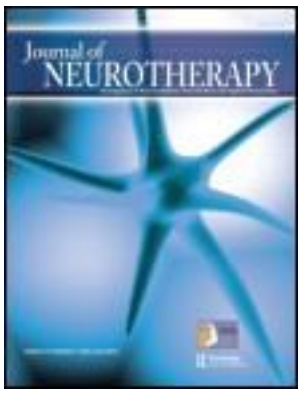


Journal of Neurotherapy: Investigations in Neuromodulation, Neurofeedback and Applied Neuroscience

Logistic Discriminant Functions in Electroencephalography

Marco Congedo PhD [a]

[a] Department of Psychology , University of Tennessee

Published online: 08 Sep 2008.

To cite this article: Marco Congedo PhD (2003) Logistic Discriminant Functions in Electroencephalography, Journal of Neurotherapy: Investigations in Neuromodulation, Neurofeedback and Applied Neuroscience, 7:2, 5-23, DOI: 10.1300/J184v07n02_02

To link to this article: http://dx.doi.org/10.1300/J184v07n02_02

SCIENTIFIC ARTICLES

Logistic Discriminant Functions in Electroencephalography

Marco Congedo, PhD

ABSTRACT. *Introduction*. Neurofeedback treatments in clinical settings are designed to normalize abnormal quantitative EEG (QEEG) features. Typically the patient's electrophysiological features are compared to a database in order to quantify the patient's deviance from normative values. A clear diagnosis is crucial in deciding the most appropriate protocol for the case. Combining the initial diagnosis with the database information offers a reliable procedure for deciding the most appropriate neurofeedback protocol.

Method. Logistic regression is a powerful model for performing discriminant analysis. Discriminant functions are a tool used to quantify the probability that the patient's QEEG features are typical of either of

Marco Congedo is affiliated with the Department of Psychology, University of Tennessee.

Address correspondence to: Marco Congedo, University of Tennessee, Department of Psychology, Austin Peay Building, Knoxville, TN 37996 (E-mail: mcongedo@utk.edu).

The author thanks Dr. Noland White, who patiently corrected and revised the original manuscript.

Journal of Neurotherapy, Vol. 7(2) 2003

10.1300/J184v07n02_02

two groups. One of the two groups is generally chosen to be normative, while the other is a homogeneous clinical group. While normative databases can detect deviation from normality, they cannot indicate if the QEEG pattern is typically observed in a particular clinical condition. Discriminant functions, however, address these types of issues. Consequently, normative databases and discriminant functions can be considered as complementary tools. With the widespread availability of software which is able to compute the maximum likelihood estimations, logistic regression has recently became a popular tool in biology. However, to date it has received little attention in electrophysiology. With this article we hope to stimulate the application of logistic regression model in this field. The logistic regression discriminant model with a single quantitative predictor is illustrated.

Results. The construction of a discriminant function is presented in a step-by-step fashion, stressing methodological, practical and theoretical concerns that should be addressed in order to achieve valid and useful results. A working example illustrates, in a step-by-step fashion, how to implement the discriminant process.

Discussion. The approach taken in this article aims to highlight the intuitive appeal and simplicity of this technique. It is shown that the discriminant process can easily be automated with little effort on the part of the clinician.

KEYWORDS. Discriminant function, logistic regression, LOGIT, EEG, qEEG, norms, normative database

INTRODUCTION

Normative database comparisons and discriminant functions are among the most elaborate and powerful data analysis procedures available in modern clinical electroencephalography (EEG). Indeed they are complementary. With a comparison to a normative database, we aim to assess the subject's deviance from the norm. With a discriminant function, we aim to assess the probability that the subject belongs to a specified clinical group, as opposed to the probability that the subject belongs to the normative group. Both procedures may prove useful in the evaluation process, but only the discriminant function can narrow the focus to a specific clinical category.

One of the major advantages of computerized data analysis is that it can be automated, requiring little effort on the part of the clinician. The disadvantages arise when it is used improperly. The evaluation process cannot be entirely automated and the expertise of the clinician cannot be replaced by the logic underlying numerical computations.

Nuwer and Hauser (1994) report a case of blind trust in the "decision" of a discriminant function. A 51 year-old man experienced episodes of weakness and right-side numbness. The EEG discriminant function provided strong evidence of chronic schizophrenia ($p < 0.025$). The clinician interpreted the symptoms of the patient as somatization. The patient worsened in the next two months. Finally, he underwent a MRI, which revealed a large, deep left-parietal mass extending across the corpus callosum. The patient died days later. The biopsy confirmed the presence of a grade 3 astrocytoma.

Schizophrenia has been found to be associated with increased Delta (< 3 Hz) and/or Theta (4-7 Hz) activity (Hughes & John, 1999). However, abnormal Delta oscillation is also associated with white matter tumors and generation of Theta activity is associated with the edema surrounding it (Fernandez-Bouzas et al., 1999; Nunez, Wingeier, & Silberstein, 2001). The traditional scalp potential recording is an oversimplified representation of the underlying complex modulation of cortical synaptic activity (Nunez et al., 2001). For this reason the distribution of scalp potentials may appear similar in a broad range of clinical conditions. Under those circumstances a computerized algorithm may be deceived, since the distribution of scalp potentials is the only information it is provided. However, the clinician is not subject to such a restriction and only a comprehensive clinical evaluation will lead to the correct diagnosis.

