BIOMETRICS 28, 125–36
March 1972

PLOTS OF HIGH-DIMENSIONAL DATA

D. F. ANDREWS[1]

Bell Telephone Laboratories, Murray Hill, New Jersey, and Princeton University

SUMMARY

A method of plotting data of more than two dimensions is proposed. Each data point, $\mathbf{x} = (x_1, \cdots, x_k)$, is mapped into a function of the form

$$f_{\mathbf{x}}(t) = x_1/\sqrt{2} + x_2 \sin t + x_3 \cos t + x_4 \sin 2t + x_5 \cos 2t + \cdots,$$

and the function is plotted on the range $-\pi < t < \pi$.

Some statistical properties of the method are explored. The application of the method is illustrated with an example from anthropology.

1. INTRODUCTION

Plotting has long been one of the most useful tools in the analysis of data. Much of model building is aided by the plotting of residuals. Distributional assumptions are frequently based on probability plots.

Unfortunately, most graphical methods are restricted to displaying univariate or bivariate data. Some progress has been made in the use of different symbols on a two-dimensional plot to give some idea of a third or fourth dimension, but these methods lack precision and seem limited to a small number of dimensions.

However, one is accustomed to examining plots of functions, and they may be infinite-dimensional. This suggests imbedding high dimensional data in a higher dimensional but easily visualized space of functions, and then plotting the functions. One way of doing this is described here, and some examples are given.

2. THE PLOTTING PROCEDURE

The proposed plotting procedure is very simple. If the data is k-dimensional each point $\mathbf{x}' = (x_1, \cdots, x_k)$ defines a function

$$f_{\mathbf{x}}(t) = x_1/\sqrt{2} + x_2 \sin t + x_3 \cos t + x_4 \sin 2t + x_5 \cos 2t + \cdots.$$

This function is then plotted over the range $-\pi < t < \pi$. A set of points will appear as a set of lines drawn across the plot. The programming effort

[1] Present address: Department of Mathematics, University of Toronto, Toronto 181, Canada.

required for this plot is small, but an output device with relatively high precision such as a CALCOMP or microfilm plotter is required.

3. A BIOLOGICAL EXAMPLE

Ashton *et al.* [1957] used graphical techniques when comparing measurements on the teeth of fossils with those of different 'races' of men and apes. From the data, group means were calculated for men and apes and canonical variables determined to maximize the between sum of squares relative to the within sum of squares. The first two such variates had the largest eigenvalues and these variables were used in the plotting procedure. The group means were plotted and 90% confidence contours were drawn about them. The fossil values were then plotted and assessed relative to the different groups.

Ashton *et al.* analyzed several teeth. Here we present the data for one tooth—the permanent first lower premolar (Table 1). In Figure 1 the group means are plotted together with approximate 90% confidence contours. The fossil measurements are also plotted. From this plot those authors concluded that *Proconsul africanus* is very like a chimpanzee while the other fossils are more like humans.

Some group means have some moderately large values of the third and fourth canonical variates, while large values occur for all 8 variates for the fossils. Plotting all the variates permits the examination of the effect of these large values.

In Figure 2 the group means of all the canonical variables have been plotted as functions. From this plot it can be seen that the humans and the apes form fairly distinct groups. Among the apes, the chimpanzees stand out from the gorillas and orang-outangs. Certain values of t seem of particular interest: at both t_2 and t_4 the function values for humans have a fairly precise value which is much different from the corresponding values for any ape. The point t_1 has an analogous property for chimpanzees, gorillas, and orang-outangs. The point t_3 corresponds to the widest separation among the group means.

In Figure 3 the fossils have been added. Immediately *Proconsul africanus* stands out as quite different from any group. For some values of t it is very like a chimpanzee, gorilla, orang-outang, or man and has certain characteristics in common with all of these! The remaining fossils share with man the characteristic human quality exhibited by t_2 and t_4 . However, for some other values of t they are quite different from men and are more like apes, particularly chimpanzees. This is not apparent from Figure 1. Fortunately it is possible to calibrate these observations with significance tests and confidence sets.

