824. Epub 2009 jun 24.
- Encyclopædia Britannica (2010). / electroencephalography (physiology)/ (Hitting the headlines article) [online]. (Updated 24 Sep 2009) Available at: <<http://www.britannica.com/>> [Accessed 29 dec 2010].
- Fisher, R. A. (1936). The use of multiple measurements in taxonomic problems. *Annals of Eugenics*, 7, 179-188.
- Georgiev, S., Minchev, Z., Christova, C. and Philipova, D. (2009) EEG Fractal Dimension Measurement before and after Human Auditory Stimulation. *Bioautomation*, 12, 70-81.
- Gómez, C., Mediavilla, Á., Hornero, R., Abásolo, D. and Fernández, A. (2009). Use of the Higuchi's fractal dimension for the analysis of MEG recordings from Alzheimer's disease patients. *Medical Engineering and Physics*, 31, 306-313.
- Higuchi, T. (1988). Approach to an irregular time series on the basis of the fractal theory. *Physica D*, 31, 277-283.
- Huberty, C.J. and Curry, A. R. (1978). Linear versus quadratic multivariate classification. *Multivariate Behavioural Research*, 13, 237-245.
- Johnson, R. A., Wichern, D. W. (2007). *Applied Multivariate Statistical Analysis*, 6th edition, USA, Pearson Education, Inc.
- Klonowski, W. (2007). From conformons to human brains: an informal overview of nonlinear dynamics and its applications in biomedicine. *Nonlinear Biomedical Physics*, doi:10.1186/1753-4631-1-5. Epub 2007 jul 5.
- Lachenbruch, P. A. (1967). An almost unbiased method of obtaining confidence intervals for the probability of misclassification in discriminant analysis. *Biometrics*, 23, 639-645.
- Lachenbruch, P. A. (1975). Some unsolved Practical Problems in Discriminant Analysis, *Institute of Statistics Mimeo Series*, No. 1050.
- Mardia, K.V. (1970). Measures of Multivariate Skewness and Kurtosis with Applications, *Biometrika*, 57 (3), 519-530.
- Mardia, K. V. (1974). Applications of some measures of multivariate skewness and kurtosis in testing normality and robustness studies. *Sankhyā*, Ser B, 36(2), 115-128.

Michaelis, J. (1973). Simulation experiments with multiple group linear and quadratic discriminant analysis. In T.Cacoullus (ed.), *Discriminant Analysis and Applications*. New York, Academic Press.

Raykov, T. and Marcoulides, G. A. (2008). *An Introduction to Applied Multivariate Analysis*, New York, Taylor and Francis Group.

Rencher, C. A. (1998). *Multivariate Statistical Inference and Applications*, Canada, John Wiley and Sons, Inc.

