Alternative Methods for the Analysis of Curves

Randi Christine Martin
Portland State University
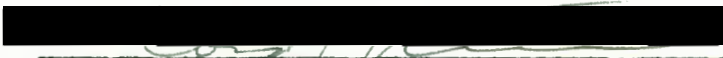
AN ABSTRACT OF THE THESIS OF Randi Christine Martin for the Master of Science in Psychology presented May 23, 1975.

Title: Alternative Methods for the Analysis of Curves.

APPROVED BY MEMBERS OF THE THESIS COMMITTEE:

_____
James Paulson, Chairman

_____
Cord B. Sengstake

_____
Hugo Maynard

The interpretation of data from psychological experiments often involves the analysis and comparison of curves, like the electrocardiogram, spectral sensitivity, or dose-response curves. Existing statistical techniques are often inadequate for making this type of analysis. This thesis presents alternative methods for handling the data comprising curves and discusses the advantages of these techniques against those of existing methods.

Four types of analysis are discussed:

1) finding a confidence band around one curve,

2) finding a confidence band around the difference between two curves, or making an overall comparison between two curves,

3) decomposing a curve into its components,

4) making a component by component comparison of two curves.

The alternative methods presented for the first two types of analysis involve finding simultaneous confidence intervals on the points making up a curve or on the difference between two curves. The first type of analysis, that of finding a confidence band around one curve, is not presently used in psychological studies. The second type of analysis, that of making an overall comparison between two curves, is presently accomplished by visual inspection, analysis of variance, or Hotelling's T^2.

For the third and fourth types of analysis the use of a set of orthogonal functions known as Haar functions is presented as an alternative to the use of orthogonal polynomials.

For each type of analysis, two cases are presented, each of which makes different assumptions about the shape of the covariance matrix. Case I assumes that the covariance matrix is unknown, while case II assumes that the covariance matrix is known except for a scalar.

To illustrate the alternative techniques for case I, data from an actual experiment on the effect of Δ^9-THC (tetrahydrocannabinol) on the p-wave of the electrocardiogram is used. To illustrate case II, simulated data is used involving a dose-response curve generated by testing independent groups on a memory task at four levels of a synthetic analogue of Δ^9-THC. This is compared to a dose-response curve testing independent groups on the same task at the same dosage levels of Δ^9-THC.

As was the case for the examples, the p-wave and the dose-response curves, the alternative techniques provide adequate means

for the analysis and comparison of curves. In some cases, as in the use of Haar functions for the decomposition of the p-wave, the alternative techniques provide advantages over existing methods. It is hoped that the new techniques will be incorporated into the analysis of psychological data.

ALTERNATIVE METHODS FOR THE ANALYSIS OF CURVES

by

RANDI CHRISTINE MARTIN

A thesis submitted in partial fulfillment of the
requirements for the degree of

MASTER OF SCIENCE
in
PSYCHOLOGY

Portland State University
1975

TO THE OFFICE OF GRADUATE STUDIES AND RESEARCH:

The members of the Committee approve the thesis of Randi Christine Martin presented May 23, 1975.

James Paulson, Chairman
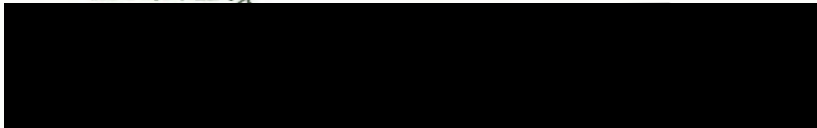
Cord B. Sengstake

Hugo Maynard

APPROVED:

Ronald E. Smith, Head, Department of Psychology

David P. Clark, Dean of Graduate Studies and Research

July 23, 1975

ACKNOWLEDGMENTS

I would like to express my thanks to the members of my com-
mittee for the assistance they have given me in the preparation of
this thesis. Through the several months of working on this thesis,
Dr. James Paulson has given many helpful suggestions and criticisms
as to the statistical content and organization of ideas. Dr. Cord
Sengstake and Dr. Hugo Maynard were helpful in suggesting changes
which would make the material more accessible to the general body
of psychologists who lack an extensive mathematical background. I
would also like to thank Dr. Sengstake for the use of his data on
the electrocardiogram and for his assistance in the use of the
PDP-11 computer to transform this data into usable form.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

I. INTRODUCTION

Many problems within psychology deal with the relationship of a dependent variable to an independent variable which varies quantitatively along some continuum. For example, a learning curve shows the relationship between number of correct responses and number of trials. Another example is a waveform such as an electrocardiogram, thought of in its digital representation, which shows voltage over time. A third example would be spectral sensitivity measured over various wavelengths of light. Typically this relationship of independent variable to dependent variable is charted as some form of curve and this curve compared to another curve obtained under another experimental condition, or to one predicted by some theoretical relation between the two variables.

The previously mentioned examples would most likely involve repeated measures on the same subject, that is, the same subject tested at each level of the independent variable. However, experiments involving independent groups for the different levels could also give rise to data in the form of curves. An example would be a study involving performance on a certain task after administration of differing doses of a drug with independent groups at each dosage level. These results could be graphed as a dose-response curve.

Although this type of data is common in psychology, the existing statistical techniques for dealing with curves are often inadequate and for some types of data non-existent. The intent of this thesis will be

to propose techniques for dealing with the analysis of curves. A first
objective will be to review existing methods of curve analysis point-
ing out their assumptions and types of problems to which they are most
applicable. Secondly, I will propose alternative techniques which for
some kinds of problems will provide advantages over existing methods.

Four kinds of analysis are typically made when considering the
data comprising curves:

1) finding a confidence interval around each of the points making up a
curve, or equivalently, finding a confidence band around the entire curve,

2) finding an overall difference between two or more curves, or the
confidence band for the difference between two or more curves,

3) decomposing a curve into its components,

4) making a component by component comparison of two or more curves.
These four kinds of analysis will be discussed in sections II through
V of this paper.

Before beginning the discussion of techniques some explanation of
the notation used will be necessary:

lower case letters will represent scalars: a, b, x, y, α

lower case letters which are underlined will represent vectors: $\underline{x}$, $\underline{y}$, μ

capital letters will represent matrices: S, X, Σ

The most commonly used vectors will be $\overline{y}$ and μ. $\overline{y}$ is the sample
mean vector which is a set of measurements showing the average value
of the dependent variable at each level of the independent variable.
For example if the following values of the dependent variable y
were found for the following values of x (x = x_1, x_2, x_3, x_4):

	x			
	x_1	x_2	x_3	x_4
	98	110	120	126
y	100	112	113	130
	102	107	115	140
average	100	109.6	116	132

$\bar{y}$ would equal (100, 109.6, 116, 132). The elements of $\bar{y}$ will be labelled $\bar{y}_i$ where $\bar{y} = (\bar{y}_1, \bar{y}_2, \bar{y}_3, \bar{y}_4, \ldots \ldots, \bar{y}_p)$ given p levels of the independent variable. In the previous example:

$$\bar{y}_1 = 100, \quad \bar{y}_2 = 109.6, \quad \bar{y}_3 = 116, \quad \bar{y}_4 = 132.$$

μ is the population mean vector and is not a random vector. $\bar{y}$ is an estimate of μ.

The most commonly used matrics will be Σ and S. Σ is the variance-covariance matrix of μ and S is the estimate of this matrix. In all cases $\bar{y}$ will be assumed to have a multivariate normal distribution with mean μ and covariance Σ.

Also, the mathematical symbols $\in$ and $\forall$ will be used. $\in$ means 'is contained in' and $\forall$ means 'for every.'

Vector multiplication and matrix multiplication will be used to simplify notation. The appendix gives an explanation of how to perform vector and matrix multiplication.

II.   FINDING A CONFIDENCE BAND AROUND ONE CURVE

Returning to the types of analysis made when dealing with curves, the first to be considered will be that of finding a confidence band around a curve. This is similar to the univariate problem of finding a confidence interval around the mean of a sample. In the univariate case, one finds the mean and uses tables of the t-distribution to find the interval:

$$\mu \in y + t_{n-1}^{\alpha/2} \frac{s}{\sqrt{n-1}}$$

In the case of curves, one wishes to arrive at a confidence band for the entire curve. For example, suppose that an experimenter has fifty observations of an electrocardiogram and wishes to find a confidence band for the average of these fifty. A natural inclination is to calculate the interval described above for each of the points making up the curve using an α of .05 for each point This is incorrect, however. Taking an α of .05 means that there is a probability of .05 that the confidence interval will not contain the true mean, μ_i, for that point. The probability that any part of the curve could lie outside the confidence band for the entire curve will be greater than .05 because we are concerned with the joint probability that one or more confidence intervals will not contain the true mean for that point while the remaining confidence intervals could cover their μ_i. The value of the joint probability will depend on the degree to which the points are correlated. Only if the points are all perfectly correlated will the α level for the

entire curve be as small as that for each point. If the points are all independent, the probability that the confidence band will cover all the points would be the product of the probabilities for the individual confidence intervals. For a curve comprised of four points with each confidence interval having a probability of .95 of covering the true mean, the joint probability for independent points would be

$$(.95) \times (.95) \times (.95) \times (.95) = .815$$

not an acceptable confidence level.

Since in actual data the points will most likely not be perfectly correlated, the simultaneous probability will be less than $1-\alpha$. An alternative technique should provide confidence intervals for each of the points all of which hold simultaneously with probability $1-\alpha$. The simultaneous multivariate techniques to be described in this paper were originally developed by Roy and Bose (1953) and are summarized in Miller's Simultaneous Statistical Inference (1966).

Miller presents two cases in the discussion of finding a confidence interval around a curve. One must determine if his data fit the assumptions of the first case or the second. In the first case, the covariance matrix is completely unknown. Most data encountered in the comparison of curve will fall under this case due to lack of knowledge about the shape of the covariance matrix. The second case assumes that the covariance matrix is known except for a scalar.

CASE I: COVARIANCE MATRIX UNKNOWN

Either of two expressions may be used to determine confidence intervals about the co-ordinate means:

(1) $\quad \mu_i \in \overline{y}_i \pm t_{n-1}^{\alpha/2p}\left(\frac{s_i^2}{n}\right)^{1/2}$ $\qquad i = 1,\ldots,p$

where $s_i^2 = \frac{1}{n-1}\sum_{j=1}^{n}(y_{ij} - \overline{y}_i)^2$, or

(2) $\quad \mu \in \overline{y}_i \pm \left(pF_{\rho,n-p}^{\alpha}\right)^{1/2}\left(\frac{n-1}{n-p}\right)^{1/2}\left(\frac{s_i^2}{n}\right)^{1/2}$ $\qquad i = 1,\ldots,p.$

Expression (1) is based on the Bonferroni inequality. If an overall α level of significance is desired, using an $\alpha/2p$ confidence level for each point (where p is the number of points) will, according to the Bonferroni inequality, assure that the p statements in expression (1) hold simultaneously with probability greater than $1-\alpha$.

Expression (2) is derived from the results of Roy and Bose(1953) and is presented in Miller (1966). Expression (1) will generally give smaller confidence intervals. For p=1, expression (1) is equal to expression (2) since

$$\left(t_{n-1}^{\alpha/2}\right)^2 = F_{1,n-1}^{\alpha} \qquad .$$

For p 2,

$$\left(t_{n-1}^{\alpha/2p}\right)^2 < F_{\rho,n-p}^{\alpha}$$

for small n. However, even when the F value is less than t value squared, the fact that expression (2) requires multiplication of the F value by $p\left(\frac{n-1}{n-p}\right)$, results generally in the Bonferroni intervals being smaller. If in doubt one should compare $t_{n-1}^{\alpha/2p}$ to $\left(pF_{\rho,n-p}^{\alpha}\left(\frac{n-1}{n-p}\right)\right)^{1/2}$ to determine whether (1) or (2) will give the smaller intervals.

An example of Case I taken from an actual experiment involves ob-servations of the electrocardiogram. The ECG was recorded on magnetic tape and transformed by analogue to digital converter to numerical data. The mean and variance of the thirty-two points making up the

p-wave are shown in Table I. Figure 1 shows a drawing of the average p-wave with confidence band derived using the formula for the Bonferroni intervals.

CASE II: COVARIANCE MATRIX KNOWN EXCEPT FOR A SCALAR

The second case is a special case which assumes that the co-variance matrix is known except for a scalar, that is,

$$\Sigma = \sigma^2 \Sigma'$$

where Σ' is known and σ^2 is unknown. For this case, Miller shows that the following expression gives Bonferoni intervals around each point such that the simultaneous probability that each point lies within the confidence band is greater than $1-\alpha$:

(3) $\mu_i \in \bar{y}_i \pm t_{p(n-1)}^{\alpha/2p} \dfrac{s}{\sqrt{n}} \sqrt{\sigma_{ii}}$ $i = 1,\ldots,p$

where $s^2 = \dfrac{1}{p-1} \displaystyle\sum_{j=1}^{n} (\underline{y}_j - \bar{\underline{y}})' (\Sigma')^{-1} (\underline{y}_j - \bar{\underline{y}})'$

and σ_{ii} is the ith diagonal element of the known matrix Σ'.