In sections 4 and 5, the variance of a function value var $[f(t)]$ is derived and is shown to be almost independent of t. The standard deviation σ_f is shown on the graph, as is the width of a 90% confidence band for $f_y(t_0)$. Applying these at the values t_2 or t_4 it can be seen that

TABLE 1

PERMANENT FIRST LOWER PREMOLAR
coefficients of canonical variates

	trans.-bas lab.	+10.2	+ 3.7	- 6.6	-18.1	- 6.7	+19.3	- 6.4	+ 9.9
	trans.-mid. lab.	- 1.4	+ 0.9	+ 0.1	+14.7	-16.4	- 0.8	+13.8	+19.7
	trans.-max. lab.	+24.6	- 3.8	+ 2.7	+24.2	+18.0	- 6.0	- 0.7	-30.5
	ht.-lab,	+ 0.8	+ 0.9	- 8.1	- 1.8	+ 5.0	- 9.6	- 5.9	+ 7.5
	ht.-prox. seg.	- 5.0	+ 2.4	+ 2.8	+ 4.6	+ 2.1	+ 8.9	- 2.8	0.0
	thick.-bas.	- 2.0	+14.0	- 5.3	- 8.4	+17.7	+ 3.3	+25.6	- 1.3
	thick.-mid.lab.	-13.1	+11.5	-19.9	+ 2.7	-10.3	- 1.2	- 7.6	-13.5
	thick.-max.	- 1.2	+ 0.6	+42.0	-10.7	-15.5	-15.5	-19.4	+ 7.7
	constant term	-31.78	-54.53	-16.55	-14.23	+10.98	+ 7.97	+ 7.76	+ 3.16
					means of groups				
A.	West African	-8.09	+0.49	+0.18	+0.75	-0.06	-0.04	+0.04	+0.03
B.	British	-9.37	-0.68	-0.44	-0.37	+0.37	+0.02	-0.01	+0.05
C.	Australian aboriginal	-8.87	+1.44	+0.36	-0.34	-0.29	-0.02	-0.01	-0.05
D.	gorilla: male	+6.28	+2.89	+0.43	-0.03	+0.10	-0.14	+0.07	+0.08
E.	female	+4.82	+1.52	+0.71	-0.06	+0.25	+0.15	-0.07	-0.10
F.	orang-outang: male	+5.11	+1.61	-0.72	+0.04	-0.17	+0.13	+0.03	+0.05
G.	female	+3.60	+0.28	-1.05	+0.01	-0.03	-0.11	-0.11	-0.08
H.	chimpanzee: male	+3.46	-3.37	+0.33	-0.32	-0.19	-0.04	+0.09	+0.09
I.	female	+3.05	-4.21	+0.17	+0.28	+0.04	+0.02	-0.06	-0.06
					fossils				
J.	Pithecanthropus pekinensis	-6.73	+3.63	+1.14	+2.11	-1.90	+0.24	+1.23	-0.55
K.		-5.90	+3.95	+0.89	+1.58	-1.56	+1.10	+1.53	+0.58
L.	Paranthropus robustus	-7.56	+6.34	+1.66	+0.10	-2.23	-1.01	+0.68	-0.23
M.	Paranthropus crassidens	-7.79	+4.33	+1.42	+0.01	-1.80	-0.25	+0.04	-0.87
N.	Meganthropus palaeojavanicus	-8.23	+5.03	+1.13	-0.02	-1.41	-0.13	-0.28	-0.13
O.	Proconsul africanus	+1.86	-4.28	-2.14	-1.73	+2.06	+1.80	+2.61	+2.48