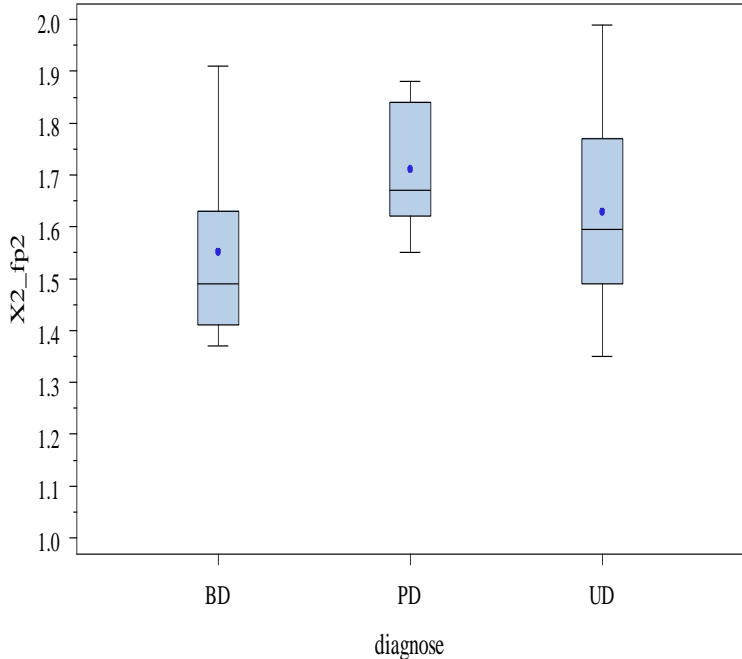
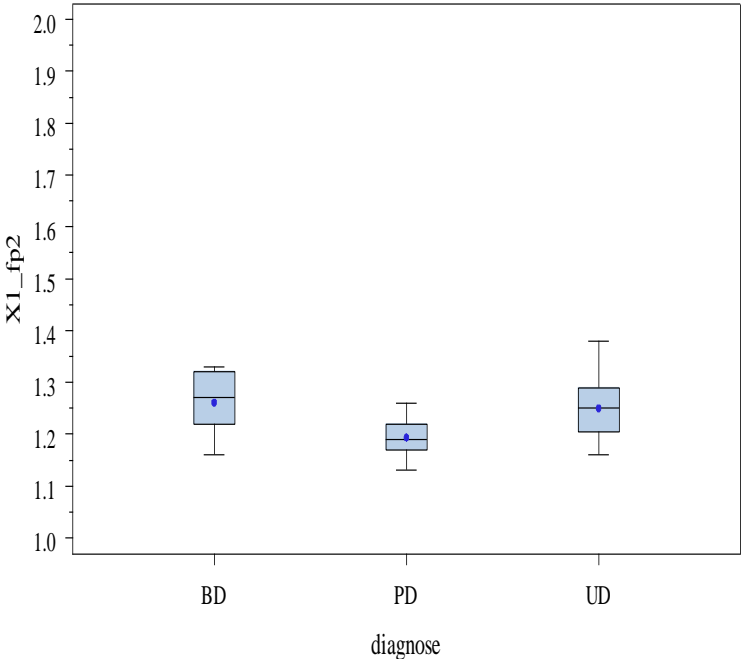
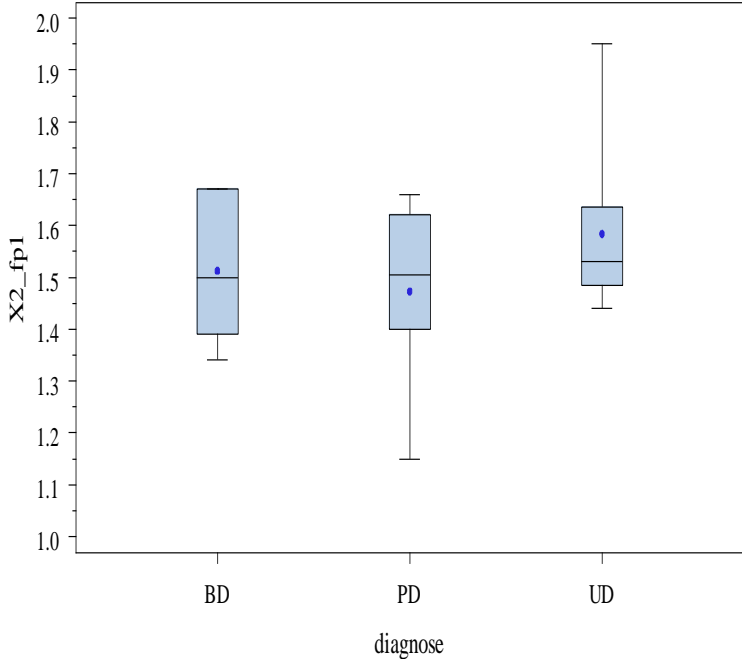
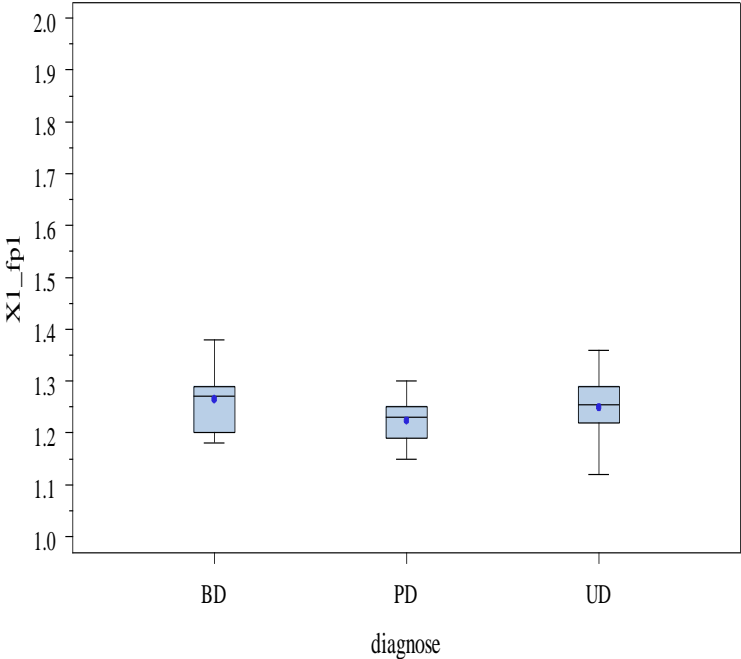
SAS (2010). /Macro to test multivariate normality: %MULTINORM MACRO / [online]. Available at: <<http://support.sas.com/kb/24/983.html>> [Accessed 29 dec 2010]

