DISCRIMINATION AND CLASSIFICATION

A COMPARISON OF TECHNIQUES USED IN DISCRIMINATION AND CLASSIFICATION

By

OWEN MICHAEL GRANT HAMILTON, B.Sc.

A Project

Submitted to the School of Graduate Studies

in Partial Fulfilment of the Requirements

for the Degree

Master of Science

McMaster University

August 1976

MASTER OF SCIENCE (1976) McMASTER UNIVERSITY
(Statistics) Hamilton, Ontario

TITLE: A Comparison of Techniques used in Discrimination and Classification

AUTHOR: Owen Michael Grant Hamilton

B.Sc. (University of Western Ontario)

SUPERVISOR: Dr. Michael A. Stephens

NUMBER OF PAGES: x,96

ABSTRACT

Application of four statistical techniques of discrimination is made to a set of multivariate data. The techniques, proposed by R.A. Fisher [6], C.R. Rao [14], D.F. Andrews [1] and H. Chernoff [4], are reviewed, applied and criticized in an intercomparison of the four methods. Graphic illustrations are also utilized to aid in the classification of sampling units.

ACKNOWLEDGEMENTS

My involvement with Dr. M.A. Stephens in preparing this project has been entirely enjoyable. His direction and motivating influence throughout the year have been invaluable. I thank him.

In times of mutual frenzy, my fellow colleagues George, Sam, Cuong and Eric were reassuring comforts from the various pressures. Their friendship will never be forgotten.

Lastly, I would like to express my appreciation to McMaster University for allowing me the opportunity to study in such a pleasant environment.

To my wife

Marion

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

1. INTRODUCTION

The purpose of this project was to compare classical methods of discrimination with two recently developed modern approaches. Discrimination in statistics generally implies a technique to facilitate the classification of a sampling unit to its parent population. The two classical procedures that we employed are Fisher's discriminant functions and canonical analysis of discriminance. The modern computer-aided approaches are Andrews' fourier series and Chernoff's faces.

The data under question had been divided into four groups but the grouping criterion was unknown. A preliminary analysis of the four groups revealed that the groups were too similar in location (i.e. means) to supply any striking results. However, to illustrate the application and potential of the statistical methods, a subset of the original data was selected and analyzed. Sections 3 and 4 deal with the selected subset while section 5 examines the entire set of data to illustrate difficulties that may arise in application of these procedures.

2. DATA

The datawere collected and compiled by a team of doctors at the Royal Victoria Hospital, Montreal, Canada. The study was concerned with investigating anemia in pregnant women. Of twenty-one measurements that were made per patient, six important measurements of the blood were used in this analysis. These were (a) Hemoglobin (Hb): an iron-containing compound found in red blood cells that carries oxygen from the lungs to the body tissues. The units of measurement are grams per 100 ml. of blood.

(b) Packed Cell Volume (PCV): also termed the hematocrit. This is a measurement of the percentage by volume that the red blood cells occupy in a sample of blood.

(c) Mean Corpuscular Hemoglobin Concentration (MCHC): this quantity refers to the amount of hemoglobin per volume of red blood cells.

(d) Serum Folate (SF): the amount of folic acid found in the serum or plasma of the blood. Measured in nanograms per millilitre. (10^{-9} g/ml.).

(e) Whole Folate (WF): the amount of folic acid found in the total volume of a sample of blood. (ng/ml.).

(f) Total Iron Binding Capacity (TIBC): transferrin is a carrier protein for iron which is necessary in the synthesis of hemoglobin. The iron binding capacity of the blood is therefore the amount of transferrin in milligrams per 100 millilitre volume of blood.

The above six measurements will be referred to as variables X_1, X_2, ..., X_6 respectively.

The women were classified into four groups D, E, F and G and forty from each group were involved in the study. The selected subset of the data was comprised of ten patients from each group.

Table 2.1 supplies a complete listing of the six-variate data for each group and table 2.2, the subset of the data.

Table 2.1. Measurements of the blood from 160 pregnant women

Group D

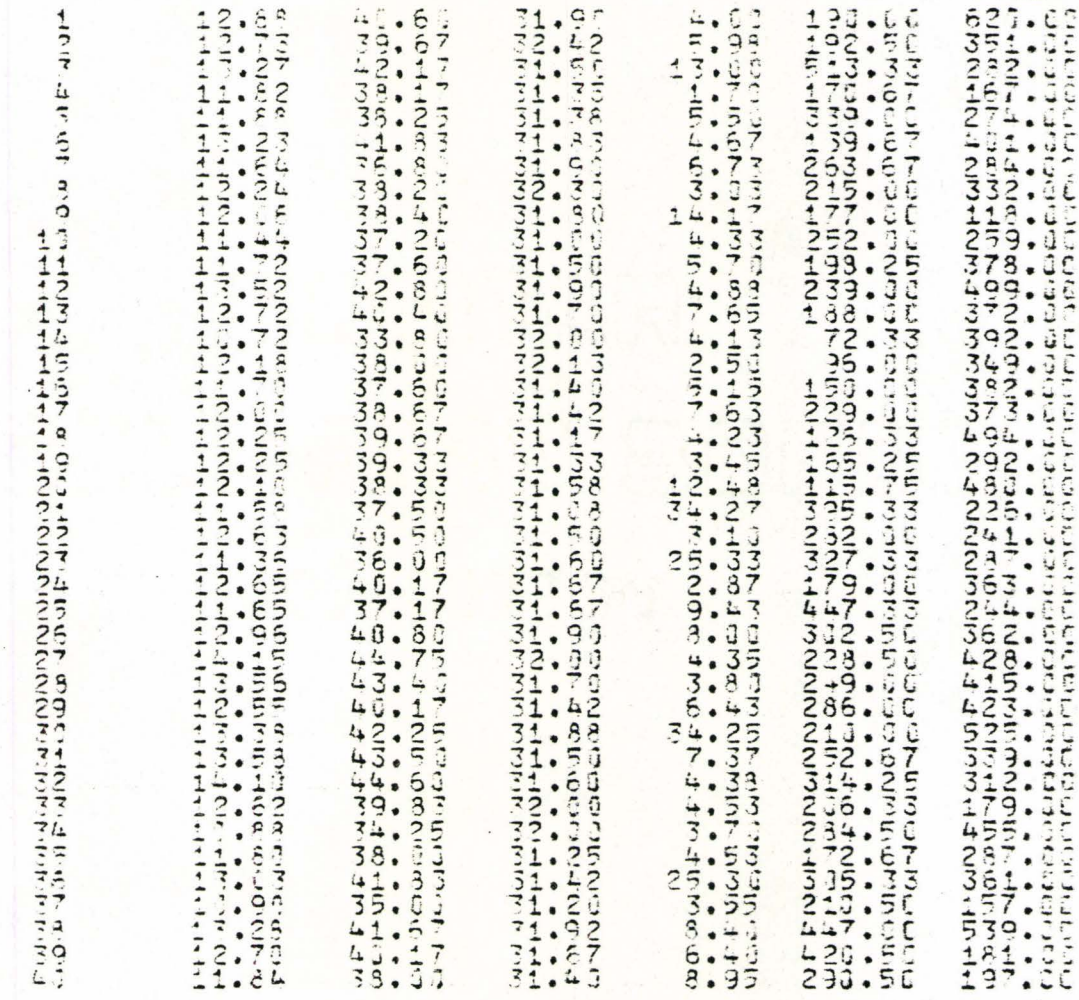


Table 2.1. (continued)

Group E

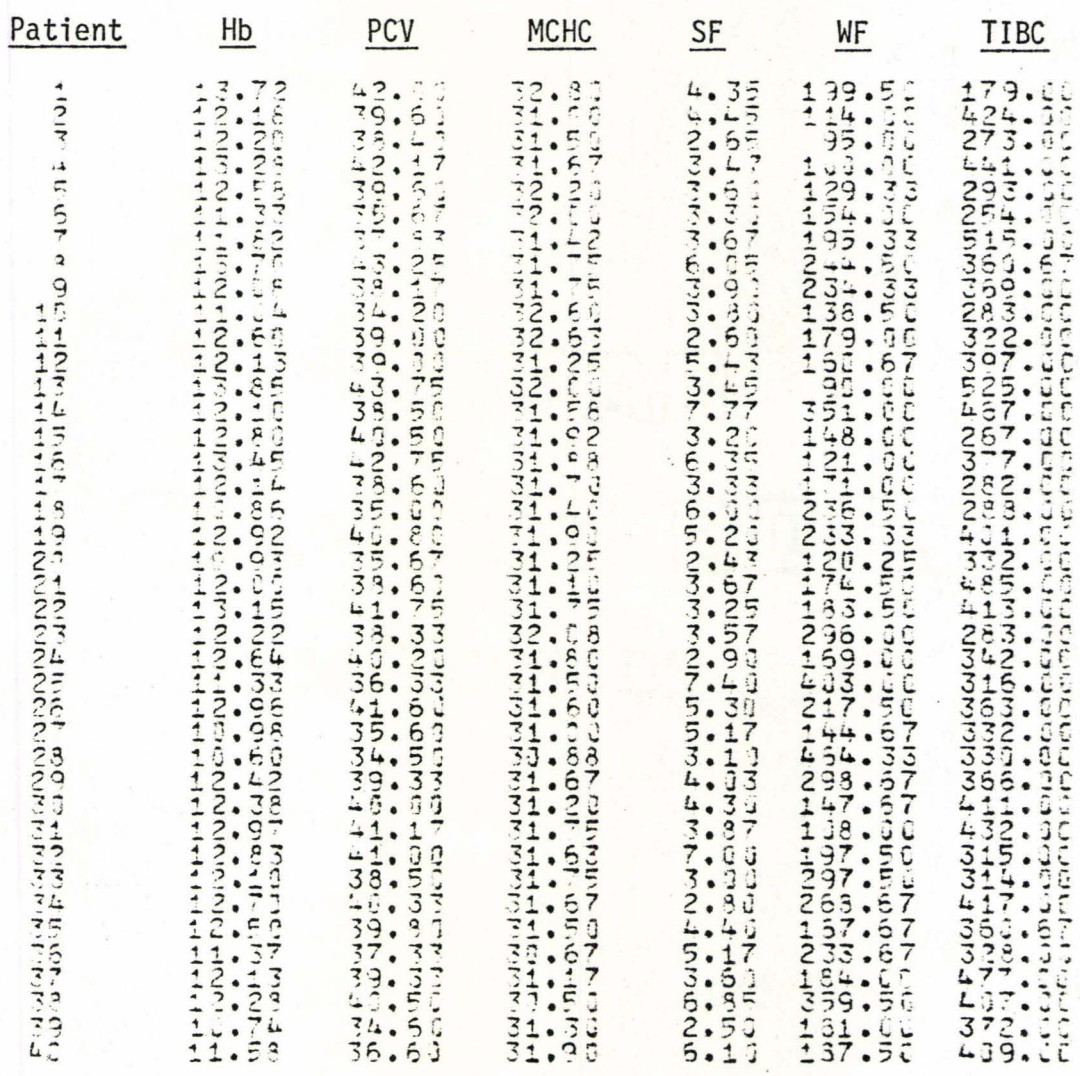


Table 2.1. (continued)

Group F

Patient	Hb	PCV	MCHC	SF	WF	TIBC

Table 2.1. (continued)

Group G

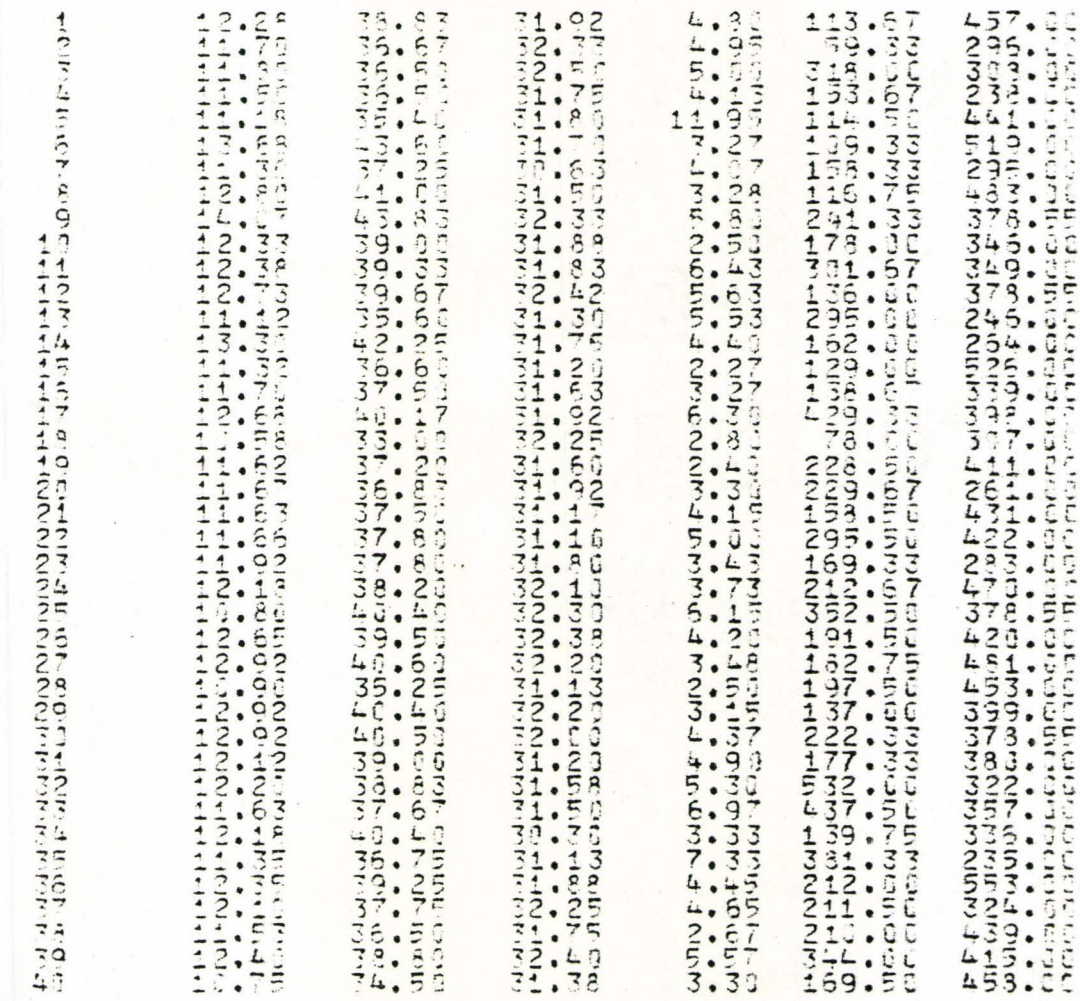


Table 2.2. Subset of table 2.1 to be analyzed

Patient	Hb	PCV	MCHC	SF	WF	TIBC
			Group D			
3	13.27	42.17	31.50	13.08	513.33	282.00
9	12.06	38.40	31.80	14.17	177.00	118.00
21	11.58	37.50	31.08	34.27	395.33	226.00
23	11.36	36.00	31.60	25.33	327.33	287.00
30	13.38	42.25	31.86	34.25	210.00	525.00
31	13.58	43.50	31.50	7.37	252.67	339.00
32	14.10	44.60	31.60	4.38	314.25	312.00
33	12.62	39.83	32.00	4.53	286.33	179.00
36	13.03	41.83	31.42	25.30	385.33	361.00
40	11.84	38.00	31.40	8.95	290.50	197.00
			Group E			
2	12.18	39.60	31.00	4.45	114.00	424.00
4	13.28	42.17	31.67	3.47	130.00	441.00
12	12.13	39.00	31.25	5.43	160.67	397.00
13	13.85	43.75	32.00	3.45	90.00	525.00
22	13.15	41.75	31.75	3.25	183.50	413.00
27	10.98	35.60	31.00	5.17	144.67	332.00
30	12.38	40.00	31.20	4.30	147.67	411.00
31	12.97	41.17	31.75	3.87	108.00	432.00
35	12.50	39.80	31.50	4.40	157.67	360.67
36	11.37	37.33	30.67	5.17	233.67	328.00
			Group F			
2	12.68	39.40	32.30	6.83	518.00	406.00
14	12.45	38.67	32.50	6.13	295.67	291.00
16	12.84	39.80	32.50	7.70	487.67	191.00
17	11.84	37.60	31.68	14.77	386.67	422.00
19	11.60	35.20	33.20	13.25	328.00	277.00
21	11.74	37.40	31.70	9.33	359.00	350.00
28	12.60	40.25	31.50	8.13	511.67	281.00
35	12.48	40.00	31.38	9.67	314.00	552.00
36	11.26	35.40	32.10	15.10	409.00	314.00
39	12.48	39.60	31.90	8.47	357.67	377.00
			Group G			
10	12.33	39.00	31.88	2.50	178.00	346.00
18	10.58	33.00	32.25	2.80	78.00	327.00
19	11.62	37.20	31.60	2.40	228.50	411.00
22	11.65	37.80	31.10	5.00	295.50	422.00
24	12.18	38.20	32.10	3.73	212.67	470.00
27	12.92	40.60	32.20	3.49	182.75	491.00
28	10.90	35.25	31.13	2.50	187.50	433.00
35	12.35	39.25	31.88	4.45	212.00	555.00
38	11.53	36.50	31.75	2.67	210.00	439.00
43	10.75	34.50	31.38	3.30	169.50	458.00

3. DISCRIMINATION AND CLASSIFICATION

3.1 *General*

Suppose we have k p-variate populations of a similar kind. If we must allocate a member to one of the k populations but are uncertain from which population it came, by what rule should we proceed? Questions of this type give rise to discrimination and classification, the general object of which is to find rules to assign individuals to predetermined classes.

3.2 *Fisher's Discriminant Function*

In 1936, R.A. Fisher introduced a linear discriminant function which distinguished two p-variate normal populations from one another [14]. His function was of the form

$$V = d'X$$

where $X' = (X_1, X_2, \ldots, X_p)$ is a p-dimensional random variable vector and $d' = (d_1, d_2, \ldots, d_p)$, a vector of coefficients.

Fisher's purpose in constructing this linear combination of the p variates was to separate the populations as much as possible. He therefore chose the coefficient vector, d, which maximized the difference in mean values of each population subject to the constraint that the within population variance was constant. That is, if $\mu_i' = (\mu_{i1}, \mu_{i2}, \ldots, \mu_{ip})$ is the mean vector of the ith population and Σ is the common covariance matrix for each population, then we

must find a vector of coefficients, d, such that

$$\{d'(\mu_1-\mu_2)\}^2$$

is maximized while

$$\begin{aligned} \text{var } (V) &= E\{(V-E(V))(V-E(V))'\} \\ &= E\{(d'(X-\mu))(d'(X-\mu))'\} \\ &= d'E\{X-\mu)(X-\mu)'\}d \\ &= d'\Sigma d \end{aligned}$$

is constant. This is equivalent to maximizing the ratio

$$\frac{\{d'(\mu_1-\mu_2)\}^2}{d'\Sigma d} \quad .$$

The solution by differentiating with respect to d yields

$$d \propto \Sigma^{-1}(\mu_1-\mu_2) \ .$$

Therefore, our linear discriminant function is

$$V = (\mu_1-\mu_2)'\Sigma^{-1}X.$$

Recalling that the motive of the discriminant function was to determine a classification rule, the rule is to assign the member x to the population which results in a mean discriminant value

$$\mu(V_i) = (\mu_1-\mu_2)'\Sigma^{-1}\mu_i \quad , \qquad i = 1, 2$$

closer to the discriminant value

$$v = (\mu_1-\mu_2)'\Sigma^{-1}x$$

evaluated for x [12].

An approximate calculation of the probability of misclassification may be made. Since we allocate x according to which $\mu(V_i)$ the evaluated v is closer, the critical value is

$$\tfrac{1}{2}(\mu(V_1)+\mu(V_2)).$$

The variance of V is

$$\begin{aligned} \text{var } (V) &= d'\Sigma d \\ &= d'\Sigma\{\Sigma^{-1}(\mu_1-\mu_2)\} \\ &= d'(\mu_1-\mu_2) \end{aligned}$$

which may be estimated by

$$\begin{aligned} \hat{\text{var}}\ (V) &= d'(\bar{x}_1-\bar{x}_2) \\ &= \bar{v}_1-\bar{v}_2 \end{aligned}$$

(without loss of generality, assume $\bar{v}_1>\bar{v}_2$). Because V is a linear combination of normal variates and is itself therefore normal, the distribution of V is approximately $N(\mu(V_i), [\mu(V_1)-\mu(V_2)])$ which may be estimated by

$$V \sim N(\bar{v}_i, (\bar{v}_1-\bar{v}_2)).$$

Therefore, the probability of misclassification, conditional on the sampling unit being from one of the assumed normal populations, is approximately

$$\Pr(V>\tfrac{1}{2}(\bar{v}_1+\bar{v}_2)) = \Pr\left(Z>\frac{|\tfrac{1}{2}(\bar{v}_1-\bar{v}_2)|}{(\bar{v}_1-\bar{v}_2)^{\frac{1}{2}}}\right)$$

where $Z \sim N(0,1)$.

Then given two multinormal populations with a common covariance matrix, we may classify any sampling unit into one of the two populations with a calculable probability.

3.2.1 *Application of Fisher's Discriminant Function, k = 2*

Hereafter, until section 5, the data under consideration will be the selected subset of the original data.

Groups F and G were chosen to illustrate this application.

The parameter estimates of the covariance matrix Σ and the population mean μ_i were the pooled dispersion matrix S and sample mean $\bar{x}_i$, respectively. To justify our assumption of a common covariance matrix Σ, we tested the hypothesis

$$H : \Sigma_1 = \Sigma_2 = \ldots = \Sigma_k = \Sigma$$

where Σ_i is the covariance matrix of the ith population. The required test criterion is

$$h = -2\left[1-\left\{\sum_{i=1}^{k}\frac{1}{n_i-1}-\frac{1}{n-k}\right\}\frac{2p^2+3p-1}{6(p+1)(k-1)}\right]\ln\left[\left\{\prod_{i=1}^{k}\frac{|S_i|}{|S|}\right\}^{\frac{1}{2}(n_i-1)}\right]$$

where $|S_i|$ is the determinant of the estimate of the ith population's covariance matrix, $|S|$ the determinant of the pooled estimate of the common covariance matrix, n_i is the ith population's sample size and $n = n_1 + n_2 + \ldots + n_k$ the total sample size [15]. The test statistic h is distributed as chi-square with degrees of freedom equal to $[(k-1)p(p+1)\div 2]$, [14]. Therefore with the dispersion matrices

Group F (S_F)

	1	2	3	4	5	6
1	.29	.94	-.04	-1.49	17.21	-7.00
2		3.48	-.49	-4.68	54.43	-1.88
3			.31	.04	-4.43	-17.99
4				10.67	-57.83	45.57
5					6827.60	-810.15
6						4916.40

Group G (S_G)

	1	2	3	4	5	6
1	.60	1.78	.13	.22	15.35	22.47
2		5.50	.23	.79	65.16	78.02
3			.17	-.05	-13.32	-3.90
4				.80	25.17	30.74
5					2979.17	1774.71
6						4797.11

Pooled (S)

	1	2	3	4	5	6
1	.44	1.36	.05	-.63	16.28	7.73
2		4.49	-.13	-1.95	59.80	38.07
3			-.24	-.003	-8.88	-10.94
4				5.73	-16.33	38.15
5					4903.38	482.28
6						4856.76

and their determinants

$$|S_F| = 2.457 \times 10^4$$

$$|S_G| = 5.944 \times 10^2$$

$$|S| = 5.364 \times 10^4$$

our test statistic becomes

$$h = -2[1-(\tfrac{2}{9} - \tfrac{1}{18})\tfrac{89}{42}] \ln \{(.458)(.011)\}^{4.5}$$

$$= 30.755$$

The critical value of the chi-square distribution at $\alpha = .05$ and d.f. = 21 is $\chi^2_{.05,21} = 32.671$. The non-significance of our test statistic suggests that the assumption of the homogeneity of dispersions may not be violated. This encourages further analysis.

If we consider the assumption that the populations have identical dispersions upheld, we now test the hypothesis

$$H : \mu_1 = \mu_2 = \ldots = \mu_k = \mu$$

that the populations are identical in location. If this were the case, then discrimination among the populations would not yield impressive results. In fact, this was the situation with the original data.

This test of the dispersion of the populations from one another employs a test statistic based on Wilks criterion [14]

$$\Lambda = \frac{|W|}{|P+W|}$$

where W and P are the sum of squares and cross-product matrices representing the within group dispersions and the between group dispersions, respectively. The actual statistic which is distributed as chi-square with p(k-1) degrees of freedom is

$$u = -[(n-1)-(\frac{p+k}{2})] \ln (\Lambda).$$

From the sum of products matrices

W(within S.P.)

	1	2	3	4	5	6
1	3.99	12.23	.42	-5.70	146.52	69.60
2		40.41	-1.18	-17.54	538.16	342.66
3			2.19	-.03	-79.94	-98.50
4				51.58	-146.97	343.37
5					44130.46	4340.53
6						43710.80

P(between S.P.)

	1	2	3	4	5	6
1	.13	.31	.09	1.71	51.52	-27.50
2		.72	.21	4.00	120.26	-64.19
3			.06	1.16	34.92	-18.64
4				22.14	665.81	-355.38
5					20018.60	-10684.97
6						5703.12

and the necessary determinants

$$|W| = 2.851 \times 10^{10}$$

$$|P+W| = 7.285 \times 10^{10}$$

the test statistic is

$$u = -[(39)-(\frac{6+2}{2})]\ln(.3914)$$

$$= 15.013$$

The critical chi-square value, $\chi^2_{.05,6} = 12.592$, indicates that the populations are significantly different at $\alpha = .05$.