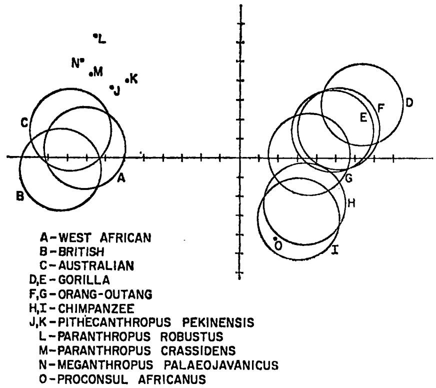

FIGURE 1

PERMANENT FIRST LOWER PREMOLAR GROUP MEANS, 90% CONFIDENCE CONTOURS AND FOSSIL VALUES

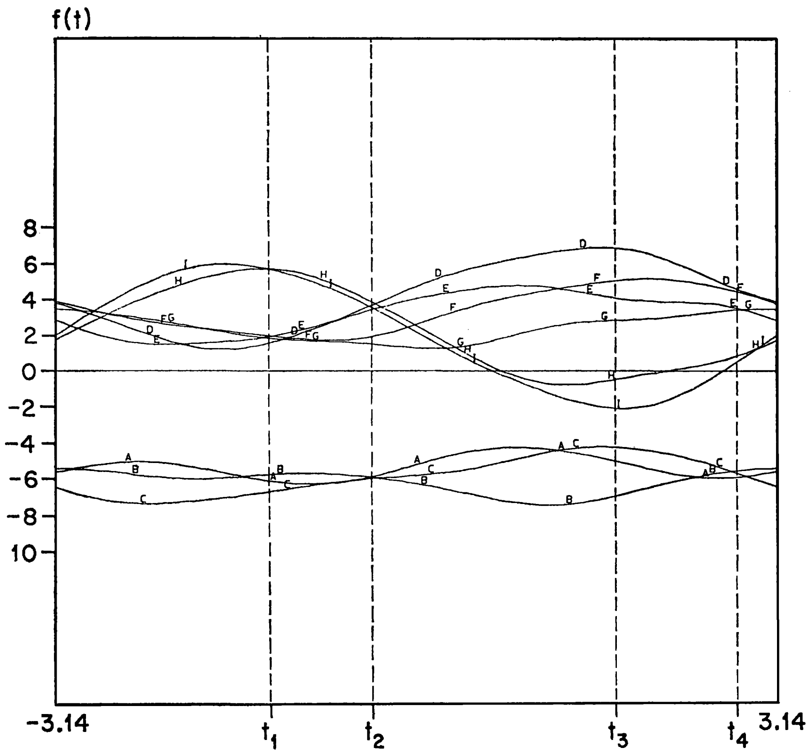

FIGURE 2

8-DIMENSIONAL DATA

PERMANENT FIRST LOWER PREMOLAR GROUP MEANS

A-WEST AFRICAN B-BRITISH C-AUSTRALIAN

D, E-GORILLA F, G-ORANG-OUTANG H, I-CHIMPANZEE

$$f(t) = x_1/\sqrt{2} + x_2 \sin(t) + x_3 \cos(t) + \cdots$$

(i) The means of all human 'races' lie well outside the 90% band for a particular test centered at *Proconsul africanus*, providing evidence against the hypothesis that this fossil is human.

(ii) The remaining fossils do not provide such evidence. Indeed, at these values, t_2 and t_4, the fossils lie within one standard deviation of each human.

In section 5 an overall test is constructed and the corresponding 90% confidence band is also shown on Figure 3. If a band of this width is centered at *Proconsul africanus* and all values of t are examined, then it can be seen that the functions corresponding to the means of *all* the known species studied fall outside this band for some value of t, thus providing evidence that this fossil does not belong to any known species. The same can be said of *Paranthropus robustus*. The remaining fossils may reasonably be associated with West African or Australian humans but less reasonably with the British. The points t_2 and t_4 yield interesting clusters of some fossils and humans. This aspect of the data certainly bears further study.