Alternatively, one could use:

(4) $\mu_i \in \bar{y}_i \pm \left(pF_{p,p(n-i)}^{\alpha}\right)^{\frac{1}{2}} \dfrac{s}{\sqrt{n}} \sqrt{\sigma_{ii}}$ $i = 1,\ldots,p.$

However, as before, the Bonferroni intervals will usually be smaller.

Examples of this case are much less common. One example would be where the curve was derived from observations on independent groups at each level of the independent variable. In this case the covariance matrix would be diagonal:

TABLE I

MEAN, VARIANCE, AND CONFIDENCE INTERVAL
FOR AVERAGE P-WAVE

point	mean	variance	Bonferroni intervals*
1	168.92	3.34	168.92 + .87
2	169.54	3.03	169.54 + .83
3	169.58	5.02	169.58 + 1.06
4	170.02	5.24	170.02 + 1.08
5	171.48	4.70	171.48 + 1.03
6	173.62	6.73	173.62 + 1.23
7	176.60	6.82	176.60 + 1,24
8	179.30	6.99	179.30 + 1.26
9	180.72	4.98	180.72 + 1.06
10	181.34	4.80	181.34 + 1.04
11	180.40	7.84	180.40 + 1.33
12	179.46	5.40	179.46 + 1.10
13	180.52	10.91	180.52 + 1.57
14	148.72	28.49	184.72 + 2.54
15	190.92	23.63	190.92 + 2.31
16	195.22	9.97	195.22 + 1.50
17	196.46	6.99	196.46 + 1.26
18	194.40	10.24	194.40 + 1.52
19	189.54	34.42	189.54 + 2.79
20	182.94	51.57	182.94 + 3.41
21	176.32	47.65	176.32 + 3.28
22	170.32	33.80	170.32 + 2.76
23	165.14	26.69	165.14 + 2.45
24	160.32	21.40	160.32 + 2.20
25	157.18	10.32	157.18 + 1.53
26	155.08	5.50	155.08 + 1.11
27	154.32	3.81	154.32 + .93
28	154.22	3.24	154.22 + .88
29	154.96	3.75	154.96 + .92
30	155.56	3.64	155.56 + .91
31	155.90	2.62	155.90 + .77
32	156.42	2.66	156.42 + .77

*α = .05
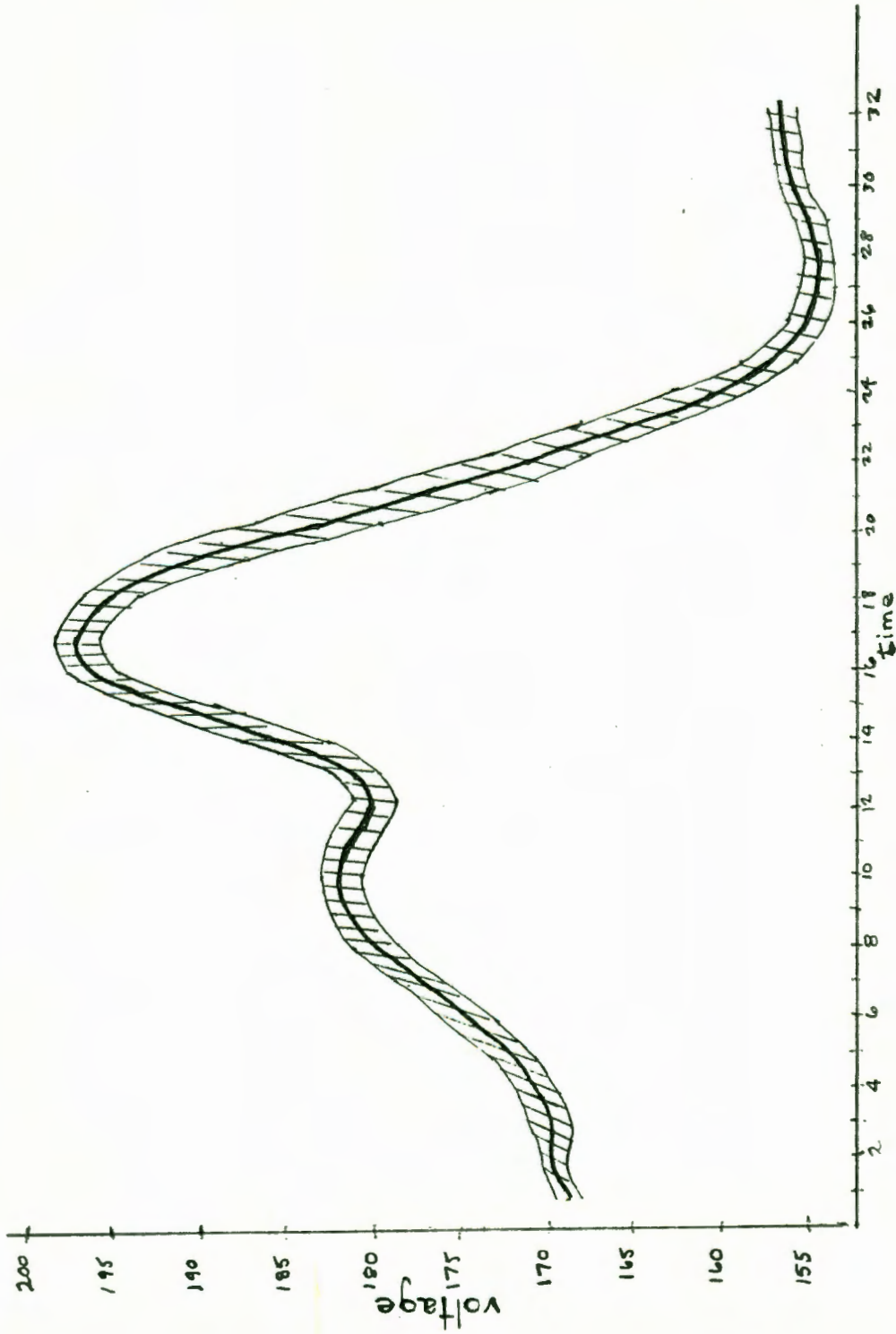
Figure 1. Confidence band around average p-wave.

$$\sum \quad = \quad \begin{pmatrix} a_1 & 0 & 0 & . & . & . & . & . & 0 \\ 0 & a_2 & 0 & . & . & . & . & 0 \\ 0 & . & . & a_i & . & . & . & 0 \\ 0 & 0 & . & . & . & . & . & a_p \end{pmatrix}$$

Assuming that the a_i were equal or that they were all multiples of some common value ($a_i = c_i \sigma^2$ where the c_i are known), then the assumptions of case II would be met:

$$\sum = a \begin{pmatrix} 1 & 0 & 0 & . & . & . & . & 0 \\ 0 & 1 & 0 & . & . & . & . & 0 \\ 0 & . & . & 1 & . & . & . & 0 \\ 0 & 0 & . & . & . & . & . & 1 \end{pmatrix} \qquad (a = \sigma^2)$$

or

$$\sum = \sigma^2 \begin{pmatrix} c_1 & 0 & 0 & . & . & . & . & 0 \\ 0 & c_2 & 0 & . & . & . & . & 0 \\ 0 & . & . & c_i & . & . & . & 0 \\ 0 & 0 & . & . & . & . & . & c_p \end{pmatrix}$$

For instance, suppose that the $\underline{y}$ represented performance measures on a memory task at four equally spaced dosages (0.0 mg/kg, 0.5 mg/kg, 1.0 mg/kg, 1.5 mg/kg) of a drug which is a synthetic analogue of Δ^9- tetrahydrocannabinol (syn. Δ^9-THC). If the experimenter had considerable experience with data derived from these dosage levels of this drug, he could possibly arrive at estimates of the a_i. Suppose that he noted twice as much variance in the perform-

ance at the upper two levels of the drug as at the lower two. Then,

$$\Sigma = \sigma^2 \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 2 & 0 \\ 0 & 0 & 0 & 2 \end{pmatrix}$$

assuming that independent groups were used at each dosage level.

Table II shows the raw scores for each of ten subjects used at each dosage level (data simulated). Table III gives the values of $\overline{y}_i$ and the confidence interval around each point which was computed using expression (3) (s^2 was calculated to be 56.43 and α was set at .05). Figure 2 shows the confidence band around the mean curve.

TABLE II

PERFORMANCE OF INDEPENDENT GROUPS
ON MEMORY TASK AT FOUR LEVELS
OF SYN. Δ^9-THC

0.0 mg/kg	0.5 mg/kg	1.0 mg/kg	1.5 mg/kg
58	62	48	43
60	60	39	39
63	65	38	27
55	61	45	46
52	64	52	43
58	60	55	35
57	55	37	31
59	61	32	41
58	63	47	28
60	62	51	36
Total 580	613	444	369

TABLE III

AVERAGE SCORE AND CONFIDENCE INTERVAL
AROUND EACH POINT FOR
SYN. Δ^9-THC

point	$\overline{y}_i$	Bonferroni confidence interval
1	58.0	58.0 + 6.23
2	61.3	61.3 + 6.23
3	44.4	44.4 + 8.85
4	36.9	36.9 + 8.85
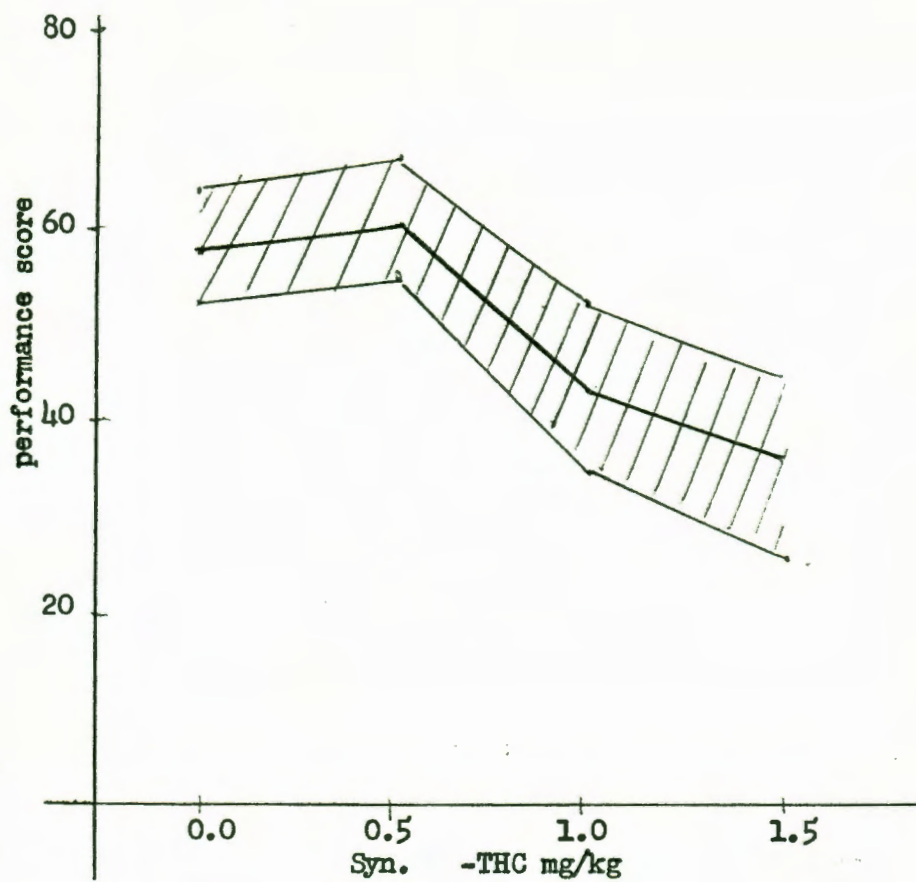
Figure 2. Confidence band around performance scores at four dosage levels of a synthetic analoque of Δ-THC.

III. FINDING A CONFIDENCE BAND FOR THE DIFFERENCE BETWEEN TWO CURVES

The second type of analysis to be considered is that of finding a confidence interval around the difference between two or more curves. This analysis has as its univariate analogue the comparison of the means of two groups, or finding a confidence interval for the difference between two groups. In the univariate case, one finds the mean for each of the two samples and then finds the confidence interval for the difference using the following formula:

$$\mu_d \in (\bar{y}_1 - \bar{y}_2) \pm t^{\alpha/2}_{n_1+n_2-2} \, \bar{\bar{\sigma}}_d$$

where $\bar{\bar{\sigma}}_d = \sqrt{\dfrac{(n_1-1)s^2 + (n_2-1)s^2}{n_1 + n_2 - 2}} \sqrt{\dfrac{1}{n_1} + \dfrac{1}{n_2}}$

In the case of curve comparisons, one wishes to find a confidence band on the difference between two average curves. An example would be comparing the average ECG of an animal before administration of a drug to the average ECG after administration (see below).