These two preliminary tests signify that our underlying assumptions of the data are upheld and that good discrimination between the populations will be possible.

The coefficients of the discriminant function between groups F and G were determined by

$$d = S^{-1}(\bar{x}_F - \bar{x}_G) .$$

Mean Vectors ($\bar{x}_i'$)

	1	2	3	4	5	6
Group F	12.2	38.33	32.1	9.9	396.5	327.2
Group G	11.7	37.1	31.7	3.3	196.4	434.0

Inverse of Pooled Cov. Matrix (S^{-1})

	1	2	3	4	5	6
1	1340.96	-413.24	-486.78	6.83	-.219	-.0196
2		127.71	149.90	-1.97	.079	.0022
3			181.59	-2.57	.107	.0188
4				.27	-.002	-.0032
5					.003	.0001
6						.0003

The result d' = (-17.51, 4.44, 4.87, -2.00, -.039, .046) would supply us with a discriminant function of the form

$$V = d'X$$

However, it is convenient to choose the coefficients of the linear combination of the p variates such that the variance within the groups of the discriminant function is one. Since our function is used only to separate the two populations, not to measure the distance between them, we may multiply it by any suitable constant. The necessary factor to standardize the discriminant function is the reciprocal of the square root of the variance of the functions. This variance may be determined since the sample dispersion of X is known. As discussed in section 3.2.

$$\hat{var}\,(V) = d'Sd$$
$$= [S^{-1}(\bar{x}_F - \bar{X}_G)]'\ S[S^{-1}(\bar{x}_F - \bar{x}_G)]$$
$$= (\bar{x}_F - \bar{x}_G)'S^{-1}(\bar{x}_F - \bar{x}_G)$$

This value is also referred to as Mahalanobis D^2, a measure of the distance between two populations [14]. Note that the estimated values

of Σ and μ_i were used in the determination of the variance of the function. This is because the discriminant function is entirely sample dependent.

The standardized coefficients are therefore

$$d^* = \left[\frac{1}{(\text{Mahal. } D^2)^{\frac{1}{2}}}\right] d$$

$$= (.1890)d$$

$$= (-3.31, .84, .92, -.38, -.007, .0086)$$

which gives us the desired discriminant function

$$V^* = -3.31X_1 + .84X_2 + .92X_3 - .38X_4 - .007X_5 + .0086X_6.$$

The rule of classification of a sampling unit to the populations is then to assign the patient to group F if her discriminant value is closer to

$$\bar{v}_F^* = d^{*\prime}\,\bar{x}_F = -17.445$$

or the group G if it is closer to

$$\bar{v}_G^* = d^{*\prime}\,\bar{x}_G = -22.736$$

Since we have standardized the discriminant function, then $V^* \sim N(\bar{v}_i^*, 1)$. Therefore the probability of misclassification of a new patient, conditional that the patient be from either population F or G, is approximately

$$\Pr(Z > |\tfrac{1}{2}(\bar{v}_F^* - \bar{v}_G^*)|) = \Pr(Z > 2.645)$$

$$= .004$$

which means that we may expect over 99% of our assignments of a patient to either group F or G to be correct.

To test this result, patient numbered 40 of group G was to be assigned to either F or G according to Fisher's discriminant function. Her p-variate vector was

$$x^{40} = (10.8, 34.5, 31.4, 3.3, 169.5, 458.0)$$

and so her discriminant value was

$$v^{*40} = d^{*\prime} x^{40} = -23.695$$

We therefore allocate her to group G where she belongs.

The negligible probability of misclassification with this discriminatory technique favours strongly its application to problems of two population discrimination.

3.2.2 *Application of Fisher's Discriminant Function, k = 3*

When the number of populations is greater than or equal to three, Fisher's discriminating function becomes a system of functions which must be satisfied. That is, if we are considering the groups D, F and G, then to classify a patient into group D her discriminant value must satisfy appropriately the two discriminant functions

$$V_1 = [\Sigma^{-1}(\mu_1 - \mu_2)]'X$$

and

$$V_2 = [\Sigma^{-1}(\mu_1 - \mu_3)]'X$$

where μ_1, μ_2, μ_3 represent the mean vectors of groups D, F and G, respectively. V_1 will make the comparison between groups D and F and V_2 between groups D and G. The patient's discriminant values in each case must be closer to group D's mean discriminant values in order for her to be classified as a group D patient. Similarly, allocation into group F requires that the new patient suitably satisfies the discriminant functions

$$V_1 = [\Sigma^{-1}(\mu_1 - \mu_3)]'X$$

and

$$V_3 = [\Sigma^{-1}(\mu_2 - \mu_3)]'X$$

whereas if she is to be classified as a group G patient her criterion would be determined by V_2 and V_3.

We therefore have $\binom{k}{2} = 3$ discriminant functions with which to determine her proper parent population. Each allocation will have a probability of misclassification which will be somewhat more complicated to compute because of the non-zero covariances of the discriminant functions with one another.

We illustrate with groups D, F and G.

The initial tests of homogeneity of dispersions and multivariate analysis of variance were applied with the following results.

To be Tested	Hypothesis	Test statistic	d.f.	$\chi^2_{\alpha=.05}$
homogeneity of dispersions	$H_i: \Sigma_1 = \ldots = \Sigma_k$	h = 71.508	42	58.124
equality of means	$H_2: \mu_1 = \ldots = \mu_k$	u = 40.03	12	21.026

The significance of h leads us to a rejection of the H_1 hypothesis that the dispersions may be represented by a common covariance matrix. However, this test is based on the assumption that the p variates are multinormally distributed whose violation may be the cause of the significant test value. We therefore transformed the data into logarithms in order to normalize the observations but to no avail. The test statistic was still significant at $h_{log} = 59.584$. We have no alternative but to conclude that there exists a difference in the sizes and/or the orientations of the density ellipsoids among the three groups. The abnormality will be in group D since groups F and G were shown to be homogeneous in section 3.2.1.

A highly significant u statistic in the multivariate analysis of dispersion implies that a strong difference appears among the group means. Even though there is heterogeneity among the covariance matrices the test is robust enough to survive a certain amount of disparity [9]. We can then expect good separation of our groups by the discriminant functions.

The following inverse of the pooled covariance matrix and the mean vectors will supply us with the required discriminant functions.

Mean Vectors ($\bar{x}_i'$)

	1	2	3	4	5	6
Group D	12.7	40.4	31.6	17.2	299.2	282.6
Group F	12.2	38.3	32.1	9.9	396.5	327.2
Group G	11.7	37.1	31.7	3.3	196.4	434.0

Inverse of Pooled Cov. Matrix (S^{-1})

	1	2	3	4	5	6
1	854.57	-265.05	-321.27	1.327	-.2416	-.1269
2		82.53	99.48	-.353	.0726	.0341
3			127.06	-.519	.1026	.0550
4				.036	-.0009	-.0019
5					.0003	.0001
6						.0003

The three standardized linear discriminant functions are

$$V_1^* = 14.30X_1 - 4.06X_2 - 6.53X_3 + .14X_4 - .012X_5 - .012X_6$$

$$V_2^* = 8.01X_1 - 2.06X_2 - 3.29X_3 + .16X_4 - .003X_5 - .014X_6$$

$$V_3^* = -4.34X_1 + 1.56X_2 + 2.53X_3 + .09X_4 - .009X_5 - .008X_6$$

between groups D and F, Dand G and groups F and G respectively. The

mean discriminant values for each group determined by

$$\bar{v}_i^* = d^{*\prime}\bar{x}_j \qquad i = 1, 2, 3; \quad j = D, F, G.$$

are

	$\bar{v}_1^*$	v_2^*	v_3^*
Group D	-193.575	-87.833	90.035
Group F	-197.979	-91.146	90.089
Group G	-198.145	-93.626	86.670

The first discriminant function gives the rule for distinguishing group D from group F when

$$V_1^* \geq \frac{-193.575-197.979}{2} = -195.777$$

Similarly, group D is distinguished from group G by the second discriminant function when

$$V_2^* \geq \frac{-87.833-93.626}{2} = -90.730$$

Therefore, the method of classification for group D is to assign the new patient to this group when her discriminant values satisfy $v_1^* \geq -195.777$ and $v_2^* \geq -90.730$. Correspondingly, the rule for group F is $v_1^* \leq -195.777$ and $v_3^* \geq 88.330$, and for group G her values must satisfy $v_2^* \leq -90.730$ and $v_3^* \leq 88.380$.

A test patient, number 31 of group D, will illustrate the classification. Her p measurements are

$$x^{31} = (13.6, 43.5, 31.5, 7.4, 252.7, 339.0)$$

which gives her the three discriminant values

$$v_1^{*31} = -194.219$$

$$v_2^{*31} = -88.993$$

$$v_3^{*31} = 89.018$$

Then according to the previously defined rules, this individual is correctly assigned to group D.

The determination of the errors of classification [14] were obtained once the variances and covariances of the discriminant functions were known. Since the ith discriminant function was standardized by that pair of populations' Mahalanobis D^2, then the variances of each V_i^* is one. The covariance

$$\begin{aligned}
\text{cov}(V_1^*, V_2^*) &= E(V_1^* - E(V_1^*))(V_2^* - E(V_2^*))' \\
&= E(d_1^{*\prime}(X-\mu_1))(d_2^{*\prime}(X-\mu_2))' \\
&= d_1^{*\prime}\Sigma d_2^* \\
&= d_1^{*\prime}\Sigma\left(\frac{d_2}{\sqrt{D_2^2}}\right) \\
&= \left(\frac{1}{\sqrt{D_2^2}}\right) d_1^{*\prime}\ \Sigma(\Sigma^{-1}(\mu_1-\mu_3)) \\
&= \left(\frac{1}{\sqrt{D_2^2}}\right) d_1^{*\prime}(\mu_1-\mu_3)
\end{aligned}$$

may be estimated by

$$\begin{aligned}
\hat{\text{cov}}(V_1^*, V_2^*) &= \left(\frac{1}{\sqrt{D_2^2}}\right)(\bar{v}_{1D}^{\,*} - \bar{v}_{1G}^{\,*}) \\
&= \left(\frac{1}{\sqrt{33.563}}\right)(-193.575 + 198.145) \\
&= .789
\end{aligned}$$

Similarly, $\hat{\text{cov}}(V_1^*, V_3^*) = \left(\frac{1}{\sqrt{D_3^2}}\right)(\bar{v}_{1P}^{\,*} - \bar{v}_{1G}^{\,*})$

$$\begin{aligned}
&= \left(\frac{1}{\sqrt{14.565}}\right)(.166) \\
&= .044
\end{aligned}$$

and $\hat{cov}(V_2^*, V_3^*) = \left(\frac{1}{\sqrt{D_3^2}}\right)(\bar{v}_{2F}^* - \bar{v}_{2G}^*)$

$= \left(\frac{1}{\sqrt{14.565}}\right)(2.48)$

$= .650$

The correlation matrix of the three discriminant functions is then

	V_1^*	V_2^*	V_3^*
V_1^*	1	.789	.044
V_2^*		1	.650
V_3^*			1

The probability of an incorrect classification of a group D patient is therefore

$Pr(V_1^* < -195.777 \text{ or } V_2^* < -90.730) = Pr\left(Z < \frac{-195.777-(-193.575)}{1}\right) +$

$Pr\left(Z < \frac{-90.730-(-87.833)}{1}\right) - Pr(Z<(195.777+193.575) \text{ and } Z<(-90.730+$

$87.833))_{r=.789}$

$= .0138 + .0019 - .0013$

$= .0144$

where the first two probabilities are obtained from univariate normal tables and the third from a tabulated function for computing bivariate normal probabilities found in Owen [13].

Similarly, the misclassification probability of a group F individual is

$Pr(V_1^* > -195.777) + Pr(V_3^* < 88.380) - Pr(V_1^* > -195.777 \text{ and } V_3^* < 88.380)_{r=.044}$

$= Pr(Z>2.202) + Pr(Z<-1.709) - Pr(Z>2.202 \text{ and } Z<-1.709)_{r=.044}$

$= .0138 + .0437 - .0033$

$= .0542$

And the probability of the wrong assignment of a group G patient is

$Pr(V_2^*>-90.730) + Pr(V_3^*>88.380) - Pr(V_2^*>-90.730 \text{ and } V_3^*>88.380)_{r=.650}$

$= .0019 + .0437 - .0012$

$= .0444$

Each of the above probabilities is conditional on the assumption that the patient to be classified is either a group D, group F or group G individual and no other possibility. Given this, we may then expect 98% of the patients we classify as group D patients to be correct, 94% assigned to group F to be correct and 95% of the allocations to group G to be correct.

However, we feel it necessary to treat these results with caution since there existed heterogeneity among the dispersions of the populations.

3.3 *Canonical Analysis of Discrimination*

A direct extension of the reasoning employed in the development of Fisher's discriminant function is the canonical analysis of discriminance. The general problem is to determine m(<p) linear combinations of the p variates which best separate the populations. Again this is the stipulation of maximizing the between populations dispersion with

respect to the within populations dispersion. Rao [14] has shown that the coefficient vectors

$$d^i = (d_1{}^i, d_2{}^i, \ldots, d_p{}^i) \qquad i = 1, 2, \ldots, m$$

of the m linear combinations of the random variables are the solutions of

$$(B-\lambda_i\Sigma)d^{i'} = 0$$

where B is the dispersion matrix of the k population means, Σ is the common covariance matrix of the p variates and λ_i is a root of the equation

$$|B-\lambda_i\Sigma| = 0$$

Thus the coefficient vectors, d^i's, are the eigenvectors associated with the eigenvalues, λ_i's, of the asymmetric matrix $\Sigma^{-1}B$. These eigenvalues and eigenvectors are termed canonical roots and canonical vectors, respectively. The linear combinations

$$T_i = d^{i'}X$$

they produce are called canonical variates.

Geometrically, the test space is transferred into a space defined by new axes whose directions are indicated by the canonical vectors. Each direction or axis is perpendicular to the others in the sense that any two canonical variates, $T_i = d^{i'}X$ and $T_j = d^{j'}X$ $i \neq j$, are independent. That is,

$$\begin{aligned} \text{cov}(T_i, T_j) &= E(T_i-E(T_i))(T_j-E(T_j))' \\ &= E(d^{i'}(X-\mu))(d^{j'}(X-\mu))' \\ &= d^{i'}\Sigma\, d^j \\ &= 0 \end{aligned}$$

However, there is an upper limit to the number of dimensions that can be represented by the canonical axes. When the number of populations is less than or equal to the number of random variables ($k \leq p$), the transformed space is restricted to (k-1) dimensions. This is exemplified by there only being (k-1) non-zero canonical roots of the $\Sigma^{-1}B$ matrix. When $k>p$, the dimensionality of the canonical space can extend only to the full test space of p dimensions.

Each canonical root is directly proportional to the amount of variation among the groups in the direction of its associated canonical vector. Therefore our best discriminating linear combination is the one associated with the largest canonical root. That is,

$$T_{(1)} = d^{(1)'}X$$

where a bracketed subscript or superscript denotes ordering with respect to magnitude of the canonical roots. Similarly, the second canonical variate

$$T_{(2)} = d^{(2)'}X$$

supplies the next best discriminating function, and so on. Therefore the zero roots of the $\Sigma^{-1}B$ matrix (if they exist) indicate that no further variability is left to be explained and hence do not introduce further functions.

Using Fisher's discriminant function, the discriminating boundary

$$[\Sigma^{-1}(\mu_1-\mu_2)]'X = \text{constant}$$

where the constant is

$$\tfrac{1}{2}\{[\Sigma^{-1}(\mu_1-\mu_2)]'\mu_1 + [\Sigma^{-1}(\mu_1-\mu_2)]'\mu_2\}$$

defines a hyperplane which bisects the line joining the means μ_1 and μ_2 of the two populations. On the other hand, with canonical analysis we discriminate along the lines of closest fit to the k means rather than the line joining each pair of means. Indeed, the first canonical axis defines the "best" line fitting the means, the second, orthogonal to the first, defines the next optimum line, and continuing in the same manner until the canonical space is filled. Therefore, when k = 2, the discriminating variate derived from the canonical analysis will be identical to Fisher's discriminator because the line of closest fit between two populations is that line which joins their means. However, when $k \geq 3$ the two techniques, in general, are not equivalent.

3.3.1 *Application of Canonical Analysis of Discriminance*, k = 2

We apply the analysis to groups F and G.

Mean Dispersion Matrix ($\hat{B}$)

	1	2	3	4	5	6
1	.13	.31	.09	1.71	51.52	-27.50
2		.72	.21	4.00	120.26	-64.19
3			.06	1.16	34.92	-18.64
4				22.14	665.81	-355.38
5					20018.60	-10684.97
6						5703.12

Inverse of Pooled Cov. Matrix ($\hat{\Sigma}^{-1}$)

	1	2	3	4	5	6
1	1340.96	-413.24	-486.78	6.83	-.2700	-.0196
2		127.71	149.90	-1.97	.0794	.0022
3			181.59	-2.57	.1068	.0188
4				.27	-.0021	-.0032
5					.0003	.0003
6						.0003

Since $k < p$, we may construct only $(k-1) = 1$ discriminant function. The non-zero canonical root of $\hat{\Sigma}^{-1}\hat{B}$ is

$$\lambda_{(1)} = 14.001$$

with its canonical vector

$$d^{(1)} = (-.931, .236, .259, -.106, -.002, .0024).$$

We wish to standardize the canonical variate so that its within groups variance is one as was done with Fisher's coefficient vector in section 3.2.1. The variance of the canonical variate is

$$\text{var}(T_{(1)}) = d^{(1)'} \Sigma\, d^{(1)}$$

which is estimated to be .079. Therefore, the standardized canonical vector is

$$d^{(1)*} = \left(\frac{1}{d^{(1)'} \hat{\Sigma}\, d^{(1)}} \right)^{\frac{1}{2}} d^{(1)}$$

$$= (-3.31, .84, .92, -.38, -.007, .0086)$$

which yields the canonical variate

$$T_{(1)}^{*} = -3.31X_1 + .84X_2 + .92X_3 - .38X_4 - .007X_5 + .0086X_6$$

identical to Fisher's V* of section 3.2.1.

Thus a parallel analysis to that section would follow.

3.3.2 *Application of Canonical Analysis of Discriminance*, k = 3

Groups D, F and G are the groups under consideration.

Mean Dispersion Matrix ($\hat{B}$)

	1	2	3	4	5	6
1	.25	.82	-.035	3.47	26.43	-38.01
2		2.75	-.184	11.41	62.55	-119.54
3			.065	-.56	17.16	1.25
4				48.13	342.46	-522.14
5					10011.77	-5431.29
6						6052.89

Inverse of Pooled Cov. Matrix ($\hat{\Sigma}^{-1}$)

	1	2	3	4	5	6
1	854.57	-265.05	-321.27	1.327	-.2416	-.1269
2		82.53	99.48	-.354	.0726	.0341
3			127.06	-.519	.1026	.0550
4				.036	-.0009	-.0019
5					.0003	.0001
6						.0003

The (k-1) = 2 non-zero canoncial roots of the $\hat{\Sigma}^{-1}\hat{B}$ matrix are

$$\lambda_{(1)} = 8.4739$$

$$\lambda_{(2)} = 2.7711$$

with their associated canonical vectors

$$d^{(1)} = (.896, -.235, -.375, -.016, -.0004, -.0014)$$

$$d^{(2)} = (-.865, .264, .427, -.0008, .0011, .00004) .$$

The estimated variances of the two canonical variates are

$$\hat{var}(T_{(1)}) = d^{(1)'}\hat{\Sigma}\, d^{(1)} = .0102$$

and $$\hat{var}(T_{(2)}) = d^{(2)'}\hat{\Sigma}\, d^{(2)} = .0053$$

Therefore, the standardized canonical vectors (such that the within groups variance of the canonical variates is one) are

$$d^{(1)*} = \left(\frac{1}{\sqrt{\hat{var}(T_{(1)})}}\right) d^{(1)}$$

$$= (8.859, -2.322, -3.709, .162, -.0039, -.0141)$$

and $$d^{(2)*} = \left(\frac{1}{\sqrt{\hat{var}(T_{(2)})}}\right) d^{(2)}$$

$$= (-11.932, 3.649, 5.890, -.0105, -.0145, .0006)$$

Thus the two canonical variates best separating the groups are

$$T_{(1)}^* = d^{(1)*'}X$$

$$= 8.859X_1 - 2.322X_2 - 3.709X_3 + .162X_4 - .0034X_5 - .0141X_6$$

and $$T_{(2)}^* = d^{(2)*'}X$$

$$= -11.932X_1 + 3.649X_2 + 5.890X_3 - .0105X_4 - .0145X_5 + .006X_6$$

A test of the significance of the dispersion of group means along each dimension was made. That is, we may test whether all of the roots after the jth say, can be given zero values. Bartlett's (1954) [15] test of this hypothesis,

$$H : \lambda_{(j+1)} = \lambda_{(j+2)} = \cdots = \lambda_{(f)} = 0 , \quad f = \min\{p, k-1\}$$

is based on the fact that under the null hypothesis

$$\{(n-1) - (\frac{p+k}{2})\} \ln \{\prod_{i=j+1}^{f} (1+\lambda_{(i)})\}$$

is approximately a chi-square variable with (p-j)(k-j-1) degrees of freedom. Therefore, we have

j+1	$\lambda_{(j+1)}$	$\prod_{i=j+1}^{f}(1+\lambda_{(i)})$	$\{24.5\}\ln\{\Pi(1+\lambda_{(i)})\}$	χ^2df	Critical Value $\chi^2_{\alpha=.05}$
1	8.4739	35.7270	87.6097	12	21.026
2	2.7711	3.7711	32.5205	5	11.071

We conclude that both canonical variates explain significant amounts of variation among the groups.

Then to classify an individual, we compare her canonical values to each group mean's canonical values.

The mean canonical variates for groups D, F and G are

Group	$\bar{t}_{(1)}^* = d^{(1)*'}\bar{x}$	$\bar{t}_{(2)} = d^{(2)*'}\bar{x}$
D	-100.9656	186.4475
F	-104.4725	189.1078
G	-106.7450	186.0397

These values represent coordinates in the transformed space with axes $T_{(1)}^*$ and $T_{(2)}^*$. We may therefore depict these three groups in a two-dimensional diagram to illustrate their relationship with one another. It was convenient to centre the groups about the grand mean canonical variates,

$$\bar{\bar{T}}_{(i)}^* = d^{(i)*'}\bar{\bar{X}} \quad , \qquad i = 1, 2$$

to emphasize the degree of separation attributed to each variate.

Let the transformed standardized canonical variates be denoted by

$$\bar{Z}_{(i)j} = (\bar{T}_{(i)j}^{*} - \bar{\bar{T}}_{(i)}^{*}) \qquad i = 1, 2; \quad j = D, F, G.$$

The coordinates to be considered are now

Group	$\bar{z}_{(1)}$	$\bar{z}_{(2)}$
D	3.0954	-.7508
F	-.4115	1.9095
G	-2.6840	-1.1587

where $(\bar{\bar{t}}_{(1)}^{*}, \bar{\bar{t}}_{(2)}^{*}) = (-104.0610, 187.1984)$.

Furthermore, since the canonical variates were constructed to have unit variance, a one standard deviation contour encircles each group mean canonical variate. The representation of groups D, F and G in this manner is found in Figure 3.3.2.1.

A classification procedure of a new patient is to assign the individual to the group in which her canonical coordinates, $(T_{(1)}^{*}, T_{(2)}^{*})$, and the group's mean canonical coordinates $(\bar{\bar{T}}_{(1)}^{*}, \bar{T}_{(2)}^{*})$, are closest in terms of Euclidean distance.