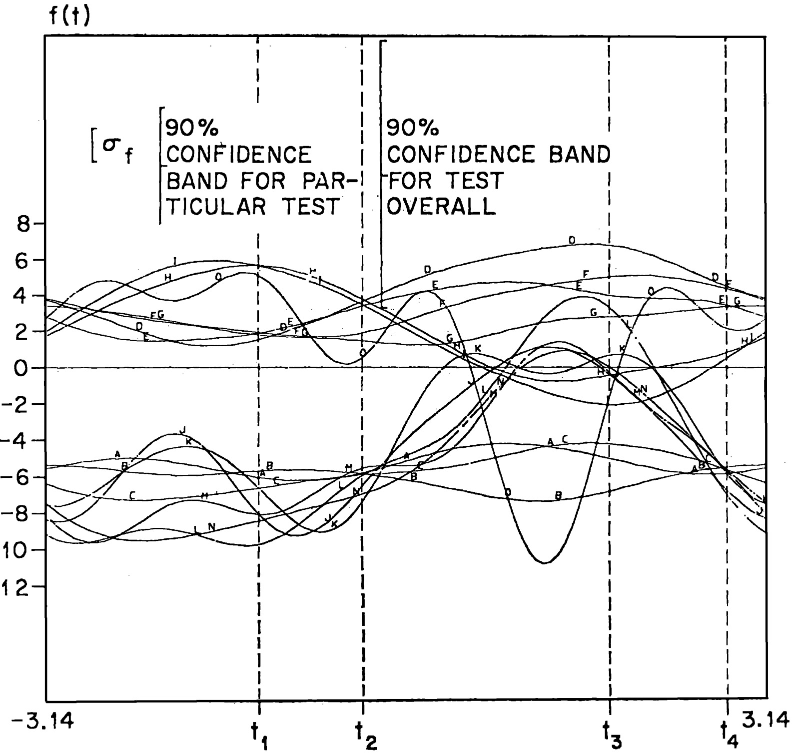

FIGURE 3

8-DIMENSIONAL DATA

AS IN FIGURE 2 BUT WITH FOSSILS ADDED

J, K- *Pithecanthropus pekinensis*
L-*Paranthropus robustus*
M- *Paranthropus crassidens*
N-*Meganthropus palaeojavanicus*
O- *Proconsul africanus*

Ashton *et al.* investigated several teeth. In Table 2 the largest canonical variates for 6 different teeth are recorded as given in their Tables 2, 3, 4, 5, 6, 8 for men, apes, and two fossils. The group means for humans and apes are plotted in Figure 4; again it is possible to identify a value t_1 at which the functions for humans form a tight cluster distinct from the function values for apes. In Figure 5 the fossils have been added; for the value of particular interest at t_1 the fossils have this human characteristic. Curiously *Pithecanthropus pekinensis* seems to lie between humans and apes: in general this fossil is closer to man than *Paranthropus crassidens*. Both fossils differ considerably from the apes, and the nature of this departure is in the direction of man.

The broad inferences drawn from these plots are similar to those of Ashton *et al.* However, the peculiarities of *Proconsul africanus* are more forcefully presented here. In addition, the precise 'human' value of certain linear combinations (t_i) has been detected from the plots. These aspects of the data warrant further study.

4. PROPERTIES OF THE PLOTS

The particular function proposed in section 2 has many useful properties relevant for the study of data. Some of these are outlined below.

TABLE 2

LARGEST CANONICAL VARIABLE FOR 6 TEETH

	Source	Tooth Type					
		1	2	3	4	5	6
A.	British	-5.35	-7.07	-9.37	-4.28	-2.15	-2.93
B.	Australian aboriginal	-3.93	-6.04	-8.87	-2.16	-0.50	-1.09
C.	gorilla male	3.12	6.66	6.28	4.96	4.13	4.60
D.	gorilla female	1.45	1.73	4.82	3.96	3.35	3.63
E.	orang-outang male	2.83	5.10	5.11	2.70	1.21	1.49
F.	orang-outang female	1.49	1.63	3.60	1.29	-0.17	0.05
G.	chimpanzee male	0.38	3.82	3.46	-1.65	-2.32	-1.92
H.	chimpanzee female	0.01	0.23	3.05	-2.25	-2.65	-2.15
I.	Paranthropus crassidens	-4.52	-6.49	-7.79	3.45	4.91	3.72
J.	Pithecanthropus pekinensis	-1.81	-2.94	-6.73	-0.36	-1.32	1.09

Tooth Types

1 - Permanent lower second incisor
2 - Permanent lower canine
3 - Permanent lower first molar
4 - Permanent upper second premolar
5 - Permanent lower first molar
6 - Permanent lower third molar