Sharma, S. (1996). *Applied Multivariate Techniques*, USA, John Wiley and Sons, Inc.

Wahlund, B., Klonowski, W., Stepien, P., Stepien, R., von Rosen, D. and von Rosen, T. (2010). EEG Data, Fractal Dimension and Multivariate Statistics, *Journal of Computer Science and Engineering*, 3(1), 1-7.

Wahlund, B., Piazza P., von Rosen, D, Liberg, B. and Liljenström, H. (2009). 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*, 2009: 965209 Epub 2009 jun 15.

Appendix 1. Box Plots of Individual Variables by Diagnose



The vertical axes show estimated fractal dimension of an EEG signal.

Appendix 2. Univariate and Multivariate Significance Tests

Table A2.1. Univariate Significance Tests

Univariate Test Statistics							
F Statistics, Num DF=2, Den DF=25							
Variable	Total Standard Deviation	Pooled Standard Deviation	Between Standard Deviation	R-Square	R-Square / (1-RSq)	F Value	Pr > F
X1_fp1	0.0587	0.0578	0.0225	0.1015	0.1129	1.41	0.2625
X2_fp1	0.1367	0.1382	0.0383	0.0543	0.0574	0.72	0.4977
X1_fp2	0.0616	0.0590	0.0286	0.1495	0.1758	2.20	0.1321
X2_fp2	0.1745	0.1729	0.0632	0.0908	0.0999	1.25	0.3042

Table A2.2. Various Multivariate Significance Tests

Multivariate Statistics and F Approximations					
S=2 M=0.5 N=10					
Statistic	Value	F Value	Num DF	Den DF	Pr > F
Wilks' Lambda	0.53247297	2.04	8	44	0.0638
Pillai's Trace	0.49567328	1.89	8	46	0.0839
Hotelling-Lawley Trace	0.82517009	2.21	8	29.2	0.0558
Roy's Greatest Root	0.75517360	4.34	4	23	0.0092
NOTE: F Statistic for Roy's Greatest Root is an upper bound.					
NOTE: F Statistic for Wilks' Lambda is exact.					