For example, we employed patient number 31 of group D to illustrate this procedure. Her canonical variates are

$$t_{(1)}^{*31} = d^{(1)*'} x^{31} = -102.0991$$

and

$$t_{(2)}^{*31} = d^{(2)*'} x^{31} = 186.0171$$

which give the transformed values of

$$(z_{(1)}^{31}, z_{(2)}^{31}) = (1.9619, -1.1813).$$

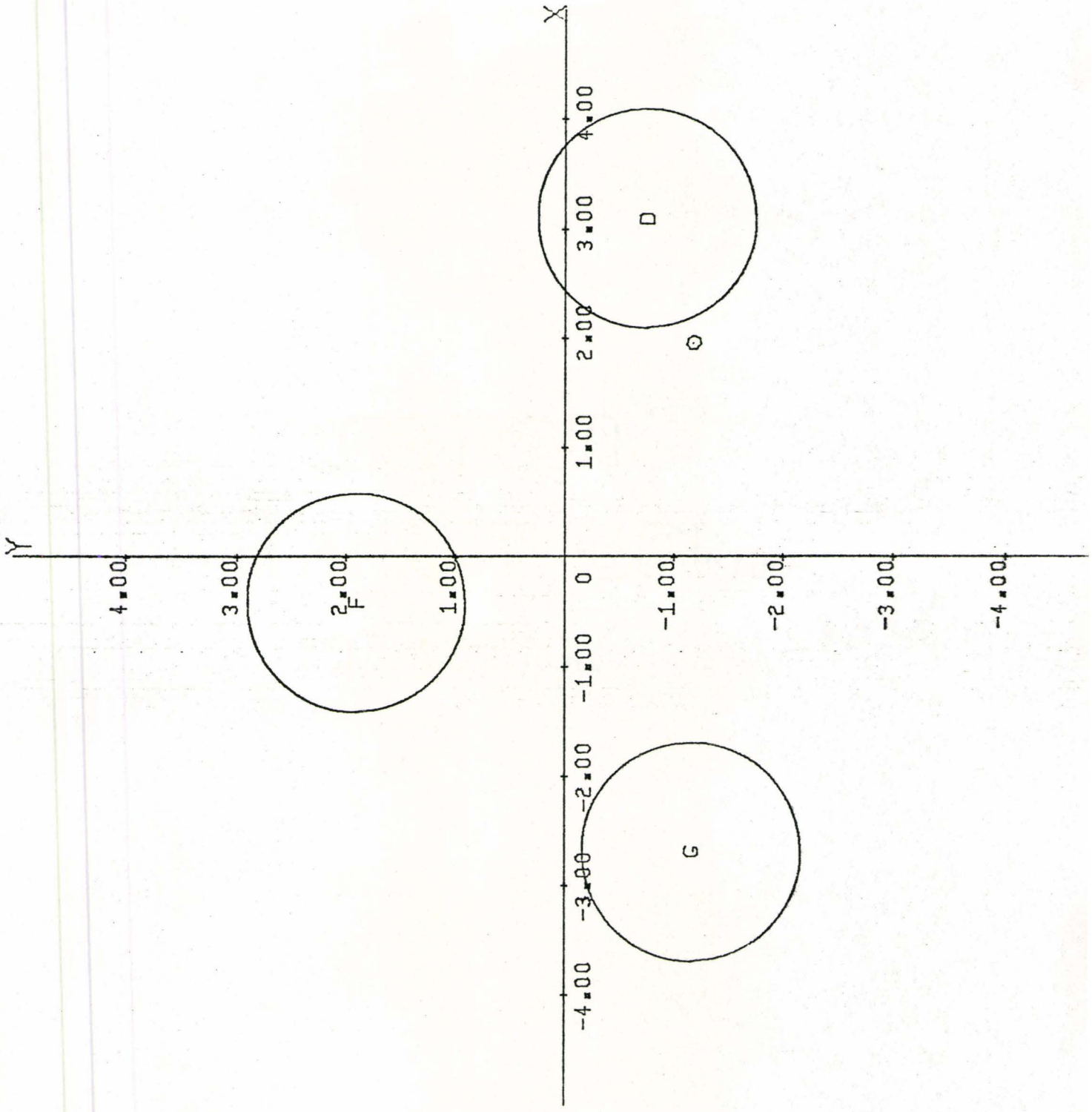


Figure 3.3.2.1. The group means of D, F and G along the canonical axes with one standard deviation contours. X=first canonical variate, Y=second canonical variate.

This coordinate point is represented in Figure 3.3.2.1 by a small circular point. Her distance from each of the three groups is determined from the well known planar distance between two points

$$\{(z_{(1)}-\bar{z}_{(1)j})^2 + (z_{(2)}-\bar{z}_{(2)j})^2\}^{\frac{1}{2}} \qquad j = D, F, G$$

From group D, patient number 31 lies

$$\{(1.9619-3.0954)^2 + (-1.1813+.7508)^2\}^{\frac{1}{2}} = 1.2125 \quad \text{units}$$

From group F, she is

$$\{(1.9619+.4115)^2 + (-1.1813-1.9095)^2\}^{\frac{1}{2}} = 3.8969 \quad \text{units,}$$

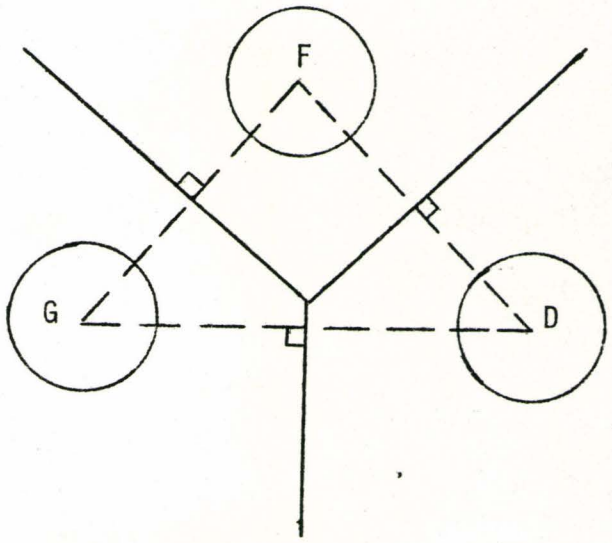
and from group G, this patient is

$$\{(1.9619+2.6840)^2 + (-1.1813+1.1587)^2\}^{\frac{1}{2}} = 4.646 \quad \text{units.}$$

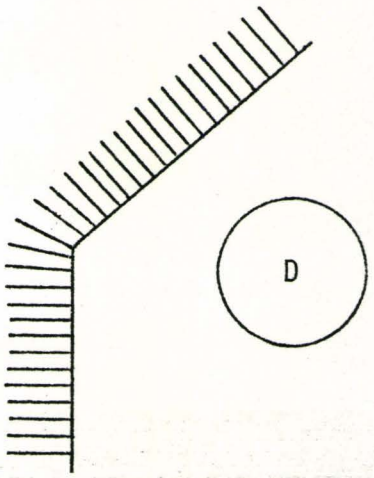
We therefore classify this individual as a group D patient.

In essence, we have divided the canonical coordinate plane into three regions. Each region, associated with a given group, has the perpendicular bisectors of the lines joining its mean to the other means as its boundaries.

Diagrammatically:



The probability of misclassification is then an integral of a bivariate normal distribution over a complex region. Namely, the probability of misclassifying a group D patient is the integral over the region denoted below.



Rather than evaluate this integral, it is reasonable to accept a classification of an individual only if her coordinates are within a predetermined confidence contour. If we decided to specify a 90% classification probability then the 90% confidence contour has a radius of 1.645. With the previous distances determined for patient number 31, we may rightly allocate her to group D because she falls within this contour (1.2125<1.645).

Using this approach, we will not classify a patient to any of the groups if all three distances to the group means are greater than 1.645. This is justifiable because a patient whose distances are very large from all groups should not be blindly allocated to a group even though she is within that groups region. This individual, in all likelihood, would not be from any of these populations at all.

This discriminatory approach did not result in a different classification of patient number 31 than Fisher's function's classification. However, if the computational facilities are available this technique is favoured due to its more informative nature with respect to variability among the groups.

3.3.3 *Application of Canonical Analysis of Discriminance*, k = 4

All four groups are now to be analyzed.

The preliminary testing yielded the following results.

Hypothesis	Test Statistic	Degrees of freedom	$\lambda^2_{\alpha=.05}$
$H:\Sigma_1=\ldots=\Sigma_k$	h = 141.041	63	82.529
$H:\mu_1=\ldots=\mu_k$	u = 78.187	18	28.869

Additional heterogeneity among the dispersions is introduced by group E. Caution must be applied in the interpretation of the remaining analysis. However, we can expect good discrimination of the groups with the (k-1) = 3 canonical variates because of the highly significant difference among the means.

The necessary matrices for the computations are

Mean Dispersion Matrix ($\hat{B}$)

	1	2	3	4	5	6
1	.19	.65	-.054	1.89	6.50	-21.07
2		2.32	-.267	5.57	-11.40	-59.34
3			.087	.23	27.24	-5.22
4				40.58	450.42	-433.25
5					12481.69	-5847.31
6						4888.88

Inverse of Pooled Cov. Matrix ($\hat{\Sigma}^{-1}$)

	1	2	3	4	5	6
1	955.10	-295.92	-360.60	1.406	-.2687	-.1251
2		92.05	111.29	-.363	.0805	.0315
3			142.96	-.575	.1177	.0554
4				.046	-.0011	-.0024
5					.0003	.0001
6						.0003

The non-zero canonical roots of the $\hat{\Sigma}^{-1}\hat{B}$ matrix are

$$\lambda_{(1)} = 7.2103$$

$$\lambda_{(2)} = 5.0296$$

$$\lambda_{(3)} = .0751$$

with their associated canonical vectors

$$d^{(1)} = (.8982, -.2181, -.3810, .0239, -.0003, -.0020)$$

$$d^{(2)} = (-.8658, .2583, .4286, -.0012, .0010, .0002)$$

$$d^{(3)} = (-.9045, .1775, .3856, .0408, -.0015, .0004)$$

The variances of the three canonical variates are estimated to be

$$\hat{var}(T_{(1)}) = .0169$$

$$\hat{var}(T_{(2)}) = .0039$$

$$\hat{var}(T_{(3)}) = .1803.$$

Therefore the three standardized canonical variates which will act as our discriminating functions are

$$T_{(1)}^{*} = d^{(1)*'}X$$

$$= 6.912X_1 - 1.679X_2 - 2.932X_3 + .184X_4 - .0024X_5 - .0156X_6$$

$$T_{(2)}^{*} = d^{(2)*'}X$$

$$= -13.794X_1 + 4.116X_2 + 6.828X_3 - .019X_4 + .0161X_5 + .0026X_6$$

and $$T_{(3)}^{*} = d^{(3)*'}X$$

$$= -2.130X_1 + .418X_2 + .908X_3 + .096X_4 - .0036X_5 + .0010X_6.$$

The test of the significance of each canonical root yields the following:

j+1	λ	$\prod_{i=j+1}^{f}(1+\lambda_i)$	{34}$\ln\{\prod_{j+1}^{f}(1+\lambda_i)\}$	df. for χ^2	Critical $\chi^2_{.05}$
1	7.2103	53.2226	135.1324	18	28.869
2	5.0296	6.4824	63.5491	10	18.307
3	.0751	1.0751	2.4621	4	9.488

The third canonical root is non-significant which implies that, when the persons and group means are projected onto this canonical variate, the differences among the group means are small relative to the differences among persons within a group. The first two canonical roots are highly significant which indicates that good separation among the groups will occur along these dimensions.

To illustrate the separation of the data imposed by each canonical variate, the group mean canonical variates, with their unit standard deviation contours, were plotted against each pair of possible axes combinations. Figure 3.3.3.1. is the first canonical variate (X-axis) versus the second canonical variate (Y-axis). The larger degree of separation along the X-axis reveals the greater discriminatory power of the first canonical variate. The transformed mean canonical variates (transformed by the grand mean canonicalvariate)

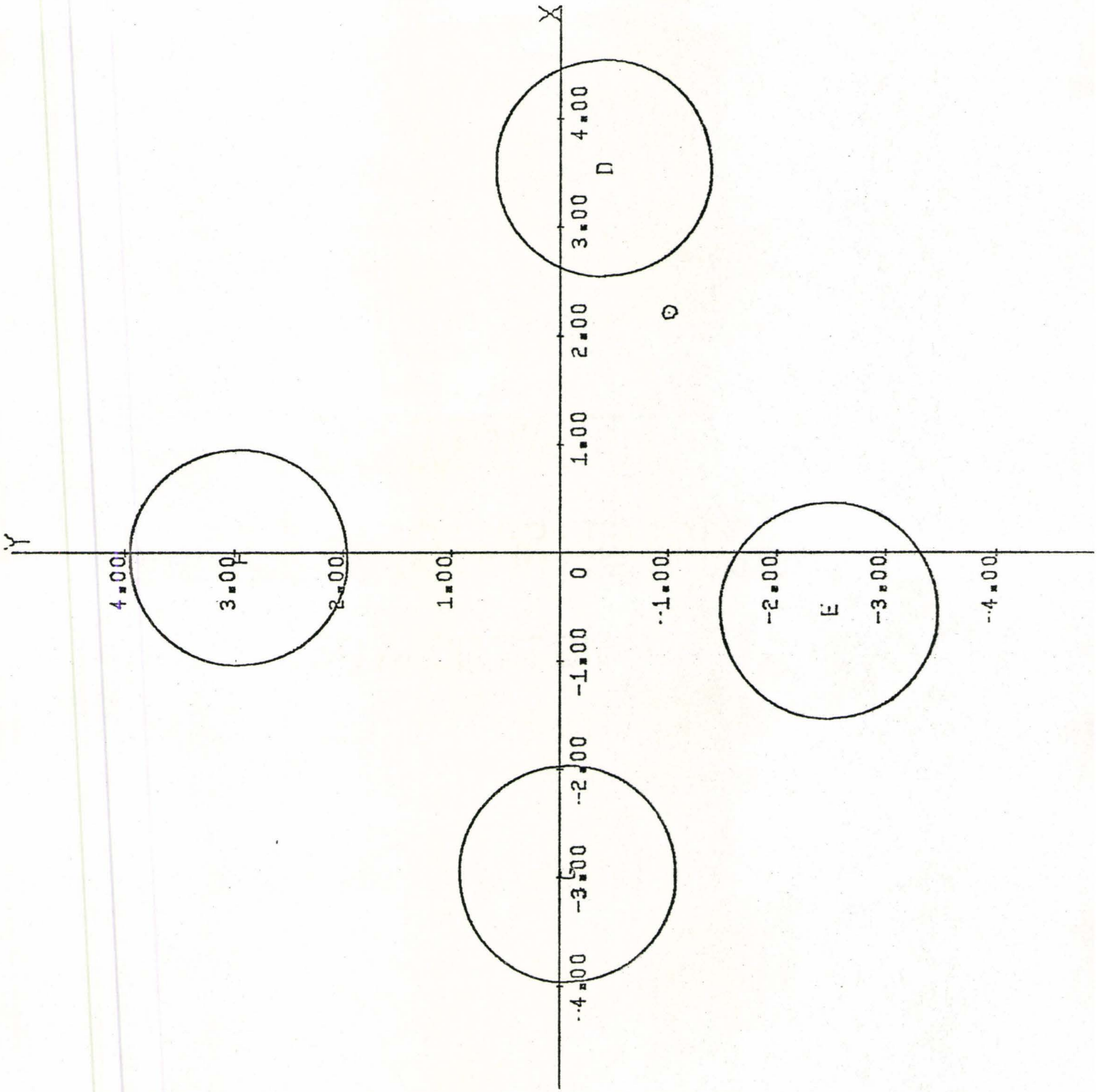


Figure 3.3.3.1. The group means of D, E, F and G along the canonical axes with one standard deviation contours. X=first canonical variate, Y=second canonical variate.

are,

	$\bar{z}_{(1)}$	$\bar{z}_{(2)}$	$\bar{z}_{(3)}$
Group D	+3.5438	-.4039	+.1890
Group E	-.5382	-2.4812	-.2722
Group F	-.0428	+2.9603	-.1952
Group G	-2.9628	-.0752	+.2784
Constant Term (Grand Mean)	-78.2939	212.6007	19.2126

Figure 3.3.3.2 is the projection of the mean canonical variates onto the $T_{(1)}$ - $T_{(3)}$ plane where the first canonical variate, $T_{(1)}$ is the abscissa. This clearly indicates the insignificant variation explained along this third canonical variate. Figure 3.3.3.3 is the representation of the $T_{(2)}$ - $T_{(3)}$ plane with the second canonical variate being the abscissa.

The apparent function of the first canonical variate is to alienate groups D and G to the fullest extent, while groups E and F are separated by the second canonical variate. The unsuccessful role of the third canonical variate appears to be to distinguish between the pairs of groups D, G and E, F. This is a good illustration of the capabilities of the canonical variates.

Test patient, number 31 of group D, has the canonical values

$$(z_{(1)}^{31}, z_{(31)}^{31}, z_{(3)}^{31}) = (2.2229, -1.0114, -1.2156)$$

We then determined her three dimensional distance from each of the group mean canonical variates, in order to classify her, by the

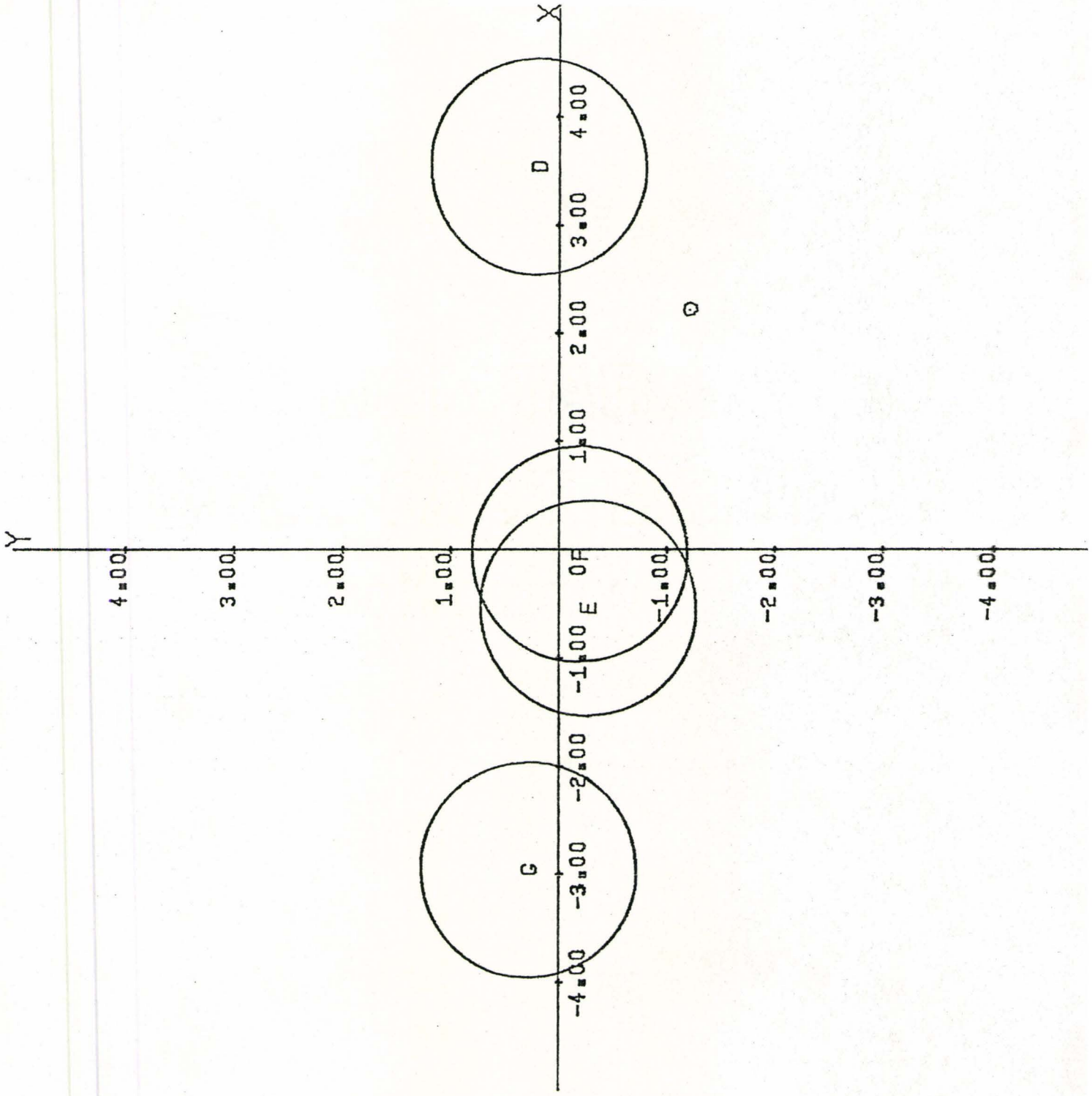


Figure 3.3.3.2. As in Fig. 3.3.3.1 except X=first canonical variate, Y=third canonical variate.

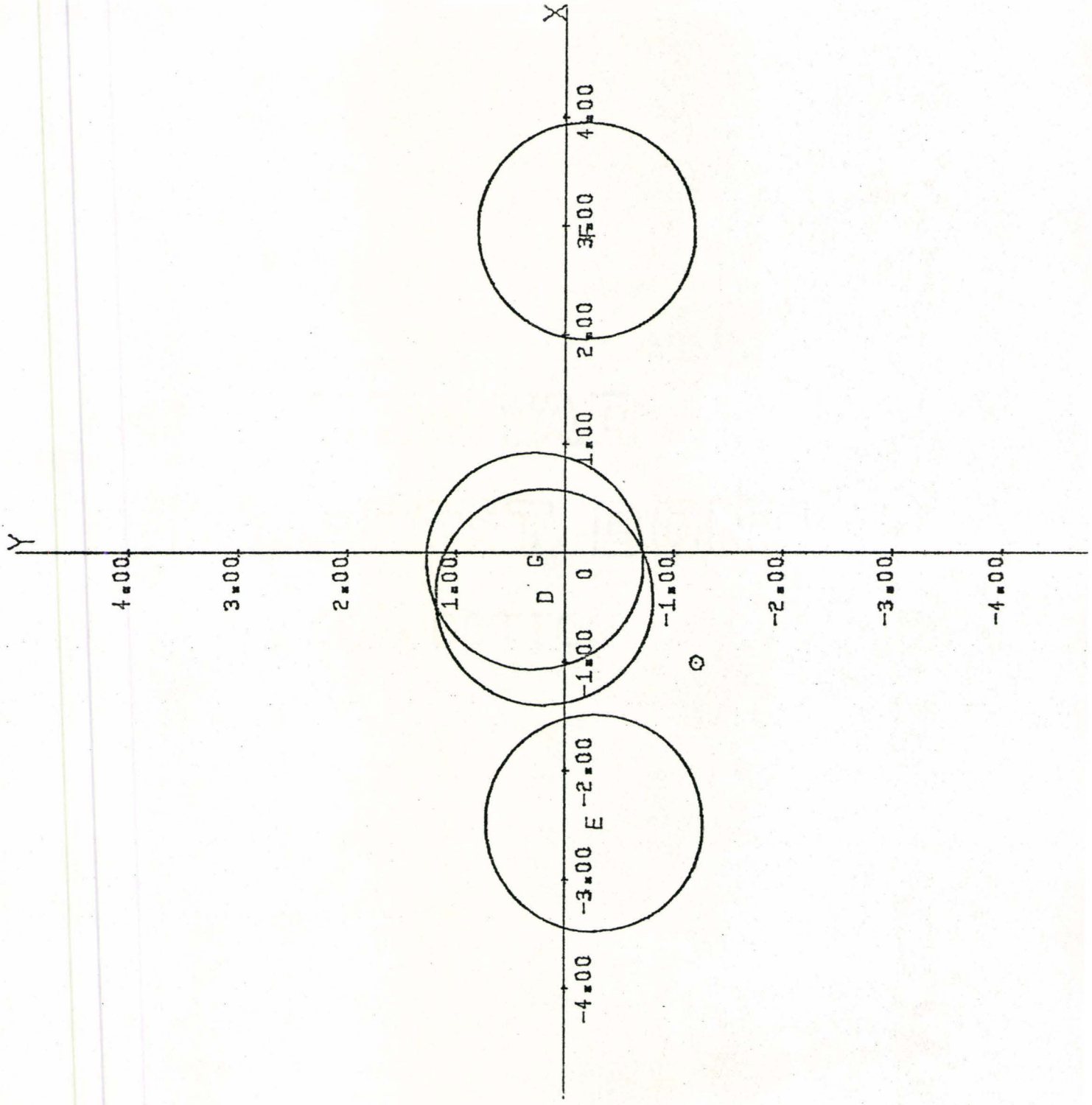


Figure 3.3.3.3. As in Fig. 3.3.3.1 except X=second canonical variate, Y=third canonical variate.

equation

$$\{(T_1^*-\bar{T}_{1i}^*)^2 + (T_2-\bar{T}_{2i}^*)^2 + (T_3-\bar{T}_{3i}^*)^2\}^{\frac{1}{2}} = \text{distance from group } i$$

$i = D, E, F, G.$

These are:

Distance from Group D = 2.0216 units

Group E = 3.2642 units

Group F = 4.6850 units

Group G = 5.4772 units

She is then correctly classified as a group D patient.

The 90% contour about these trivariate normal points is a sphere of radius 2.5003. This value was obtained from the critical values tabulated forthe spherical normal distribution in Owen [13].

The test patient falls only within group D's 90% contour which leaves no ambiguity to which group she belongs.

When the dimension of the canonical space is greater than or equal to four (i.e. $\min\{p,k-1\}\geq 4$), one would restrict the classification rule to the first three canonical variates so that the confidence contours are easily determined. In practice, a great proportion of the variation is usually explained by the first three variates so that this approach would supply accurate results.

This analysis has a distinct advantage over Fisher's when $k = 4$ populations. With Fisher's discriminant functions we need $\binom{4}{2} = 12$ discriminators while canonical analysis accomplishes the same with three. Furthermore, the properties of the canonical variates make the computations easier (i.e. $\text{cov}(T_i, T_j) = 0$) and the graphical

representation of these variates enhances the understanding of the data dispersion.

However, when the dimensions do become large, the projection of the variates onto two-dimensional axes will be inadequate. The next section will attempt to supply us with alternatives.

4. HIGH-DIMENSIONAL REPRESENTATION OF DATA

4.1 *General*

Graphical portraying of data has long been a useful aid in its analysis. The plotting of the residuals of a time series model reveals a great deal of information concerning the model. Histograms often are the basis of underlying distribution assumptions.

The problem is not with univariate or bivariate data, but with data that may not be depicted by a planar coordinate plotting. To this end, the following procedures apply.

4.2 *Fourier Series*

This procedure is attributed to D.F. Andrews [1] who attempted to visualize high dimensional data in a space of functions. That is, each data point $X = (x_1, \ldots, x_k)$, is mapped into a fourier series function of the form

$$f_x(\tau) = (2)^{-\frac{1}{2}}x_1 + x_2 \sin\tau + x_3 \cos\tau + x_4 \sin 2\tau + x_5 \cos 2\tau + \ldots$$

and the function is plotted on the range $0 \leqslant \tau \leqslant 2\pi$.