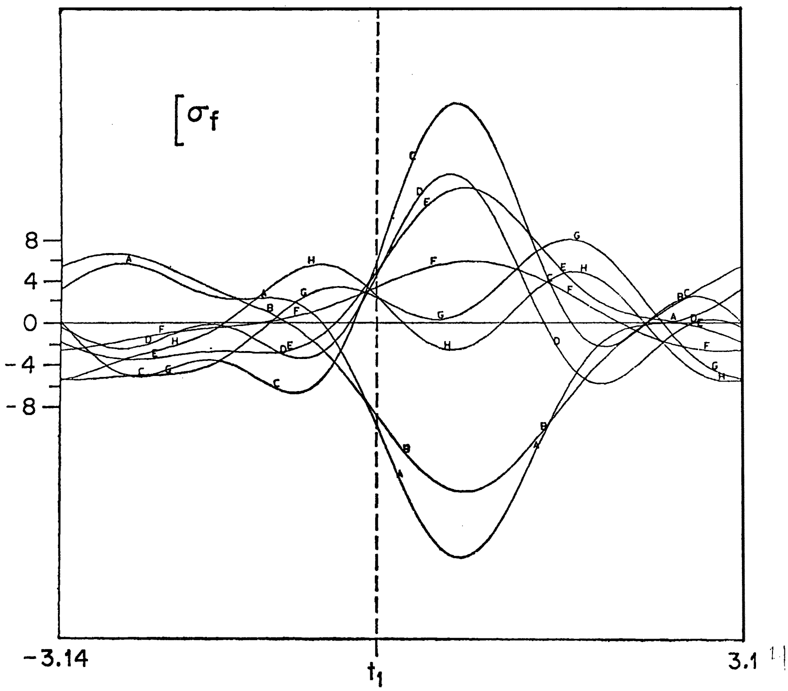

FIGURE 4

6-DIMENSIONAL DATA

THE FIRST CANONICAL VARIATES FOR 6 TEETH

GROUP MEANS

A-BRITSH B-AUSTRALIAN C, D-GORILLA

E, F-ORANG-OUTANG G, H-CHIMPANZEE

$f_{\mathbf{x}}(t) = x_1/\sqrt{2} + x_2 \sin t + x_3 \cos t + x_4 \sin 2t + \cdots$

(i) The function representation preserves means. If $\bar{\mathbf{x}}$ is the mean of a set of n multivariate observations $\mathbf{x}_i$, then the function corresponding to $\bar{\mathbf{x}}$ is the pointwise mean of the functions corresponding to the n observations:

$$f_{\bar{\mathbf{x}}}(t) = \frac{1}{n} \sum_{i=1}^{n} f_{\mathbf{x}_i}(t).$$

As a result the average will appear like an average in this plot.

(ii) The function representation preserves distances. One measure of the distance between two functions that seems in accord with distance as judged by the human eye is

$$||f_{\mathbf{x}}(t) - f_{\mathbf{y}}(t)||_{L_2} = \int_{-\pi}^{\pi} [f_{\mathbf{x}}(t) - f_{\mathbf{y}}(t)]^2 \, dt.$$

Moreover, this distance is proportional to the familiar Euclidean distance between the corresponding points since

$$||f_{\mathbf{x}}(t) - f_{\mathbf{y}}(t)||_{L_2} = \pi \, ||\mathbf{x} - \mathbf{y}||^2 = \pi \sum_{i=1}^{k} (x_i - y_i)^2.$$

Because of this relation, close points will appear as close functions and distant points as distant functions. Thus multivariate clusters and outliers may be identified visually from the plot of the functions. It is this distance-preserving property that determines the coefficient $1/\sqrt{2}$ and restricts the coefficients of t to integers.

(iii) The representation yields one-dimensional projections. For a particular value of $t = t_0$ the function value $f_x(t_0)$ is proportional to the length of the projection of the vector $(x_1, \cdots, x_k)$ on the vector

$$\mathbf{f}_1(t_0) = (1/\sqrt{2}, \sin t_0, \cos t_0, \sin 2t_0, \cos 2t_0, \cdots)$$

since

$$f_x(t_0) = \{\mathbf{x}'\mathbf{f}_1(t_0)/[\mathbf{f}_1'(t_0)\mathbf{f}_1(t_0)]\} \cdot [\mathbf{f}_1'(t_0)\mathbf{f}_1(t_0)].$$

The projection on this one-dimensional space may reveal clusterings, outlier patterns, or other peculiarities that occur in this subspace and which may be otherwise obscured by other dimensions. The advantage of this plot is that a continuum of such one-dimensional projections is plotted on one graph.