ECG before   ECG after

Current methods of making such comparisons are visual inspection of the graphs, analysis of variance or Hotelling's T^2.

Visual inspection has been commonly used in the comparison of waveforms like the ECG. In such waveforms there are often regularly occurring phenomena which can easily be identified by visual inspection as in the above drawings. The difficulty with visual inspection for curve comparison is similar to inspection of single values in the uni-

variate case: the variance about each point may be so large, that apparent differences are not statistically significant; that is, the differences could be due to chance alone.

Existing statistical techniques, on the other hand, may obfuscate real differences, hiding differences that would be significant by the alternative techniques to be presented. Use of the analysis of variance for the comparison of two curves may often hide real differences.

Analysis of variance can be used for the comparison of curves by considering each point as one column in the analysis, and each sample of curves as a row. The design would be interpreted as a two-way ANOVA, with repeated measures on the row variables. Only an overall difference will be apparent. If a difference exists in a small portion of the total number of points, it may be that there will be no significant difference between rows. Differences in different segments will show up as significant interaction, but will not indicate at which points the differences occur, thus obscuring a difference that would be obvious to visual inspection.

Another difficulty with the use of ANOVA presents itself when the points of the curve represent repeated measures on the same subjects. In using ANOVA, it is assumed that the variance-covariance matrix has compound symmetry, that is, that:

$$\Sigma = \sigma^2 \begin{pmatrix} 1 & \rho & \rho & \cdot & \cdot & \cdot & \rho \\ \rho & 1 & \rho & \cdot & \cdot & \cdot & \rho \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ \rho & \rho & \cdot & \cdot & \cdot & \cdot & 1 \end{pmatrix}$$

In the case of independent groups, $\rho = 0$, and this assumption amounts to saying that the treatment means have equal variance. Box (1954) has shown that in the usual ANOVA (not repeated measures), the F-test is relatively insensitive to violations of the assumption of compound symmetry. However, in the case of repeated measures (or whenever observations are correlated) violation of this assumption will result in a positive bias (an F value which is too low).

Therefore, before performing ANOVA with repeated measures, it is necessary to test that Σ has the above form. This involves testing the hypothesis that the covariance matrices for each sample are equal and then testing that the pooled matrix has variances equal to the average variance of the pooled estimate and covariances equal to the average covariance of the pooled estimate (Winer, 1971). The computations involve finding the matrix of the average variance and covariances, the determinant of the pooled estimate, and the natural log of the determinant. If the matrix is found to have the required form, then one can proceed with the ANOVA. This procedure of determining if Σ has the proper form is already more difficult computationally than proceeding with Miller's simultaneous techniques to be presented below. Also, if one can proceed with ANOVA, this will still involve the limitations described above regarding the analysis of differences in different portions of the curves.

If the matrix does not have the required shape, the usual procedure is to use Hotelling's T^2 to determine if there are column differences (Winer, 1971). However, this is a test of column effects only, not of row differences or interaction. When considering differences between two or more curves like the ECG or spectral sensitivity

curves, one already knows that there are differences in level between the points of the curve inherent in the shape of the curve. One would be interested in differences in shape between different groups, that is, in row differences and interaction.

The T^2 test can still be used to test for differences between curves by using the procedure described in Morrison's Multivariate Statistical Methods (1967). Computation of T^2 is difficult, involving the calculation of a matrix inverse. The confidence regions that can be determined are p-dimensional ellipsoids. With $p > 3$, these regions become difficult to visualize. Also, this test, like ANOVA, gives just an overall difference rather than a difference at each point.

An alternative technique presented by Miller provides for the determination of a confidence interval for the difference between the means of each co-ordinate of the curve, allowing one to distinguish at which points the curves differ and at which they are the same. Roy and Bose (1953) developed this method by determining the point by point implication of the p-dimensional ellipsoid derived from Hotelling's T^2. The computations involved in this technique are much simpler than those required for Hotelling's T^2 (T^2 does not have to be calculated) or those required for ANOVA. All that is required is an estimate of the variance about each point for case I or an estimate of the unknown scalar for case II. Since case I makes no assumption about the shape of the covariance matrix, there is no need to make the tests for compound symmetry.

CASE I: COVARIANCE MATRIX UNKNOWN

In the general case where the covariance matrix is unknown, the formula for finding the confidence interval about the difference between two curves based on the Bonferoni inequality is:

$$(5) \qquad \mu_{1i} - \mu_{2i} \in \bar{y}_{1i} - \bar{y}_{2i} = t_{2(n-1)}^{\alpha/2p} \left(\frac{2s_i^2}{n} \right)^{1/2}$$

where s_i^2 is the pooled estimate of the variance at each point and where the $\bar{y}_{1i}$ are co-ordinate means of the first sample and the $\bar{y}_{2i}$ are the co-ordinate means of the second sample.

An example of this case would be the comparison of the p-wave before administration of a drug to the p-wave after administration. Data from an actual experiment gave the values shown in Table IV for the mean and variance of the p-waves before and after administration of Δ^9-THC. A sample of fifty was taken in each case. The difference between the two curves and the confidence band about the difference is also shown in table IV. Figure 3 shows the average curves of the two curves being compared. Shown in figure 4 is the confidence band about the difference between the two curves. The two curves are different at all points except 24 and 25 where the confidence bands include zero, with the largest differences occurring between points 14 and 20. This analysis provides statistical validation of what would be obvious to visual inspection.

CASE II: COVARIANCE MATRIX KNWON EXCEPT FOR A SCALAR

The second case assumes that the covariance matrix is known except for a constant ($\Sigma = \sigma^2 \Sigma'$ where Σ' is known and σ^2 unknown).

TABLE IV

CONFIDENCE BAND ABOUT DIFFERENCE BETWEEN
AVERAGE P-WAVES BEFORE AND AFTER
ADMINISTRATION OF   -THC

pt.	$\bar{y}_{1i}$	$\bar{y}_{2i}$	$\bar{y}_{1i} - \bar{y}_{2i}$	s_{1i}^2	s_{2i}^2	pooled s_i^2	confidence interval*
1	168.92	159.54	9.48	3.34	8.58	5.96	9.48 + 1.59
2	169.54	159.94	9.60	3.03	13.98	8.50	9.60 + 1.90
3	169.58	159.58	10.00	5.02	10.29	7.66	10.00 + 1.80
4	170.02	159.42	10.60	5.24	6.53	5.89	10.60 + 1.58
5	171.48	159.68	11.80	4.70	5.20	4.95	11.80 + 1.45
6	173.62	159.78	13.84	6.73	6.54	6.64	13.84 + 1.68
7	176.60	158.98	17.62	6.82	7.98	7.40	17.62 + 1.77
8	179.30	157.96	21.34	6.99	9.39	8.19	21.34 + 1.87
9	180.72	156.82	23.90	4.98	18.88	11.93	23.90 + 2.25
10	181.34	155.66	25.68	4.80	22.72	11.84	25.68 + 2.24
11	180.40	154.46	25.94	7.84	28.87	18.35	25.94 + 2.79
12	179.46	152.52	26.94	5.40	42.05	23.72	26.94 + 3.17
13	180.52	149.46	31.06	10.91	51.27	31.09	31.06 + 3.63
14	184.72	146.70	38.02	28.49	85.89	57.19	38.02 + 4.93
15	190.92	143.88	47.04	23.63	93.70	58.66	47.04 + 4.99
16	195.22	142.14	53.08	9.97	93.80	51.89	53.08 + 4.70
17	196.46	141.36	55.10	6.99	61.95	34.47	55.10 + 3.82
18	194.40	142.16	52.24	10.24	43.44	26.84	52.24 + 3.37
19	189.54	143.30	46.24	34.42	50.91	42.67	46.24 + 4.26
20	182.94	145.24	37.70	51.57	53.86	52.71	37.70 + 4.73
21	176.32	148.16	28.16	47.65	50.14	48.89	28.16 + 4.56
22	170.32	150.46	19.86	33.80	43.40	38.60	19.86 + 4.05
23	165.14	153.64	11.50	26.69	38.32	32.51	11.50 + 3.72
24	160.32	157.10	3.22	21.40	43.23	32.31	3.22 + 3.71
25	157.18	160.34	-3.15	10.32	48.15	29.23	-3.15 + 3.53
26	155.08	162.66	-7.58	5.50	63.54	34.52	-7.58 + 3.83
27	154.32	164.64	-10.32	3.81	58.97	31.39	-10.32 + 3.65
28	154.22	166.88	-12.66	3.24	47.74	25.49	-12.66 + 3.29
29	154.96	168.84	-13.88	3.75	28.72	16.24	-13.88 + 2.63
30	155.56	169.80	-14.24	3.64	22.86	13.25	-14.24 + 2.37
31	155.90	170.22	-14.32	2.62	20.79	11.71	-14.32 + 2.23
32	156.42	170.06	-13.64	2.66	26.87	14.76	-13.64 + 2.50

* α = .05
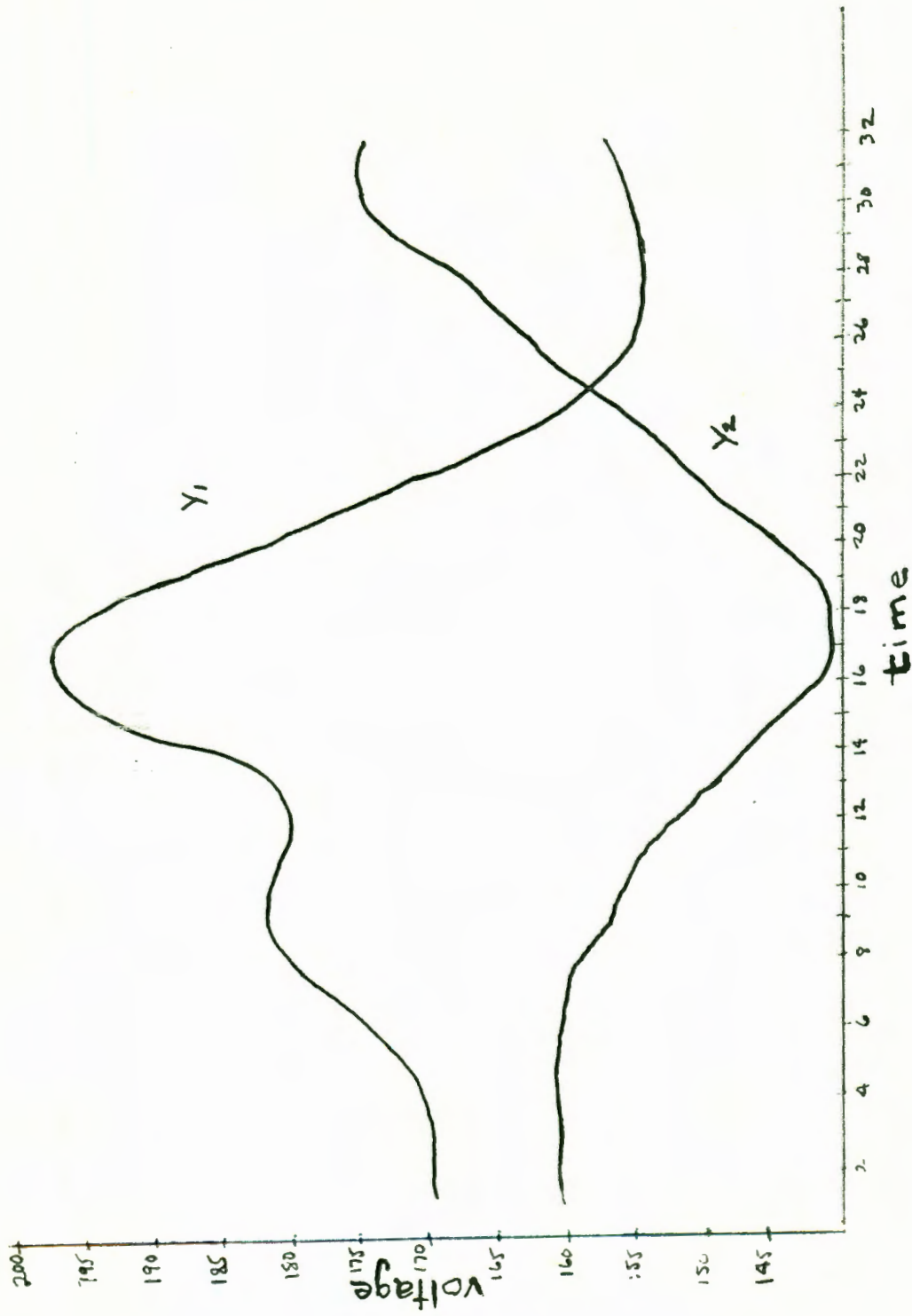
Figure 3. Average p-wave before (y_1) and after (y_2) administration of Δ^9- THC.
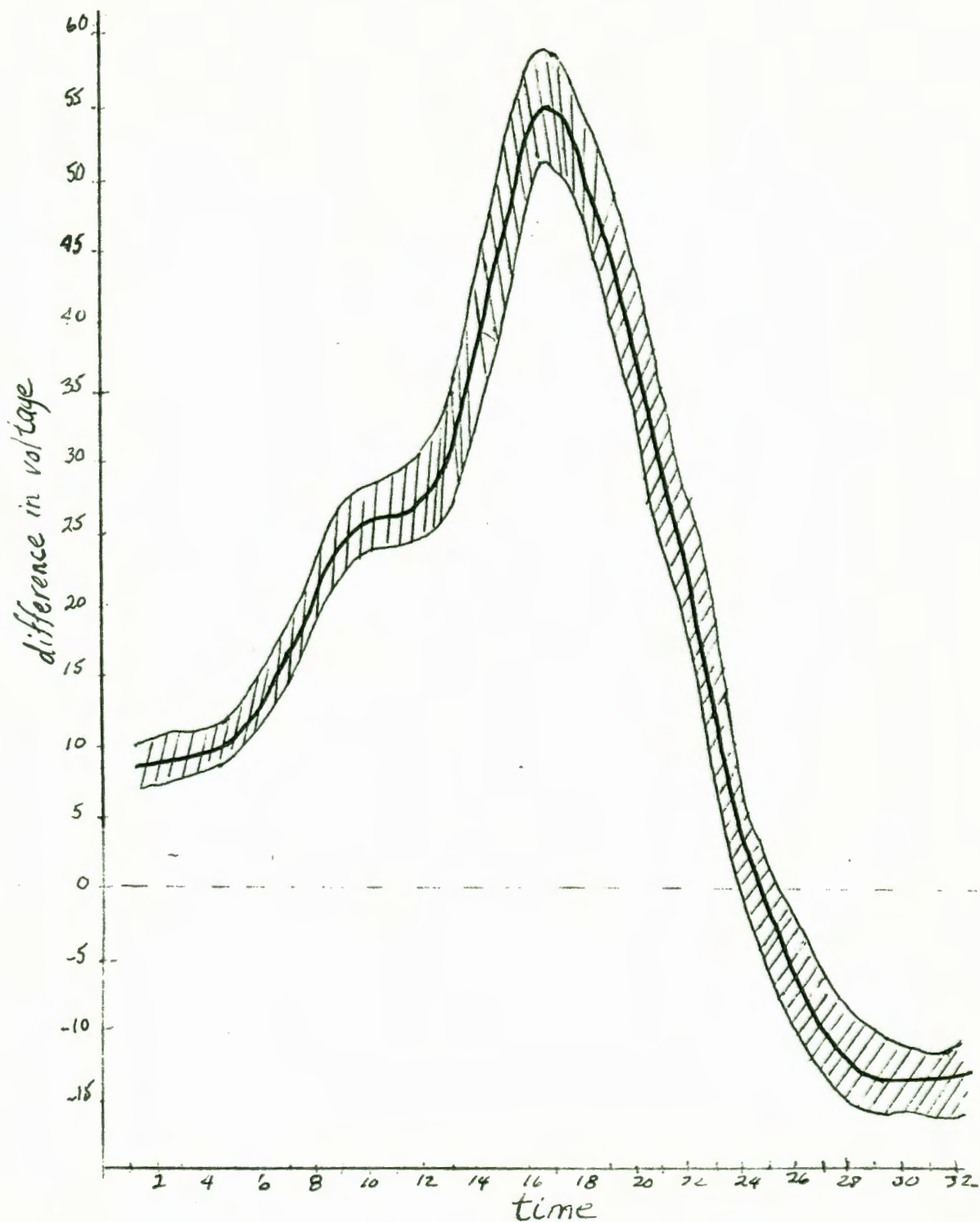
Figure 4. Confidence band about difference between average p-waves shown in figure 3.