So as to compare this approach with canonical analysis, the data points that were plotted were the canonical variates

$$T = (T_{(1)}, T_{(2)}, \ldots, T_{(f)}) \quad , \qquad f = \min\{p, k-1\}.$$

Andrews notes that the choice of which numbers to examine, here canonical variables, is based on the nature of the data and the objectives of the analysis. Our objective is differentiating among

the groups, so this selection of variables to use is exemplary.

4.2.1. *Properties of the Fourier Series Plots*

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

$$f_{\bar{t}}(\tau) = \frac{1}{n} \sum_{i=1}^{n} f_t{}^i(\tau)$$

As a result the average will appear like an average in this plot [1].

(ii) An analagous concern to canonical analysis is the distance between two functions to imply their "closeness" to one another. One measure of distance is

$$|| f_t{}^i(\tau) - f_t{}^j(\tau) || = \int_0^{2\pi} \{f_t{}^i(\tau) - f_t{}^j(\tau)\}^2 \, d\tau$$

One may then evaluate a test patient's distance from each of the groups by determining her functional distance from each of the group mean fourier series plots [1].

(iii) If the components of the data are uncorrelated with common variance σ^2, then

$$\text{var}(f_T(\tau)) = \sigma^2[\tfrac{1}{2}+\sin^2\tau+\cos^2\tau+\sin^2 2\tau+\cos^2 2\tau+\ldots].$$

If $f = \min\{p,k-1\}$ is odd this reduces to a constant, $\frac{1}{2}\sigma^2 f$; if f is even the variance lies between $\frac{1}{2}\sigma^2(f-1)$ and $\frac{1}{2}\sigma^2(f+1)$. In the first case the variance does not depend on τ and in the second the dependence on τ is slight. Thus the variability of the plotted

function is almost constant across the graph [1].

4.2.2. *Application of Andrews Technique, k = 4*

The group mean canonical variates,

$$(\bar{Z}_{(1)j}, \bar{Z}_{(2)j}, \bar{Z}_{(3)j}) \quad j = D, E, F, G$$

which are centred about the origin of the canonical axes, give the following four fourier series:

$$f_{\underline{Z}_D}(\tau) = 3.5438/\sqrt{2} - .4039 \sin\tau + .1890 \cos\tau$$

$$f_{\underline{Z}_E}(\tau) = -.5382/\sqrt{2} - 2.4812 \sin\tau - .2722 \cos\tau$$

$$f_{\underline{Z}_F}(\tau) = -.0428/\sqrt{2} + 2.9603 \sin\tau - .1952 \cos\tau$$

$$f_{\underline{Z}_G}(\tau) = -2.9628/\sqrt{2} - .0752 \sin\tau + .2784 \cos\tau$$

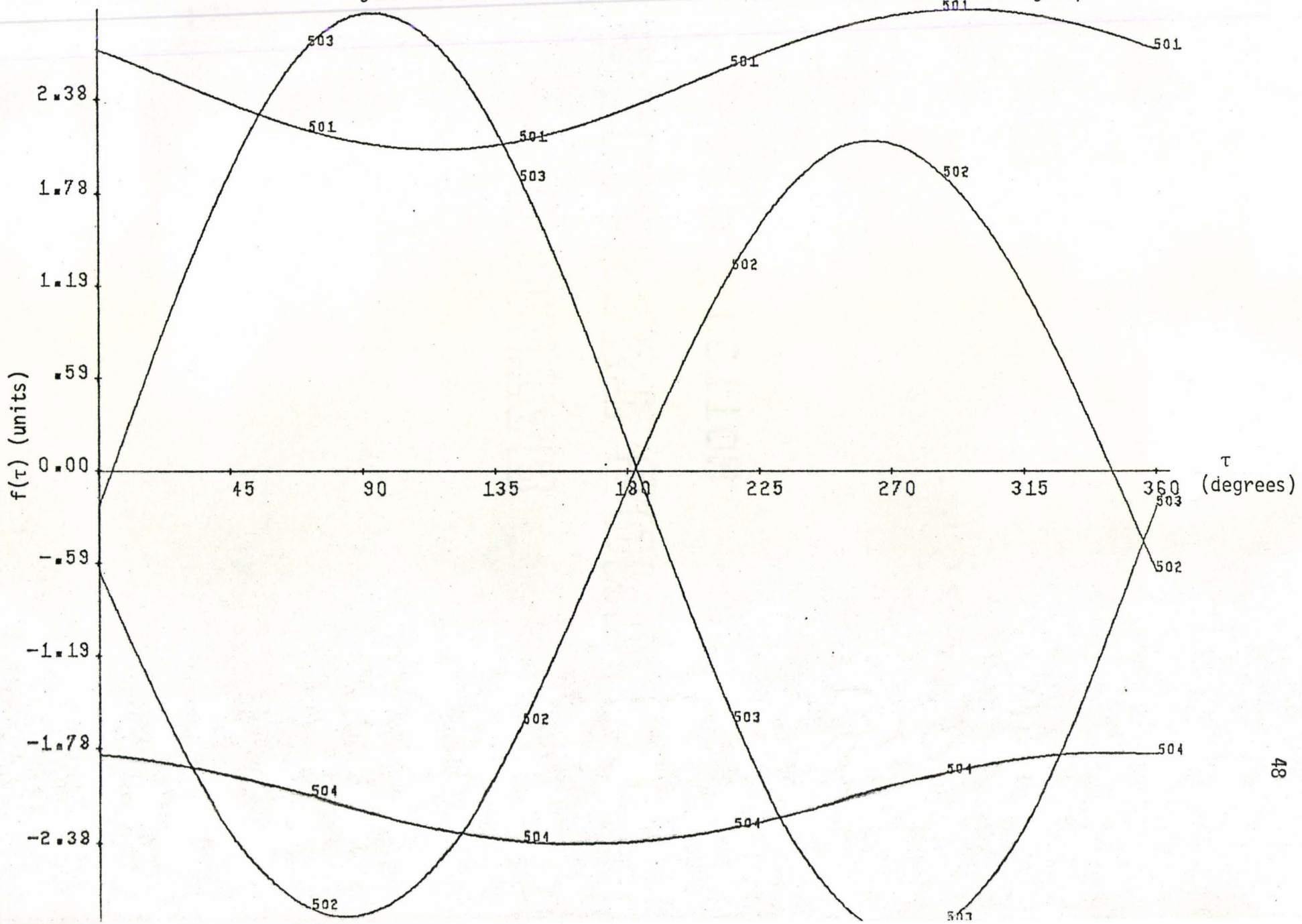
$$0 \leq \tau \leq 2\pi .$$

Figure 4.2.2.1 is a plotting of these functions. The functions in all subsequent plots are numbered by the following scheme:

(i) the group mean plots are the 500 series where 501, ..., 504 represents the groups D, E, F, G, respectively.

(ii) all other functions are the individual patient's fourier series where the first digit of the function number implies the group from which the patient originated and the second and third digit her patient number. For example, function number 131 is the plotting of patient

Figure 4.2.2.1. Fourier Series of three dimensional data. All groups.



number 31 of group D; function number 315 would be patient number 15 of group F; etc.

Recalling how the canonical variates separated the data, it is evident that they continue to do so in these plots. The first canonical variate which is the coefficient of the term $(2)^{-\frac{1}{2}}$ widely alienates groups D (501) and G (504) by separating their Y-axis intercepts. The second canonical variate accomplishes its separation of groups E and F (502, 503) by imposing a large positive sine term on group E versus a large negative sine term on group F. The third term of the fourier series representing the third canonical variate offers little influence on the functions which was expected because of its insignificant canonical root.

We plotted patient number 31 of group D among these group mean functions. Her fourier series is

$$f_{z31}(\tau) = 2.229(2)^{-\frac{1}{2}} - 1.0114 \sin\tau - 1.2156 \cos\tau, \quad 0 \leq \tau \leq 2\pi.$$

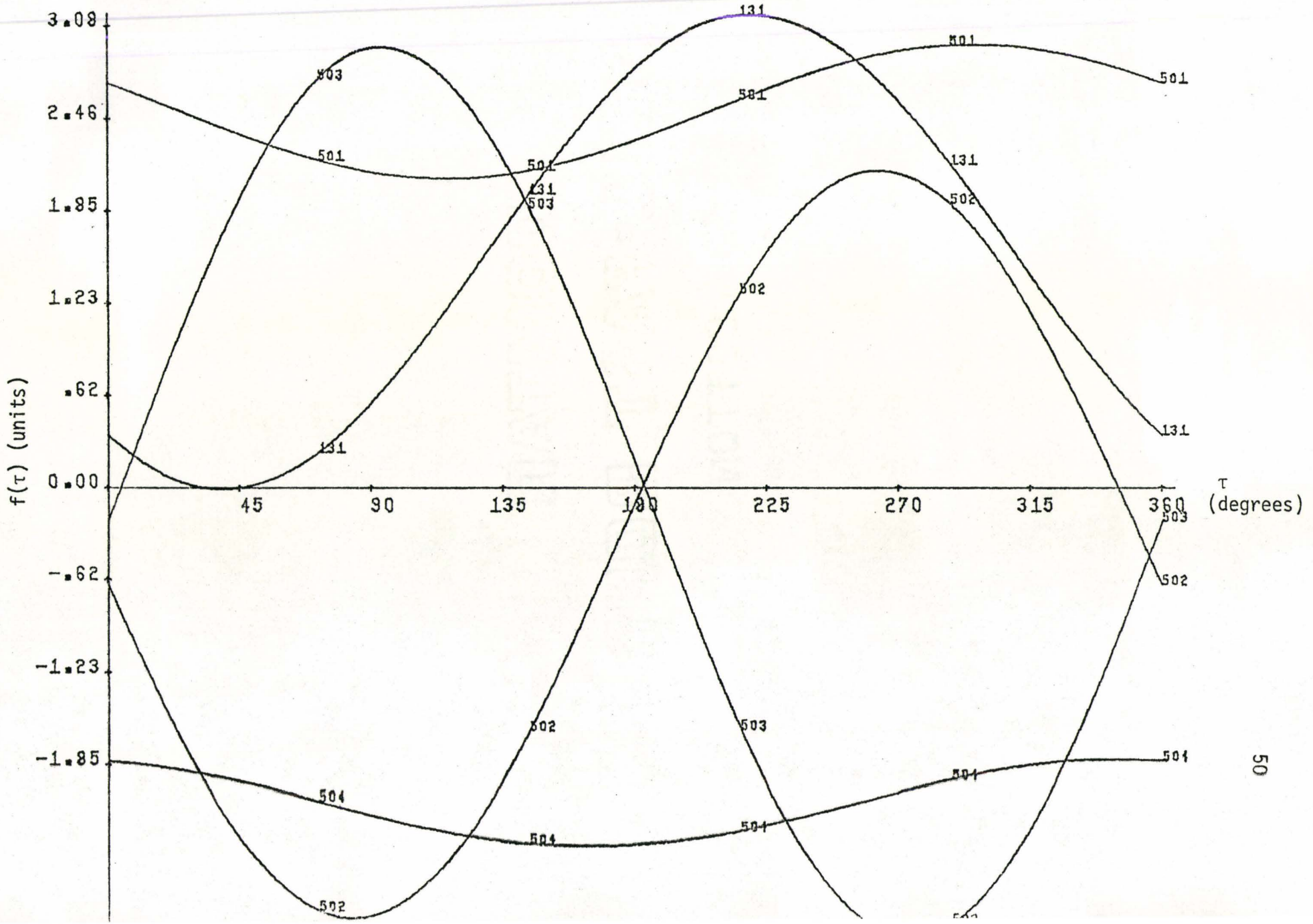
Figure 4.2.2.2 contains the five functions.

By inspection, it appears that patient number 31 belongs to either group D or group E; group D because for all τ values her function is positive as is group D's, and perhaps group E because the shape of her function is very close to group E's but not as severe.

We took a closer look at the individual plots. The variance of the plots may be estimated by

$$\hat{var}(f(\tau)) = \tfrac{1}{2}\,\hat{\sigma}_Z^2 f = (.5)(1)(3) = 1.5 \text{ units}$$

Figure 4.2.2.2. Fourier Series of all groups but with patient 131 added. Three dimension data.



since the variance of the standardized canonical variates is one. As was done for the plotting of the canonical variates against their canonical axes, a one standard deviation contour or band is plotted with each group mean fourier series. That is, for any τ, $0 \leq \tau \leq 2\pi$, the vertical distance between the band and the function is

$$(\tfrac{1}{2}\ \hat{\sigma}^2 f)^{\frac{1}{2}} = 1.2447 \text{ units.}$$

The test patient was then plotted along with these one standard deviation bands against each group mean function. Figures 4.2.2.3, 4.2.2.4, 4.2.2.5, 4.2.2.6 are these representations.

An overall 90% confidence band can be determined to emcompass the function. Andrews [1] shows that, for all values of τ, the probability of

$$|f_T(\tau) - f_{\mu_T}(\tau)|^2 \leq (\tfrac{f+1}{2})\ \sigma^2\ \chi^2_{\alpha,f}$$

where $\chi^2_{\alpha,f}$ denotes the upper α point with f degrees of freedom of the chi-square distribution, is approximately $1-\alpha$. Our 90% confidence band then becomes

$$\pm\ \{(\tfrac{f+1}{2})\ \hat{\sigma}^2\ \chi^2_{\alpha,f}\}^{\frac{1}{2}} = \pm\ \{(\tfrac{4}{2})(1)(6.251)\}^{\frac{1}{2}} = \pm\ 3.5358 \text{ units} .$$

Referring to Figure 4.2.2.3 it is easy to see that the patient completely lies within a band width of $\pm$ 3.5358 about group D. However, in Figure 4.2.2.4 the patient function extends outside this band about the group E function at approximately $100° < \tau < 140°$. In the other two plots, Figures 4.2.2.5 and 4.2.2.6, the patient exits the confidence band in many regions. Therefore an appropriate class-

Figure 4.2.2.3. Fourier Series of group D with one standard deviation band and patient 131. Three dimensional data.

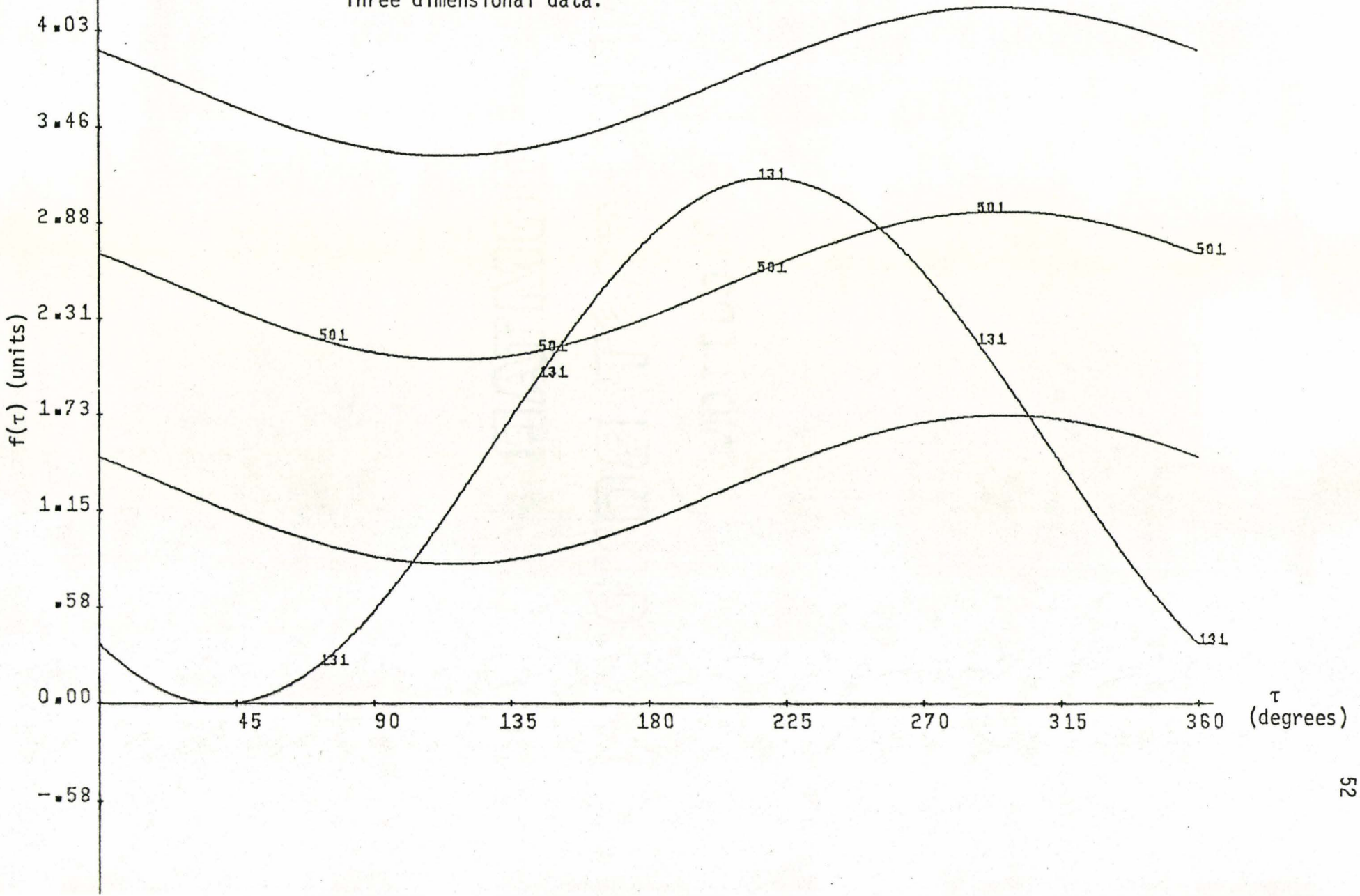


Figure 4.2.2.4. Fourier Series of group E with one standard deviation band and patient 131. Three dimensional data.

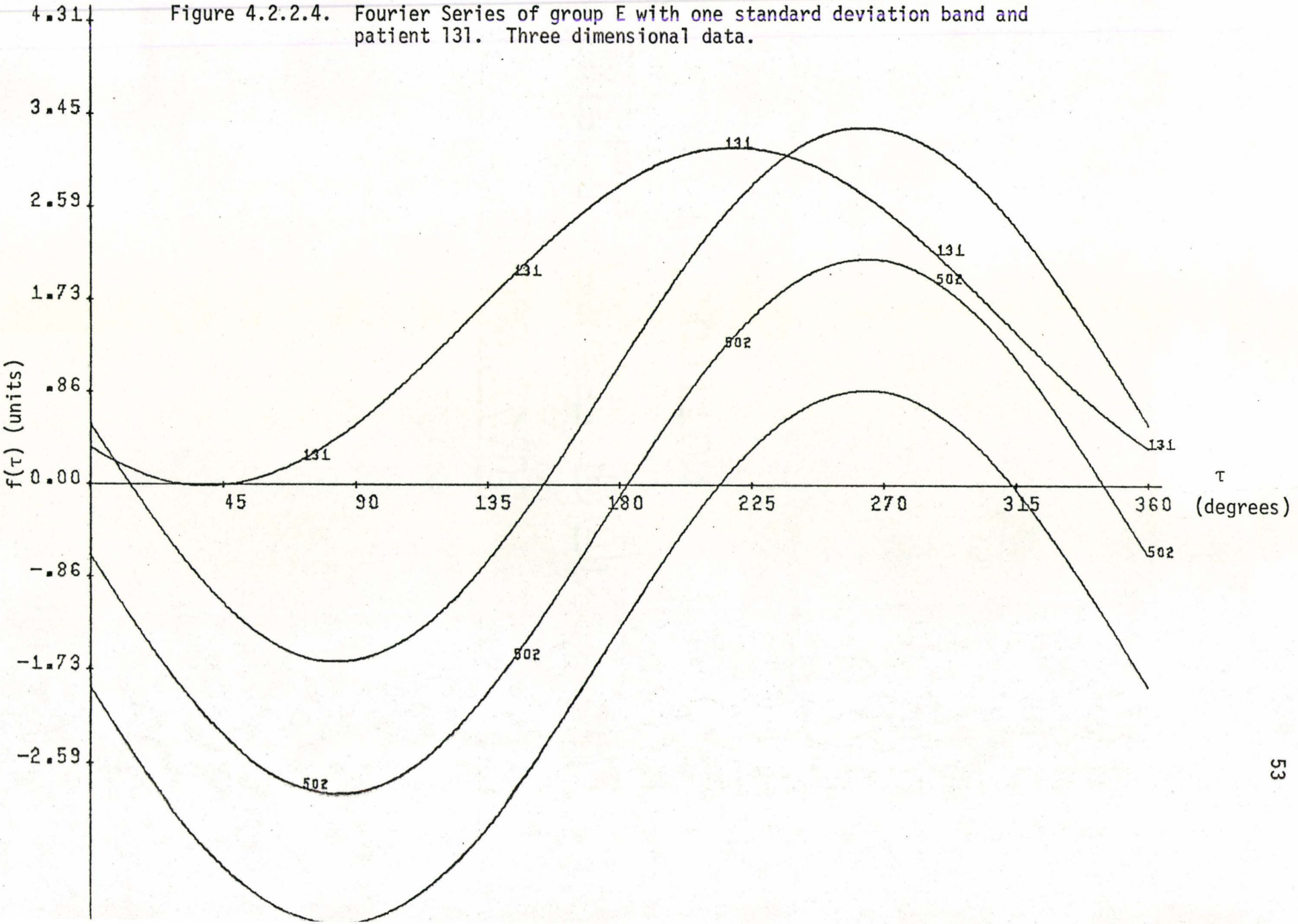


Figure 4.2.2.5. Fourier Series of group F with one standard deviation band and patient 131. Three dimensional data.