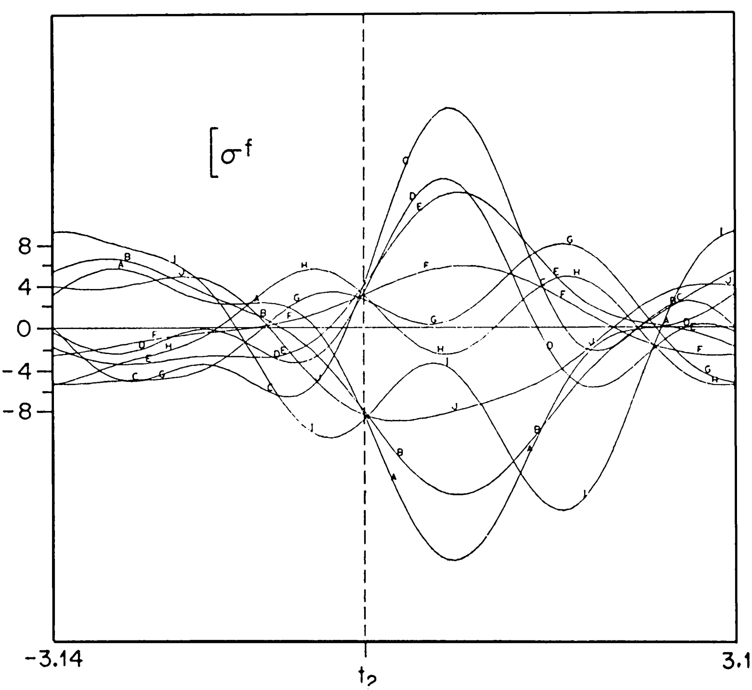

FIGURE 5

6-DIMENSIONAL DATA

AS IN FIGURE 4 WITH FOSSILS ADDED

I-*Paranthropus crassidens*

J-*Pithecanthropus pekinensis*

(iv) The representation preserves variances. If the components of the data are uncorrelated with common variance σ^2 then the function value at t, $f_{\mathbf{x}}(t)$, has variance

$$\text{var}\,[f_{\mathbf{x}}(t)] = \sigma^2(\tfrac{1}{2} + \sin^2 t + \cos^2 t + \sin^2 2t + \cos^2 2t + \cdots).$$

If k is odd this reduces to a constant, $\frac{1}{2}\sigma^2 k$; if k is even the variance lies between $\frac{1}{2}\sigma^2(k-1)$ and $\frac{1}{2}\sigma^2(k+1)$. In the first case the variance does not depend on t and in the second the relative dependence on t is slight and decreases as k increases. Thus the variability of the plotted function is almost constant across the graph. This standard deviation of f is denoted by σ_f where it appears on the plots. This facilitates the interpretation of the plot as outlined in the next section.

5. INTERPRETATION OF THE PLOTS

(i) *Clustering.* If some plotted functions form a band by remaining close together for all values of t, then the corresponding points are close together in the Euclidean metric (see 4(ii)). Such a band represents a cluster of data points. If a group of functions come close together for only one or a few values of t, then the corresponding points are close in the directions defined by the corresponding vectors $\mathbf{f}_1(t)$. Thus it is possible to identify clusters of points even when some additional variables are present.

(ii) *Tests of significance at particular values of t.* In the example it is possible to identify *a priori* from Figure 2 certain values of t for which it is of interest to test the hypothesis that the expectation of $f_{\mathbf{x}}(t) = f_{\boldsymbol{\mu}}(t)$ for some hypothesized $\boldsymbol{\mu}$. Since the variance of $f_{\mathbf{x}}(t)$ is known (4(iv)) a test of this hypothesis may be constructed by evaluating the significance level of

$$z = [f_{\mathbf{x}}(t) - f_{\boldsymbol{\mu}}(t)]/\{\text{var}\,[f_{\mathbf{x}}(t)]\}^{\frac{1}{2}}.$$

If the components x_i are assumed to be independent normal variables, then z has a standard normal distribution under the hypothesis $\boldsymbol{\mu} = \boldsymbol{\mu}_0$. This distribution may be used to assess the hypothesis $\boldsymbol{\mu} = \boldsymbol{\mu}_0$ or to construct a confidence interval for $\boldsymbol{\mu}$ as the set of values not rejected by this test at a level α. The width of this confidence set is shown on Figure 3. Note, however, that this test is exact if the value of t is chosen *a priori.*

(iii) *Overall tests.* It is also possible to construct tests suggested by the data. If the x_i are independent normal variates with variance σ^2 and expectation μ_i, then $||\mathbf{x} - \boldsymbol{\mu}||^2/\sigma^2$ has a χ^2 distribution with k D.F. The squared length of the projection of $\mathbf{x} - \boldsymbol{\mu}$ on any one-dimensional space is not larger than $||\mathbf{x} - \boldsymbol{\mu}||^2$. Furthermore, the squared length of the projection of $\mathbf{x} - \boldsymbol{\mu}$ on the vector

$$\mathbf{v} = \mathbf{f}_1(t)/[\mathbf{f}_1'(t)\mathbf{f}_1(t)]^{\frac{1}{2}}$$

is just

$$|(\mathbf{x} - \boldsymbol{\mu})'\mathbf{v}|^2 = |f_{\mathbf{x}}(t) - f_{\boldsymbol{\mu}}(t)|^2/[\mathbf{f}_1'(t)\mathbf{f}_1(t)].$$