Appendix 3. Discriminant Analysis

Table A3.1. Significance Tests of Discriminant Functions

Test of H0: The canonical correlations in the current row and all that follow are zero					
	Likelihood Ratio	Approximate F Value	Num DF	Den DF	Pr > F
y₁	0.53247297	2.04	8	44	0.0638
y₂	0.93458251	0.54	3	23	0.6619

Table A3.2. Standardized Coefficients of the Discriminant Functions

Pooled Within Canonical Structure		
Variable	y₁	y₂
X1_fp1	0.368845	-0.381924
X2_fp1	0.067274	0.878220
X1_fp2	0.478130	-0.212184
X2_fp2	-0.29747	0.687186

Table A3.3. Various Measures of Practical Significance of the Discriminant Functions

	Canonical Correlation	Adjusted Canonical Correlation	Approximate Standard Error	Squared Canonical Correlation	Eigenvalues of the $W^{-1}B$ matrix			
					Eigenvalue	Difference	Proportion	Cumulative
y₁	0.655939	0.595233	0.109647	0.430256	0.7552	0.6852	0.9152	0.9152
y₂	0.255768	0.138687	0.179860	0.065417	0.0700		0.0848	1.0000

Appendix 4. Test of Multivariate Normality with SAS

The following is an extract from SAS online Documentation giving some details concerning the macro to test multivariate normality (SAS, 2010).

Purpose

The %MULTNORM macro provides tests and plots of multivariate normality. A test of univariate normality is also given for each of the variables. A chi-square quantile-quantile plot of the observations' squared Mahalanobis distances can be obtained allowing a visual assessment of multivariate normality. Univariate plots with overlaid normal curves are also available.

Usage

Follow the instructions in the Downloads tab of this sample to save the %MULTNORM macro definition. Replace the text within quotes in the following statement with the location of the %MULTNORM macro definition file on your system. In your SAS program or in the SAS editor window, specify this statement to define the %MULTNORM macro and make it available for use:

```
%inc "<location of your file containing the MULTNORM macro>";
```

Following this statement, you may call the %MULTNORM macro. See the Results tab for an example.

The options and allowable values are:

DATA=

SAS data set to be analyzed. If the DATA= option is not supplied, the most recently created SAS data set is used.

VAR=

REQUIRED. The list of variables to use when testing for univariate and multivariate normality. Individual variable names, separated by blanks, must be specified. Special variable lists may NOT be used (such as VAR1-VAR10, ABC--XYZ, _NUMERIC_).

PLOT=BOTH | MULT | UNI | NONE

PLOT=MULT requests a high- or low-resolution chi-square quantile-quantile (Q-Q) plot of the squared Mahalanobis distances of the observations from the mean vector. PLOT=UNI requests high-resolution univariate histograms of each variable with overlaid normal curves and additional univariate tests of normality. Note that the univariate plots cannot be produced if HIRES=NO. PLOT=BOTH (the default) requests both of the above. PLOT=NONE suppresses all plots and the univariate normality tests.

HIRES=YES | NO

Ignored if PLOT=NONE. HIRES=YES (the default) requests that high-resolution graphics be used when creating plots. You must set the graphics device (GOPTIONS DEVICE=) and any other graphics-related options before invoking the macro. HIRES=NO requests that the multivariate plot be drawn with low-resolution. The univariate plots are not available with HIRES=NO.

Details

If a set of variables is distributed as multivariate normal, then each variable must be normally distributed. However, when all individual variables are normally distributed, the set of variables may not be distributed as multivariate normal. Hence, testing each variable only for univariate normality is not sufficient. Mardia (1974) proposed tests of multivariate normality based on sample measures of multivariate skewness and kurtosis.