For this case the confidence bands for the difference between two curves can be found by the use of the following:

$$(6) \qquad \mu_{1i} - \mu_{2i} \in \bar{y}_{1i} - \bar{y}_{2i} \pm t^{\alpha/2p}_{2p(n-1)} \left(\frac{2s^2 \sigma_{ii}}{n} \right)^{1/2}$$

where $s^2 = \dfrac{1}{2p(n-1)} \left(\sum\limits_{j=1}^{n} (\underline{y}_{1j} - \underline{\bar{y}}_1)' (\Sigma')^{-1} (\underline{y}_{1j} - \underline{\bar{y}}_1) \right.$

$$\left. + \sum\limits_{j=1}^{n} (\underline{y}_{2j} - \underline{\bar{y}}_2)' (\Sigma')^{-1} (\underline{y}_{2j} - \underline{\bar{y}}_2) \right)$$

and where σ_{ii} is the ith diagonal element of the known matrix Σ'.

An example of this case would be the comparison of the two dose-response curves shown in figure 4 where syn. Δ^9-THC was the drug used in obtaining curve 1 and Δ^9-THC was used in obtaining curve 2. Assuming that independent groups were used at each dosage level and that the variance was known to be twice as large at the two higher levels as at the two lower levels for both drugs, the covariance matrix would have the form:

$$\Sigma = \sigma^2 \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 2 & 0 \\ 0 & 0 & 0 & 2 \end{pmatrix}$$

The raw data for the scores obtained using the first drug is shown in table II, while table V gives the raw scores for drug 2 (data simulated in both cases). Table VI gives the mean difference between the curves at each dosage level and a confidence band about this difference. Figure 5a shows the two average curves and figure 5b the confidence band for the difference between the two curves. The two curves differ significantly at dosage levels 1.0 and 1.5 mg/kg. There was no significant difference at 0.0 and 0.5 mg/kg.

TABLE V

PERFORMANCE OF INDEPENDENT GROUPS ON MEMORY TASK AT FOUR LEVELS OF Δ-THC

	0.0 mg/kg	0.5 mg/kg	1.0 mg/kg	1.5 mg/kg
	61	60	32	19
	64	62	30	28
	65	63	38	17
	68	61	42	15
	63	58	27	31
	69	57	26	31
	60	59	39	15
	58	61	31	27
	66	56	43	16
	59	59	24	19
Total	633	596	332	218

TABLE VI

CONFIDENCE BAND AROUND DIFFERENCE BETWEEN AVERAGE SYN. -THC SCORES AND AVERAGE -THC SCORES

point	y_{1i}	y_{2i}	$y_{1i} - y_{2i}$	Bonferroni confidence interval
1	58.0	63.3	-5.3	-5.3 + 7.29
2	61.3	59.6	1.7	1.7 + 7.29
3	44.4	33.2	11.2	11.2 + 10.27
4	36.9	21.8	15.1	15.1 + 10.27
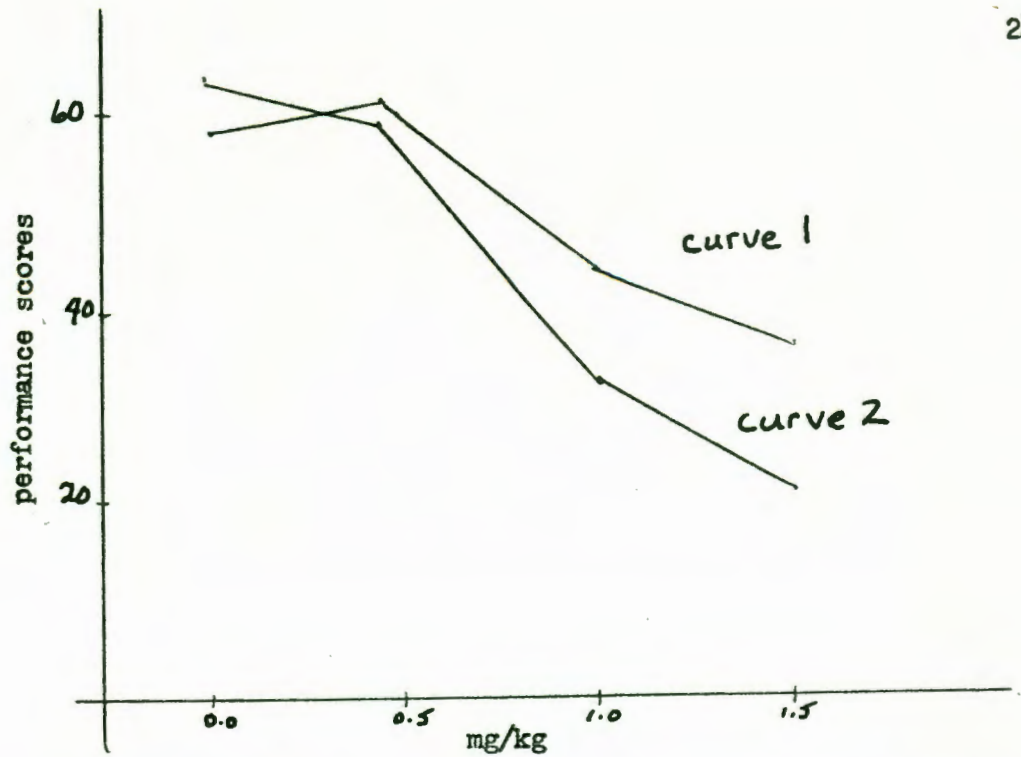
Figure 5a. Curve 1 shows average performance score at four dosage levels of syn. Δ^9-THC. Curve 2 shows the same for Δ-THC.
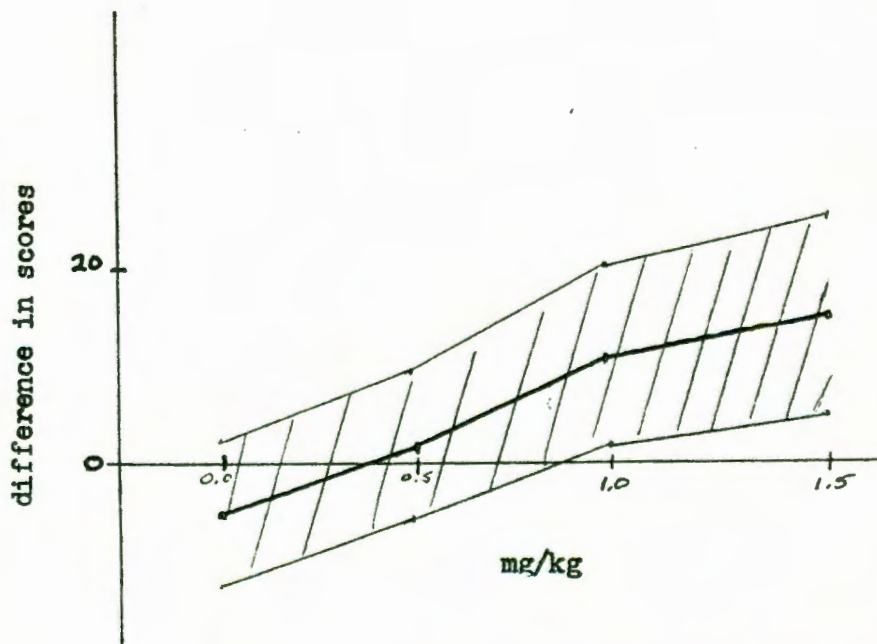
Figure 5b. Confidence band about the difference between the two curves shown in figure 5a.

IV.   DECOMPOSITION OF A CURVE INTO ITS COMPONENTS

This section deals with a type of analysis which attempts to
define the shape of a curve or the form of relationship between
the independent variable and the dependent variable. First to be
discussed will be the general idea of a component of a curve as being
a linear combination of co-ordinate means. Next to be presented
will be the determination of orthogonal components and two special
sets of orthogonal components known as orthogonal polynomials and
Haar functions. Finally, techniques for making simultaneous tests
on the significance of a set of components will be outlined.