With the Fourier analysis, the EEG is shifted from the time domain to the frequency domain. The results of a comprehensive quantitative analysis of the EEG include a plethora of measurements. These measurements may include both absolute and relative measures of signal energy (amplitude and power), amplitude or power difference between pairs of electrode locations (asymmetry), and the degree of phase consistency and lag between pairs of locations (coherence and phase). Additionally, these measures may vary along the frequency spectrum and according to the spatial location.

The term "descriptor" refers to a measure derived at a particular location for a particular frequency or frequency band. For example "Theta power at FP1" may refer to the energy of the EEG signal as recorded on the left frontal pole and averaged between 4 and 7 Hz. If a descriptor (or

the combination of a set of descriptors) proves to assume values systematically different between a clinical and normal sample, we wish to evaluate the descriptor with regard to a new individual in order to predict his/her group membership. The theoretical background justifying this inference is the same as the one justifying comparison to normative databases. The human EEG is assumed to possess sufficient intra-subject and inter-subject reliability, strong genetic components, and inter-racial/inter-cultural consistency. In addition, brain pathology is assumed to have peculiar features that are detectable with the EEG (John, Prichep, & Easton, 1987). Studies justifying these assumptions along with fundamental methodological consideration have been reviewed elsewhere (Congedo & Lubar, in press). See Hughes and John (1999) for a recent review of the electrophysiological correlates of psychiatric disorders.

Traditionally, discriminant functions have followed linear theory models (Kutner, Nachtsheim, Wasserman, & Neter, 1996). With the introduction of algorithms for maximum likelihood iterative fit, logistic regression has become increasingly popular in the biological area (Agresti, 1990; Hosmer & Lemeshow, 1989). Among others, linear discriminant functions have been applied to dementia of the Alzheimer's type (Knott, Mohr, Mahoney, & Ilivitsky, 2001), mild brain injury (Thatcher et al., 2001; Thornton, 1999), and in migraine and headache (de Tommaso et al., 1999).

Other applications of the logistic regression discriminant function include a study of violence in schizophrenic inpatients by Arango, Calcedo Barba, González-Salvador, and Calcedo Ordónez (1999), a sleep study aimed to discriminate elderly depressed and demented patients by Houck, Reynolds, Mazumdar, and Kupfer (1991), and a study on alpha relative power and magnesium levels in athletes by Delorme, Bourdin, Viel, Simon Rigaud, and Kantelip (1992). Besides these isolated applications, the use of logistic regression discriminant function has never received a systematic review in the EEG community.

In this article we introduce the logistic regression model and will illustrate how to construct a discriminant function in the case of a single quantitative predictor. The goal is to show the intuitive appeal of the technique and the simplicity of its implementation. We follow the practical approach of Agresti (1996) and the implementation is facilitated by the use of common statistical packages. We also discuss the use of logistic regression discriminant functions in electrophysiology, stressing its advantages and highlighting its flexibility and potential. We believe that a conscientious use of discriminant functions may be of great

utility in both the clinical and research setting. In sum, we hope to stimulate applications of the technique and to provide a practical guide for those interested researchers that have not been previously exposed to logistic regression.

METHOD

This section is divided in four parts. First, we highlight general issues implicated in the design of a discriminant function. Second, we review the basic theory of logistic regression and show how to automate the actual discriminant process in the case of a single predictor. Third, we propose a suitable framework for the interpretation of the outcome of a discriminant function. Finally, a working example shows in a step-by-step fashion how to actually perform the discriminative process. While the construction of a logistic regression discriminant function requires the expertise of a statistician, the discriminant process of new individuals can easily be automated.

General Issues

The construction of a discriminant function is not limited to its actual implementation. Table 1 summarizes the steps that should be undertaken in order to make the most effective use of this technique.

The first step in Table 1, Clarify Objectives, should be evaluated carefully. Much of the success of the discriminant function is determined at this stage. First, we should define the goal we want to achieve. The literature on the target disorder should be studied carefully and potential candidate variables to be used as predictors of the disorder should be identified. The discrimination of a specific clinical condition in a broad nosological category may prove elusive since there may be too much variability among sub-categories of the disorder. Therefore, the clinical category should be as specific and homogeneous as possible and potential predictors should have some a priori justification.

In the statistical community, the practice of blindly testing all possible hypotheses and repeatedly "torturing" the data until some significant effect pops out is known as *fishing*. It is well known that false significant effects can be found in totally random data if one keeps manipulating the numbers a sufficient number of times. For a given data set, among the thousands of descriptors gained from the quantitative electroencephalographic (qEEG) analysis, some may in fact display

TABLE 1. Steps Involved in the Construction of a Discriminant Function

Steps	Description
1) Clarify Objectives	Decide on the purpose of the discriminant function. Assess theoretical justification of the discriminant function. Select good candidates for predictors. Collect Data.
2) Model Selection	Select the best discriminant model. Check the model by goodness-of-fit and residual diagnostic.
3) Implementation	Automate the discriminant process.
4) Validation	Assess the validity of the model with external data.
5) Maintenance	Up-date the discriminant function with new data. Periodically repeat steps 2 to 4. If needed, repeat all steps.

discriminant power due to chance alone. A discriminant function conceived by fishing will likely fail Step 4, Validation.