Hence with probability $1 - \alpha$, for all values of t,

$$|f_{\mathbf{x}}(t) - f_{\boldsymbol{\mu}}(t)|^2 \leq \sigma^2 |f_1(t)| \chi_k^2(\alpha) \leq \left(\frac{k+1}{2}\right) \sigma^2 \chi_k^2(\alpha),$$

where $\chi_k^2(\alpha)$ denotes the upper α point of the χ^2 distribution.

Thus with probability $1 - \alpha$ the function $f_{\mathbf{x}}(t)$ lies in a band with fixed width about $f_{\boldsymbol{\mu}}(t)$. If $\boldsymbol{\mu}$ is known, outliers will fall outside this band and may be easily identified visually. Note that the level of the test is roughly n times the rejection probability if n points are examined. A band of the same width centered at $f_{\mathbf{x}}(t)$ is a $1 - \alpha$ confidence region for $f_{\boldsymbol{\mu}_0}(t)$. If $f_{\boldsymbol{\mu}}(t)$ falls outside this band there is evidence against the hypothesis $\boldsymbol{\mu} = \boldsymbol{\mu}_0$.

(iv) *Linear relationships.* If a point $\mathbf{y}$ lies on a line joining $\mathbf{x}$ and $\mathbf{z}$, then for all values of t, $f_{\mathbf{y}}(t)$ is between $f_{\mathbf{x}}(t)$ and $f_{\mathbf{z}}(t)$. This fact may be readily observed as in Figure 5 where the fossil *Paranthropus crassidens* lies for the most part between the apes and man.

6. FURTHER REMARKS

In this section are summarized a number of observations based on experience with this type of plot.

(i) *Plotting many points.* Only a limited amount of information may be absorbed from one plot. If each point is to be examined in detail, only about 10 points may be plotted in this way on the same graph. A procedure that proved useful in cases involving more points was first to plot all the points on one graph from which general characteristics of the sample could be noted. Separate plots of 10 points were then used to assess individual points in relation to the whole.

(ii) *Other functions.* The properties of sections 2 and 3 will also hold for functions of the form

$$f_{\mathbf{x}}(t) = x_1 \sin n_1 t + x_2 \cos n_1 t + x_3 \sin n_2 t + x_4 \cos n_2 t + \cdots,$$

where the n_i are different integers. Where more than one plot is desired, function plots based on such a series will provide more information about the data. It was found that plots with small n_i were easier to look at than those with large n_i. The statistical properties of the function plots depend heavily on the orthogonality of the trigonometric functions and on the approximate constancy of the length of the vector $\mathbf{f}_1(t)$. Other families of orthogonal functions such as orthogonal polynomials may be tried. Typically these do not have all the properties of the proposed function. Trigonometric functions involving a pure sine or cosine series will also lack many of the properties of the mixed series proposed.

(iii) *The one-dimensional projections.* As noted previously the function values $f_{\mathbf{x}}(t)$ are the projections of the vector $\mathbf{x}$ on the vector

$$\mathbf{f}_1(t) = (1/\sqrt{2}, \sin t, \cos t, \sin 2t, \cdots).$$

When normalized this vector defines a point on the k-sphere. As t varies this point describes a periodic curve on the k-sphere.

Some idea of how close this curve comes to points on the sphere may be obtained by calculating the distance of points to the closest point on the curve or on its reflection through the origin. This distance may be measured in terms of the Euclidean metric or alternatively in terms of the size of the angle, measured at the origin, between points on the sphere. Table 3 gives average values of these two measures for two curves: the curve defined by the function of section 2 and by the more complicated function

$$f_1(t) = \sin 2t, \cos 2t, \sin 4t, \cos 4t, \sin 8t \cdots .$$

The averages were calculated by Monte Carlo assuming a uniform distribution on the sphere.

From Table 3 we note that (a) as the dimension increases points tend to be further from the curve on the average; and (b) on the average points lie closer to the more complicated curve.