Besides finding a confidence band about a mean curve or mean
difference between curves, one may wish to analyze the shape of a
curve. This type of analysis has no univariate analogue since one
cannot talk about shape or form in one dimension. In the case of
curves, one would like to be able to say, for instance, that curve
6a (figure 6) shows a difference in level between first half and
second half, while curve 6b shows a linear trend. The means of making
this type of analysis depend on making comparisons between weighted
combinations of the co-ordinate means. For example, to compare the
first half of curve 6a to the second half one could look at the
difference between the sum of means of the first half and the sum of
means of the second half:

$$(7) \qquad \left(1 \cdot y_1 + 1 \cdot y_2\right) - \left(1 \cdot y_3 + 1 \cdot y_4\right)$$
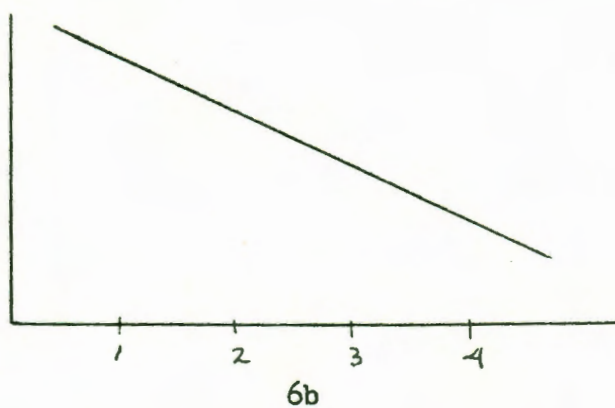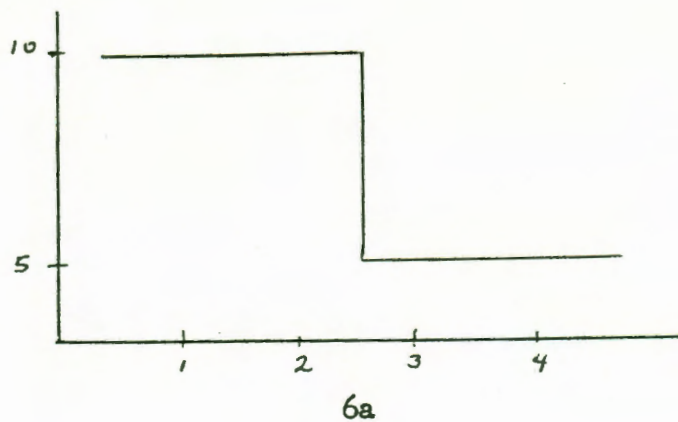
Figure 6. Curve 6a shows a difference in level between the first half and the second half. Curve 6b shows a linear trend.

Then this difference can be tested for statistical significance. Thinking of the co-ordinate means as being column means in an analysis of variance, the statistical test for the significance of combinations of co-ordinate means has close relationship to the test of column effects in one-way ANOVA.

A component of the sum of squares for co-ordinate means is defined by

$$SS_c = \frac{(c_1 y_1 + c_2 y_2 + c_3 y_3 + \ldots\ldots\ldots + c_p y_p)^2}{n(c_1^2 + c_2^2 + c_3^2 + \ldots\ldots + c_p^2)} \quad \text{where} \quad \sum_{i=1}^{p} c_i = 0.$$

or, in vector notation where $\underline{c}' = (c_1, c_2, c_3, \ldots\ldots, c_p)$

$$SS_c = (\underline{c}'\underline{y})^2 / n\underline{c}'\underline{c} \ .$$

In the foregoing example, $c_1 = c_2 = 1$, $c_3 = c_4 = -1$, and $SS_c = 25/n$. A component of the sum of squares has one degree of freedom because it represents the squared difference between two basic observations each of which is a weighted sum of treatment means (Winer, 1971).

The expected value of a component of sum of squares is:

$$E(SS_c) = \sigma_\epsilon^2 + n/2 \ (\underline{c}'\underline{y})^2 \ .$$

Under the hypothesis that $\underline{c}'\underline{y} = 0$, that is, that the weighted difference equals zero,

$$E(SS_c) = \sigma_\epsilon^2$$

Therefore, the ratio $\dfrac{SS_c}{MS \text{ error}}$ has an F distribution with 1 and $p(n-1)$ degrees of freedom.

Many different combinations of the co-ordinate means could be

derived meeting the requirement that $\sum c_i = 0$. However, only certain sets of these combinations will provide non-overlapping information about the shape of the curve. Returning to curve 6a, one could compare the first co-ordinate mean to the fourth by taking the combination:

$$(8) \qquad 1\,\bar{y}_1 + 0\,\bar{y}_2 + 0\,\bar{y}_3 - 1\,\bar{y}_4$$

However, the information obtained from doing so would be redundant with that obtained in taking the combination in (7). On the other hand, one could look at the difference between the first co-ordinate and the second by taking the combination:

$$(9) \qquad 1\,\bar{y}_1 - 1\,\bar{y}_2 + 0\,\bar{y}_3 + 0\,\bar{y}_4$$

This would not overlap with (7) since (7) tells us nothing about relationships in the first half, only about relationships between the first half and the second.

An easy means of determining whether two linear combinations are uncorrelated and therefore provide non-overlapping information is to take the vector product $\underline{c}'_1\underline{c}_2$, where $\underline{c}'_1$ is the set of weights for one combination and $\underline{c}_2$ is the set of weights for the other. If $\underline{c}'_1\underline{c}_2 = 0$, the combinations are orthogonal (i.e., uncorrelated). For example,

$$(1,\ 1,\ -1,\ -1) \quad (1,\ 0,\ 0,\ -1) = 2,$$

and these two combinations would not be orthogonal, while

$$(1,\ 1,\ -1,\ -1) \quad (1,\ -1,\ 0,\ 0) = 0,$$

and these combinations would be orthogonal.

If the combinations are orthogonal, then the sums of squares for

the components will also be orthogonal. For a curve comprised of p points (or a design with p columns), one can determine p-1 components all of which are orthogonal to each other. Each component of the sums of squares for these p-1 components covers a different portion of the total variation of the co-ordinate means. The sum of these p-1 components of the sum of squares will account for all the variation between groups. Non-orthogonal components do not have this additive property. (Winer, 1971)

Therefore, by taking orthogonal linear combinations of the co-ordinate means one can arrive at k-1 non-overlapping bits of information about the shape of the curve. Two sets of components which provide useful information about the shape of a curve are orthogonal polynomials and Haar functions. The first to be discussed will be orthogonal polynomials.

ORTHOGONAL POLYNOMIALS

The only method discussed in most texts on psychological statistics for the charaterization of the components of a curve is the method of orthogonal polynomials (eg. Hays, 1973). The weights to be used in taking linear combinations of means for finding orthogonal components have been tabled. The weights have been chosen so that each of the p-1 sets of coefficients in combination with the co-ordinate means represent exactly one polynomial trend in the data. The first set of coefficients allows one to test for a linear trend, the second for a quadratic trend, the third for a cubic trend, the fourth for a quartic trend and so on. (See figure 7.)
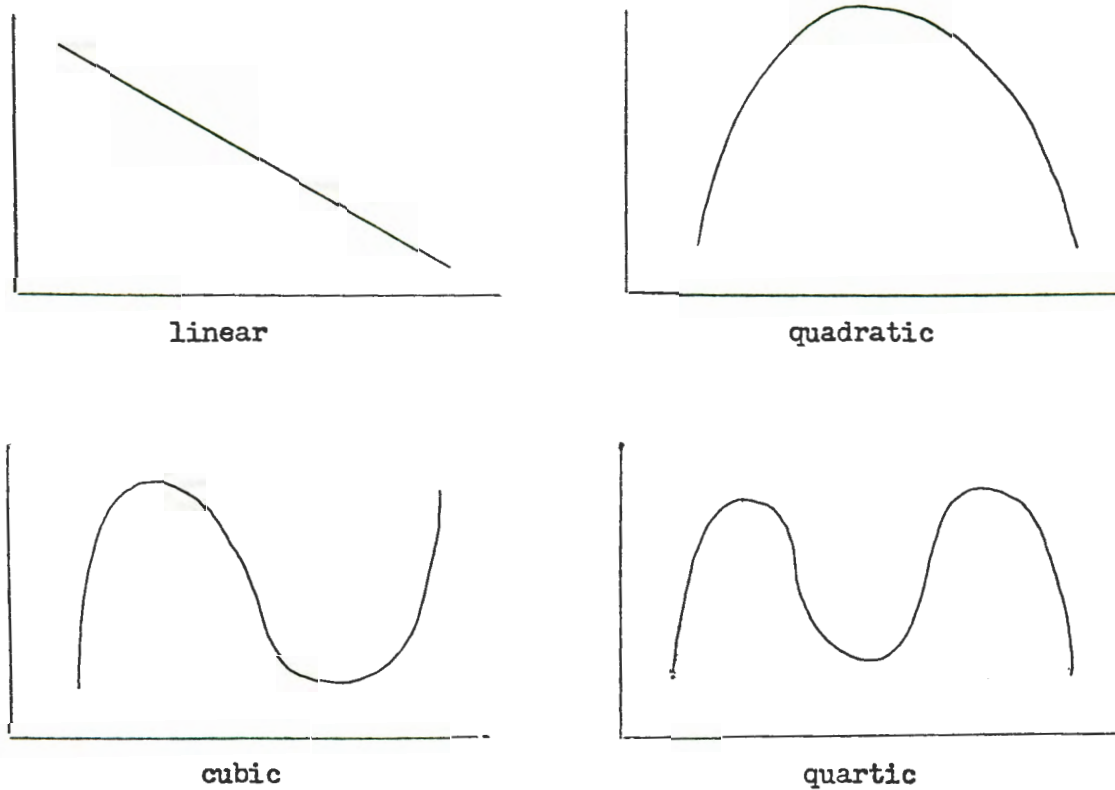
Figure 7. Curves show the first four polynomial trends.

A curve may have one or more significant trends, up to p-1, or it may have no significant trends. For example, a curve of shape 8a (see figure 8) would show a small linear trend; a curve of shape 8b would show a larger linear trend; curve 8c would show a quadratic trend and no other; and curve 8d would show a significant linear trend as well as a significant quadratic trend.

Difficulty with this method arises if the curve involves a number of significant trends. This can occur even if the curve is relatively simple but non-symmetric as in the case of figure 9. Suppose that for this curve there were 10 observations at each point and that the mean squares for error (MS error) was equal to 40.0. The symbol ψ_{p_i} will be defined as follows:

(10)
$$\psi_{p_i} = \underline{p}_i' \, \overline{Y}$$

where $\underline{p}_i$ is the set of orthogonal polynomial coefficients for the ith component. The SS_c's for the ψ_{p_i} were found to be:

$$SS(\psi_{p_1}) = 2881 \qquad SS(\psi_{p_4}) = 0$$
$$SS(\psi_{p_2}) = 0 \qquad SS(\psi_{p_5}) = 468$$
$$SS(\psi_{p_3}) = 969$$

With an α of .05 and with 1 and 72 degrees of freedom, the ratio of $SS(\psi_{p_i})$/ Ms error would be compared to the F value of 3.98. Trends 1, 3, and 5 would be significant. It is also quite likely that the seventh trend would also be significant. However, coefficients for trends beyond the fifth are not usually tabled.

The problem with this method lies in the physical meaning to accord to the various significan trends beyond quadratic. This
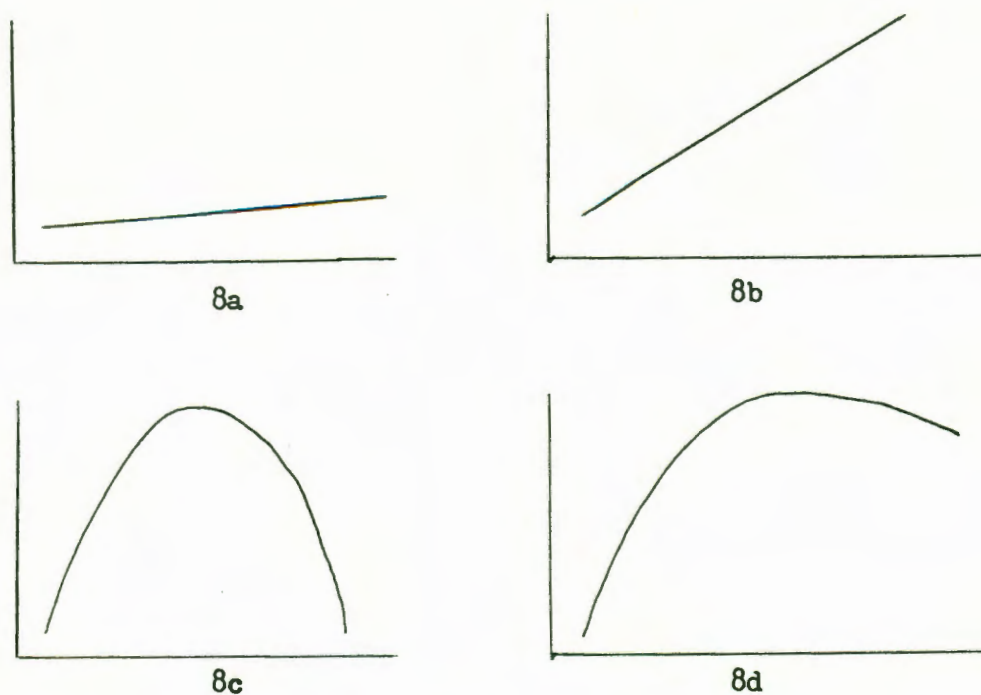
Figure 8. Curve 8a shows a small linear trend. Curve 8b shows a larger linear trend. 8c shows a quadratic trend. 8d shows a linear trend as well as a quadratic trend.
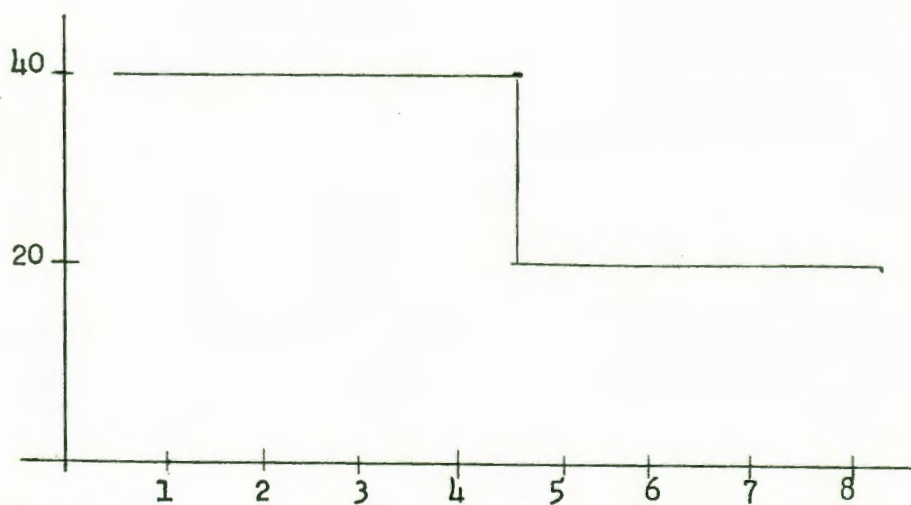
Figure 9. The first, third, and fifth polynomial trends would be significant in this simple, but non-symmetric curve.

difficulty in interpretation of the higher trends is evidenced by the fact that tables do not usually give coefficients beyond the fifth. Even stopping at the fifth it would be difficult to guess that the curve had the given shape from the knowledge that the first, third, and fifth components were significant. Using orthogonal polynomials for this example does not clarify the shape of the curve, but rather confuses the issue.