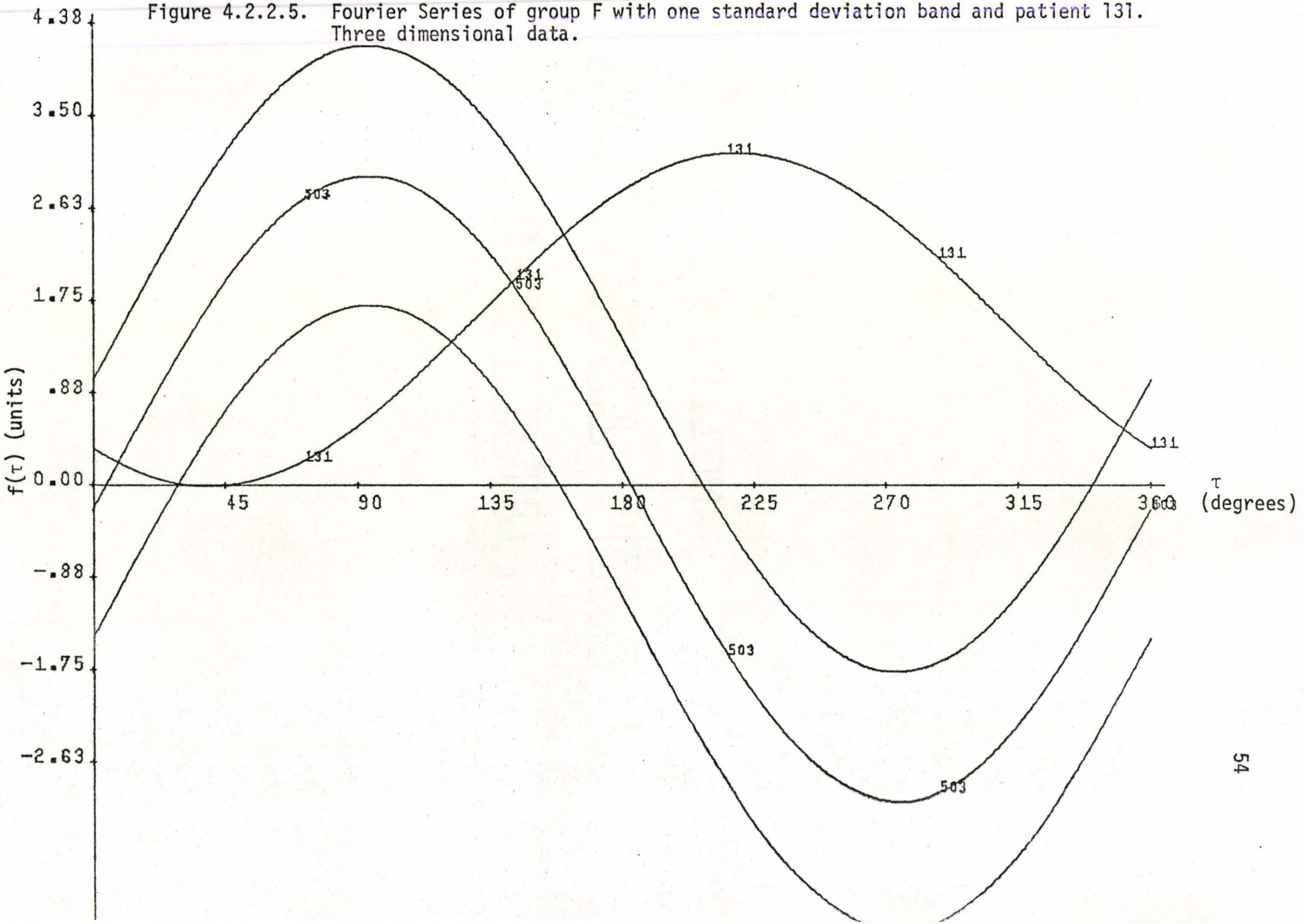
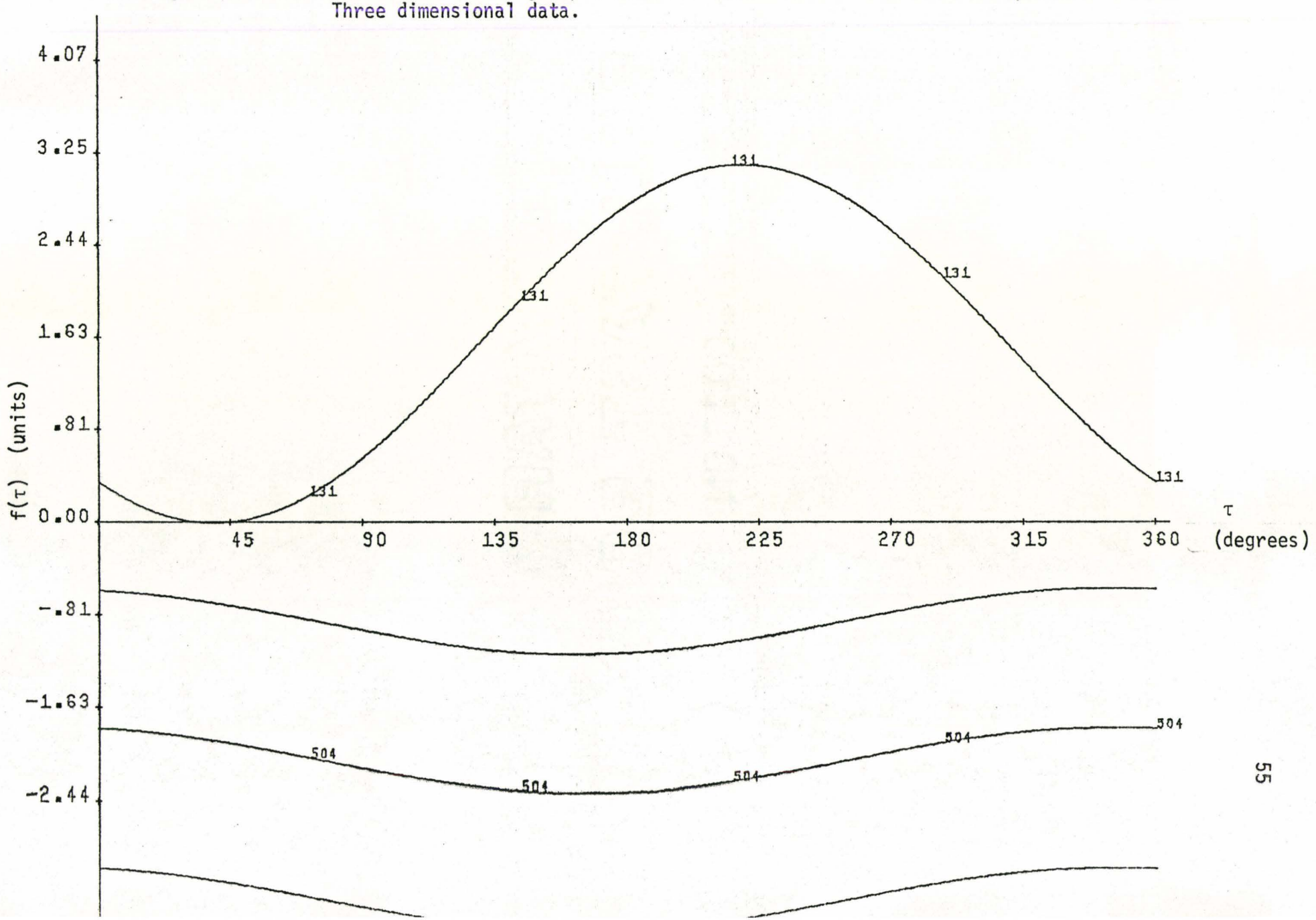
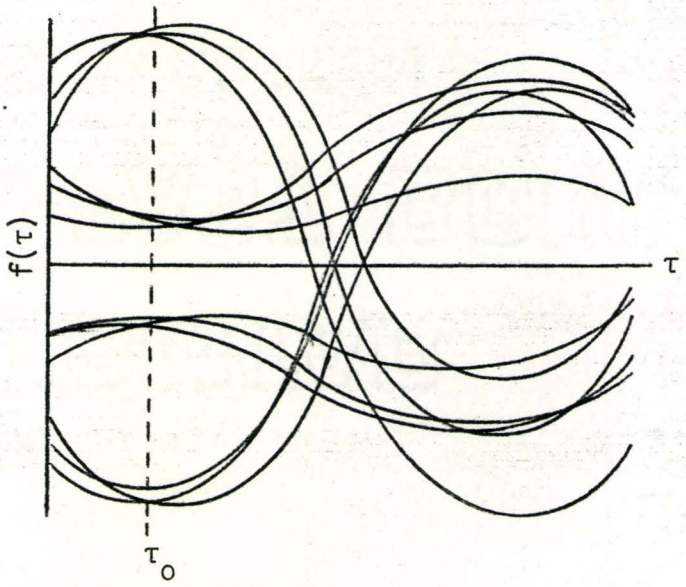


Figure 4.2.2.6. Fourier Series of group G with one standard deviation band and patient 131. Three dimensional data.



ification of patient number 31 is to assign her to group D.

It was of interest to try to find a single discriminant value associated with each group. That is, can we come up with a single statistic which will indicate the group to which a patient belongs? This problem was considered by looking at each patient's fourier series in the hopes of finding a specific τ_o where all the patients within a group clustered but each group clustered at different $f_T(\tau_o)$'s. Diagrammatically, the τ_o we were looking for would look like:



We can then determine each group's mean discriminant fourier value $f_{\bar{t}_i}(\tau_o) = f_{\mu_i}(\tau_o)$, i = D, E, F, G. The hypothesis that the expectation of $f_T(\tau_o) = f_\mu(\tau_o)$ for some hypothesized μ may be tested. This is accomplished by evaluating the significance level of

$$\{f_T(\tau_o) - f_\mu(\tau_o)\} \Big/ \{\mathrm{var}(f_T(\tau_o))\}^{\frac{1}{2}}$$

as a standard normal variate.

In effect, we are determining a linear combination of the canonical variates which themselves are linear combinations of the original data, to obtain a representative quantity associated with each group. Given a test patient, her discriminant fourier value, $f_t(\tau_o)$, is calculated and tested against each of the $f_{\bar{t}_i}(\tau_o)$, i = D, E, F, G, to determine her allocation.

Figure 4.2.2.7 to Figure 4.2.2.10 are the plottings of each group's individual patients' fourier functions. All of group D's fourier series (Fig. 4.2.2.7) are quite variable and do not cluster well at any τ. The only striking feature of these plots is that essentially all values of $f_{t_D}(\tau)$ are positive. Group E on the other hand appears to have a very characteristic plot (Fig. 4.2.2.8) with good clustering at $\tau = 0°$ and $\tau = 140°$ perhaps. A noteworthy point is that for $\tau \lesssim 180°$ we find $f_{t_E}(\tau) < 0$ and if $\tau \gtrsim 180°$, $f_{t_E}(\tau) > 0$. Group F (Fig. 4.2.2.9) approximates a mirror image of group E in the X-axis. Therefore the reverse statements are true; for $\tau \lesssim 180°$, $f_{t_F}(\tau) > 0$, and $\tau \gtrsim 180°$ implies $f_{t_F}(\tau) < 0$. Similarly group G (Fig. 4.2.2.10) imitates group D but below the X-axis.

It is evident that a single τ_o, as discussed previously, cannot be found to construct our discriminator among these groups. However, another testing procedure applies. If we select $\tau_o = 90°$ and test whether or not the expected value of $f_t(\tau_o)$ is greater than zero, we may determine into which pair of groups this function belongs. That is, if the test supports the hypothesis that $E(f_t(\tau_o)) > 0$ then the patient is a member of group D or group F. If the hypothesis is

Figure 4.2.2.7. Fourier Series of all of the individual patients of group D.. Three dimensional data.

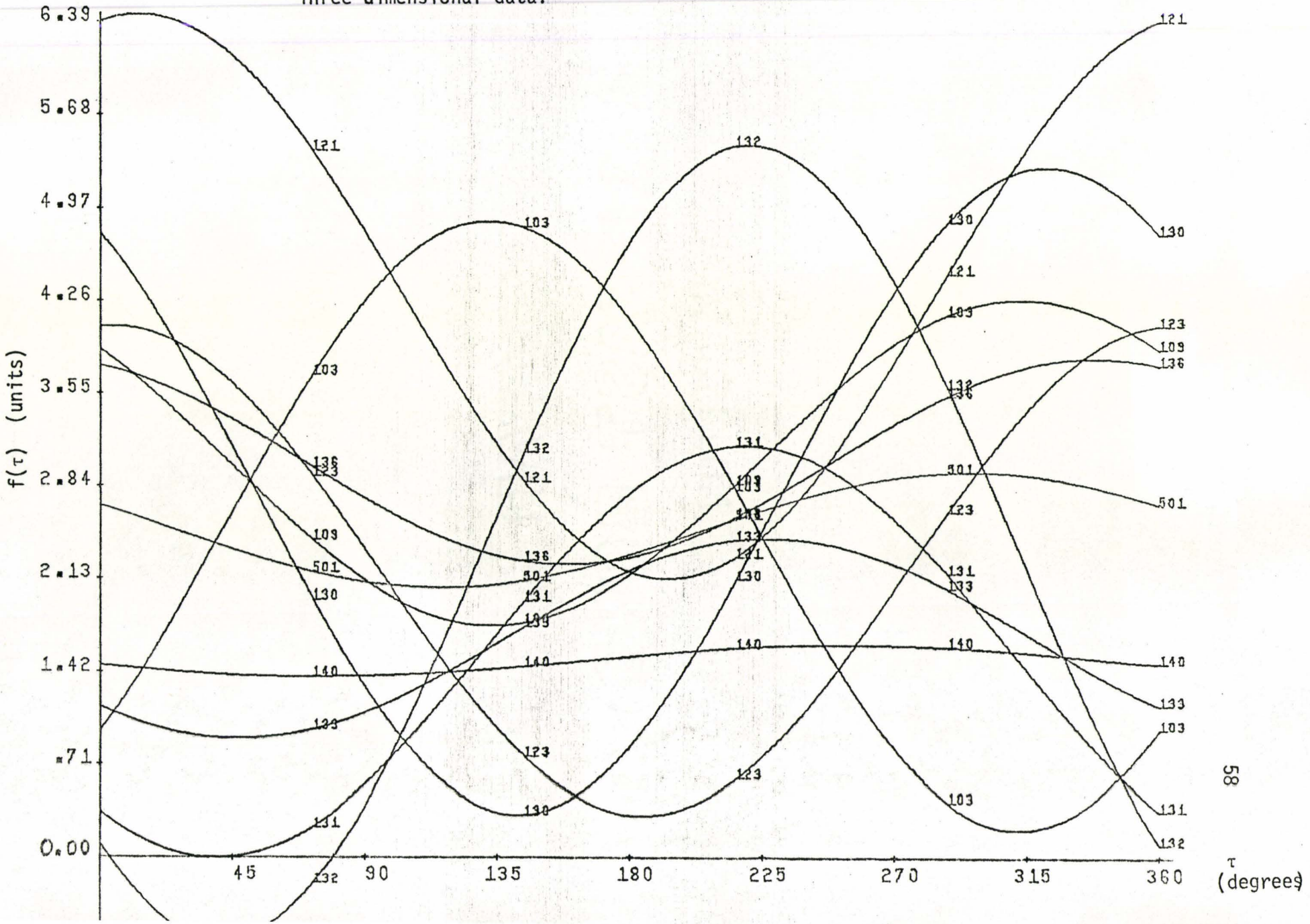


Figure 4.2.2.8. Fourier Series of all of the individual patients of group E.
Three dimensional data.

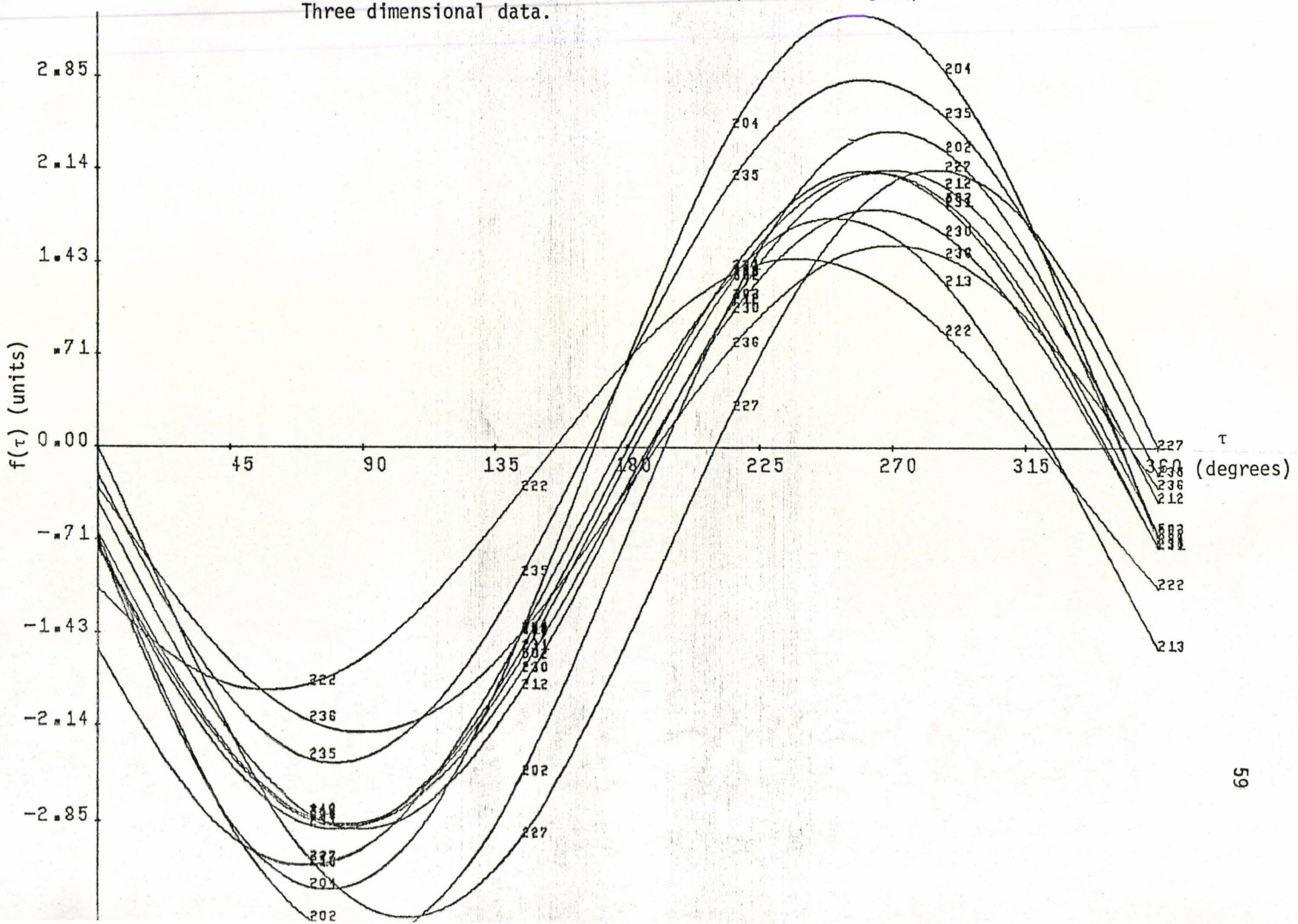


Figure 4.2.2.9. Fourier Series of all of the individual patients of group F.
Three dimensional data.

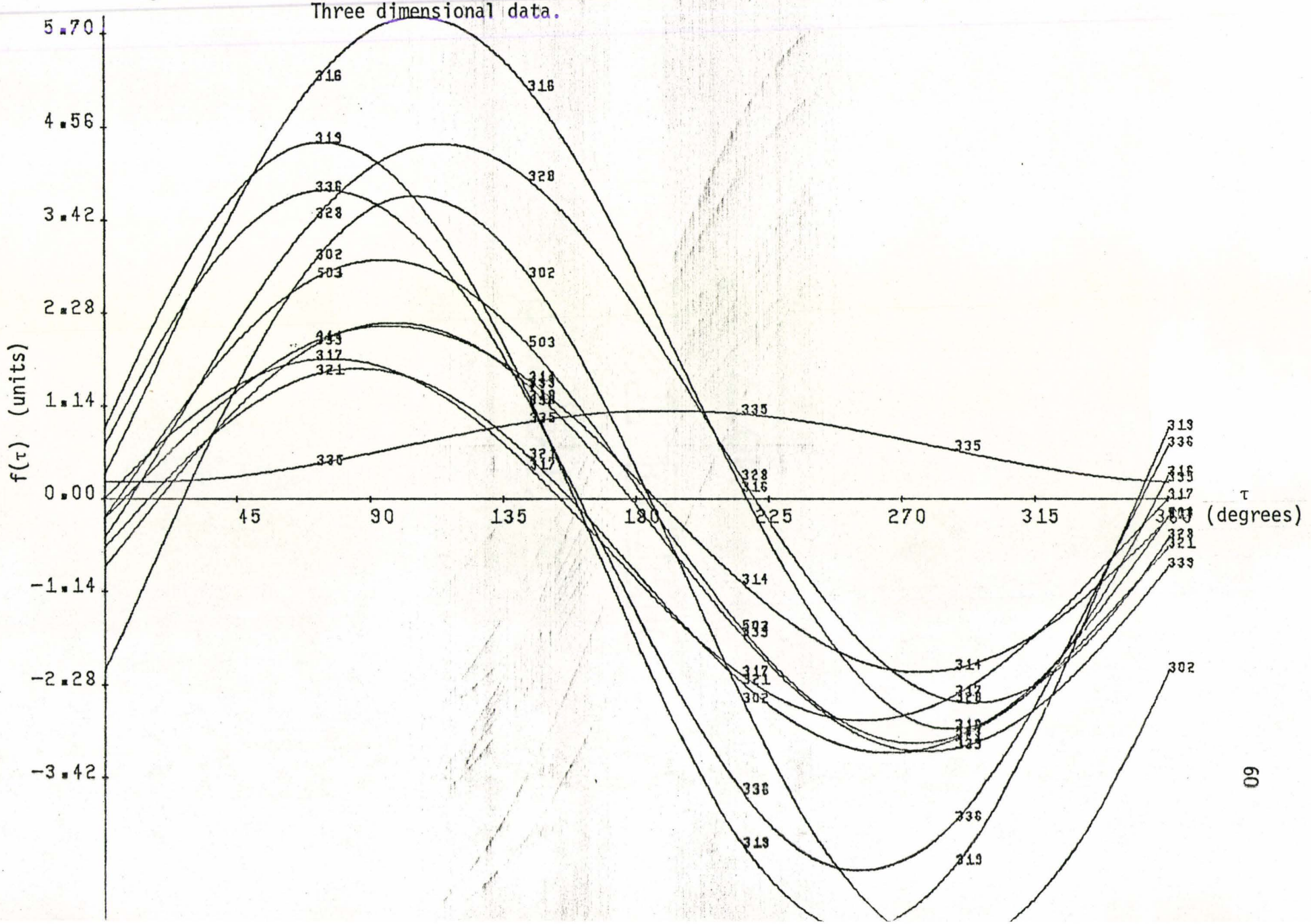
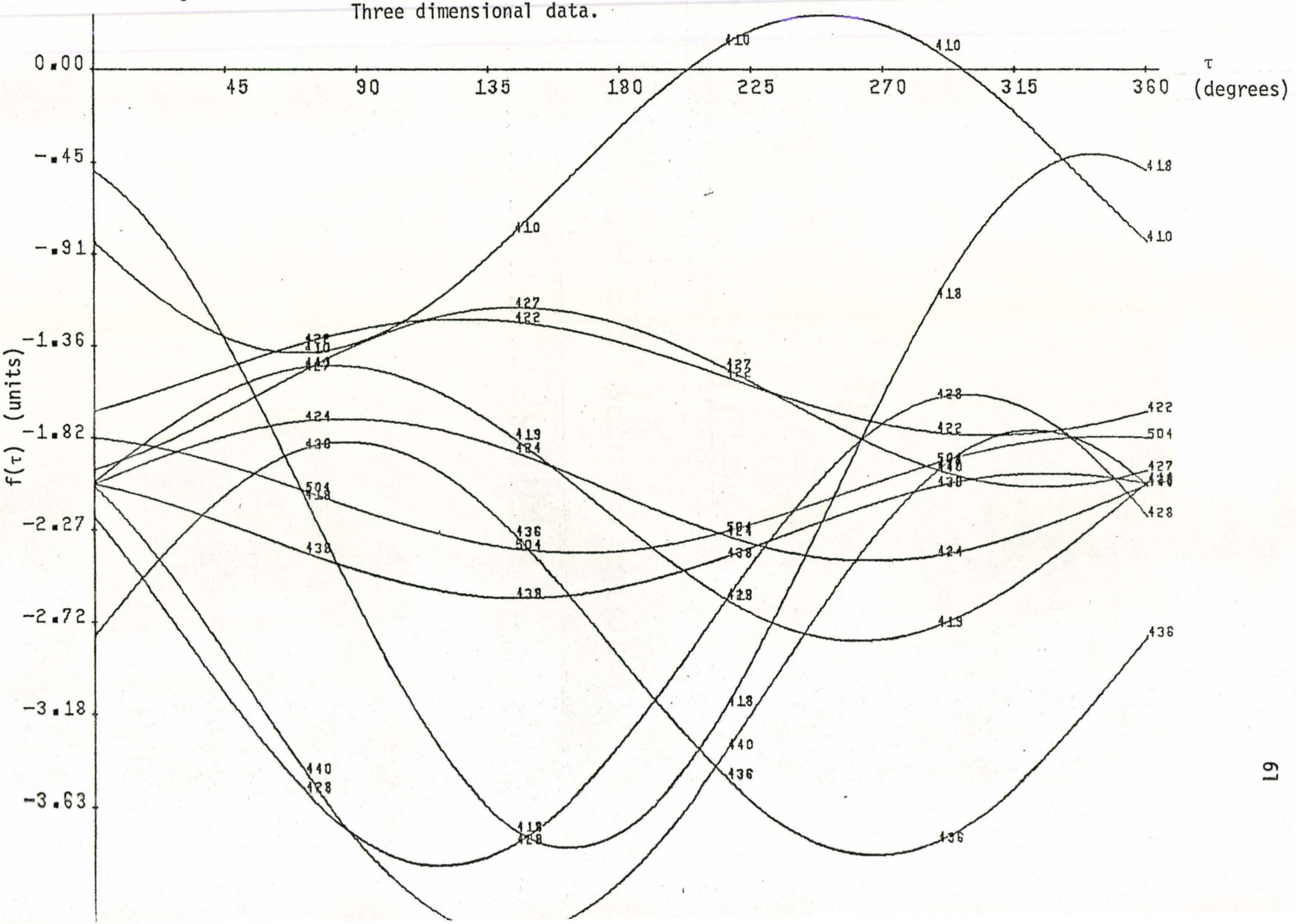


Figure 4.2.2.10. Fourier Series of all of the individual patients of group G.
Three dimensional data.



rejected, then she belongs to groups E or G. Once the pair of groups to which she belongs has been decided we may make the analagous test at $\tau_o = 270°$ to distinguish between these two groups and allocate her accordingly.

The allocation of patient number 31 of group D is as follows. Her two test criterions are:

$$f_{z31}(\tau_o=90°) = 2.2229/\sqrt{2} - 1.0114 \sin(90°) - 1.2.56 \cos(90°) = .5604$$

$$f_{z31}(\tau_o=270°) = 2.2229/\sqrt{2} - 1.0114 \sin(270°) - 1.2156 \cos(90°) = 2.5832$$

The test of whether $f_{t31}(\tau_o=90°)$ is significantly greater than zero reduces to the determination of whether or not the standard normal value of

$$(.5604-0.0) \; / \; \sqrt{1.5}) = .4572$$

lies in the upper $\alpha = .05$ tail of the normal distribution. It does not and in fact this value is non-significant approximately up to $\alpha = .10$. We therefore must conclude that we do not have adequate information to classify this patient. However, we can determine the pair of groups that she belongs to because $f_{t31}(\tau_o=270°)$ is significantly positive at $\alpha = .05$ (i.e. a standard normal value of 2.1092). Thus she is either a member of group D or group E.

Admittedly this procedure may accomodate a great number of variables but the interpretations and testings leave ambiguous results. The existence of a solitary τ_o to discriminate among the groups would be extremely convenient. With the restriction that the τ's be integers, this linear combination is very unlikely to be found.

Furthermore the plotting of the canonical variates in the fourier series did not illuminate any hidden features butappeared to camouflage those discovered by the canonical axes. For example, the attempt of the third canonical variate to separate the pair of groups D, G from the pair E, F as illustrated by the two-dimensional canonical axes plot was nowhere evident in the fourier plot.

