Brit. J. Psychiat. (1966), 112, 1165–1171

The Establishment of a Psychiatric Syndrome

By P. A. P. MORAN

Recently several papers have appeared which apply the techniques of multivariate analysis in attempts to show that patients suffering from depression can be divided into two groups, often called "endogenous" and "neurotic", which tend to have different groups of symptoms and different outcomes with E.C.T. or other treatment. The purpose of the present paper is to look at this problem from the point of view of a statistician in an attempt to clarify the logic involved in the various procedures.

The two papers which will be used as a basis for the discussion are those by Kiloh and Garside (1963), and by Carney, Roth, and Garside (1965).

Kiloh and Garside studied 143 patients suffering from depression and recorded whether they did or did not show each of a set of 35 symptoms. Scoring these as 0, 1, they calculated the product-moment correlation between these scores and carried out a factor analysis, extracting two factors. The first factor appeared to be descriptive of "depression" as a whole, and the second factor was closely related to "diagnosis", i.e. the judgment by the psychiatrist as to whether the depression was "endogenous" or "neurotic". This diagnosis was considered by the psychiatrists involved to be reasonably certain in 92 of the cases.

More recently, Carney, Roth and Garside have carried out a more detailed analysis on 129 patients; these were again examined for 35 symptoms (similar but not identical with those of Kiloh and Garside) which were also scored 0 and 1 (except for "guilt" which was scored 0, 1, and 2). A principal component analysis was carried out on the resulting correlation matrix, and the three largest orthogonal components extracted. These were then compared with the diagnosis ("neurotic" or "endogenous") made by the psychiatrists concerned, and with the outcome, judged at 6 months, of treatment (mainly E.C.T.). The largest component was very closely correlated with the subjective diagnosis, whilst the second component appeared to be one measuring general depression.

The most striking feature of this analysis appeared when a frequency distribution was made of the scores resulting from a multiple regression analysis of diagnosis on 18 symptoms, selected as being most important, after a multiple regression analysis of diagnosis on the whole 35 symptoms had been done. It was then found that the resulting frequency distribution was strongly bimodal, from which it was concluded that the patients could be divided into two distinct groups.

This type of statistical investigation raises problems which lie somewhat outside the scope of present day statistical theory, for what is being attempted here (except for the prediction of E.C.T. success) is to construct a discriminant function on the evidence provided by the structure of a single sample and not on two samples known beforehand to come from two different populations. The logic of the procedure is therefore different and deserves careful scrutiny.

In fact procedures of this kind can be used for three quite different purposes:

(1) To provide evidence that the patients belong to two different groups tending to show different collections of symptoms (this might best be described as a problem of Internal Discrimination or Grouping);

(2) To construct a discriminating function, obtained by weighting symptom scores, which is as closely descriptive of the subjective judgment of the psychiatrists as possible;

(3) To construct a numerical function of the symptom scores which will predict the outcome of various types of treatment as closely as possible.

To achieve these aims requires different types of analysis and logic.

Internal Discrimination into Two Groups

Consider first what might happen if a frequency distribution is made of the results of measuring some continuously variable quantity on a large number of randomly selected individuals from some population. If such a frequency distribution was not unimodal but showed two (or more) peaks corresponding to well separated groups it would be reasonable to deduce that in the population studied there is more than one underlying group of individuals. On the other hand, if the distribution was unimodal, only one group would be postulated. If the measurement was the result of some medical procedure on a patient it might or might not be relevant to the particular disease being investigated, but if there are other reasons for believing that it is, the existence of bimodality strongly suggests that there are two different groups involved, with possibly different aetiologies.

In psychiatric diagnosis the problem is much more complicated because of two further difficulties: (a) the symptoms are often of an "either-or" type (which can be scored as 0 or 1), or of such a kind that they are not measured on a continuous scale by an instrument but are the result of applying a "scoring system" consisting of a small number of discrete values such as 0, 1, 2, 3, 4 (in fact in the two above quoted investigations the symptoms were, with one exception, all represented by the scores 0 and 1, which greatly simplifies the representation of a set of symptoms in a binary computer); (b) instead of a single measurement a whole group of symptoms have to be used. If there are k of these and each is scored with the value X_i ($i=1, \ldots, k$) then to demonstrate that there are two underlying groups, we want to find constants a_i such that the "linear discriminator" $T=\Sigma a_i X_i$ has a frequency distribution which is bimodal. It is clear that a single either-or symptom could not be used to provide internal evidence that the observed group of patients consists of two different groups, with different aetiologies or different diseases.