HAAR FUNCTIONS

Orthogonal polynomials are not, however, the only set of orthogonal functions which can be used to analyze a curve into its components. For example, Fourier analysis is a technique for analyzing periodic functions into simple trigonometric components. Another set of orthogonal functions which are easy to use computationally and also easy to interpret are Haar functions (Harmuth, 1972).

Haar functions are step functions which take on one of three values - 1, -1, and 0 - at any value of the independent variable. Figure 10a through 10d show the first four Haar functions. The first Haar function is a constant function having the value one at every point. The second Haar function equals one for the first half and minus one for the second half. A curve which had a positive component of the second Haar function would be higher in the first half than in the second. See figure 11a through 11c.

The third Haar function is one for the first quarter of the curve, minus one for the second quarter and zero for the second half. A curve having a positive component of the third Haar function would be higher in its first quarter than the second.
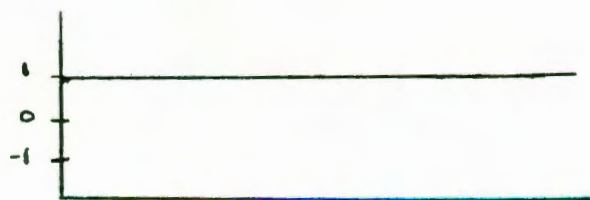
10a. The first Haar function
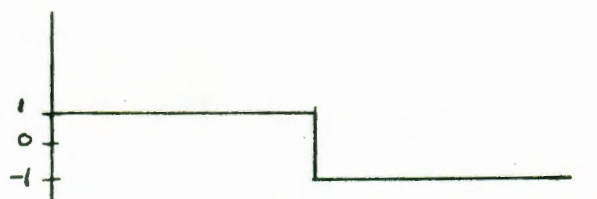
10b. The second Haar function
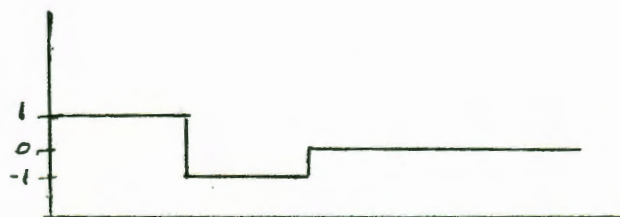
10c. The third Haar function
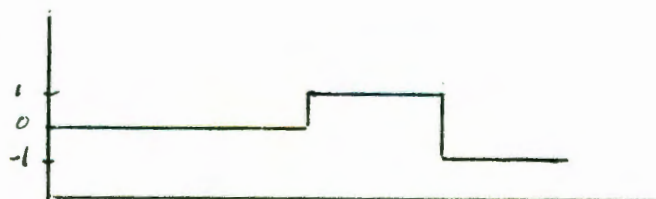
10d. The fourth Haar function

Figure 10.   The first four Haar functions.

11a

11b

11c
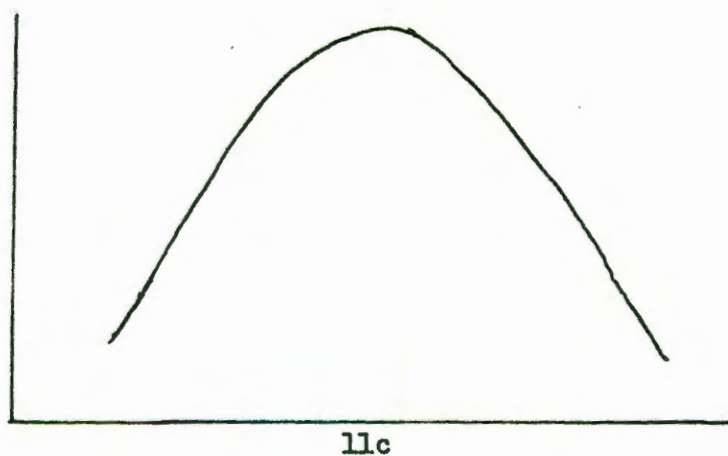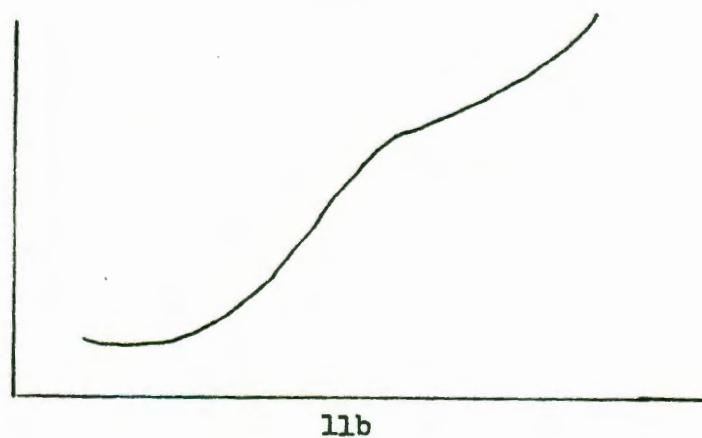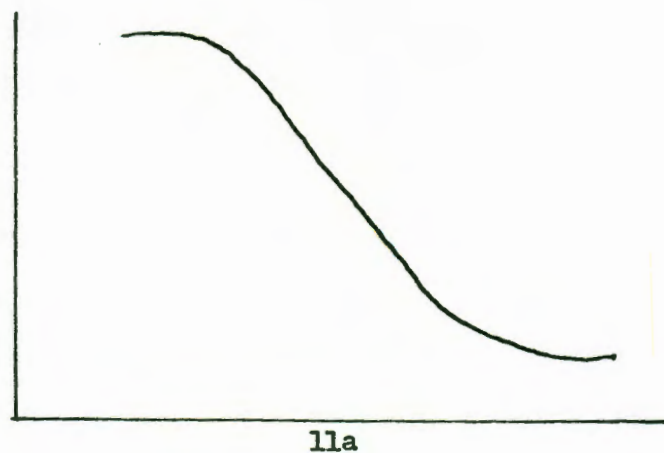
Figure 11. Figure 11a would show a significant positive component of the second Haar function, while 11b would show a significant negative component. Curve 11c would not show a significant component of the second Haar function since it is the same height in both halves on the average.

The shape of the second half of the curve does not influence whether or not this third component is significant. See figure 12a and 12b.

The fourth Haar function is equal to zero for the first half, one for the third quarter and minus one for the fourth quarter. A curve having a significant positive component of the fourth Haar function would have a third quarter which is higher than the fourth.

The fifth through eighth Haar functions are equal to one for one-eighth, minus one for the adjacent one-eighth, and zero for the rest of the curve. Curves which have a significant component of the fifth Haar function have a first eighth higher than the second eighth; a significant sixth Haar function implies a third eighth which is higher than the fourth eighth; and so on. The ninth through sixteenth Haar functions are used to determine if adjacent sixteenths of the curve differ in height. Similarly, the 17th through 32nd Haar functions are used to determine if neighboring thirty-seconds are different in height.

To determine which Haar components are significant one must take the linear combination of the appropriate Haar coefficients with the co-ordinate means:

$$\psi_{h_i} = \underline{h}'_i \, \overline{Y}$$

where $\underline{h}_i$ is the set of Haar coefficients for the ith component. The Haar coefficients are not tabled in statistical text, but can easily be determined. Table VII. shows the Haar coefficients to use for a curve comprised of four points, eight points, and sixteen points. For a number of points equal to 2^n the coefficients can be determined analogously.
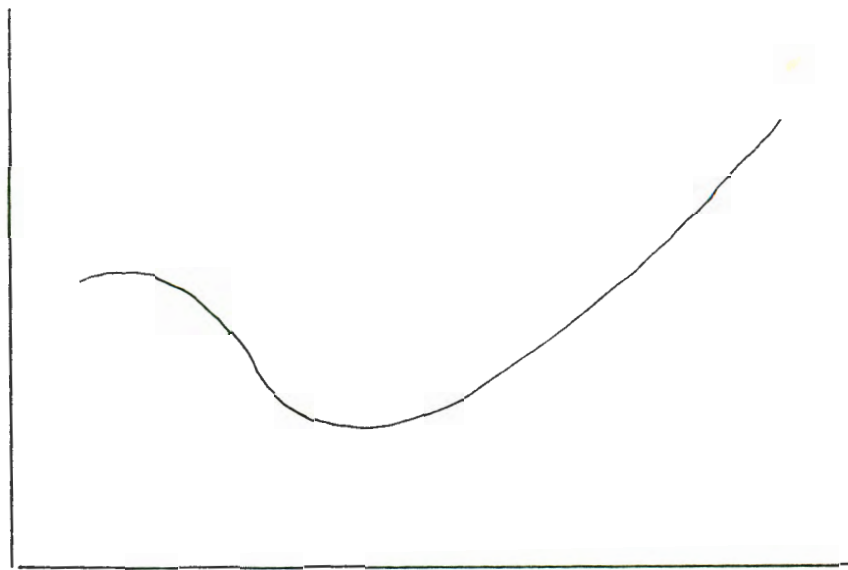
12a
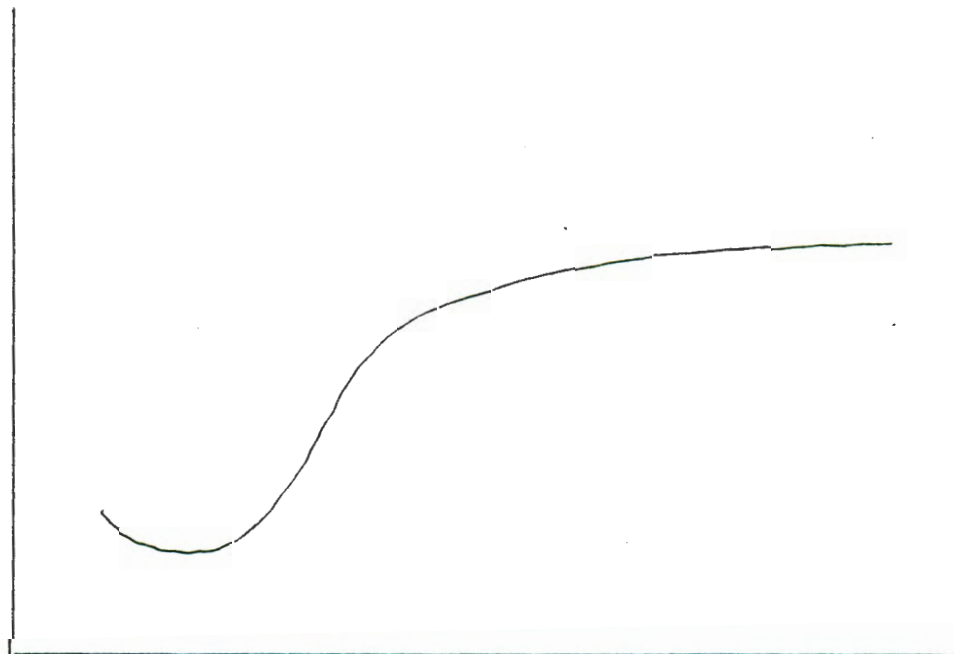
12b

Figure 12. Curve 12a would have a significant positive component of the third Haar function. Surve 12b would have a significant negative component of the third Haar function.