An equitable comparison of the two graphical representations was made by plotting the fourier series with only the first two canonical variates as coefficients, in the form

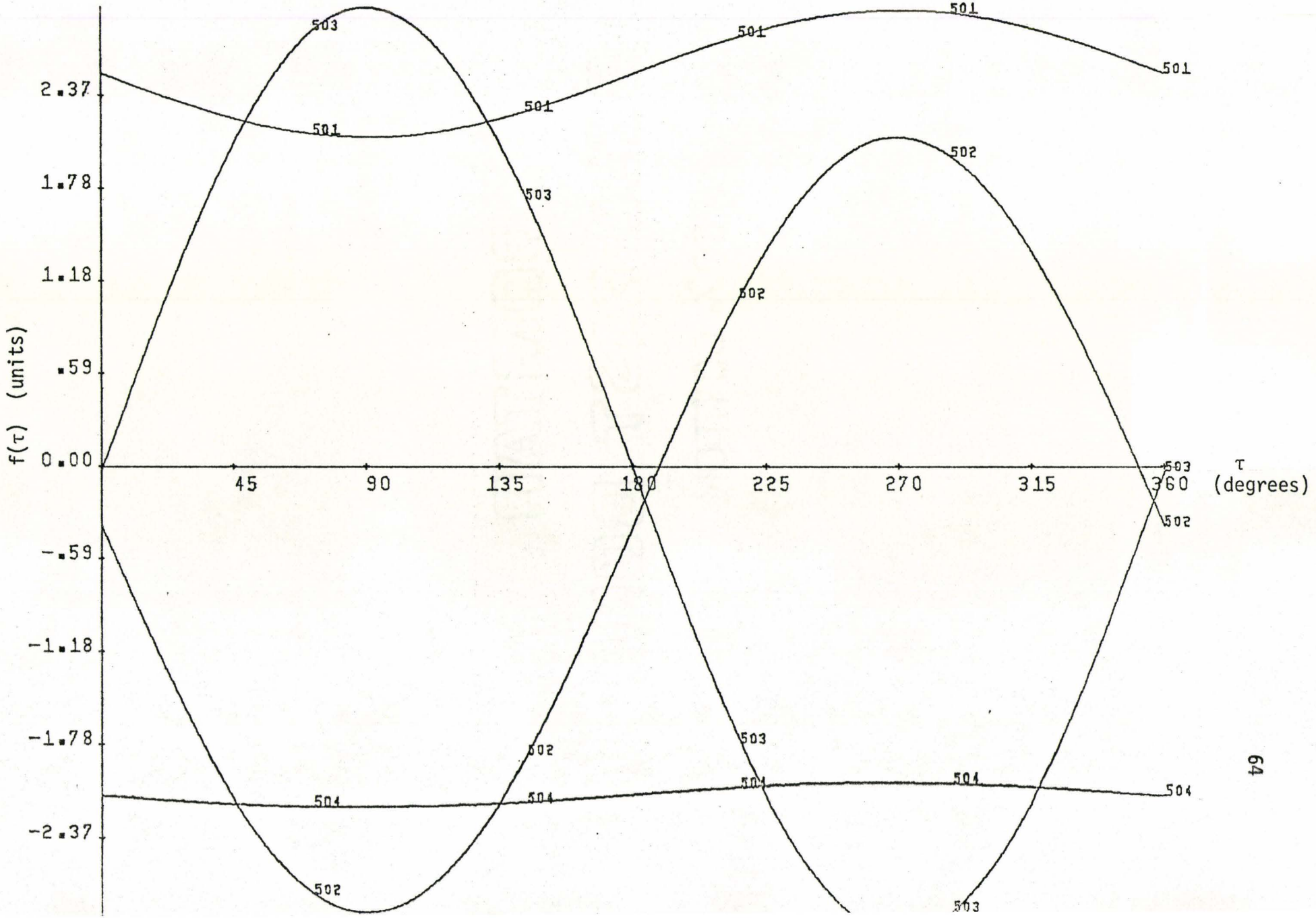
$$f_Z(\tau) = (2)^{-\frac{1}{2}} Z_1 + Z_2 \sin\tau \quad , \qquad 0 \leq \tau \leq 2\pi.$$

Then this plotting (Fig. 4.2.2.11.) is equivalent to the canonical axes plot in Figure 3.3.3.1. Both plots illustrate the same features, that is, group D and G have large (in magnitude) first canonical variates and relatively small second variates and vice versa for groups E and F. Nonetheless the canonical axes supplies this information immediately while one has to interpret the fourier functions to obtain the same conclusions. This further interpretation of an already complex set of data (canonical variates) may well be a source of error.

4.3. *Chernoff's Faces*

This graphical method of displaying points in p-dimensions was developed by Herman Chernoff in 1973 [4]. The method consists of representing a point in p-dimensional space by a drawing of a face

Figure 4.2.2.11. Fourier Series of all groups. Two dimensional data.



whose characteristics are determined by the position of the point.

The object was to enable the investigator to quickly comprehend relevant information of the data and then apply appropriate statistical analysis. No quantitative results are produced but a better "feel" of the data is obtained.

At present, the number of variables that may be accommodated is less than or equal to eighteen. Each variable is associated with a given characteristic by the following scheme: if $Y = (y_1, y_2, \ldots, y_{18})$ is an 18-dimensional data point then the corresponding facial characteristics are

Variable	Characteristic
y_1	radius r to corner of face
y_2	angle of r to horizontal
y_3	vertical size of face
y_4	eccentricity of upper face
y_5	eccentricity of lower face
y_6	length of nose
y_7	vertical position of mouth
y_8	curvature of mouth
y_9	width of mouth
y_{10}	vertical position of eyes
y_{11}	separation of eyes
y_{12}	slant of eyes
y_{13}	eccentricity of eyes

y_{14}	size of eyes
y_{15}	position of pupils
y_{16}	vertical position of eyebrows
y_{17}	slant of eyebrows
y_{18}	size of eyebrows

Two proposed advantages of the facial representation of data, as suggested by Chernoff, are (i) enhancing the user's ability to detect and comprehend important phenomena.

People are in constant contact with varying faces each day. They subconsciously filter out repetitive and common features and focus their attention to the most striking characteristic of a person. If he be oriental, then the notable distinction is his eyes; his eyebrows and mouth perhaps would leave little impression. In this way, the relevant data is detected and comprehended.
(ii) serving as a mnemonic aid for remembering major conclusions.

If numerical data is inspected, preliminary separation or distinctions may be made. However, as these features of the data become more numerous the ability to retain the information is poor. With the data represented as faces, certain major characteristics of the faces are instantly observed and easily remembered in terms of emotions and appearance.

The major advantage to be derived from using the faces should be in the heightened qualitative awareness of which numerical calculations are relevant [4].

4.3.1. *Application of Chernoff's Faces, k = 4*

As a direct comparison to the previous graphical representations of the canonical variates, these set of variables were employed as the data points. Hence we have three-dimensional points representing each group.

$$\bar{z}_D = (3.5438, -.4039, .1980)$$
$$\bar{z}_E = (-.5382, -2.4812, -.2722)$$
$$\bar{z}_F = (-.0428, 2.9603, -.1952)$$
$$\bar{z}_G = (-2.9628, -.0752, .2784)$$

There are (18-3) = 15 characteristics which had no defining variable and therefore were fixed for each group. The traits that were defined by the three canonical variates were

$\bar{z}_1$ = vertical size of face
$\bar{z}_2$ = eccentricity of lower face
$\bar{z}_3$ = curvature of mouth

The method of plotting converts the range of each variable into a range of suitable proportions for the faces. Thus the maximum and minimum of each variable become the extremes in the range adopted for the faces. This ensuresthat each variable is scaled relative to a measure of variability (i.e. range).

Figures 4.3.1.1. through 4.3.1.4 are the computer drawn faces for groups D, E, F and G. The vertical size of the faces clearly emphasizes the effect of the first canonical variate. However the second canonical variate does not appear to separate groups E and F

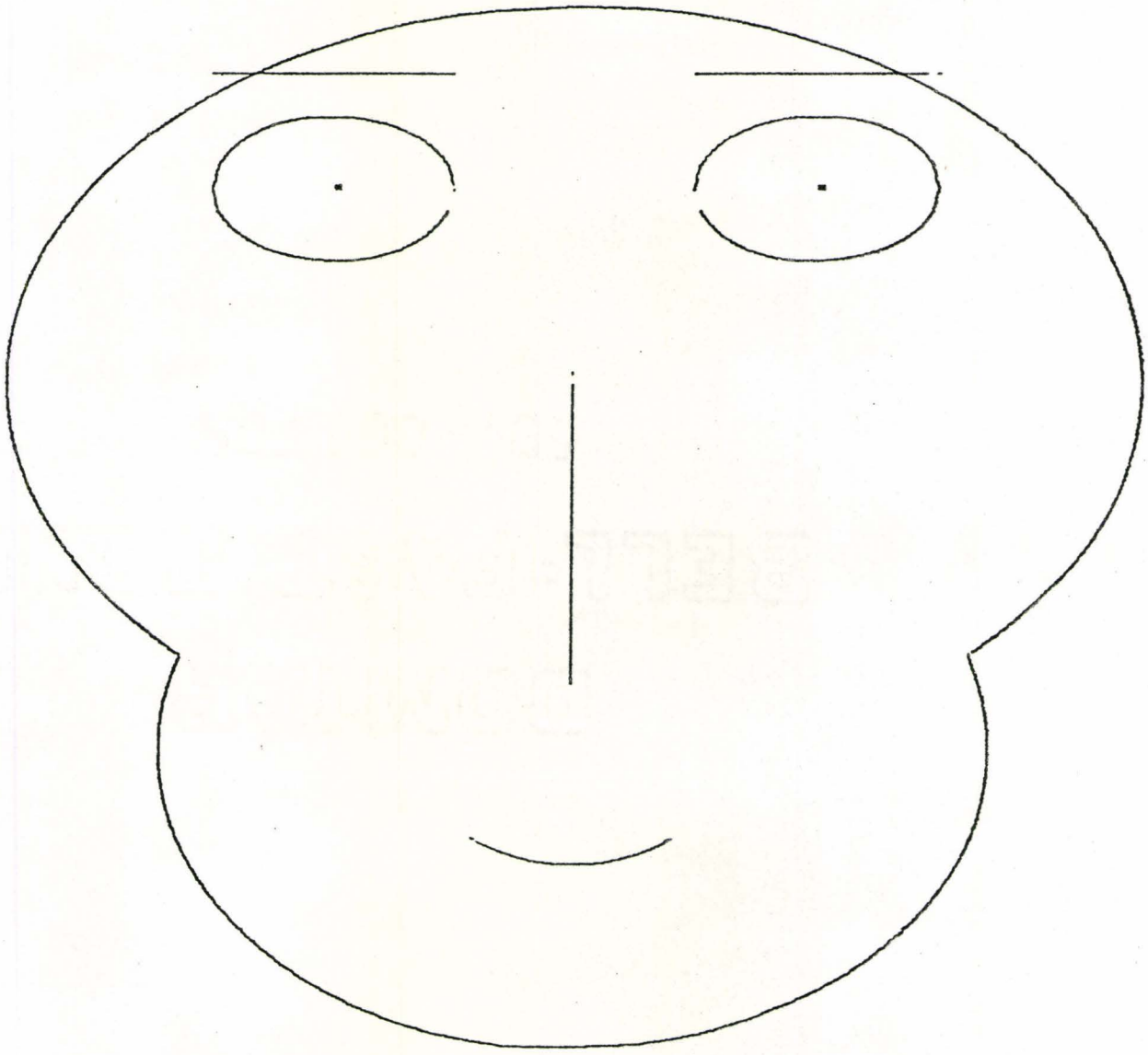


Figure 4.3.1.1. Face of group D. Derived from three canonical variates.

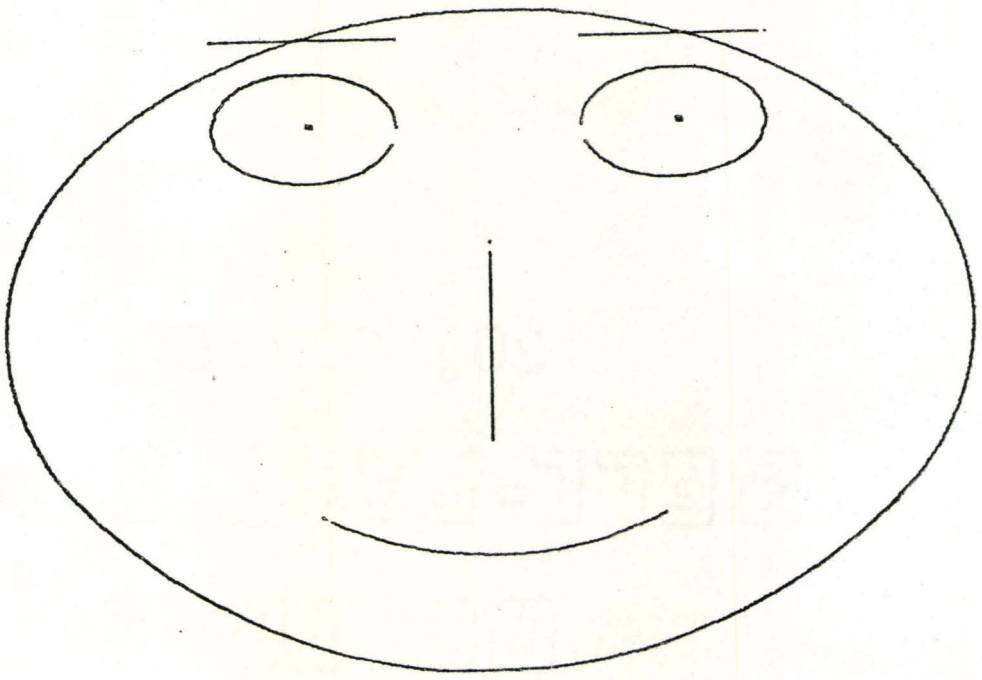


Figure 4.3.1.2. Face of group E. Derived from three canonical variates.

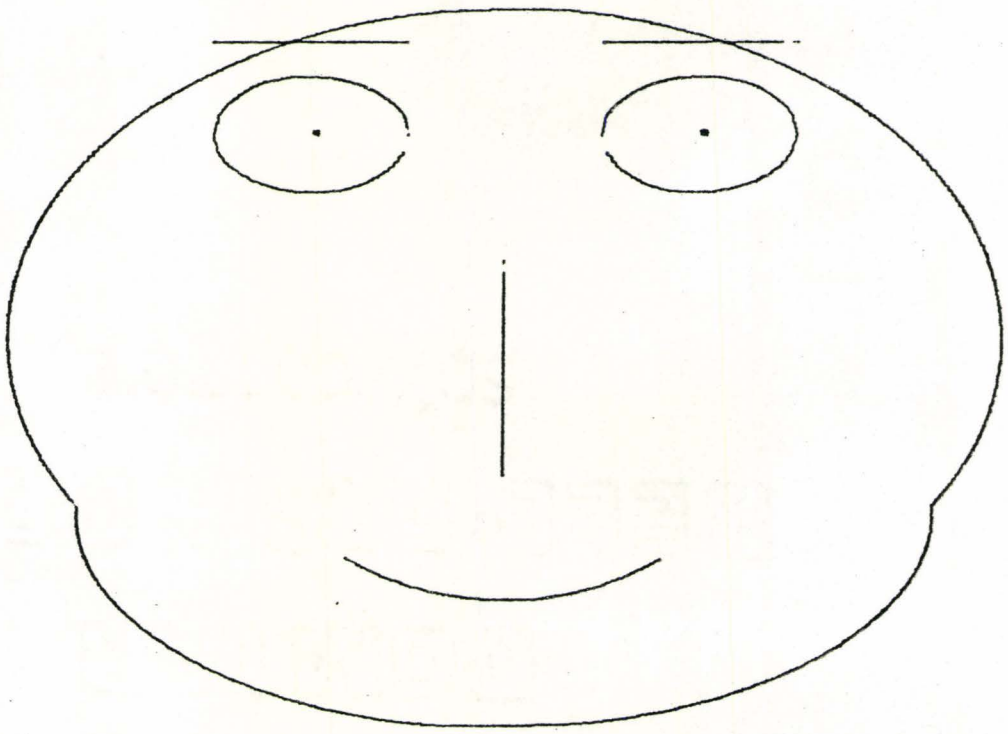


Figure 4.3.1.3. Face of Group F. Derived from three canonical variates.

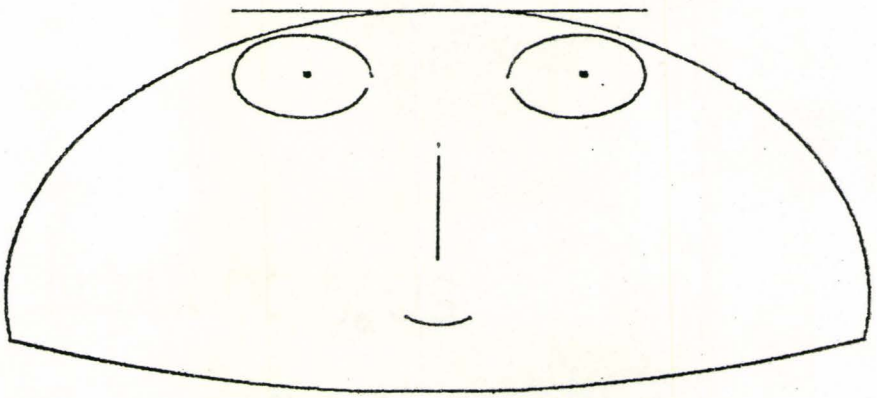


Figure 4.3.1.4. Face of group G. Derived from three canonical variates.

well at all. This implies that the significance of a variable depends to a large extent on the facial characteristic that the variable is assigned. The eccentricity of the lower face is not a good discriminating feature. The third canonical variate controlling the curvature of the mouth offers little in separating the groups.

The test patient number 31 of group D had her caricature drawn with the same defined characteristics. Figure 4.3.1.5 reveals her facial features. On inspection, we classify her as a group D patient even though her mouth is uncharacteristic of group D. Her features of size of face and eccentricity of lower face convincingly suggest she is from group D.

Another application of this procedure is to initially scrutinize the data to determine which statistical analysis to employ. We therefore looked at the faces constructed from the raw data and not the canonical variates. This enabled us to increase our defining variables to six. The point $X = (x_1, \ldots, x_6)$ was assigned to the features as follows;

x_1 = vertical size of face

x_2 = eccentricity of upper face

x_3 = eccentricity of lower face

x_4 = curvature of mouth

x_5 = slant of eyes

x_6 = size of eyes

The remaining characterisitcs were fixed. The mean faces for each group were plotted using the data

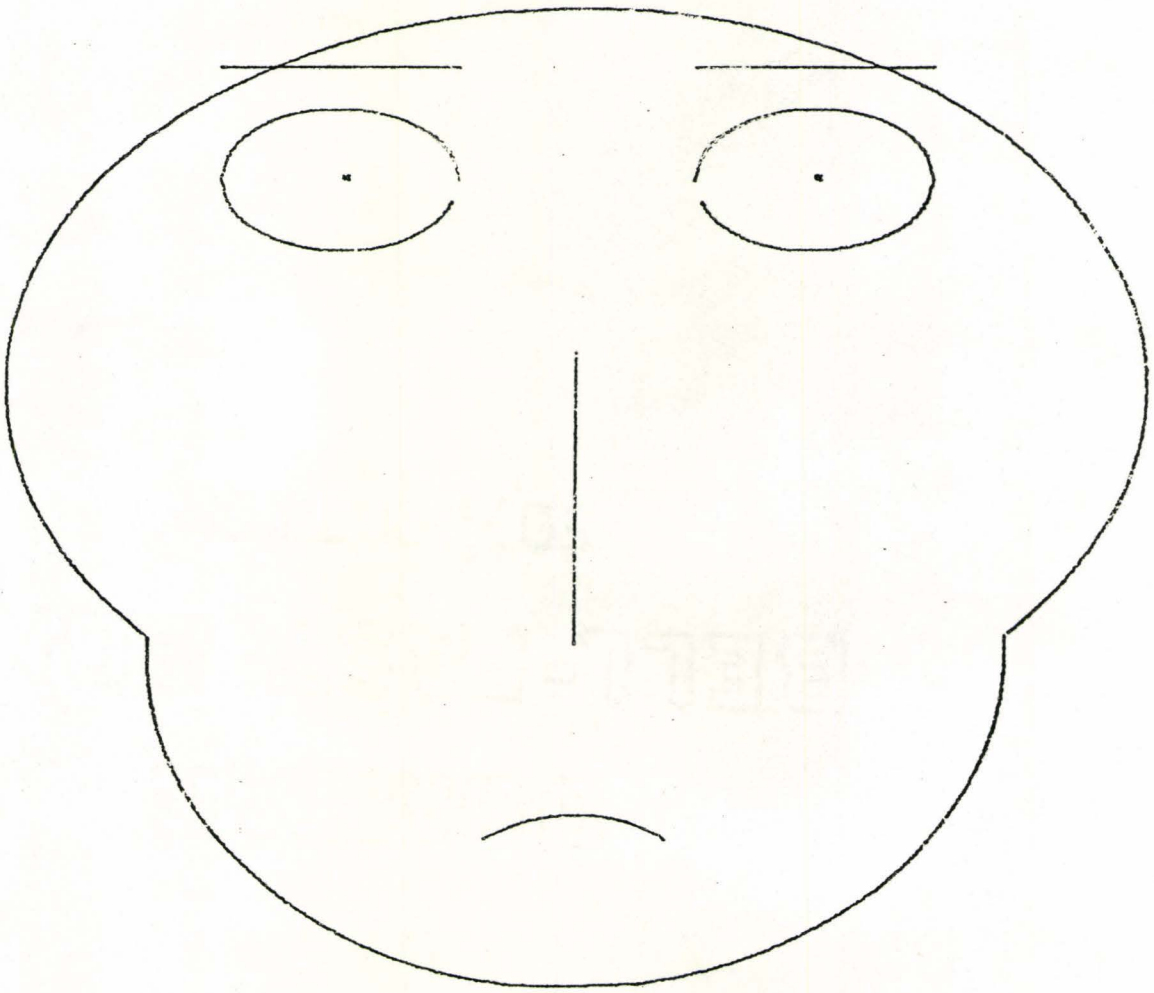


Figure 4.3.1.5. Face of test patient 31 of group D. Derived from three canonical variates.

$\bar{x}_D$ = (12.682, 40.408, 31.578, 17.155, 299.207, 282.600)

$\bar{x}_E$ = (12.479, 40.017, 31.379, 4.296, 144.985, 406.3667)

$\bar{x}_F$ = (12.197, 38.332, 32.076, 9.938, 396.535, 327.200)

$\bar{x}_G$ = (11.682, 37.130, 31.727, 3.283, 196.442, 434.000)

Figure 4.3.1.6 through 4.3.1.9 are their portraits. The most obvious differences among the plots are the shape of the heads and the size of the eyes. This would suggest that close investigation should be aimed at variables X_2, X_3 and X_6. This brings to light an interesting question. Are the eyes and the shape of a head the most notable features one sees in a face? Under closer examination of the faces the mouths also are very distinct in each group, however this was unnoticed in first impressions. Similarly, the slant of the eyes were of secondary importance. Chernoff [4] admits these shortcomings and is attempting methods of counteracting these psychological effects. He suspects that a series of faces, representing a single multivariate point, constructed by permuting the variables associated with given characteristics may remedy this.

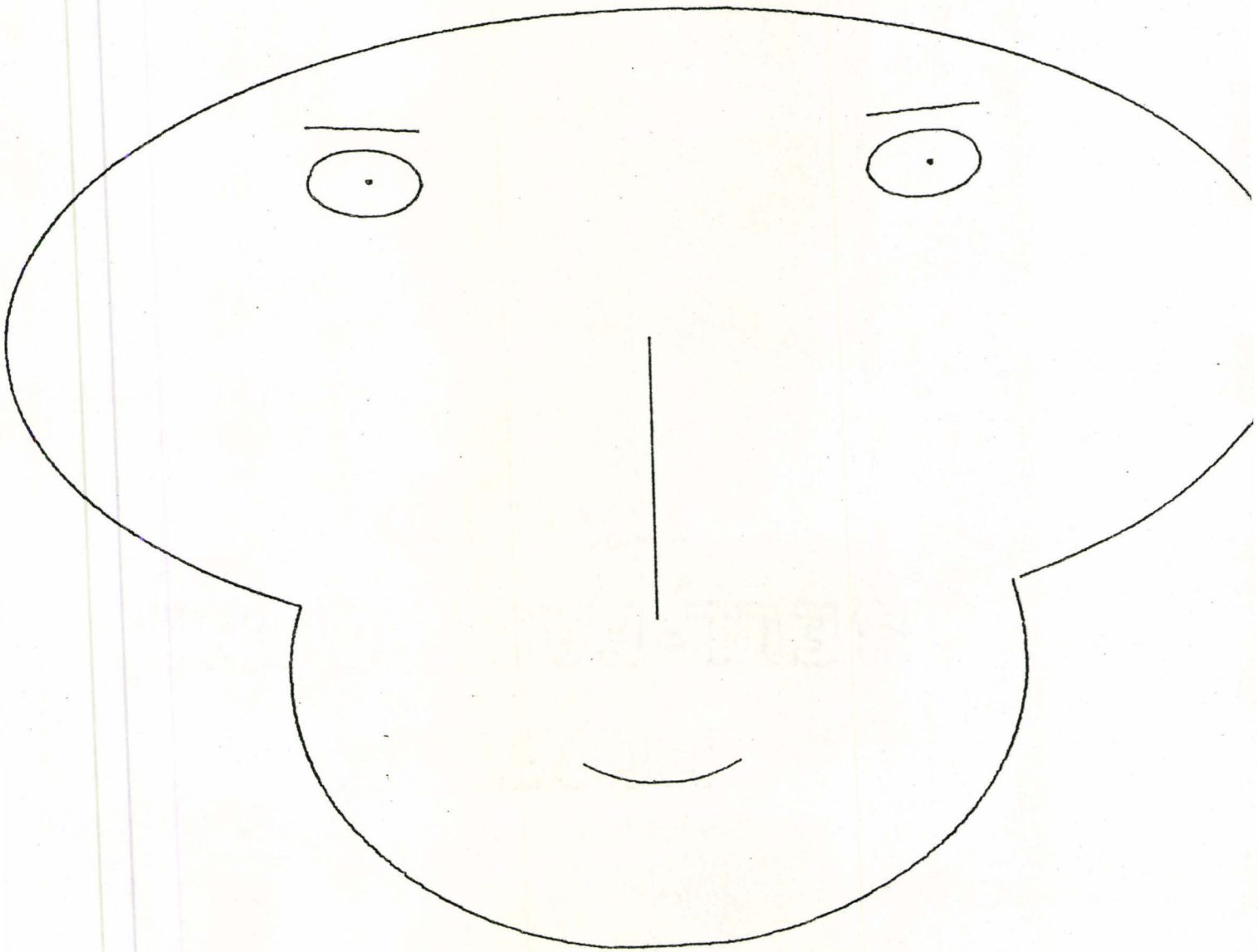


Figure 4.3.1.6. Face of group D. Derived from the six original variates.