The ordinary theory of multivariate analysis (and in particular the use of factor, component and discriminatory analysis) assumes that the random variables involved, or at any rate some of them, are distributed in normal or multivariate normal distributions. This assumption is partly made in order to justify the methods of estimation used, and partly to enable significance tests to be easily carried out. All the analyses with which we are here concerned are, however, done on variables which take only a discrete set of values. It is therefore rather difficult to justify, in a strictly rigorous manner, some of the mathematical procedures. However, we can reasonably expect that large multivariate samples, each of many 0–1 variables, will behave in approximately the same manner as they would if they were normally distributed. This would therefore seem to be one of the least of the difficulties involved in this sort of work. We also have to remember that "bimodality" has a rather different meaning in discrete distributions than in continuous ones. In a sense a random variable which can only take two values is "bimodal" but what we want here is to find frequency distributions in which the values tend to be concentrated at two points which are well separated by other possible values at which the frequencies are significantly smaller.

In "Internal Discrimination" the problems raised by the fact that we are dealing with a relatively large number of symptom scores are much more difficult. These problems have already been encountered in other branches of science. In particular Dr. A. J. Fabens of Boston College, Massachusetts, has studied these problems in connection with the classification into groups, from internal evidence alone (e.g. measurements of length and breadth), of sets of objects such as Polynesian adze-heads. I am indebted to Dr. Fabens for access to his as yet unpublished paper and the following discussion of internal discrimination follows substantially similar lines to his.

It is simplest to illustrate the problems involved by considering a situation in which each individual is subject to only two measurements, X_1 and X_2, both of which are on a continuous scale.

If either of these have frequency distributions which are markedly bimodal we can legitimately conclude that the group of objects studied is basically composed of two different groups. Even if this is so, however, a better discrimination between these groups can often be obtained by using both measurements together in a discriminating function of the form

$$T = a_1 X_1 + a_2 X_2,$$

where the constants a_1, a_2 are chosen to maximize the discriminating power.

However, it is easy to construct examples in which either measurement by itself has a unimodal distribution, but T is bimodal and discriminates between two sharply distinguished groups. This can be illustrated by a diagram (Fig. 1).

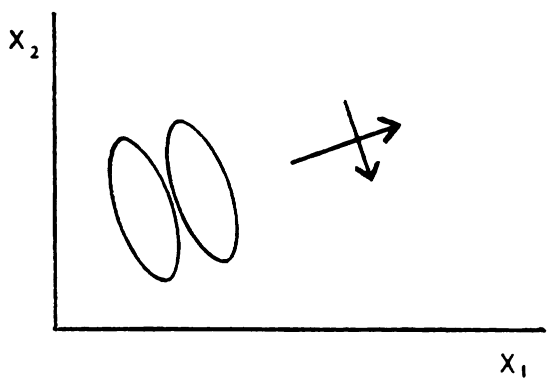

FIG. 1.

X_1 and X_2 are the two measurements. The population is supposed to consist of individuals roughly uniformly scattered inside the two ellipses (or alternatively it may be supposed that these represent levels of equal probability density in the components of a mixture of two bivariate normal distributions each of which has a strong negative correlation). Then either X_1 or X_2 is, by itself, useless as a discriminator between the two groups, However, taken together in the form of a linear discriminant function the separation is highly accurate.

The Practical Problem of Internal Discrimination

In the case of two continuous measurements the practical problem is relatively easy, since we can make a scatter diagram which represents all the information and we can then judge by eye whether the points fall into two (or more) well separated groups, or at any rate the direction of an optimum discriminator.

With more than two measurements some other approach has to be used, and we need some method of describing, as a whole, the position of the points in space. Suppose we have N symptoms (35, although not the same 35, in the two above cases) and n patients (143 in the Kiloh and Garside paper, 129 in the Carney, Roth and Garside paper). It is convenient to adopt the usual mathematical description and say that this sample can be represented by a cluster of n points in N-dimensional space. It is then natural to try to describe the shape of this cluster by calculating the variances and co-variances of the 35 symptoms with each other. These are, of course, the crude variances and co-variances of scores which are 0 or 1 (except for one symptom in the second study which was scored 0, 1, 2).

An attempt can now be made to describe the general position of the n points either by using factor analysis (Kiloh and Garside) or component analysis (Carney, Roth and Garside). Factor analysis is a subject of controversy amongst statisticians, and its correct application requires the truth of assumptions underlying a certain type of statistical model. It therefore seems better and more natural to use component analysis. This simply amounts to rotating the co-ordinate axes in the N-dimensional space so that the new axes coincide with the principal axes of the ellipsoid defined by the correlation matrix. Both these approaches involve a degree of arbitrariness and in particular they are here applied to the correlation matrix and not to the variance-covariance matrix. This is not unreasonable if we are not particularly interested in taking account of the fact that different symptoms will have different variances. Standard computer programmes exist for doing such calculations.