If the predictors are chosen on the basis of previously replicated studies, not only will the discriminant function likely be validated, but we are also provided with a more manageable framework for Step 2, Model Selection. One example of such a discriminant function can be found in Monastra, Lubar, and Linden (2001), where the well studied Theta/Beta ratio at Cz (Lubar, 1991) was chosen as a predictor of Attention Deficit Hyperactivity Disorder (ADHD).

Data collection should also be planned at this stage. For example, if a candidate predictor is the "power at Cz in the gamma band" we will want to make sure not to low-pass filter the data below 45 Hz. Step 2, Model Selection, involves the selection of the model with the best statistical properties. By the end of Step 1, we should have identified one or more potential predictors. If we are left with several candidates, there exist as many regression models as all the combinations of predictor main effects and terms accounting for the interaction among predictors. The chosen model has to have discriminative power and has to fit the data adequately. This step typically requires the technical expertise of a statistician. In fact, there are numerous pitfalls in which the novice analyst may fall while selecting and validating the best model.

In logistic regression, as in other similar procedures (e.g., linear regression), the model selection proceeds in both forward and reverse directions, eliminating and adding in the equation main effect and interaction terms until the optimal model is found. Note that at each stage of

the selection procedure, the model is checked and compared with alternative models. This is accomplished by means of tests of goodness-of-fit ("Wald" and "log-likelihood" statistics) and analysis of the residuals (Agresti, 1996).

The principle of parsimony requires that the selected model should be as simple (include as few terms) as possible (DeLurgio, 1998). This universal principle assumes even greater importance in qEEG. In fact, electrophysiological measures are highly correlated and the prediction power supplied by different predictors is often redundant. If there are two or more terms in the equation, the statistician will have to evaluate problems due to multicollinearity (correlation among predictors), which may invalidate the model. If an optimal model is found, and the model proves to have discriminant power, we are ready for Step 3, Implementation. Step 3 is the focus of this article and will be explored in greater detail.

Step 4, Validation, is necessary to show that the discriminant function possesses both internal and external validity. New data from both the clinical and non-clinical population should be collected. The data is entered into the discriminant function and the accuracy of the classification is assessed. Measures of accuracy typically examine both the sensitivity and specificity of a given assessment or classification method. Sensitivity is provided by the normalized percentage of subjects correctly classified as belonging to the clinical group (true positives). Specificity is examined by the normalized percentage of subjects correctly classified as belonging to the non-clinical group (true negatives). These accuracy measures are the complement of false negatives and false positives, respectively. See Congedo and Lubar (in press) for a detailed description of these measures.

Sometimes the same data used to build the discriminant function is then used for its validation. This kind of validation, sometimes referred to as a "jackknife" procedure, even it has little to do with a true jackknife procedure, is not as robust. In fact, a jackknife validation is not a validation in the strictest sense; rather, it only assesses the degree of differentiation of the two groups as evidenced by the discriminant function. Furthermore, using the same data set to both build a discriminant function and to validate it is a circular argument. The data is *found* to yield discrimination between two groups and then the validity of the discrimination function is supported showing that the data yields discrimination. This circularity is avoided by using a different data set to validate the discriminant function.

The last step, Maintenance, involves periodic checks of the discriminant function's validity. The definition and nosology of the clinical category may change over time, and the composition of the population for which the discriminant function developed may change as well. New data may also be added periodically to the discriminant function. To do so one has to refit the model (Step 2) and check that it is still the optimal solution. Step 3 and 4 should then be repeated as well.

Logistic Regression Discriminant Function

We now turn to logistic regression theory and to the implementation of the discriminant function. While the illustration below is strictly technical, the working example that follows shows in a step-by-step fashion how to perform all computations. The Appendix reports an example code for the SAS software package (SAS Institute, Inc.). This code can be used to obtain all quantities involved in the equations presented in this section. For a binary response Y and a quantitative variable X (predictor) taking on value x, let $\pi_{1(x)}$ be the probability that the Y variable takes on value "1" ("success"), and let $\pi_{0(x)}$ be the probability that the Y variable takes on value "0" ("failure"). In a typical application, "1" is the probability of belonging to the clinical group and "0" is the probability of belonging to the control group. The probability of belonging to the clinical group is the probability that we want to estimate. In this discussion X is a quantitative continuous random variable (descriptor) taking on value x. However, X might be a discrete variable as well (e.g., the presence/absence of an electrophysiological measure, a risk factor, gender, etc.). Typically in EEG applications X is a quantitative variable (e.g., power, coherence, etc.). For simplicity, hereafter we will refer to $\pi_{1(x)}$ as π_1 and to $\pi_{0(x)}$ as π_0. Define the LOGIT of the probability of success (probability of belonging to the clinical group) as the natural logarithm of the odds of success:

$$\text{LOGIT} = 1\text{n}(\pi_1/\pi_0) \tag{1.0}$$

Assuming a binomial probabilistic model with parameter π_1, the logistic regression model has linear form for the LOGIT

$$\text{LOGIT} = \alpha + \beta x \tag{2.0}$$

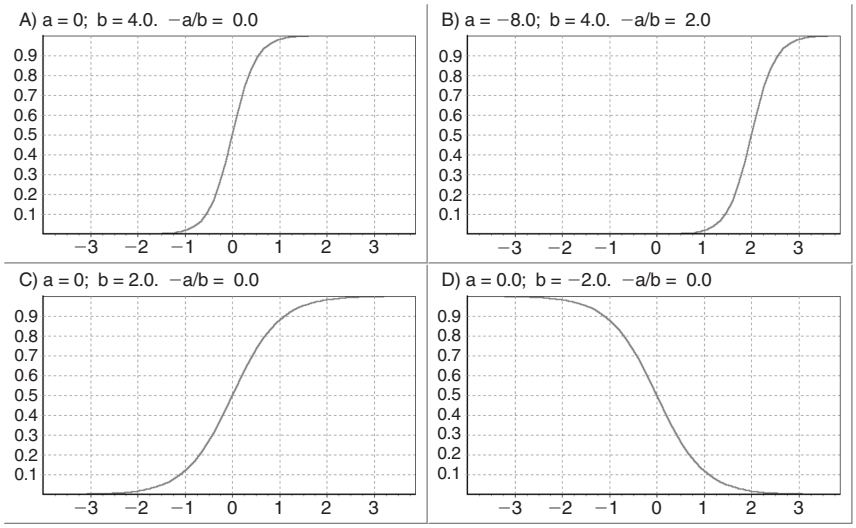
Modeling the probability of success itself and solving (2.0) for π_1, we obtain equivalently

$$\pi_1 = e^{\alpha + \beta x}/(1 + e^{\alpha + \beta x}) \qquad (3.0)$$

Equation (3.0) describes the target probability. The logistic regression model in (2.0) and its equivalent in (3.0) unfold a sigmoid function with a range (probability of success, y-axis) bounded by 0 and 1. The domain (predictor, x-axis) is unbounded (Figure 1).

The function has two parameters, α and β, called the intercept and the slope of the curve, respectively. The slope in (2.0) determines how steep the function is, with a higher slope resulting in a steeper rate of increase or decrease. The sign of the slope indicates whether the curve ascends or descends. The ratio $-\alpha/\beta$ is called the *Median Effective Level* (EL_{50}) and is the x-value for which the probability of success and failure is equivalent (50%). When the LOGIT function has a positive β, the func-

FIGURE 1. *Illustration of the LOGIT function parameters*. The y-axis represents the probability of success (e.g., belonging to the clinical group). The x-axis represents a continuous predictor. a = intercept; b = slope; −a/b = median effective level (EL_{50}).



A) EL_{50} is 0, the function has value x = 0 for y = 0.5. As an example, for x = 1 the probability of success is very close to 1.0.
B) The intercept is now negative. EL_{50} is now shifted to the right but the slope is unchanged.
C) As in A) but with smaller slope. Note that for x = 1 the probability of success is now around 0.9. The function is less steep.
D) As in C) but with the slope of opposite sign. The function is now decreasing, meaning that as x increases the probability of success decreases.

tion ascends with an increasing rate of change (first derivative, or rate of change per unit increase in x) up to the EL_{50} and then descends with a decreasing rate of change. The opposite pattern (first descending and then ascending) occurs when β is negative. If β equals zero, the function reduces to a straight horizontal line. This implies that the outcome probability is fixed and the model has no discriminant power. At the EL_{50} point, the rate of change is at its maximum with zero second derivatives (Figure 1). Thus, we see that unlike a "pure" linear model, like linear regression, the LOGIT model allows the rate of change to vary per unit change in x. This is also seen by the exponential growth of the odds. Following (2.0), the odds of success are given by

$$(\pi_1/\pi_0) = (e^{\alpha})(e^{\beta})^x \tag{4.0}$$

The odds increase multiplicatively by e^{β} for every unit increase of x. When β equals zero, e^{β} equals one, and the odds do not change over x.

A discriminant function accepts as input the predictor value of the subject (X variable) and assigns a probability of belonging to each group (Y variable). Of course, for a binary response the probability of success (i.e., the subject belongs to the clinical group) and the probability of failure (i.e., the subject belongs to the control group) are complements of each other. Hence, both probabilities are uniquely defined by one of them.

It is common practice to report the probability of success π_1, hereafter referred to as π. Depending on the number of subjects in the two groups, the amount of error tolerated, and the strength of the discriminative power achieved by the discriminant function, the point estimate π will be obtained with a variable degree of uncertainty. This is expressed by *confidence intervals*, which are low and high limits of the probability π for a given fixed amount of tolerated error (e.g., the error rate is usually set to 0.05).