For multivariate normal data, Mardia (1974) shows that the expected value of the multivariate skewness statistic is

$$p(p+2)[(n+1)(p+1)-6] / (n+1)(n+3)$$

and the expected value of the multivariate kurtosis statistic is

$$p(p+2)(n-1)/(n+1) .$$

As discussed in the *SAS/ETS User's Guide*, the Henze-Zirkler test of multivariate normality is based on a nonnegative function that measures the distance between two distribution functions and is used to assess distance between the distribution function of the data and the Multivariate Normal distribution function.

Univariate normality is tested by either the Shapiro-Wilk W test or the Kolmogorov-Smirnov test. For details on the univariate tests of normality, see **Tests for Normality** in "The UNIVARIATE Procedure" chapter in the [SAS Procedures Guide](#).

If the p-value of any of the tests is small, then multivariate normality can be rejected. However, it is important to note that the univariate Shapiro-Wilk W test is a very powerful test and is capable of detecting small departures from univariate normality with relatively small sample sizes. This may cause you to reject univariate, and therefore multivariate, normality unnecessarily if the tests are being done to validate the use of methods that are robust to small departures from normality.

When the data's covariance matrix is singular, the macro quits and the following message is issued:

ERROR: Covariance matrix is singular.

The PRINCOMP procedure in SAS/STAT Software is required for this check. If it is not found, a message is printed, nonsingularity is assumed and the macro attempts to perform the tests.

For p variables and a large sample size, the squared Mahalanobis distances of the observations to the mean vector are distributed as chi-square with p degrees of freedom. However, the sample size must be quite large for the chi-square distribution to obtain unless p is very small. Also, this plot is sensitive to the presence of outliers. So, this plot should be cautiously used as a rough indicator of multivariate normality.

Missing Values

Observations with missing values are omitted from the analysis and plot.

Macro Code

```
%macro multnorm (version,
  data=_last_ , /*          input data set          */
  var=          , /* REQUIRED:  variables for test      */
                /* May NOT be a list e.g. var1-var10  */
  plot=both    , /* Create multivar and/or univar plot? */
  hires=yes    , /* Create high-res plots?          */
                );

%let _version=1.3;
%if &version ne %then %put MULTNORM macro Version &_version;

%if %sysevalf(&sysver < 7) %then %do;
  %put MULTNORM: SAS 7 or later is required.  Terminating.;
  %let opts=;
  %goto exit;
%end;

%let opts = %sysfunc(getoption(notes))
            _last_=%sysfunc(getoption(_last_));
%if &data=_last_ %then %let data=&syslast;
options nonotes;

/* Check for newer version */
%if %sysevalf(&sysver >= 7) %then %do;
  filename _ver url
'http://ftp.sas.com/techsup/download/stat/versions.dat';
  data _null_;
    infile _ver;
    input name:$15. ver;
    if upcase(name)="&sysmacroname" then call symput("_newver",ver);
  run;
%if &syserr ne 0 %then
  %put &sysmacroname: Unable to check for newer version;
%else %if %sysevalf(&_newver > &_version) %then %do;
  %put &sysmacroname: A newer version of the &sysmacroname macro is
available.;
  %put %str(          ) You can get the newer version at this location;;
  %put %str(          ) http://support.sas.com/ctx/samples/index.jsp;
%end;
%end;

/* Verify that VAR= option is specified */
%if &var= %then %do;
  %put ERROR: Specify test variables in the VAR= argument;
  %goto exit;
%end;
```

```

/* Parse VAR= list */
%let _i=1;
%do %while (%scan(&var,&_i) ne %str() );
    %let arg&_i=%scan(&var,&_i);
    %let _i=%eval(&_i+1);
%end;
%let nvar=%eval(&_i-1);

/* Remove observations with missing values */
%put MULTNORM: Removing observations with missing values...;
data _nomiss;
    set &data;
    if nmiss(of &var )=0;
run;

/* Quit if covariance matrix is singular */
%let singular=nonsingular;
%put MULTNORM: Checking for singularity of covariance matrix...;
proc princomp data=_nomiss outstat=_evals noprint;
    var &var ;
run;
%if &syserr=3000 %then %do;
    %put MULTNORM: PROC PRINCOMP required for singularity check.;
    %put %str(          Covariance matrix not checked for singularity.);
    %goto findproc;
%end;
data _null_;
    set _evals;
    where _TYPE_='EIGENVAL';
    if round(min(of &var ),1e-8)<=0 then do;
        put 'ERROR: Covariance matrix is singular.';
        call symput('singular','singular');
    end;
run;
%if &singular=singular %then %goto exit;

%findproc:
/* Is IML or MODEL available for analysis? */
%let mult=yes; %let multttext=%str( and Multivariate);
%put MULTNORM: Checking for necessary procedures...;
proc model; quit;
%if &syserr=0 and %substr(&sysvlong,1,1)>=8 %then %do;
    %put MULTNORM: Using SAS/ETS PROC MODEL...;
    %goto model;
%end;
proc iml; quit;
%if &syserr=0 %then %do;
    %put MULTNORM: Using SAS/IML...;
    %goto iml;
%end;
%put MULTNORM: SAS/ETS PROC MODEL with NORMAL option or SAS/IML is
required;
%put %str(          to perform tests of multivariate normality.
Univariate);
%put %str(          normality tests will be done.);
%let mult=no; %let multttext=;
%goto univar;

%iml:
proc iml;