TABLE VII

HAAR COEFFICIENTS FOR CURVES COMPRISED
OF FOUR, EIGHT, OR SIXTEEN POINTS

function	point 1	2	3	4
h_1	1	1	1	1
h_2	1	1	-1	-1
h_3	1	-1	0	0
h_4	0	0	1	-1

function	point 1	2	3	4	5	6	7	8
h_1	1	1	1	1	1	1	1	1
h_2	1	1	1	1	-1	-1	-1	-1
h_3	1	1	-1	-1	0	0	0	0
h_4	0	0	0	0	1	1	-1	-1
h_5	1	-1	0	0	0	0	0	0
h_6	0	0	1	-1	0	0	0	0
h_7	0	0	0	0	1	-1	0	0
h_8	0	0	0	0	0	0	1	-1

| function	point 1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
h_1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
h_2	1	1	1	1	1	1	1	1	-1	-1	-1	-1	-1	-1	-1	-1
h_3	1	1	1	1	-1	-1	-1	-1	0	0	0	0	0	0	0	0
h_4	0	0	0	0	0	0	0	0	1	1	1	1	-1	-1	-1	-1
h_5	1	1	-1	-1	0	0	0	0	0	0	0	0	0	0	0	0
h_6	0	0	0	0	1	1	-1	-1	0	0	0	0	0	0	0	0
h_7	0	0	0	0	0	0	0	0	1	1	-1	-1	0	0	0	0
h_8	0	0	0	0	0	0	0	0	0	0	0	0	1	1	-1	-1
h_9	1	-1	0	0	0	0	0	0	0	0	0	0	0	0	0	0
h_{10}	0	0	1	-1	0	0	0	0	0	0	0	0	0	0	0	0
h_{11}	0	0	0	0	1	-1	0	0	0	0	0	0	0	0	0	0
h_{12}	0	0	0	0	0	0	1	-1	0	0	0	0	0	0	0	0
h_{13}	0	0	0	0	0	0	0	0	1	-1	0	0	0	0	0	0
h_{14}	0	0	0	0	0	0	0	0	0	0	1	-1	0	0	0	0
h_{15}	0	0	0	0	0	0	0	0	0	0	0	0	1	-1	0	0
h_{16}	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0

If the number of points does not equal 2^n, the Haar coefficients may be determined as follows. The first function is represented by all ones, as before. Suppose that 2^k is the largest power of 2 which is less than p, the number of points. The second Haar function is one for the first 2^k points and minus one for the remaining $p-2^k$ points. The coefficients for comparing the mean levels in the two segments are $p-2^k$ for the first 2^k points and $-(2^k)$ for the remainder, reflecting the unequal number of points in the two portions of the curve. Coefficients for subsequent functions within the segment with 2^k points (the first segment) can be determined in the usual way. If $p-2^k$ is an integral power of 2 (for example, if $p=12$, $p-2^k=4$), then coefficients for functions within this second segment can also be determined in the usual way. If not, repeat the process described above for the segment of $p-2^k$ points, finding the largest power of 2 which is less than $p-2^k$. For example, if $p = 7$, the coefficients would be as shown in table VIII.

After finding the linear combination of the Haar components with the co-ordinate means, the sum of squares for the linear combination will be found and this value divided by the mean squares for error.

As a numerical example, consider again the curve used in the discussion of orthogonal polynomials, where $\bar{y} = (40,40,40,40,20,20,20,20)$ and where MS error $= 40$. (See figure 9.) The sum of squares for the eight components and $SS(\psi_{h_i})/$ MS error is given in table IX. Comparing the value of the ratios SS/MS to the $F^{.05}$ value of 3.98, only components 1 and 2 would be significant.

Finding $\psi_{h_i}/(\underline{h}_i' \underline{h}_i)$ will give the magnitude and direction of

TABLE VIII

HAAR COEFFICIENTS FOR A CURVE
COMPRISED OF SEVEN POINTS

		point						
		1	2	3	4	5	6	7
	h	1	1	1	1	1	1	1
	h	3	3	3	3	-4	-4	-4
	h	1	1	-1	-1	0	0	0
function	h	0	0	0	0	1	1	-2
	h	1	-1	0	0	0	0	0
	h	0	0	1	-1	0	0	0
	h	0	0	0	0	1	-1	0

TABLE IX

SUM OF SQUARES FOR HAAR COMPONENTS
OF CURVE IN FIGURE 9

	$SS(\psi_{hi})$		$\dfrac{SS(\psi_{hi})}{MS\ error}$*
h_1	$(1\cdot40+1\cdot40+1\cdot40+1\cdot40+1\cdot20+1\cdot20+1\cdot20+1\cdot20)^2 =$	7200	180
h_2	$(1\cdot40+1\cdot40+1\cdot40+1\cdot40-1\cdot20-1\cdot20-1\cdot20-1\cdot20)^2 =$	800	20
h_3	$(1\cdot40+1\cdot40-1\cdot40-1\cdot40+0\cdot20+0\cdot20+0\cdot20+0\cdot20)^2 =$	0	0
h_4	$(0\cdot40+0\cdot40+0\cdot40+0\cdot40+1\cdot20+1\cdot20-1\cdot20-1\cdot20)^2 =$	0	0
h_5	$(1\cdot40-1\cdot40+0\cdot40+0\cdot40+0\cdot20+0\cdot20+0\cdot20+0\cdot20)^2 =$	0	0
h_6	$(0\cdot40+0\cdot40+1\cdot40-1\cdot40+0\cdot20+0\cdot20+0\cdot20+0\cdot20)^2 =$	0	0
h_7	$(0\cdot40+0\cdot40+0\cdot40+0\cdot40+1\cdot20-1\cdot20+0\cdot20+0\cdot20)^2 =$	0	0
h_8	$(0\cdot40+0\cdot40+0\cdot40+0\cdot40+0\cdot20+0\cdot20+1\cdot20-1\cdot20)^2 =$	0	0

*MS error = 40

the significant components. $\psi_{h_1}/8$ gives the overall level of the curve which is 30. $\psi_{h_2}/8 = 10$, being positive indicates that the first half is higher than the second, and that on the average, the first half is ten units above the mean while the second half is ten units below. From the knowledge of the significant Haar components, one could easily determine the shape of the curve. For this example, the Haar components provide a simple and more easily interpretable decomposition of the curve.

The procedure used above for determining which orthogonal components (Haar or polynomial) are significant is the procedure outlined in most texts (Winer, 1971; Hays, 1973). However, as Hays points out, unless one is solely interested in determining whether one particular component is significant, one should perform a simultaneous test on the components of interest. As noted in the discussion on finding a confidence band around one curve (p. 5), the probability that a set of statements will hold will be less than that of each individual statement unless the random variables (in this case linear combinations) are perfectly correlated.

In order to perform a simultaneous test with an overall level of significance, one again turns to Miller for the formula. Two cases will again be considered.

CASE I: COVARIANCE MATRIX UNKNOWN

For the general case with Σ unknown, expression (11) gives Bonferroni intervals for all possible linear combination $\underline{l}'\underline{\bar{y}}$, k of which hold simultaneously with probability $> 1-\alpha$:

(11) $\quad \underline{1}'\underline{\mu} \in \underline{1}'\underline{\bar{y}} \pm t_{n-1}^{\alpha/2k} \left(\dfrac{\underline{1}'S\,\underline{1}}{n} \right)^{1/2}$

where k is the number of combinations under consideration.

Statement (11) is written in terms of a confidence band. It could be rewritten as a significance test:

(12) $\quad p\left(\left| \dfrac{\underline{1}'\underline{\mu} - \underline{1}'\underline{\bar{y}}}{\dfrac{\underline{1}'S\,\underline{1}}{n}} \right| \le t_{n-1}^{\alpha/2k} \right) = 1-\alpha.$

One would reject the null hypothesis that $\underline{1}'\underline{\mu} = 0$ if

(13) $\quad \dfrac{\underline{1}'\underline{\bar{y}}}{\left(\dfrac{\underline{1}'S\,\underline{1}}{n} \right)^{1/2}} > t_{n-1}^{\alpha/2k} \quad .$

By substituting the $\underline{h}_i$ for the $\underline{1}$, and squaring both sides (13) becomes:

(14) $\quad \dfrac{(\psi_{h_i})^2}{\dfrac{\underline{h}_i'S\,\underline{h}_i}{n}} > (t_{n-1}^{\alpha/2k})^2$

Alternatively one could use the expression

(15) $\quad p\left(\left| \dfrac{\underline{1}'\underline{\mu} - \underline{1}'\underline{y}}{\left(\dfrac{n-1}{n-p} \right)^{1/2} \left(\dfrac{\underline{1}'S\,\underline{1}}{n} \right)^{1/2}} \right| \le (pF_{p,n-p}^{\alpha})^{1/2} \right) = 1-\alpha \quad \forall \underline{1} \in \mathcal{L}$ where $\mathcal{L}$ is the set of all possible linear combinations,

which translates into the significance test:

(16) $\quad \dfrac{(1'\underline{\bar{y}})^2}{\left(\dfrac{n-1}{n-p} \right)\left(\dfrac{\underline{1}'S\,\underline{1}}{n} \right)} = \dfrac{(\psi_{h_i})^2}{\left(\dfrac{n-1}{n-p} \right)\left(\dfrac{\underline{h}_i'S\,\underline{h}_i}{n} \right)} > pF_{p,n-p}^{\alpha} \quad .$

One would reject the null hypothesis if the left side is greater than the right. Once again the value $(t_{n-1}^{\alpha/2k})^2$ will usually be less than

than $\left(\dfrac{n-1}{n-p}\right)\left(pF_{r,n-p}^{\alpha}\right)$ unless k is very large.

As a numerical example from actual data consider again the p-wave shown in figure 1 with mean and variance shown in table I. Table X shows ψ_{h_i}', $\dfrac{(\psi_{h_i})^2}{\dfrac{h'Sh}{n}}$, Σc_i^2, and $\psi_{h_i}/\Sigma c_i^2$ values for this curve. With $\alpha = .01$,

$$(t_{49}^{.464})^2 = 15.13.$$

Using expression (14), significant components were determined, and are shown with an asterisk (*) in table X.

In order to construct the shape of the curve from knowledge of the significant Haar components, we will look at the ψ_{h_i}'/σ^2 values for these components. First for ψ_{h_1}, the overall level is found to be 172.85. $\psi_{h_2}/\Sigma c_i^2 = 5.42$ indicating that the first half is on the average 5.42 units above the average while the second half is 5.42 units below the average.

Since there are so many components we will look at quarters of the curve separately. $\psi_{h_3}/\Sigma c_i^2 = -5.09$, indicating that the first quarter is lower than the second quarter. Within the first quarter, the first eighth is 2.85 ($\dfrac{\psi_{h_5}}{\Sigma c_i^2} = -2.85$) units lower than the average for that quarter, while the second eighth is 2.85 units above indicating an increase to the right in the first quarter. Breaking this into finer details, we see that the first sixteenth and the second sixteenth do not differ significantly, while the third sixteenth is 2.70 units below the average for that eighth and the fourth sixteenth is 2.70 units above ($\psi_{h_{10}}/\Sigma c^2 = -2.70$). The pattern of the first eight thiry-seconds confirms the pattern of the sixteenths, with $\psi_{h_{17}}$ and $\psi_{h_{18}}$ not being significant while the negative values of

$\psi_{h_{15}}$ and $\psi_{h_{20}}$ show an increasing trend to the right in the second eighth of the curve.

The second quarter was found to be 5.42 units above the average for the first half. Within this quarter, the third eighth is lower than the fourth since $\psi_{h_6}/\Sigma c_i^2 = -3.68$. The comparison of the fifth sixteenth to the sixth ($\psi_{h_{11}}/\Sigma c_i^2 = .55$) indicates a small decrease in level from the fifth sixteenth to the sixth, while the comparison of the seventh sixteenth to the eighth ($\psi_{h_{12}}/\Sigma c_i^2 = -5.23$) indicates that the curve again increases to the right in the next eighth. Comparisons between neighboring thirty-seconds show no significant difference between the ninth and tenth and between the eleventh and twelfth thirty-seconds, while the values $\psi_{h_{23}}/\Sigma c_i^2 = -2.10$ and $\psi_{h_{24}}/\Sigma c_i^2 = -2.15$ reflect a generally even increase to the right in the third eighth of the curve.

Therefore in the second quarter, while the curve generally increases to the right, there is a slight downturn in the third eighth with the fourth eighth again increasing to the right.

In the third quarter, all components (ψ_{h_7}, $\psi_{h_{13}}$ $\psi_{h_{14}}$, $\psi_{h_{25}}$, $\psi_{h_{26}}$, $\psi_{h_{27}}$, $\psi_{h_{28}}$) are significant and positive indicating a relatively rapid decrease to the right in this quarter.