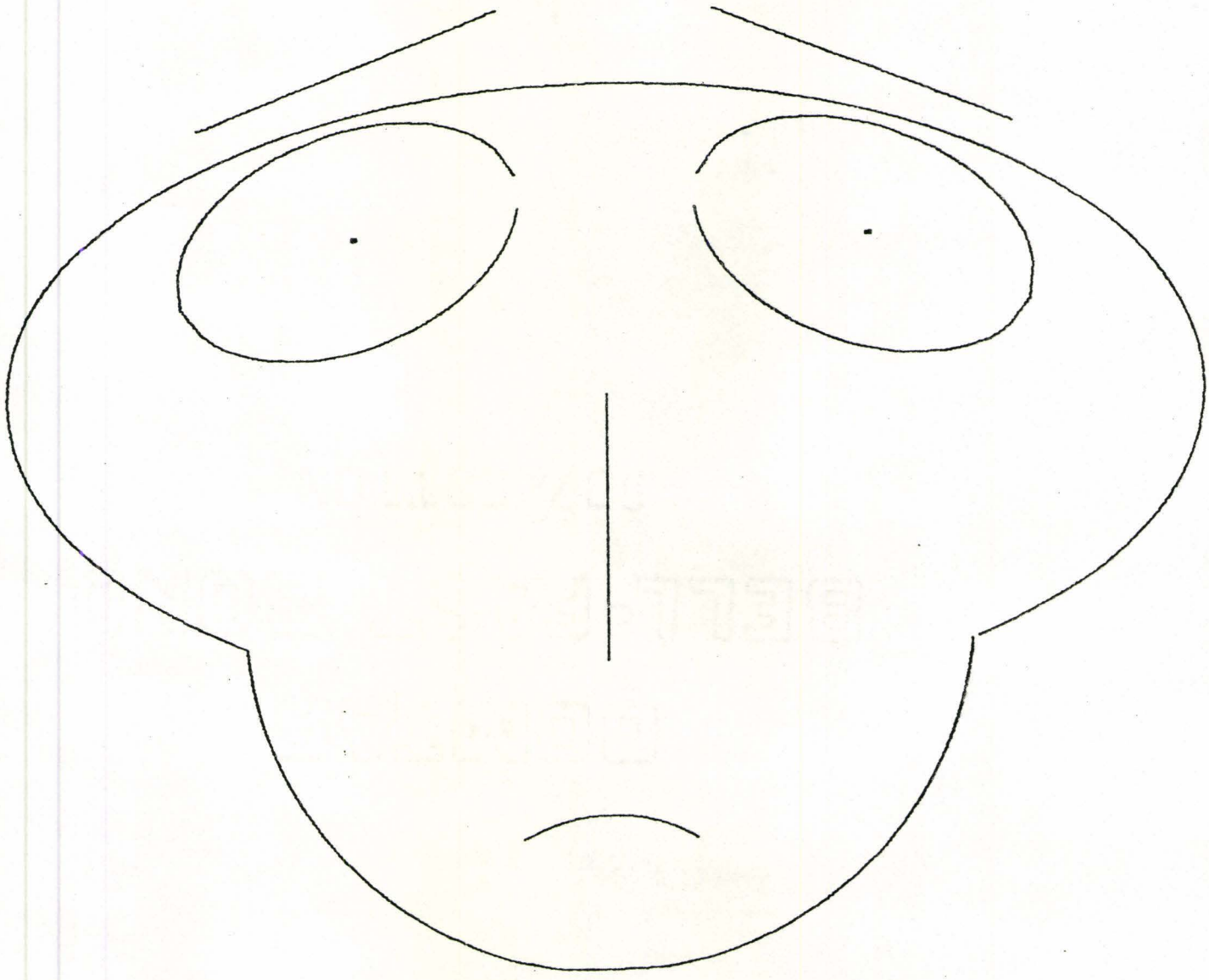


Figure 4.3.1.7. Face of group E. Derived from the six original variates.

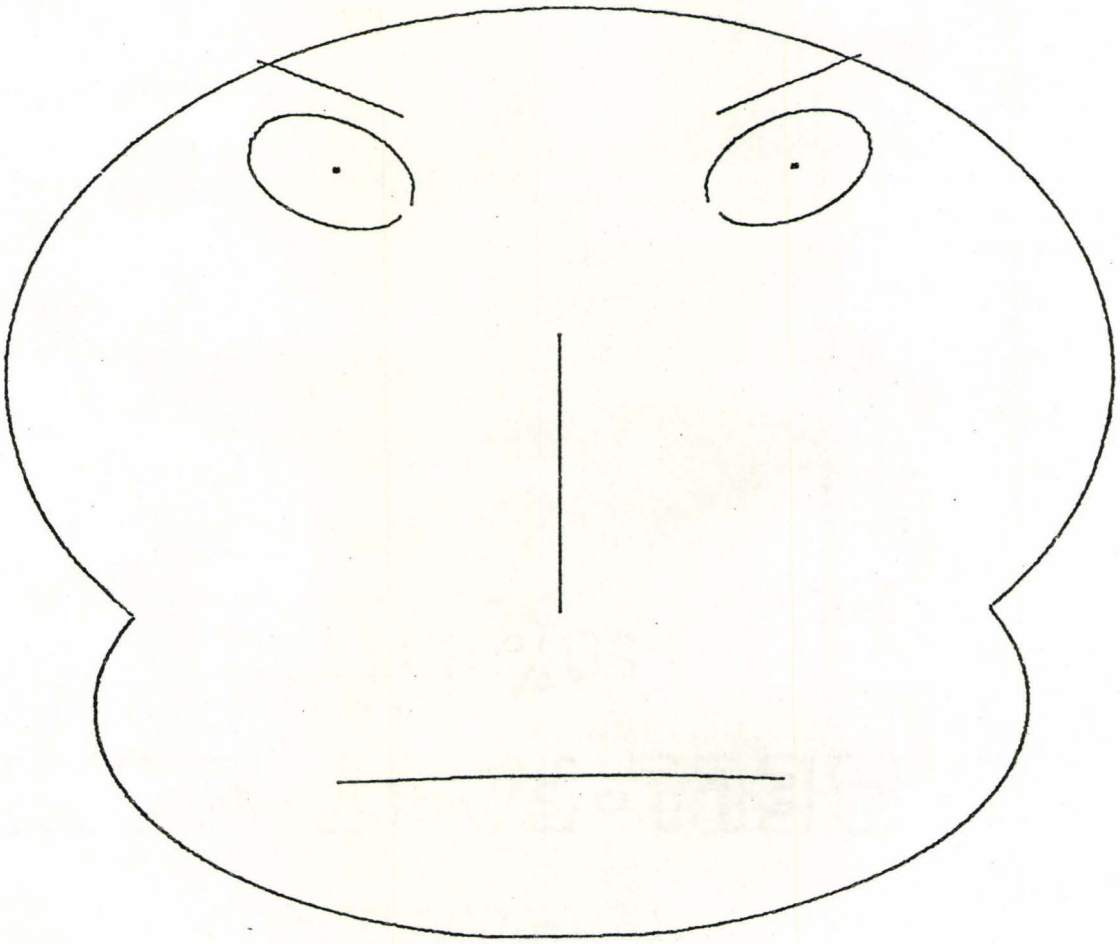


Figure 4.3.1.8. Face of group F. Derived from the six original variates.

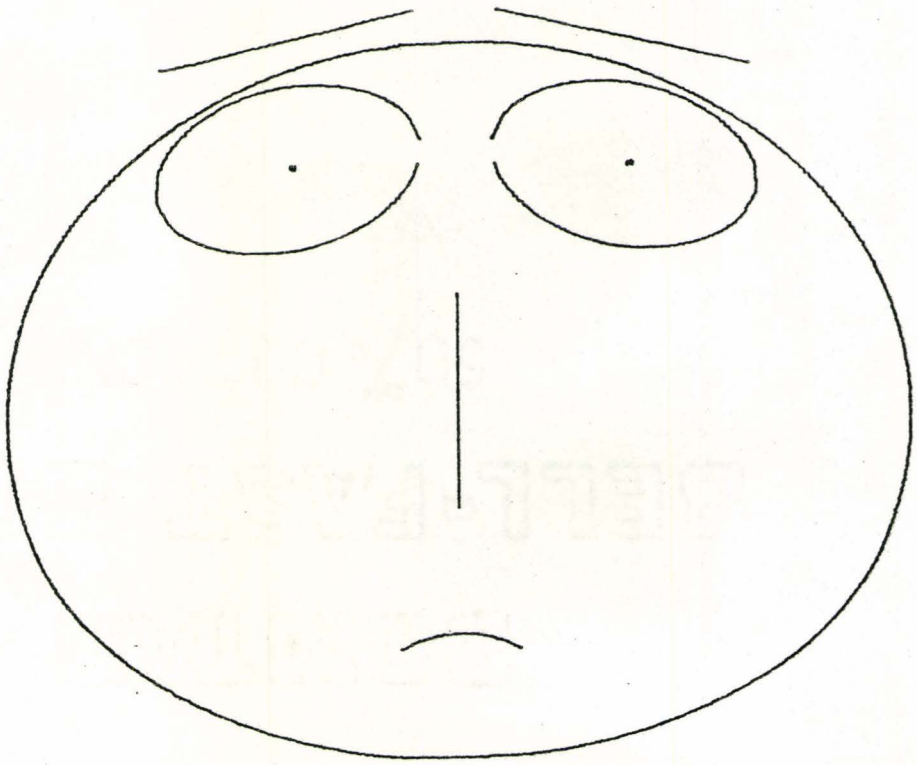


Figure 4.3.1.9. Face of group G. Derived from the six original variates.

5. REAL DATA

Anyone who has worked with real life situations and measurements is sure to admit the inconveniences that often arise. The data will seldom seem to comply to the theoretical restrictions that are necessary for underlying assumptions of analysis. However, even though these prerequisites for some statistical procedures are not upheld, information may still be extracted by the application of these techniques to the data. The important condition to remember is that the conclusions and interpretations drawn from the results must be treated with caution but not necessarily total rejection.

5.1 *Analysis of the Original Data*, k = 4

The entire set of data consisting of 160 patients is tabulated in Table 2.1.

A preliminary overview of the data was based on the tests of the homogeneity of dispersions and of the equality of means. To test the homogeneity of dispersions

$$H : \Sigma_1 = \Sigma_2 = \ldots = \Sigma_k = \Sigma$$

we evaluate the significance of the h statistic (section 3.2.1.). The necessary determinants of the matrices for its calculation are

$$|S_D| = 7.5406 \times 10^6$$
$$|S_E| = 1.8660 \times 10^5$$
$$|S_F| = 4.6512 \times 10^6$$

$$|S_G| = 2.4670 \times 10^7$$

$$|S_{pooled}| = 7.5433 \times 10^8$$

The test statistic is therefore

$$h = -2[1-\{\frac{4}{39} - \frac{1}{156}\}\frac{89}{126}]\ln[(.009996)(.000247)(.006.66)(.03270)]^{19.5}$$

$$= 778.6095$$

The critical $\chi^2_{\alpha=.05}$ value with 63 degrees of freeom is 82.529. The test statistic is highly significant and thereby forewarns us to the application of the discriminating procedures.

The test of the equality of the mean vectors of the populations,

$$H : \mu_1 = \ldots = \mu_k = \mu$$

based on the Wilks criterion (section 3.2.1.) yields the test statistic

$$u = -[(159)-(5)]\ln\{(.9640)\} = 5.6805.$$

The critical $\chi^2_{.05}$ value with 18 degrees of freedom is 28.869. This would indicate that there is no significant difference among the group means.

5.1.1. *Canonical Analysis of Discriminance*

The differences among the groups that do exist will be best exhibited by the canonical variates. The necessary matrices for their computation were the inverse of the pooled covariance matrix and the group means covariance matrix.

Mean Dispersion Matrix ($\hat{B}$)

	1	2	3	4	5	6
1	.023	.069	-.008	.204	1.109	-2.192
2		.230	-.040	.501	-3.052	-4.459
3			.015	.041	4.433	-.774
4				3.640	57.386	-33.489
5					2032.428	-743.467
6						417.485

Inverse of Pooled Cov. Matrix ($\hat{\Sigma}^{-1}$)

	1	2	3	4	5	6
1	33.22	-10.29	-11.34	.009	.0034	.00077
2		3.36	3.48	-.010	-.0009	-.00122
3			7.90	.026	-.0010	.00043
4				.054	-.0007	-.00004
5					.0001	.00002
6						.00014

The (k-1) = 3 non-zero roots of $\hat{\Sigma}^{-1}\hat{B}$ matrix are

$$\lambda_{(1)} = .28826$$

$$\lambda_{(2)} = .15898$$

$$\lambda_{(3)} = .03331$$

with their associated standardized canonical vectors

$$d^{(1)*} = (1.543, -.406, -.046, .132, .0039, -.0041)$$

$$d^{(2)*} = (1.108, -.154, -1.486, .093, -.0056, -.0032)$$

$$d^{(3)*} = (3.708, -1.046, -1.546, -.152, -.0029, -.0039).$$

Thus the three discriminating variates which best separate the groups are

$$T_{(1)}^{*} = 1.543X_1 - .406X_2 - .046X_3 + .132X_4 + .0039X_5 - .0041X_6$$

$$T_{(2)}^{*} = 1.108X_1 - .154X_2 - 1.486X_3 + .093X_4 - .0056X_5 - .0032X_6$$

$$T_{(3)}^{*} = 3.708X_1 - 1.046X_2 - 1.546X_3 - .152X_4 - .0029X_5 - .0039X_6$$

The mean canonical values for each group centred about the grand mean canonical value, $\bar{\bar{t}} = (1.900, -41.511, -45.986)$, are

Group	$\bar{z}_{(1)}$	$\bar{z}_{(2)}$	$\bar{z}_{(3)}$
D	.406	.451	-.115
E	-.426	.206	.212
F	.519	-.416	.087
G	-.499	-.241	-.185

We plotted these coordinates along their canonical axes in three plots. Figure 5.1.1.1 represents the projection of the canonical coordinates onto the $T_{(1)} - T_{(2)}$ plane and Figure 5.1.1.3 onto the $T_{(2)} - T_{(3)}$ plane. One standard deviation contours were drawn about each group mean's canonical coordinates which illustrates the "closeness" of the groups to one another relative to their variability. The coordinates of patient number 31 of group D, $z^{31} = (.516, .246, .426)$, are denoted by a small circle.

A classification of this patient would have little meaning since she falls well within each groups 90% confidence contour of radius 2.500.

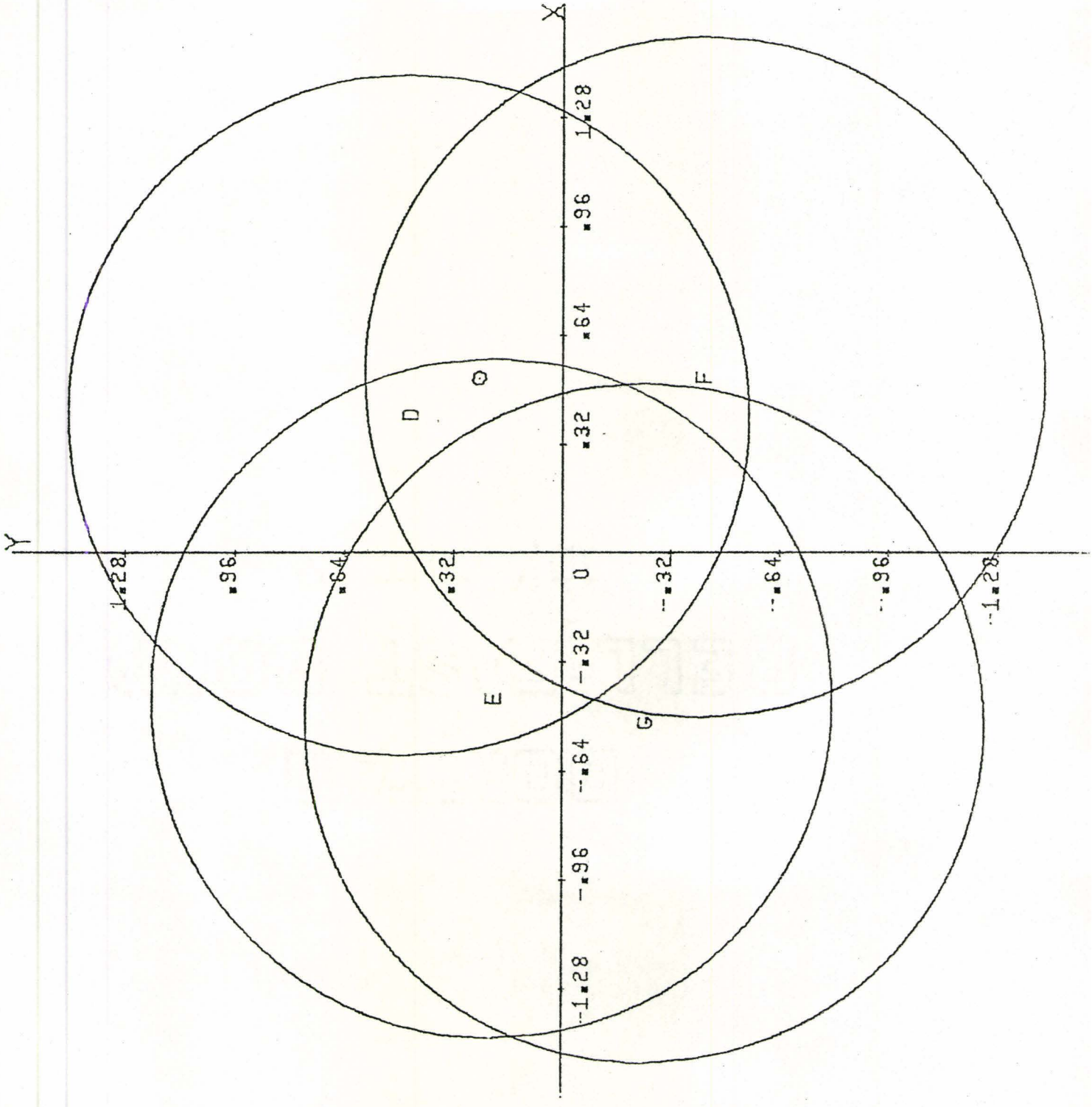


Figure 5.1.1.1. The group means of D, E, F and G along the canonical axes with one standard deviation contours X=first canonical variate, Y=second canonical variate.

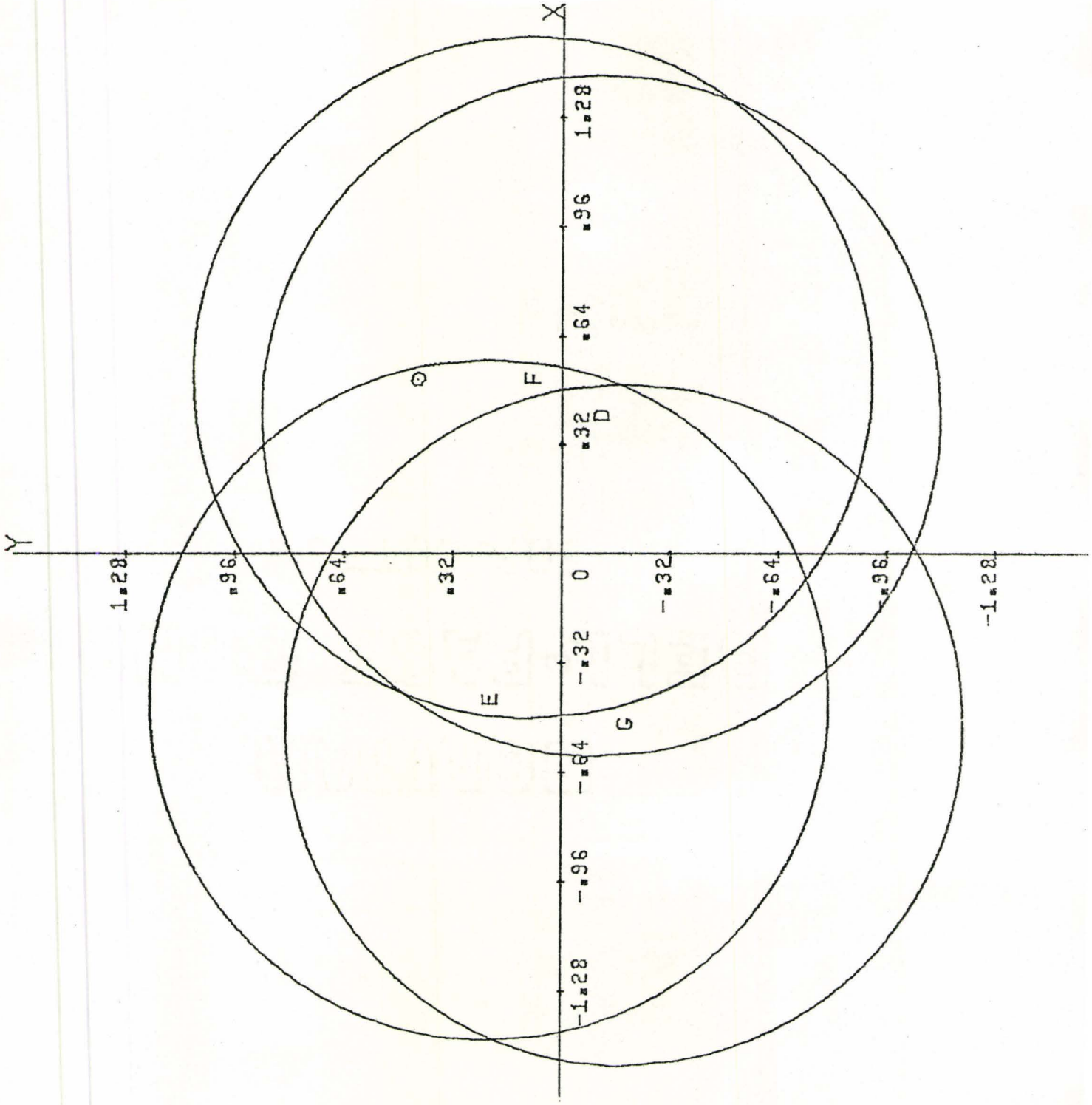


Figure 5.1.1.2. As in Fig. 5.1.1.1 except X=first canonical variate, Y=third canonical variate.

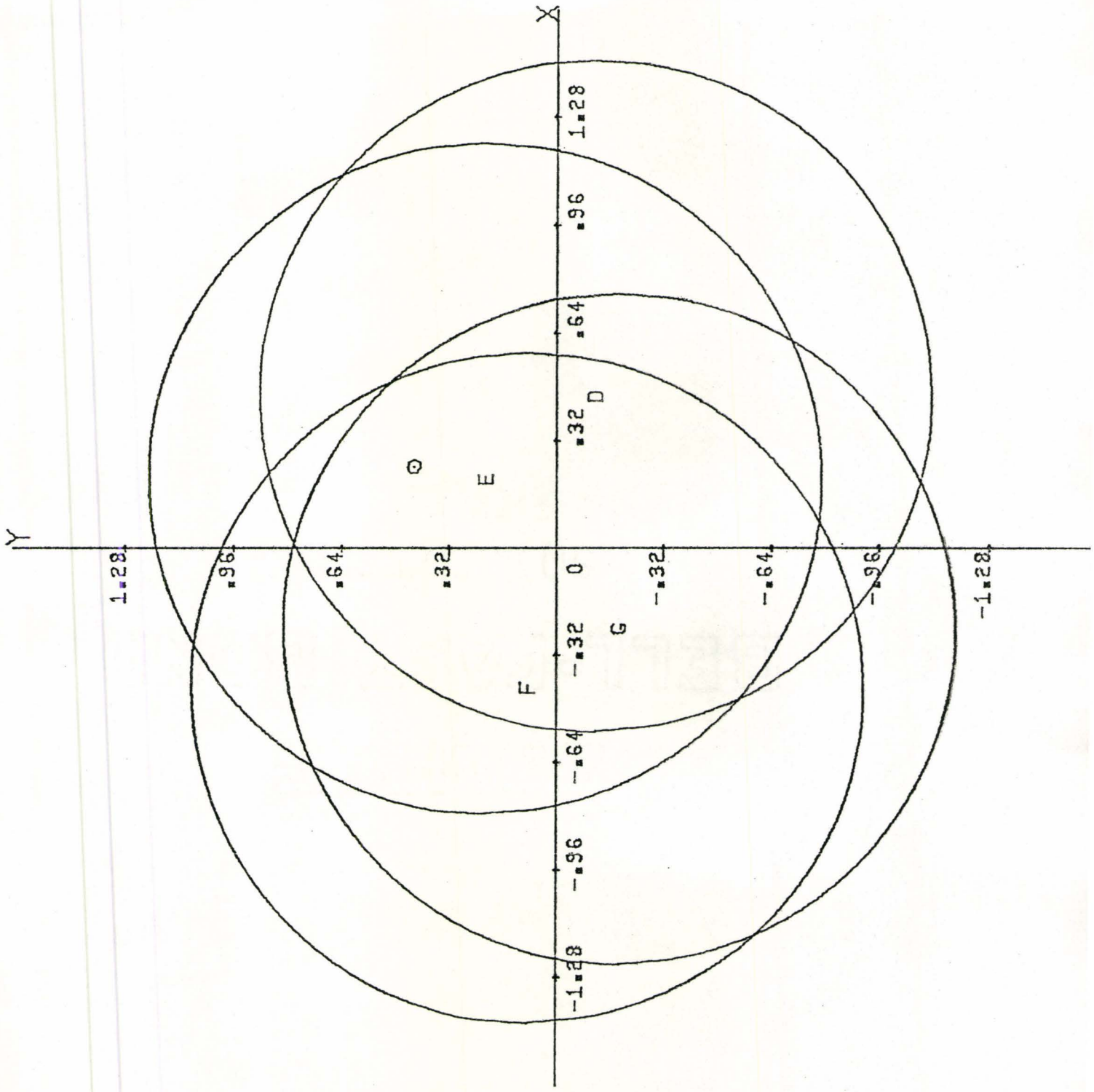


Figure 5.1.1.3. As in Fig. 5.1.1.1 except X=second canonical variate, Y=third canonical variate.