```

```

reset;
use _nomiss; read all var {&var} into _x; /* input data */

/* compute mahalanobis distances */
_n=nrow(_x); _p=ncol(_x);
_c=_x-j(_n,1)*_x[:,,]; /* centered variables */
_s=(_c`*_c)/_n; /* covariance matrix */
_rij=_c*inv(_s)*_c`; /* mahalanobis angles */

/* get values for probability plot and output to data set */
%if %upcase(%substr(&plot,1,1))=M or %upcase(%substr(&plot,1,1))=B %then
%do;
_d=vecdiag(_rij#(_n-1)/_n); /* squared mahalanobis distances */
_rank=ranktie(_d); /* ranks of distances */
_chi=cinv((_rank-.5)/_n,_p); /* chi-square quantiles */
_chiplot=_d||_chi;
create _chiplot from _chiplot [colname={'MAHDIST' 'CHISQ'}];
append from _chiplot;
%end;

/* Mardia tests based on multivariate skewness and kurtosis */
_b1p=(_rij##3)[+,+]/(_n##2); /* skewness */
_b2p=trace(_rij##2)/_n; /* kurtosis */
_k=(_p+1)#(_n+1)#(_n+3)/(_n#((_n+1)#(_p+1)-6)); /* small sample
correction */
_b1pchi=_b1p#_n#_k/6; /* skewness test
statistic */
_b1pdf=_p#(_p+1)#(_p+2)/6; /* and df
*/
_b2pnorm=(_b2p-_p#(_p+2))/sqrt(8#_p#(_p+2)/_n); /* kurtosis test
statistic */
_prob1p=1-probchi(_b1pchi,_b1pdf); /* skewness p-value */
_prob2p=2*(1-probnorm(abs(_b2pnorm))); /* kurtosis p-value */

/* output results to data sets */
_names={"Mardia Skewness","Mardia Kurtosis"};
create _names from _names [colname='TEST'];
append from _names;
_probs=(_n||_b1p||_b1pchi||_prob1p) // (_n||_b2p||_b2pnorm||_prob2p);
create _values from _probs [colname={'N' 'VALUE' 'STAT' 'PROB'}];
append from _probs;
quit;
%if &syserr ne 0 %then %do;
%put MULTNORM: Errors encountered in PROC IML. Terminating.;
%goto exit;
%end;

data _mult;
merge _names _values;
run;

%univar:
/* get univariate test results */
proc univariate data=_nomiss noprint;
var &var;
output out=_stat normal=&var ;
output out=_prob probn=&var ;
output out=_n n=&var ;
run;