Having carried out a component analysis we may examine the quantities defined by the largest factors, i.e. the variations along the axes with the largest variances.

Using factor analysis, Kiloh and Garside

https://doi.org/10.1192/bjp.112.492.1165 Published online by Cambridge University Press

picked out two factors. The first was identified by them with the intensity of the depression as a whole, and the bi-polar second factor was identified with the contrast between endogenous and neurotic depression because the factor loadings were extremely similar to the product-moment correlations with the diagnoses, into these two classes, by the psychiatrists concerned. They did not construct a frequency distribution for the values of this second factor, which they might have done with a view to seeing if it was bimodal.

Carney, Roth and Garside carried out a component analysis on the correlation matrix and picked out the three largest components. The largest was closely related to the psychiatrists' diagnosis into the two groups, the second largest to general level of depression, and the third to a "cluster of psychotic features found in a proportion of depressed patients". The largest was not used as a diagnostic score, nor was a frequency distribution of its values constructed. Instead, a diagnostic index was constructed from the results of a separate multiple regression analysis on 18 of the 35 symptoms. These 18 symptoms had been picked out of multiple regression analyses (here really discriminant analyses) of diagnosis and E.C.T. outcome on the symptoms. The index used weights which were rounded values of the regression coefficients.

A frequency distribution of these diagnostic indices was strikingly bimodal (Carney, Roth and Garside, Table VIII and Fig. 1), thus providing very strong evidence that their group of 129 patients consisted of two separate groups. It is worth emphasizing that the evidence for such heterogeneity rests on the bimodality of the frequency distribution of the score and not on the manner in which the latter has been found.

The frequency distribution does not look as if it could be fitted very closely by a "mixture" of two normal distributions. One would expect, however, that for a sufficiently large sample this would be possible, and a threshold value of the diagnostic score could then be accurately calculated using as criterion the minimization of the total probability of misclassification (this value would be about 5–6 in this data). It might be mentioned here that the problem of fitting mixtures of normal distributions to data of this kind is very complicated. A fairly satisfactory method, which is not difficult on a computer, is given by Rao (1948). However, the problem of testing for bimodality has not really been satisfactorily solved (see Haldane (1951–1952)).

Having obtained a discriminator in any of the ways considered in this paper, the heterogeneity of depression could be further verified simply by calculating the scores for some other large group of unselected patients and seeing if bimodality again occurred. It would be easy to do this with the scores given by Carney, Roth and Garside in their Tables VIII and X. Now that they have found a scoring system which appears to result in a bimodal frequency distribution, it is highly desirable that further evidence of a real division of "depression" into two syndromes should be obtained, and the ascertainment of the values of such scores could well become a routine practice requiring no elaborate calculation. It would also be instructive to compare and contrast the features which are most important in this method with those considered most important in subjective diagnosis by the psychiatrist.

Internal discrimination, as we have considered it so far, requires that we look for a discriminating function of the form $T=\Sigma a_i X_i$ whose frequency distribution shows strong evidence of bimodality. Although Carney, Roth and Garside did in the end construct such a discriminator from further evidence (the multiple regression analyses), they would almost certainly have succeeded in doing so on purely internal evidence if they had used the largest factor of their component analysis, because they showed that this was strongly correlated both with the subjective diagnosis and with the results of E.C.T.

It is therefore important that we consider why and in what circumstances component analysis alone may enable us to find such a discriminator if the data really consist of two distinct groups.

The Theoretical Basis of Internal Discrimination

The basic theoretical problem is to find a discriminating function, $T=\Sigma a_i X_i$, whose frequency distribution will demonstrate the

https://doi.org/10.1192/bjp.112.492.1165 Published online by Cambridge University Press

existence of two different groups, if such groups exist. No mathematical theory for the construction of such a function exists and we are forced into using hit or miss methods. Since the general position of the n points in the N-dimensional space can be more or less described by the variance-covariance matrix we start from this (or the correlation matrix) and use the method of component analysis to see how the points are distributed along the directions of the principal axis.