The confidence intervals (CI) are the very essence of the discrimination problem in a modern statistical framework and we will see how to make use of them in the next section. Here we show how confidence intervals are computed. We need the covariance matrix of the parameter estimates for the logistic model (Agresti, 1996). These quantities are easily computed with the aid of most modern statistical software packages. An example code for SAS programming is reported in the Appendix. A comprehensive description of SAS programs for Logistic Regression can be found in Stokes et al. (2000). Other software packages, like

LogXact (Cytel Software, Inc.) have a user-friendly spreadsheet interface and do not require programming.

Here we show how to obtain CI for discrimination without performing the analysis with the software again. In this way the model has only to be estimated once during Step 2 (Table 1). Thereafter, the discriminant function can be automated and results for each new individual submitted to the discriminant function can be easily obtained without invoking the statistical software (and the statistician) again. The asymptotic standard error (ASE) of the estimated LOGIT as defined in (2.0) is given by

$$\text{ASE} = \sqrt{[Var(\alpha + \beta x)]} = \sqrt{[Var(\alpha) + x^2\, Var(\beta) + 2xCov(\alpha, \beta)]} \quad (5.0)$$

The confidence intervals for the LOGIT are given by

$$\text{CILOGIT} = (\alpha + \beta x) \pm [(1.96)\text{ASE}] \quad (6.0)$$

Finally, substituting those intervals for the exponents $(\alpha + \beta x)$ in (3.0) we obtain the confidence interval for the probability π as

$$\text{CI}_{\pi} = e^{(\text{CILOGIT})}/[1 + e^{(\text{CILOGIT})}] \quad (7.0)$$

In (7.0) the *upper confidence interval* (UCI) is obtained by substituting the upper CILOGIT (6.0), while the *lower confidence interval* is obtained by substituting the lower CILOGIT (6.0). If one desires to tolerate a different error, then he/she has to change the constant 1.96 in (6.0) with the appropriate quintile of the normal standard distribution. Note that while these CI are obtained by means of maximum likelihood iterative fitting algorithms, exact methods based on conditional inference (permutation) theory (Mehta & Patel, 1995) are now widely available. LogXact is highly specialized for this purpose. For the implementation of the technique in SAS, see Stokes et al. (2000, page 225).

Conditional exact logistic regression is advisable when the data is sparse or unbalanced. Furthermore it is the last resource in all cases when the asymptotic iterative algorithm fails to converge. This will be the case, for example, if there is a complete separation of the two groups along the x variables. Due to computational problems with the exact methods, for large data sets (a must in discriminant function analysis) the asymptotic approach is preferable if the algorithm converges.

With equation (7.0) we are able to automate a discriminant function. Once the model has been fitted and validated all we need is to store the

ASE as in (5.0) and the model parameter estimates α and β. Then, for every subject we want to submit to the discriminant function, we enter the descriptor value x in (6.0). Finally (3.0) will return the point estimate of the target probability of belonging to the clinical group and (7.0) will return the confidence intervals for this estimate. The quantities are easily computed using a spreadsheet program or by means of a computer program. The working example below shows how to obtain confidence intervals for the estimated probability of belonging to the clinical group. How to interpret those CI is the topic of the next section.

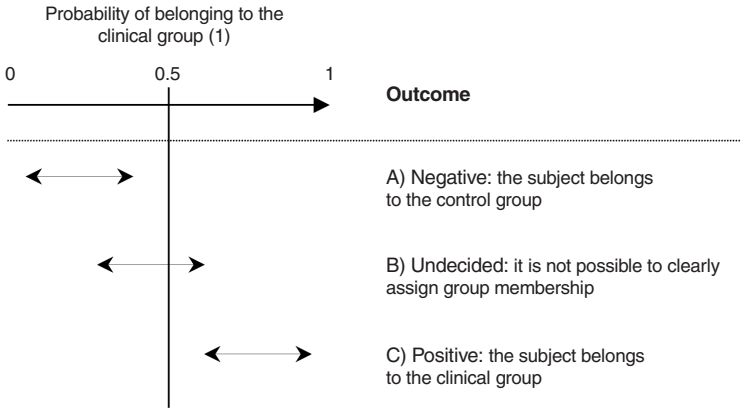
Interpretation of the Discriminant Function

A discriminant function returns the point estimate probability and the CI estimate probability that the subject belongs to the clinical group. In order to assign membership to the clinical group we may require that the CI does not include the 0.5 point, for which there is maximal indecision about the membership. Figure 2 summarizes the three possible outcomes of the discriminant function and the corresponding interpretations. See the caption of Figure 2 for details. Notice that there will be occasions in which a clear positive or negative outcome is possible, and occasions in which the outcome of the discriminant function is undecided.

A Working Example

The following example is based on simulated data and intends to show in a step-wise fashion, how to compute point estimates and confidence intervals to be used in the discriminant process. Imagine the following scenario: a researcher finds in their electronic archive a great number of old EEG recordings. Because of deterioration of the storage medium, part of the information about the data has been lost. The researcher knows that some of the recordings were eyes-closed (EC) and the remaining recordings were eyes-open (EO), but is not able to determine the respective recording condition of each file. The researcher wants to use these data files, so they decide to construct a discriminant function for EC versus EO in order to assign a recording condition to each file. They start by setting the tolerated error rate at 0.05 (i.e., decides that they can afford to incorrectly classify no more than 5 files out of every 100). From the literature they learn that the best predictor of recording condition EC vs. EO is the alpha (8-13 Hz) power (normalized

FIGURE 2. *Interpretation of the outcome of a discriminant function***.** The unidirectional arrow at the top of the figure represents the probability of belonging to the clinical group. The extremes, zero (0) and one (1), represent the certainty of belonging to the control group or the clinical group, respectively. The middle of the arrow (0.5) represents the point of maximal indecision about group membership. The bi-directional arrows represent possible outcomes of the Confidence Intervals issued by a discriminant function.