```

```

data _univ;
  set _stat _prob _n;
  run;

proc transpose data=_univ name=variable
  out=_tuniv(rename=(col1=stat col2=prob col3=n));
  var &var ;
  run;

data _both;
  length test $15.;
  set _tuniv
    %if &mult=yes %then _mult;;
  if test='' then if n<=2000 then test='Shapiro-Wilk';
    else test='Kolmogorov';
  run;

proc print data=_both noobs split='/';
  var variable n test      %if &mult=yes %then value;
  stat prob;
  format prob pvalue.;
  title "MULTNORM macro: Univariate&multttext Normality Tests";
  label variable="Variable"
    test="Test"      %if &mult=yes %then
    value="Multivariate/Skewness &/Kurtosis";
    stat="Test/Statistic/Value"
    prob="p-value";
  run;
%if %upcase(%substr(&plot,1,1))=N %then %goto exit;
%if (%upcase(%substr(&plot,1,1))=U or %upcase(%substr(&plot,1,1))=B) and
  %upcase(%substr(&hires,1,1))=Y %then %do;
  ods exclude fitquantiles parameterestimates;
  proc univariate data=_nomiss noprint;
    hist &var / normal;
  run;
%end;
%if %upcase(%substr(&plot,1,1))=M or %upcase(%substr(&plot,1,1))=B %then
  %if &mult=yes %then %goto plotstep;
  %else %goto plot;
%else %goto exit;

%model:
/* Multivariate and Univariate tests with MODEL */
ods select normalitytest;
proc model data=_nomiss;
  %do _i=1 %to &nvar;
    &&arg&_i = _a;
  %end;
  fit &var / normal;
  title "MULTNORM macro: Univariate&multttext Normality Tests";
  run;
quit;
%if &syserr ne 0 %then %do;
  %put MULTNORM: Errors encountered in PROC MODEL. Terminating.;
  %goto exit;
%end;

%if %upcase(%substr(&plot,1,1))=N %then %goto exit;
%if (%upcase(%substr(&plot,1,1))=U or %upcase(%substr(&plot,1,1))=B) and
  %upcase(%substr(&hires,1,1))=Y %then %do;

```

```

ods exclude fitquantiles parameterestimates;
proc univariate data=_nomiss noprint;
  hist &var / normal;
run;
%end;
%if %upcase(%substr(&plot,1,1))=U %then %goto exit;

%plot:
/* compute values for chi-square Q-Q plot */
proc princomp data=_nomiss std out=_chiplot noprint;
  var &var ;
run;
%if &syserr=3000 %then %do;
  %put ERROR: PROC PRINCOMP in SAS/STAT needed to do plot.;
  %goto exit;
%end;
data _chiplot;
  set _chiplot;
  mahdist=uss(of prin1-prin&nvar );
  keep mahdist;
run;
proc rank data=_chiplot out=_chiplot;
  var mahdist;
  ranks rdist;
run;
data _chiplot;
  set _chiplot nobs=_n;
  chisq=cinv((rdist-.5)/_n,&nvar);
  keep mahdist chisq;
run;

%plotstep:
/* Create a chi-square Q-Q plot
NOTE: Very large sample size is required for chi-square asymptotics
unless the number of variables is very small.
*/
%if %upcase(%substr(&hires,1,1))=Y %then
  %if %sysprod(graph)=1 %then %str(proc gplot data=_chiplot;);
  %else %do;
    %put MULTNORM: SAS/GRAPH not found. PROC PLOT will be used instead.;
    proc plot data=_chiplot;
  %end;
%else %str(proc plot data=_chiplot;);
  plot mahdist*chisq;
  label mahdist="Squared Distance"
        chisq="Chi-square quantile";
  title "MULTNORM macro: Chi-square Q-Q plot";
run;
quit;

%exit:
options &opts;
title;
%mend;

```