If heterogeneity is present, these components may or may not provide a suitable discriminator, and the latter may be directed along the largest axes or one of the others. To illustrate this we again consider the case of two variables, each varying continuously (similar comments will apply to a reasonably large number of discrete variables, i.e. to discrete variables in a space of a reasonably large number of dimensions). Thus in Fig. 1, where the ellipses are meant to represent regions in which most of the points are more or less uniformly distributed, the variation along the minor axis will provide good discrimination, whilst that along the major axis will provide no discrimination at all.

In Fig. 2 on the other hand we will get good discrimination along the major axis (the line joining the centres of the circles) given by component analysis, but no discrimination at all along the the minor axis.

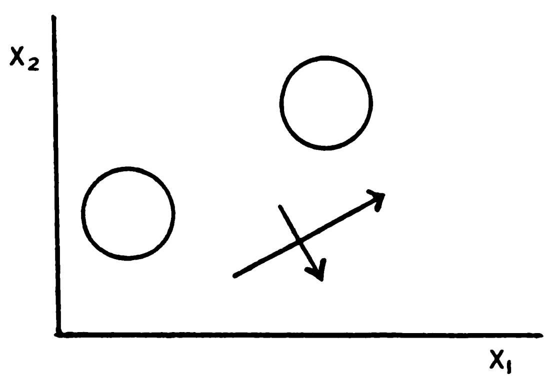

FIG. 2.

Fig. 3 demonstrates a case in which we have concentrations in two ellipses whose axes are not in the directions of the axes given by component analysis, the latter being shown by the arrows. In this case there is a good discriminator, but it will not be obtained by component analysis, nor by factor analysis; these are therefore empirical methods which may or may not work even when heterogeneity is present.

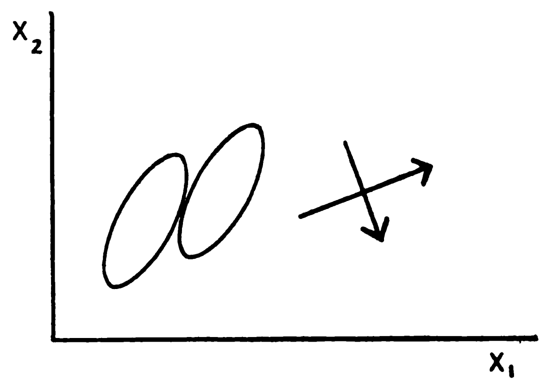

FIG. 3.

One might at first think it would be possible to programme a computer to explore systematically discriminating functions in a large number of directions more or less uniformly distributed over all possible directions. A little computation soon shows that this is impossible. Even if the weights were restricted to the values ± 1, 35 symptoms would require the testing of $2^{34} = 1 \cdot 718 \times 10^{10}$ possible discriminators, which would be beyond the capacity of any computer. Thus in this problem some empirical approach is necessary.

There is, however, one other possibility which has not so far been exploited. Mania is considered to occur in patients who are often at other times depressed and the depression they suffer is believed by many psychiatrists to be "endogenous". If, therefore, we could collect a large enough group of depressive patients who are known to have had mania at some other time we would have a sample of patients whose distribution of symptoms should be only that of the endogenous component of the mixture of distributions in an unselected group. We could therefore construct a discriminant function in the usual manner between these two groups. Even though the second group is a mixture of both the endogenous and neurotic subgroups, this discriminant function should be, apart from sampling errors, identical with the desired discriminator between the two subgroups. The selection of a sufficiently large number of manic-depressives might, however, be difficult.

Another problem which requires further consideration is the choice of the symptoms used and their number. Even with a small number of symptoms it may or may not pay to increase their number. For example suppose a symptom is added which is neither correlated with the other symptoms nor of any discriminatory value. Since the analysis has been done on the correlation matrix, and not on the variance-covariance matrix the addition of this symptom should not in general, affect the estimates of the largest components provided the sample size is large. However, the addition of a number of such symptoms when a small sample is used may weaken the efficiency of estimation of the main components.

On the other hand, if a very small number of symptoms are used we may get into trouble as a result of their discreteness making it difficult to show bimodality in the distribution of T. The choice of between 10 and 20 seems about right, and there is a good deal of room for futher experimentation in their choice.

It must also be pointed out that the above remarks on the choice of symptoms and their number applies only to the problem of internal discrimination and not to the problems involved in using regression and discriminant analysis considered in the next section.

Regression and Discriminant Analysis

Carney, Roth, and Garside also carried out regression analyses using as regressors the psychiatric diagnosis (scored as 0, 1) and scores based on the outcome after E.C.T. (combined in most cases with drug therapy) at 3 and 6 months. The exact nature of the score used to grade the latter is not quite clear.