A) Both the lower and upper CI are smaller than 0.5: the outcome is *negative*, indicating that the subject belongs to the control group.
B) The lower CI is smaller than 0.5 and the upper CI is larger than 0.5: the outcome is *undecided*, indicating that it is not possible to assign group membership.
C) Both the lower and upper CI are greater than 0.5: the outcome is *positive*, indicating that the subject belongs to the clinical group.

and log-transformed) at occipital leads O1 and O2. They then compare 90 EC EEG recordings with 90 EO EEG recordings taken from more recent studies that match all other characteristics of the recordings in the old archive. On those new files, they compute means, standard deviations (SD) and t-tests for the two conditions using alpha power at O2 as the primary variable. The normalized and log-transformed mean power for the EC condition is 1.08 with SD 0.18. For the EO condition the mean is 0.66 and the SD 0.21. The t-value comparing EC vs. EO is 14.406 (df = 178), yielding very strong evidence of a difference between the two conditions ($p < 0.0001$). Using the SAS software package and the code reported in the Appendix, they fit the logistic regression model where EC is treated as "success" and EO as "failure." The resulting estimated logistic model (2.0) is

$$\text{LOGIT} = -10.92 + 12.1(x)$$

and the estimated ASE for the β parameter is 1.72. The model is shown in Figure 3. The X-axis reports alpha normalized and log-transformed alpha power at O2 and the Y-axis reports probability of belonging to the EC group. Since β is positive, the sigmoid curve ascends (i.e., the greater the alpha power at O2, the greater the probability that the recording is EC). Observed data (N = 180), the old EEGs in the archive, are shown as blue squares along the 0 (EO) and the 1 (EC) horizontal lines. Note the shift in the means of the two distributions and how logistic regression models the shift. The Median Effective Level (EL_{50}) is given by $-\alpha/\beta$. For this model $EL_{50} = -(-10.92)/12.1 = 0.903$. This is the x-value for which the probability of success and failure is equivalent (50%) (i.e., for such a value of alpha power the model has no discriminant power). The software also reports (variance-covariance matrix of parameters) $Var(\alpha) = 2.52$, $Var(\alpha) = 2.98$, and $Cov(\alpha, \beta) = -2.70$. The researcher is now ready to implement the discriminant function. Suppose the first of the EEG files in the old archive has mean occipital power at O2 = 0.5. This value is represented in Figure 3 as a red bubble. The estimated probability of the file to be an EC recording is given by equation (3.0)

$$\pi_1 = e^{\alpha + \beta x}/(1 + e^{\alpha + \beta x}) \tag{3.0}$$

from which we obtain

$$\pi_{EC} = e^{(-10.92 + (12.1(0.5)))}/(1 + e^{(-10.92 + (12.1(0.5)))}) = 0.0076$$

The probability of the file being an EO recording is the complement of this quantity (i.e., $1 - 0.0076 = 0.992$). The file appears to be an EO recording; however, confidence intervals are required in order to reach a conclusion in probabilistic terms. A few more steps are required for this purpose. The reported ASE of the estimated LOGIT is given by (5.0)

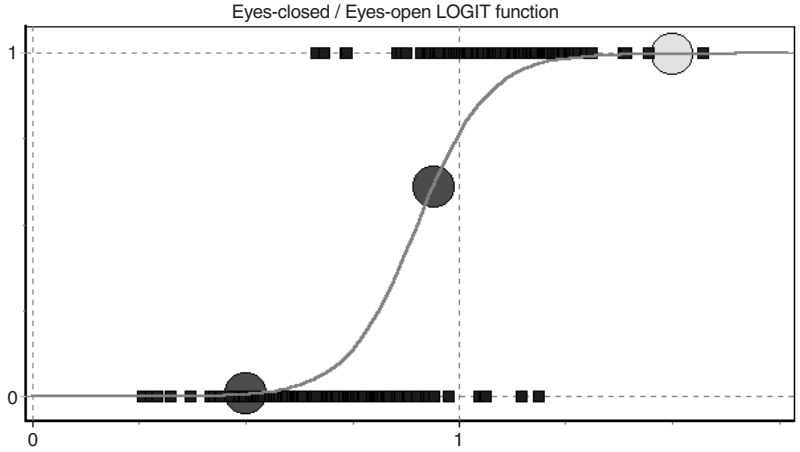
$$\text{ASE} = \sqrt{[Var(\alpha + \beta x)]} = \sqrt{[Var(\alpha) + x^2\, Var(\beta) + 2xCov(\alpha, \beta)]} \tag{5.0}$$