The functions proposed in (ii) above describe complex curves passing closer to more points on the sphere. The simplicity of the curve (and hence smoothness of the plot) and the proximity of its path to every point on the sphere define two mutually conflicting goals. If the curve is too complicated the plot contains too much detail to be readily absorbed.

(iv) *Using principal components.* The tests and confidence sets were developed in section 5 for data with independent normal components with the same variances. For these tests to be applicable, multivariate data must first be transformed to approximate those conditions. A principal component analysis was used by Ashton *et al.* to yield such variables and in general this procedure is recommended. In the plots low frequencies are more readily seen than high frequencies. For this reason it is useful to associate the most important variables with low frequencies. This may be done by associating x_1 with the first principal component, x_2 with the second, and so on.

(v) *Other applications.* This plotting procedure was used to solve a somewhat artificial problem in pattern recognition posed by Mrs. J. J. Chang. An unknown 2-dimensional pattern of 25 points was prepared. Gaussian noise with comparable second moments was added in another 3 dimensions and the resulting 5-dimensional structure was given an arbitrary orientation.

TABLE 3

AVERAGE DISTANCES TO CLOSEST POINT ON CURVE OR ON ITS REFLECTION THROUGH THE ORIGIN

Simple Curve $f_1(t) = 1/\sqrt{2}, \sin t, \cos t, \sin 2t, \cos 2t, \cdots$

Complicated Curve $f_1(t) = \sin 2t, \cos 2t, \sin 4t, \cos 4t, \sin 8t, \cdots$

	Simple		Complicated	
Dimension	Euclidean distance	Angle in degrees	Euclidean distance	Angle in degrees
3	.33	19	.25	14
5	.57	33	.49	28
7	.68	40	.61	35
9	.76	45	.66	39

The problem was to identify the pattern. These data were plotted using the function proposed in section 2 and other functions as in (ii) above.

It was assumed that patterns could be distinguished from noise by the presence of strong clustering or regularity and the absence of outliers in projections on 1-dimensional spaces that lie in the 2-dimensional space of the pattern. Such regularity was found for 2 values of t. One vector $\mathbf{f}_1(t)$ from each of 2 different plots was used to define a plane and the projection of the data on this plane revealed the pattern.

Other examples involved many measurements on several subjects. The measurements were transformed to principal components and then plotted as in section 2. This procedure gave some idea of the relation and similarities among the measurements. All of these observations based on the plots of section 3 may be determined directly by examining the data. The advantage in plotting is that these observations come more easily and more quickly.

The plotting procedure is not a substitute for data manipulation. The choice of which numbers to examine, here canonical variables, is based on the nature of the data and the objectives of the analysis. The plotting procedure is only a method of displaying whatever numbers are deemed relevant. The examination of such plots permits the detection of many abnormalities in the data. Some of these will be spurious, others may represent real effects. The existence of the overall tests permits the identification of the very significant effects. Typically this cannot be done with other plotting methods. The advantage of the plot is that peculiarities in the data are brought to the attention of the investigator.

ACKNOWLEDGMENTS

The author gratefully acknowledges many helpful suggestions by the referees and conversations with Colin Mallows. These aided in the revision of the paper and in the understanding of the nature of the geometric problems involved.

GRAPHIQUES DE DONNEES AVEC DE GRANDES DIMENSIONS

RESUME

On propose une méthode pour tracer des graphiques de données à plus de deux dimensions. Chaque point donnée, $\mathbf{x} = (x_1, \cdots, x_k)$ est transformé en une fonction de la forme

$$f_{\mathbf{x}}(t) = x_1/\sqrt{2} + x_2 \sin t + x_3 \cos t + x_4 \sin 2t + x_5 \cos 2t + \cdots,$$

et on trace le graphique de la fonction dans l'intervalle $-\pi < t < \pi$.

On explore quelques propriétés statistiques de la méthode, son application est illustrée par un exemple d'anthropologie.

REFERENCE

Ashton, E. H., Healy, M. J. R., and Lipton, S. [1957]. The descriptive use of discriminant functions in physical anthropology. *Proc. Roy. Soc. B 146*, 552–72.

Received October 1970, *Revised July* 1971

Key Words: Plotting multivariate data; Cluster analysis; Pattern recognition; Discrimination; Classification.