Although in fact the result of a regression analysis of psychiatric diagnosis on 18 symptoms was also used, as described above, for the purposes of internal discrimination by demonstrating bimodality in the frequency distribution of the predictors, the purpose and logic of this type of analysis is different from that used above.

In the case of the regression using E.C.T. outcome, the purpose is to construct a numerical predictor, based on observed symptoms, which would make possible the choice of those patients who would be likely to do well with E.C.T. The ordinary theory of regression or of discriminatory analysis is therefore applicable in spite of the approximations due to both regressor and regressands being discrete. An interesting feature of this analysis is that a better prediction of E.C.T. outcome can apparently be made directly from a score based on the symptoms than from the psychiatric diagnosis.

A further mathematical problem deserves mention. We sometimes wish to carry out a regression analysis of a regressor, Y, on a set of regressands $X_1, \ldots, X_n$. To do this we calculate all the $n+1+\frac{1}{2}n(n+1)$ variances and covariances. We can then straightforwardly calculate the regression of Y on $X_1, \ldots, X_n$. Out of this data we may, however, wish to find a regression of Y on $m(<n)$ X's, and to choose these m X's to have the greatest multiple correlation with Y. We can choose m X's out of n in $\binom{n}{m}$ ways. When m and n are not extremely small, however, $\binom{n}{m}$ is a large number. Thus, for example, in the case of Carney, Roth, and Garside, 18 regressands out of 35 can be chosen in $\binom{35}{18}$ ($=4 \cdot 538 \times 10^9$ approximately) ways. It is clearly not practicable to programme a computer to calculate all these and pick out the best. No mathematical theory yet exists which provides a systematic method of picking such optimum regressands. It might at first sight be thought that this could be done by picking out the regressand which had the highest correlation with Y, finding the partial correlations of Y with the X_i with this regressand removed, picking out the next regressand with the highest partial correlation with Y, and similarly carrying on until the required number had been obtained. Simple examples can be constructed to show that this method is not in general optimal (this fact was pointed out to me by Mr. G. A. McIntyre).

It would also be desirable to consider whether the predictor of E.C.T. effectiveness is significantly different (in the statistical sense) from the optimum discriminator between the two syndromes.

It may also easily happen that in problems of this kind there are more than two underlying groups. It may then become much more difficult to apply such methods as the above. The problem of multiple groups also becomes important if numerical diagnosis is to be extended to all mental diseases. The natural procedure might be as follows. One might make a subjective classification of a large number of patients into psychotics and neurotics and then do a conventional discriminant analysis between these two groups. If the frequency distribution of the resulting discriminator is strongly bimodal one may then conclude that there is a genuine grouping into two groups. One would try to find internal discriminators within each of these groups. In this way we could hope that from the systematic investigation of 30 to 50 symptoms, their conversion into scores, and the calculation of the numerical values of perhaps half a dozen discriminators, one might be able to do numerical diagnoses. Such diagnoses would be highly correlated with the subjective diagnoses of psychiatrists but perhaps have the advantage of being relatively more consistent and repeatable. In this way one might hope to map all mental diseases in a space of a small number of dimensions. After this, one could attempt to calculate the probabilities of success with various forms of treatment as a function of the position of the point, representing the patient, in this projection of the whole symptom space.

Summary

(1) Recent researches are reviewed which involve the use of various kinds of numerical discriminators based on scores derived from symptoms.

(2) If it is desired to establish the existence of two underlying syndromes in a group of patients, we need to find such a discriminator which is such that it has a strongly bimodal frequency distribution. One may attempt to find such a discriminator by means of component analysis, by using discriminant analysis on two subjectively chosen groups, or by regression analysis of the effectiveness of treatment on the symptoms.

(3) The logic and some of the mathematical difficulties of such procedures are discussed, and also the possibility of using such methods to establish a complete geography of mental disease in a space of a relatively small number of dimensions.

Acknowledgments

I am much indebted to Dr. Garside and Professors Kiloh and Roth for some very helpful criticism.

References

Carney, M. W. P., Roth, M., and Garside, R. F. (1965). *Brit. J. Psychiat.*, **111**, 659.

Haldane, J. B. S. (1951–52). *Ann. Eugenics*, **16**, 359.

Kiloh, L. G., and Garside, R. F. (1963). *Brit. J. Psychiat.*, **109**, 451.

Rao, C. R. (1948). *Jour. roy. Statist. Soc. B.*, **10**, 159.

P. A. P. Moran, *Professor of Statistics, Institute of Advanced Studies, Australian National University, Canberra, A.C.T., Australia*

(*Received 28 February, 1966*)