For these data ASE = $\sqrt{[2.52 + ((0.5)^2\ (2.98) + (2(0.5)(-2.70))]} = \sqrt{0.558} = 0.747$

The 95% confidence intervals for the LOGIT are given by

$$\text{CILOGIT} = (\alpha + \beta x) \pm [(1.96)\text{ASE}] \tag{6.0}$$

FIGURE 3. *Illustration of the discriminative process performed by a Logistic Regression discriminant function*. The gray curve describes the Logit model for the simulated data in the "working example" (see text for details). N = 180 (90 eyes-open (EO) and 90 eyes-closed (EC) recordings). Observed data is represented by blue squares (1 = EC, 0 = EO). The y-axis represents the probability of success. In the example, this is the probability that the EEG was recorded in the eyes-closed condition. The x-axis represents a continuous predictor. In the example, this is the Alpha (8-13 Hz) power at occipital site O2. The depicted model is LOGIT = −10.92 + 12.1(*x*). The x-value of colored bubbles represents alpha power values of EEG recordings entering the discriminant function. The y-value is their estimated probability of belonging to the EC condition.



or, solving,

$$\begin{aligned} \text{CILOGIT} &= (-10.92 + ((12.1)(0.5)) \pm [(1.96)(0.747)] \\ &= -4.87 \pm 1.4645 = [-6.334,\ -3.405] \end{aligned}$$

Finally, substituting this interval for the exponents $(\alpha + \beta x)$ in (3.0) the researcher obtains the confidence interval for the probability that the file is EC by means of equation 7.0

$$\text{CI}_{EC} = e^{(\text{CILOGIT})}/[1 + e^{(\text{CILOGIT})}] \qquad (7.0)$$

$$\text{LOWER CI}_{EC} = e^{(-6.334)}/[1 + e^{(-6.334)}] = 0.0018$$

$$\text{UPPER CI}_{EC} = e^{(-3.405)}/[1 + e^{(-3.405)}] = 0.032$$

Confidence Interval for EO is the specular complement = [1 − 0.032, 1 − 0.0018] = [0.968, 0.9982]. Since the confidence interval does not include 0.5, the researcher concludes that the file is an eyes-open recording.

For the second file, alpha power at O2 is 0.94. This value is represented in Figure 3 as a green bubble. The researcher again uses equations (3.0), (5.0), (6.0) and (7.0) as shown above. The point estimate probability that the file is EC is 0.611 and the CI is [0.48, 0.72]. In this case the CI includes 0.5 and does not allow any classification. A third file with alpha power at O2 = 1.5 is represented in Figure 3 as a yellow bubble. The probability that the file is EC is 0.999 and the CI is [0.9943, 0.9999]. The CI does not include 0.5, hence the researcher classifies it as EC. This example was conceived to show that, once the model has been fitted with a software package like SAS, and relevant quantities for a given discriminant function have been found, the computations of the point estimates and CI can be easily obtained with a computer algorithm. The estimates in this example (equations [3.0], [5.0], [6.0], and [7.0]) have been computed by a computer program written in Delphi Pascal. The program is available upon request to the author.

DISCUSSION

Logistic regression discriminant functions are a flexible tool for the evaluation process. The large amount of information provided with a quantitative electroencephalographic examination can be conveniently reduced by the careful selection of non-redundant descriptors possessing sufficient discriminative power. Typically, discriminant functions can be evoked while comparing the patient to EEG norms. The additional information provided by the discriminant function may immediately focus attention to a specific clinical category. The logistic regression model does not require normal distribution of the predictors and does not impose limitations in the choice of predictors. Predictors may be continuous or discrete variables, and of any combination. Furthermore, a discriminant function based on logistic regression is not limited to a binary response (category). The same model may be used to predict the membership in several categories, or in several degrees within the same category (ordinal response variables). For example, Thatcher et al. (2001) define a mild and a severe clinical condition of traumatic brain injury. Finally, the implementation of the technique is highly facilitated by modern statistical packages like SAS or LogXact and the actual pro-

cess of discriminating a new patient can be easily automated as shown above in the case of a single quantitative predictor.

Overall, when using logistic discriminant functions, it is important to realize the major limitation of all discriminant functions. No discriminant function is completely accurate. Both false positives and false negatives are unavoidable. Hence, the function's decision requires independent validation. Non-conclusive outcomes exist as well. Despite the progress in computerized medical technology, to date no expert system seems capable of fully substituting for human thought.

REFERENCES

Agresti, A. (1990). *Categorical data analysis*. New York: John Wiley & Sons.

Agresti, A. (1996). *An introduction to categorical data analysis*. New York: John Wiley & Sons.

Arango, C., Calcedo Barba, A., Gonzalez-Salvador, T., & Calcedo Ordonez, A. (1999). Violence in inpatients with schizophrenia: A prospective study. *Schizophrenia Bulletin*, *25*, 493-503.

Congedo, M. F., & Lubar, J. F. (2003). Parametric and non-parametric normative database comparisons in electroencephalography: A simulation study on accuracy. *Journal of Neurotherapy*, 7 (3/4), in press.