The failure of canonical analysis to obtain discriminating results is a consequence of the high variability of the data within each group relative to the variability between the groups.

5.1.2. *Fourier Series Analysis*

The canonical variates associated with each group mean were plotted in the fourier series functions,

$$f_{\bar{z}_D}(\tau) = {}^{.406}/_{\sqrt{2}} + .451 \sin\tau - .115 \cos\tau$$

$$f_{\bar{z}_E}(\tau) = {}^{-.426}/_{\sqrt{2}} + .206 \sin\tau + .212 \cos\tau$$

$$f_{\bar{z}_F}(\tau) = {}^{.519}/_{\sqrt{2}} - .416 \sin\tau + .087 \cos\tau$$

$$f_{\bar{z}_G}(\tau) = {}^{-.499}/_{\sqrt{2}} - .241 \sin\tau - .185 \cos\tau , \quad 0 \leq \tau \leq 2\pi.$$

These representations of the groups are graphed in Figure 5.1.2.1. The apparent separation of the groups appears promising, however when each fourier function is plotted with its one standard deviation band the distinction between groups diminishes. Figures 5.1.2.2 through 5.1.2.5 are these plots. The patient number 31 of group D also was represented on these graphs to illustrate the ambiguity that would result if one tried to classify her. Her function is defined by

$$f_{z_{31}}(\tau) = {}^{.516}/_{\sqrt{2}} + .246 \sin\tau + .426 \cos\tau.$$

A 90% confidence band for each group would be

$$f_{\bar{z}}(\tau) \pm 3.536 \quad , \qquad 0 \leq \tau \leq 2\pi$$

Figure 5.1.2.1. Fourier Series of three dimensional data. All groups.

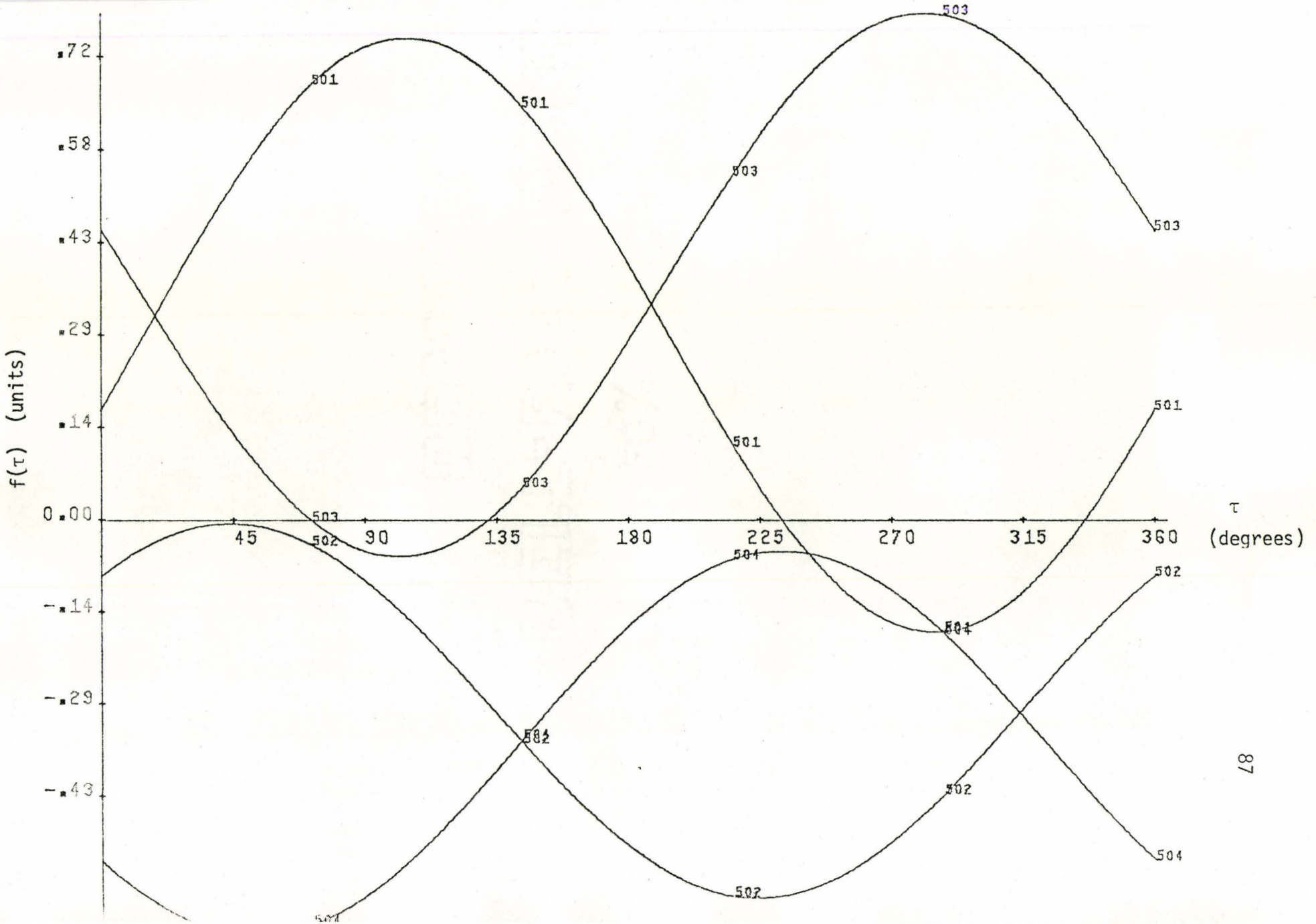


Figure 5.1.2.2. Fourier Series of group D with one standard deviation band and patient 131. Three dimensional data.

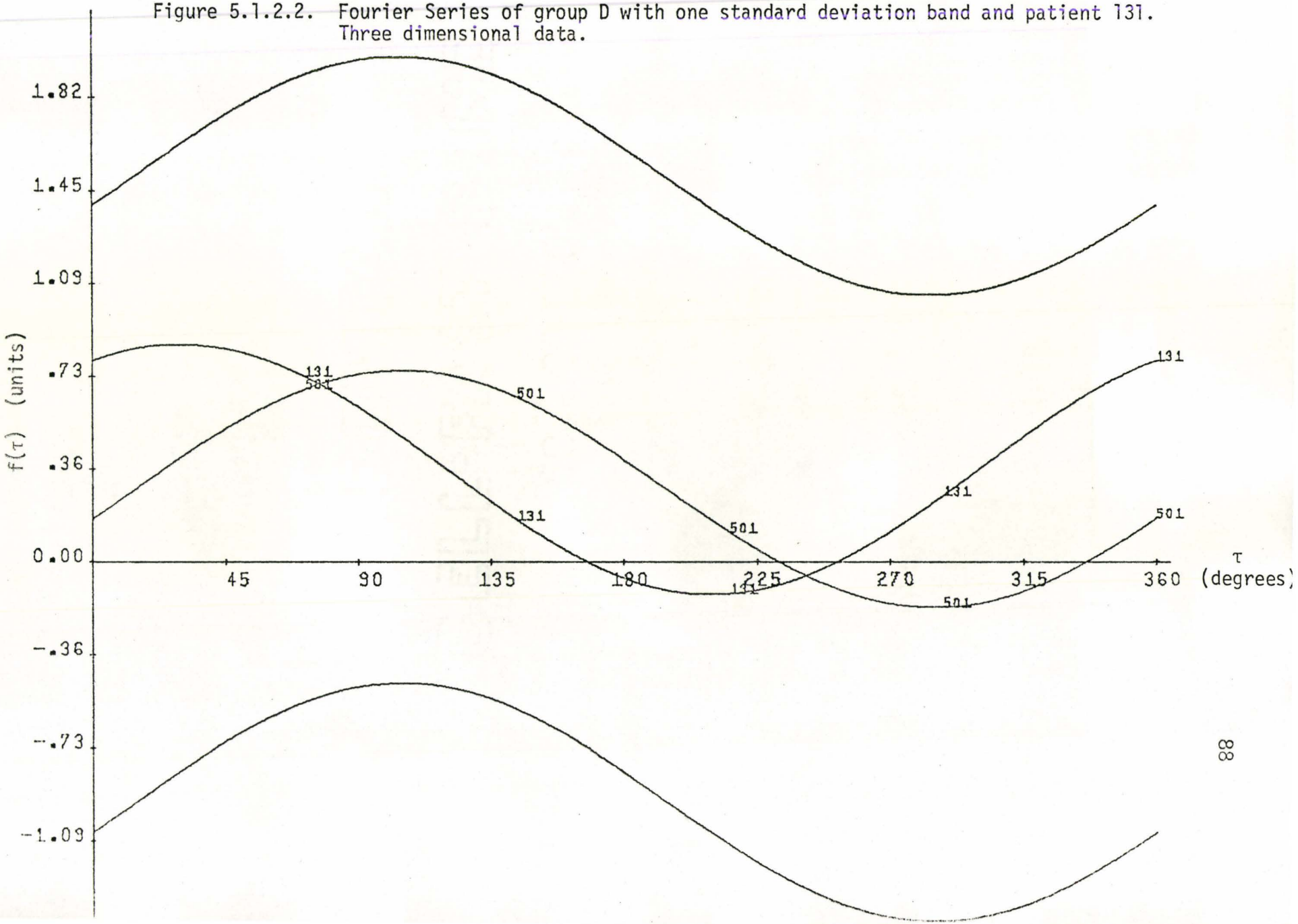


Figure 5.1.2.3. Fourier Series of group E with one standard deviation band and patient 131. Three dimensional data.

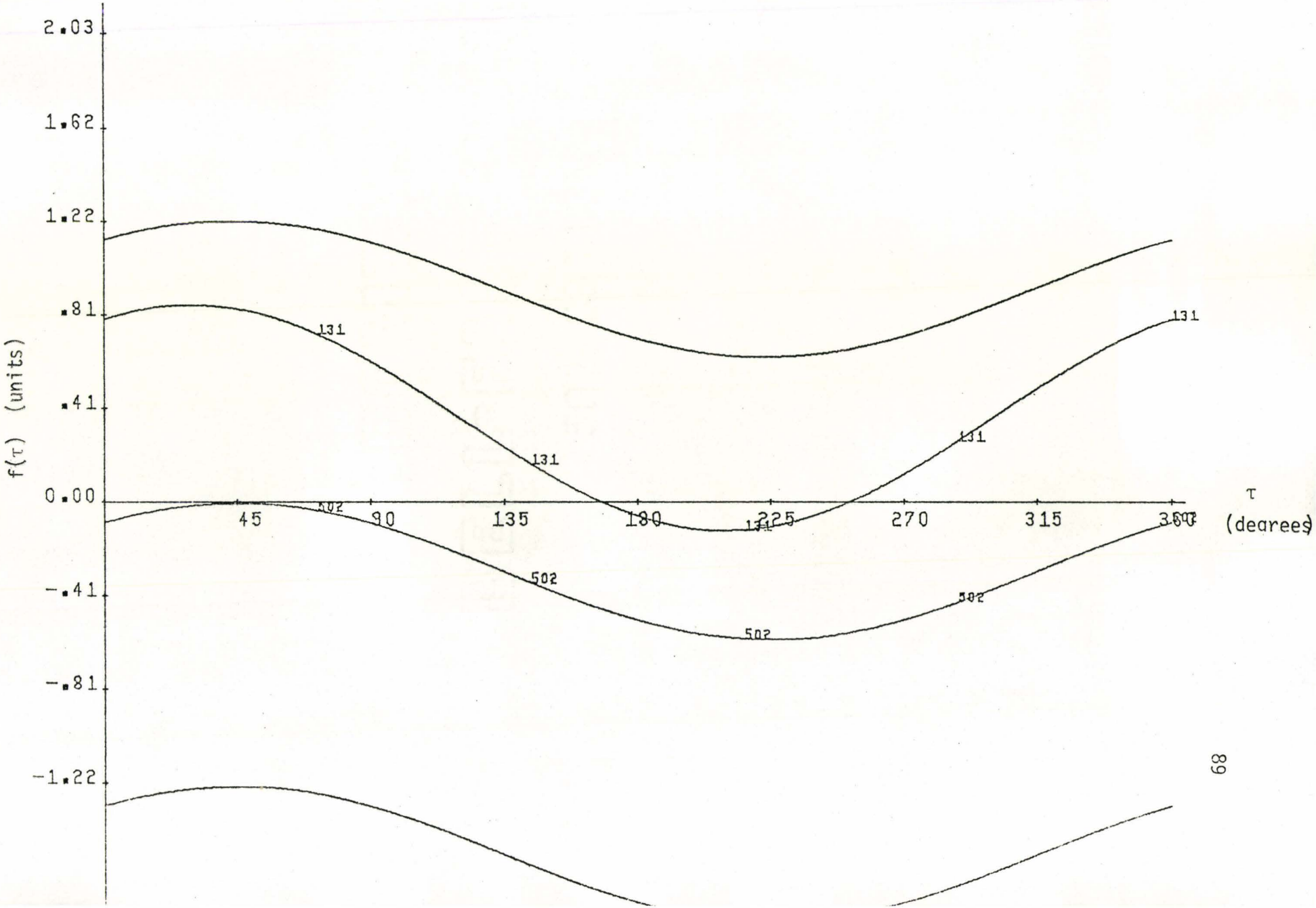


Figure 5.1.2.4. Fourier Series of group F with one standard deviation band and patient 131. Three dimensional data.

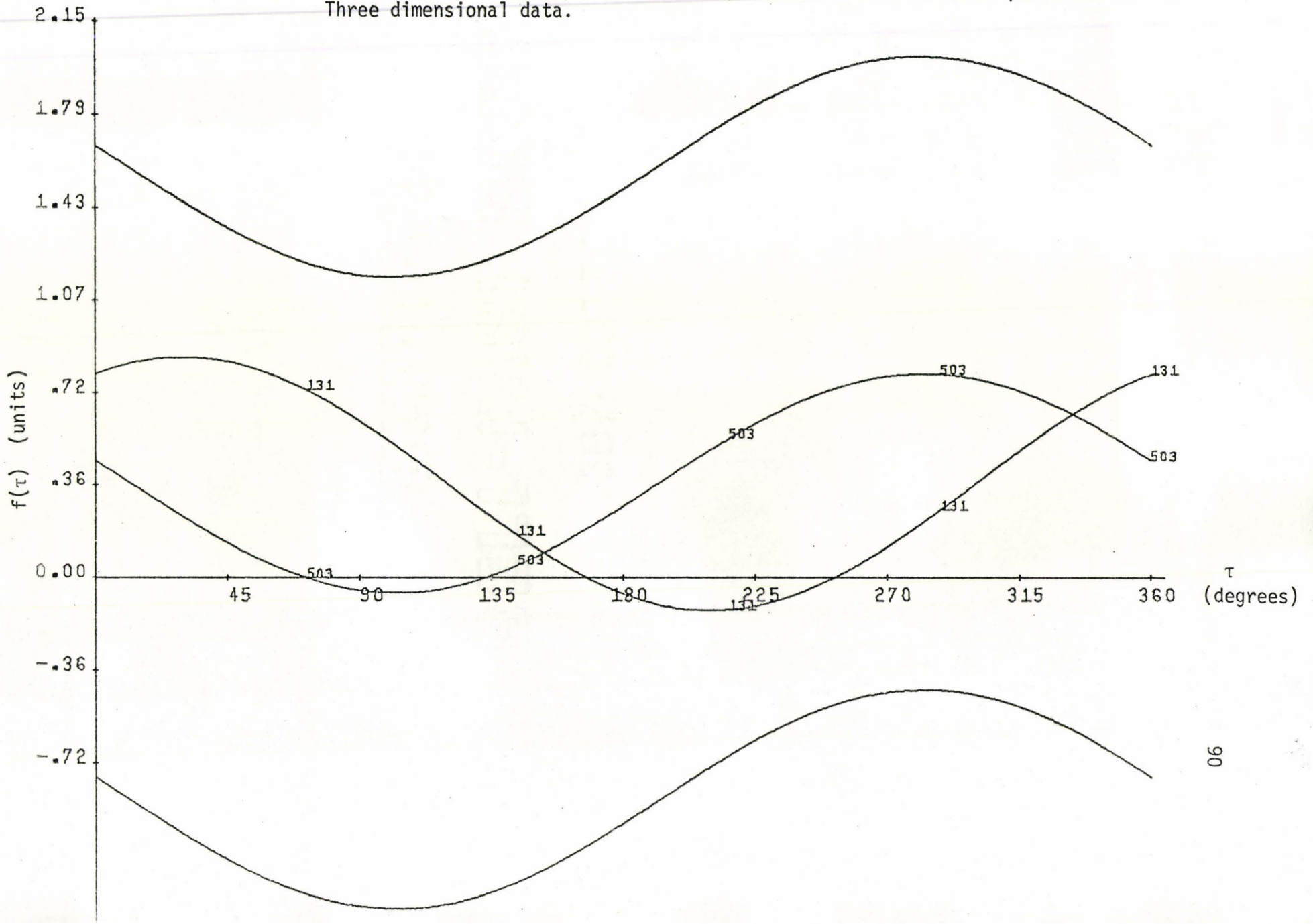
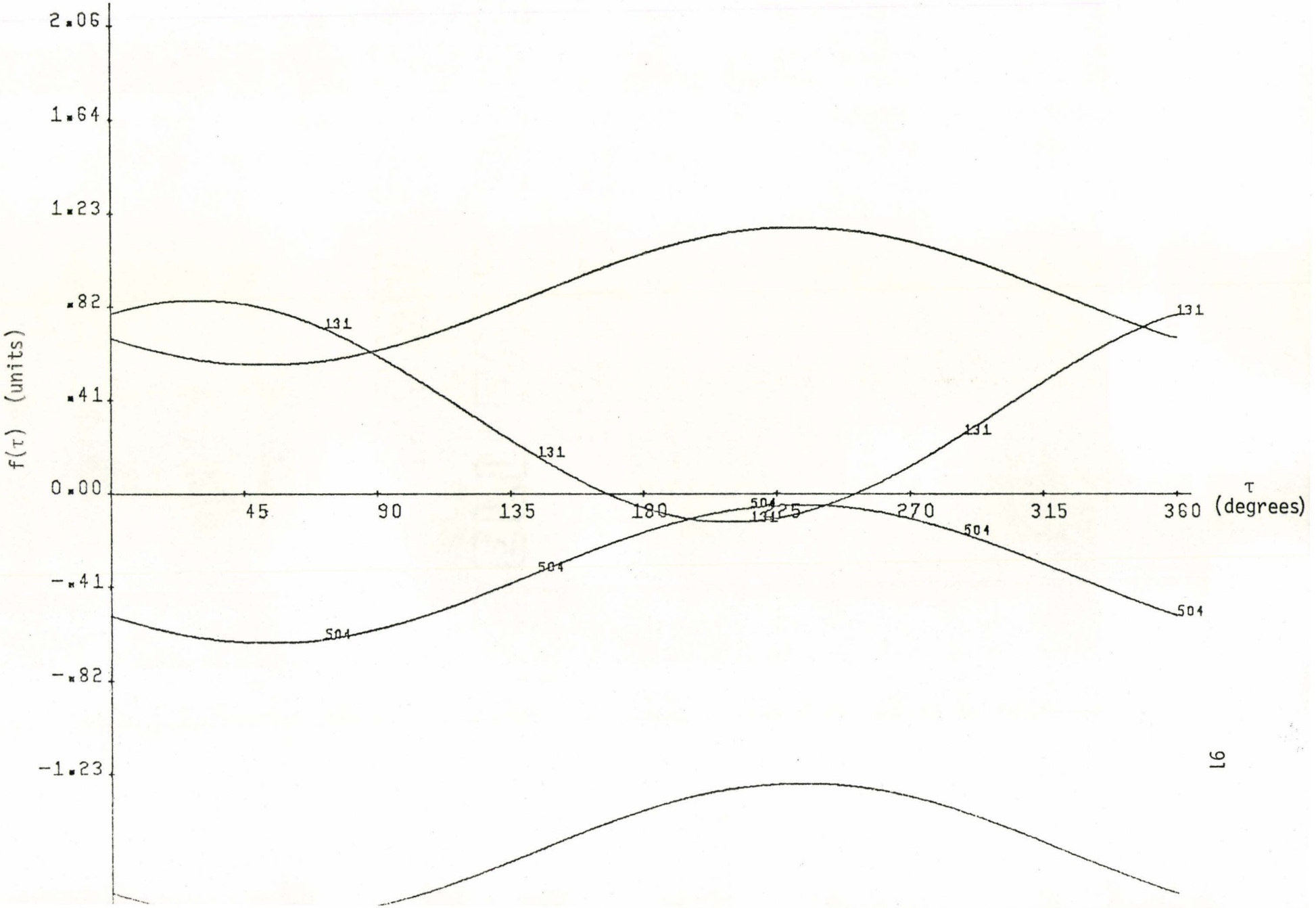


Figure 5.1.2.5. Fourier Series of group G with one standard deviation band and patient 131. Three dimensional data.



which would encompass patient number 31 in all cases.

Attempting to classify a new patient would again be impractical because of the nature of the data.

6. CONCLUSIONS AND REMARKS

For purposes of classification, Andrews' technique appears to be the most general approach. The reason being that one can construct confidence bands about the functions which cannot be done easily with canonical axes if the dimension of the space is greater than or equal to four. In this way, quantitative assignments of individuals may be made with the fourier series approach by assigning those individuals to the population in which their function is entirely encompassed by the population's confidence band.

However, the interpretations of the fourier functions are not clear at all. The complexity of the plots increases with the number of dimensions being represented and will therefore compound the task of interpreting any results of the data. Its use in terms of informative representations is limited.

A by-product of the application of Andrews' technqiue is the determination of outliers in the data. For example, patient 35 of group F (function 335 of Fig. 4.2.2.9) does not appear consistent with the other patients of that group. Inspection of the raw measurements in table 2.2 however would not have revealed this. This property of the plots is very useful in that respect.

Chernoff's faces further contribute to technqiues available in classification. Used in conjunction with an analytic method this

representation of the results can supply very convincing allocation. The classification of patient 31 of group D by the canonical analysis of discriminance is certainly corroborated by the comparison of Figure 4.3.1.5 with Figure 4.3.1.1. In this manner, these faces may be an invaluable tool when attempting to communicate results of an analysis to a client or an employer not familiar with statistics.

The heavy computations were performed by the CDC 6000 computer and the Benson Lehner Plotter was used for the graphic illustrations.

REFERENCES AND RELATED READINGS

The material listed below but not referenced in text was necessary to the understanding of the content of this project. These are marked with an asterisk.

[1] Andrews, D.F. (1972).
Plots of high-dimensional data.
Biometrics 28, 125-136.

[2]* Ashton, E.H., Healy, M.J.R. and Lipton, S. (1956)
The descriptive use of discriminant functions in physical anthropology.
Royal Society of London, B146, 552-572.

[3]* Blackith, R.E. (1960).
A synthesis of multivariate techniques to distinguish patterns of growth in grasshoppers.
Biometrics 16, 28-29.

[4] Chernoff, H. (1973).
The use of faces to represent points in k-dimensional space graphically.
Journal of the American Statistical Association 68, 361-368.

[5]* Eysenck, H.J. (1960).
Experiments in Personality, vol. ll.
London: Routledge & Kegan Paul.

[6] Fisher, R.A. (1936).
The use of multiple measurements in taxonomic problems.
Annals of Eugenics 7, 179.

[7]* Friedman, H.P. and Rubin, G. (1967).
On some invariant criteria for grouping data.
Journal of the American Statistical Association 62, 1159-1178.

[8]* Hammarling, S.J. (1970).
Latent Roots and Latent Vectors.
University of Toronto Press.

[9] Hope, K. (1968).
Plots of high-dimensional data.
Biometrics 28, 125-136.

[10]* Kendall, M.G. (1968).
A course in Multivariate Statistics.
London: Griffin.

[11]* Kendall, M.G. and Stuart, A.S. (1968).
The Advanced Theory of Statistics, vol. 3
London: Griffin.

[12] Morrison, D.F. (1967).
Multivariate Statistical Methods.
New York: McGraw-Hill.

[13] Owen, D.B. (1962).
Handbook of Statistical Tables.
Reading: Addison-Wesley.

[14] Rao, C.R. (1952).
Advanced Statistical Methods in Biometric Research.
New York: Wiley.

[15] Seal, H.L. (1964).
Multivariate Statistical Analysis for Biologists.
Methuen and Co. Ltd.