Delorme, O., Bourdin, H., Viel, J. F., Simon Rigaud, M. L., & Kantelip, J. P. (1992). Spectral analysis of electroencephalography data in athletes with low erythrocyte magnesium. *Magnesium Research*, *5*, 261-264.

DeLurgio, S. A. (1998). *Forecasting principle and applications*. New York: McGraw-Hill.

de Tommaso, M., Sciruicchio, V., Bellotti, R., Guido, M, Sasanelli, G., Specchio, L. M., et al. (1999). Photic driving response in primary headache: Diagnostic value tested by discriminant analysis and artificial neural network classifier. *Italian Journal of Neurological Science*, *20*, 23-28.

Fernandez-Bouzas, A., Harmony, T., Bosch, J., Aubert, E., Fernandez, T., Valdes, P., et al. (1999). Sources of abnormal EEG activity in the presence of brain lesions. *Clinical Electroencephalography*, *30*, 46-52.

Hosmer, D. W., & Lemeshow, S. (1989). *Applied logistic regression*. New York: John Wiley & Sons.

Houck, P. R., Reynolds, C. F., 3rd, Mazumdar, S., & Kupfer, D. J. (1991). Receiver operating characteristic analysis for validating EEG sleep discrimination of elderly depressed and demented patients. *Journal of Geriatric Psychiatry and Neurology*, *4*, 30-33.

Hughes, J. R., & John, E. R. (1999). Conventional and quantitative electroencephalography in psychiatry. *Journal of Neuropsychiatry and Clinical Neuroscience*, *11*, 190-208.

John, E. R., Prichep, L. S., & Easton, P. (1987). Normative data banks and neurometrics. Basic concepts, method and results of norm constructions. In A. S. Gevins

& A. Remond (Eds.), *Method of analysis of brain electrical and magnetic signals. EEG handbook revised series. Vol. 1*. New York: Elsevier Science (Biomedical Division).

Knott, V., Mohr, E., Mahoney, C., & Ilivitsky, V. (2001). Quantitative electroencephalography in Alzheimer's disease: Comparison with a control group, population norms and mental status. *Journal of Psychiatry and Neuroscience*, *26*, 106-116.

Kutner, M. H., Nachtsheim, C. J., Wasserman, W., & Neter, J. (1996). Applied linear statistical models (4th Ed.). New York: McGraw-Hill.

LogXact Software, Cytel Software Inc., 675 Massachusetts Avenue, Cambridge, MA 02139.

Lubar, J. F. (1991). Discourse on the development of EEG diagnostics and biofeedback for attention-deficit/hyperactivity disorders. *Biofeedback and Self-Regulation*, *16*, 201-225.

Mehta, C. R., & Patel, N. R. (1995). Exact logistic regression: Theory and examples. *Statistics in Medicine*, *13*, 2143-2160.

Monastra, V. J., Lubar, J. F., & Linden, M. (2001). The development of a quantitative electroencephalographic scanning process for attention deficit-hyperactivity disorder: Reliability and validity studies. *Neuropsychology*, *15*, 136-144.

Nunez, P. L., Wingeier, B. M., & Silberstein, R. B. (2001). Spatial-temporal structures of human alpha rhythms: Theory, micro current sources, multiscale measurements, and global binding of local networks. *Human Brain Mapping*, *13*, 125-164.

Nuwer, M. R., & Hauser, H. M. (1994). Erroneous diagnosis using EEG discriminant analysis. *Neurology*, *44*, 1998-2000.

SAS Software, SAS Institute Inc., 100 SAS Campus Drive, Cary, NC 27513-2414, USA.

Stokes, M. E., Davis, C. S., & Koch, G. G. (2000). Categorical data analysis using the SAS system (2nd Ed.). Cary, NC: SAS Institute.

Thatcher, R. W., North, D. M., Curtin, R. T., Walker, R. A., Biver, C. J., Gomez, J. F., et al. (2001). An EEG severity index of traumatic brain injury. *Journal of Neuropsychiatry and Clinical Neuroscience*, *13* (1), 77-87.

Thornton, K. E. (1999). Exploratory investigation into mild brain injury and discriminant analysis with high frequency bands. *Brain Injury*, *13*, 477-488.

RECEIVED: 06/20/02
REVISED: 09/25/02
ACCEPTED: 10/08/02

APPENDIX

The following SAS (SAS Institute, Inc., version 8.2 or above) code fits a logistic regression model, checks the model's goodness-of-fit, computes odds ratios, parameter estimates along with their Asymptotic Standard Error (ASE), and variance-covariance matrix of parameters. For a comprehensive description of SAS "Logistic" procedure, see Stokes et al. (2000). This code loads data from the text file (ASCII format) as specified in the "infile" procedure. This file will contain N lines, where N is the number of data-points. For each line there will be predictor values (called "VAR 1") followed by the indicator of the group membership (called "Group"). Indicators may be "1" for success and "0" for failure. An example data file is

```
0.23    0
0.65    0
……
0.78    1
0.97    1
…….
```

Example Code:

```
TITLE1'Example';
Data A;
  infile'C:\sasData.txt';
  input VAR1 Group;
Run;

Proc Logistic data=A order=data;
  model group= VAR1/lackfit CLODDS=both CLPARM=both COVB EXPB;
run;
```