In the fourth quarter, the fact that ψ_{h_8} is not significant indicates that the curve is rather flat in this quarter. Within eighths, however, there is a decrease between the thirteenth and fourteenth sixteenths ($\psi_{h_{15}}/\Sigma c_i^2 = .93$), while the fifteenth sixteenth is lower than than the sixteenth ($\psi_{h_{16}}/\Sigma c_i^2 = -.45$). Of the components relating to the difference between thirty-seconds, only the component

TABLE X

HAAR COMPONENTS OF P-WAVE BEFORE ADMINISTRATION OF Δ^9-THC

i	ψ_{hi}	$\dfrac{(\psi_{hi})^2}{b\sum s_{hi}}$	$\sum c_i^2$	$\dfrac{\psi_{hi}}{\sum c_i^2}$	comparison
1	5531.44	483,700.86*	32	172.85	level
2	173.28	439.63*	32	5.42	1st ½ – 2nd ½
3	-94.24	2555.42*	16	-5.09	1st ¼ – 2nd ¼
4	191.80	1706.82*	16	11.99	3rd ¼ – 4th ¼
5	-22.94	366.14*	8	-2.85	1st ⅛ – 2nd ⅛
6	-29.46	161.97*	8	-3.68	3rd ⅛ – 4th ⅛
7	91.24	2638.98*	8	11.40	5th ⅛ – 6th ⅛
8	-2.04	2.72	8		7th ⅛ – 8th ⅛
9	-1.14	8.17	4		1st 1/16 – 2nd 1/16
10	-10.80	573.02*	4	-2.70	3rd 1/16 – 4th 1/16
11	2.20	16.33*	4	.55	5th 1/16 – 6th 1/16
12	-20.90	643.80*	4	-5.23	7th 1/16 – 8th 1/16
13	18.38	130.66*	4	4.89	9th 1/16 – 10th 1/16
14	21.18	975.91*	4	5.29	11th 1/16 – 12th 1/16
15	3.72	36.75*	4	.93	13th 1/16 – 14th 1/16
16	-1.80	27.22*	4	-.45	15th 1/16 – 16th 1/16
17	-0.62	8.17	2		1st 1/32 – 2nd 1/32
18	-0.44	2.72	2		3rd 1/32 – 4th 1/32
19	-2.14	93.92*	2	-1.07	5th 1/32 – 6th 1/32
20	-2.70	118.42*	2	-1.35	7th 1/32 – 8th 1/32
21	-0.62	8.17	2		9th 1/32 – 10th 1/32
22	0.94	10.89	2		11th 1/32 – 12th 1/32
23	-4.20	63.97*	2	-2.10	13th 1/32 – 14th 1/32
24	-4.30	102.08*	2	-2.15	15th 1/32 – 16th 1/32
25	2.06	21.78*	2	1.03	17th 1/32 – 18th 1/32
26	6.60	185.11*	2	3.30	19th 1/32 – 20th 1/32
27	6.00	230.03*	2	3.00	21st 1/32 – 22nd 1/32
28	4.82	317.14*	2	1.05	23rd 1/32 – 24th 1/32
29	2.10	50.36*	2		25th 1/32 – 26th 1/32
30	0.10	0.00	2		27th 1/32 – 28th 1/32
31	-0.60	9.52	2		29th 1/32 – 30th 1/32
32	-0.52	6.81	2		31st 1/32 – 32nd 1/32

*significant at the .01 level

$\psi_{h_{l4}}$ is significant, indicating a decrease from the twenty-fifth thirty-second to the twenty-sixth ($\psi_{h_{l4}}/ c^2 = 1.05$).

Therefore, while the last quarter is relatively flat in comparison to the other quarters, the curve continues to decrease in the seventh eighth, but shows a flattening out and slight upturn in the last eighth.

Figure 13 shows a reconstruction of the curve from using the values of the $\psi_{h_i}/\Sigma c_i^2$ for the significant components. Also shown is the mean curve as shown in figure 1. While the use of all thirty-two Haar components can provide a very accurate fit to the curve, as in this case, the use of only eight or sixteen may give a good representation of the curve without bringing in the complications of fine divisions. However, even the highest order Haar component does lend itself to interpretation.

Using orthogonal polynomials, the following values were found for the $\psi_{p_i}/\frac{p_i'S\,p_i}{n}$ for the first five components:

$$\frac{(\psi_{p_1})^2}{\frac{p_1'S\,p_1}{n}} = 1747.44 \qquad \frac{(\psi_{p_2})^2}{\frac{p_2'S\,p_2}{n}} = 5950.77 \qquad \frac{(\psi_{p_3})^2}{\frac{p_3'S\,p_3}{n}} = 75.47$$

$$\frac{(\psi_4)^2}{\frac{p_4'S\,p_4}{n}} = 6735.48 \qquad \frac{(\psi_{p_5})^2}{\frac{p_5'S\,p_5}{n}} = 1049.48$$

All five orthogonal components would be significant for this curve. It would be very difficult to assign physical meaning to this fact. For this example, the Haar components provide a more meaningful decomposition of this curve.
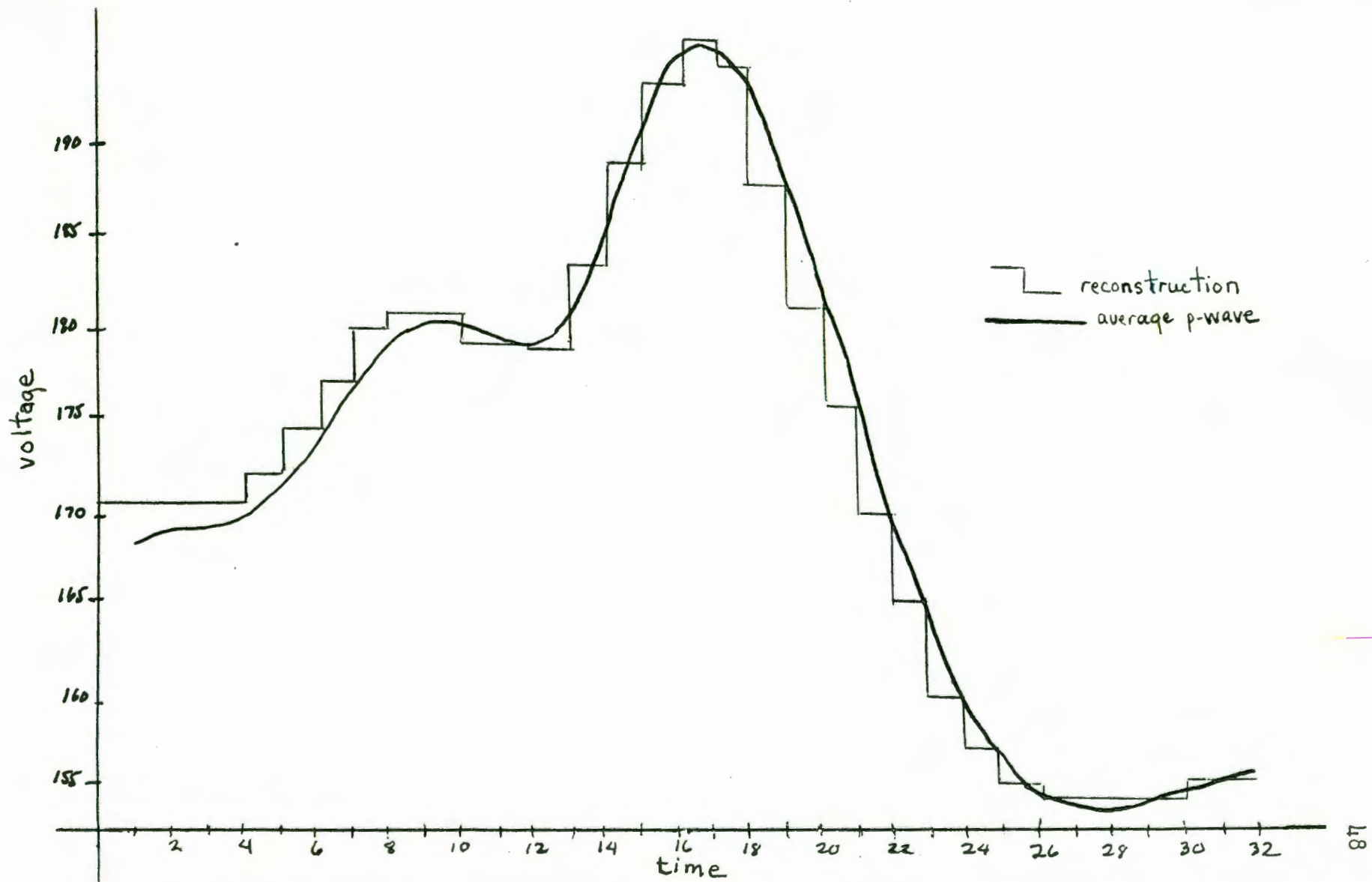
Figure 13. Average p-wave before administration of Δ^9-THC and reconstruction of the curve from the significant Haar components.

CASE II: COVARIANCE MATRIX KNOWN EXCEPT FOR A SCALAR

In the special case where the covariance matrix is known except for a scalar ($\Sigma = \sigma^2 \Sigma'$ where Σ' is known and σ^2 unknown), the formula for determining which components are significant is derived from:

$$
(17) \qquad \underline{l}'\mu \in \underline{l}'\underline{y} \pm t^{\alpha/2k}_{p(n-1)}\left(\frac{s^2 \underline{l}'(\Sigma')\underline{l}}{n}\right)^{\frac{1}{2}}
$$

where s^2 is the estimate of σ^2 and k is the number of linear combinations. Substituting the $\underline{h}_i$ for the $\underline{l}$ and re-writing in terms of a significance test this becomes —

$$
(18) \qquad P\left(\left|\frac{\underline{l}'\underline{y}}{\frac{s}{n}\underline{l}'(\Sigma')\underline{l}}\right| \leq t^{\alpha/2k}_{p(n-1)}\right) = 1-\alpha
$$

One would reject the hypothesis that $\underline{l}'\mu = 0$ if

$$
(19) \qquad \frac{(\underline{l}'\underline{y})^2}{\frac{s^2}{n}\underline{l}'\Sigma'\underline{l}} > (t^{\alpha/2k}_{p(n-1)})^2
$$

Alternatively, one could use

$$
(20) \qquad \underline{l}'\mu \in \underline{l}'\underline{y} - (pF^{\alpha}_{p,p(n-1)})^{\frac{1}{2}}\left(\frac{s^2}{n}\underline{l}'\Sigma'\underline{l}\right)^{\frac{1}{2}}
$$

rejecting the null hypothesis if

$$
(21) \qquad \frac{(\underline{l}'\underline{y})^2}{\frac{s^2}{n}\underline{l}'\Sigma'\underline{l}} > pF^{\alpha}_{p,p(n-1)}
$$

Once again the equation based on the Bonferroni inequality (17) will provide the better test, unless k is very large.

As an example, use again the data from the experiment on the effect of drug dosage on performance of a task where independent groups were used at each level of the independent variable (p.10). Here it was assumed that the covariance matrix had the form:

$$\Sigma = \sigma^2 \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 2 & 0 \\ 0 & 0 & 0 & 2 \end{pmatrix}$$

The mean vector of the curve representing the effects of the syn. Δ^9-THC was found to be

$$\bar{y} = (\ 58.0,\ 61.3,\ 44.4,\ 36.9)$$

with $s^2 = 56.43$. Table XI gives the values of ψ_{h_i}, $\underline{h}'\Sigma'\underline{h}$, and $(\psi_{h_i})^2 / \left(\frac{s^2}{n} \underline{h}'\Sigma'\underline{h} \right)$ for this example. With $p = 4$, $\alpha = .05$, $n = 10$, and $k = 4$,

$$(t^{\alpha/2k}_{p(n-1)})^2 = 6.92$$

Using expression 19, the first and second Haar components are significant, while the third and fourth would not be significant. Calculating $\psi_{h_1}/\Sigma c_{1j}^2$ and $\psi_{h_2}/\Sigma c_{2j}^2$ gives:

$$\frac{\psi_{h_1}}{\sum_{j=1} c_{1j}} = \frac{200.6}{4} = 50.12 \qquad\qquad \frac{\psi_{h_2}}{\sum_{j=1} c_{2j}} = \frac{38.0}{4} = 9.50$$

This shows that the overall level is 50.12 and that the first half is 9.5 units above this average while the second half is 9.5 units below

50.12. The fact that the third and fourth components were not sig-
nificant would indicate that the difference between the first quarter
and the second and between the third quarter and fourth quarter were
not significant. A reconstruction of the curve from this information
would appear as in figure 14. Also shown is the mean curve.

Using orthogonal polynomials for this example would give the
values shown in table XII for ψ_{p_i}, $p_i' \Sigma' p_i$ and $(\psi_{p_i})^2 / \frac{s^2}{n} p_i' \Sigma' p_i$.
Using expression 19 to determine which components are significant,
only the first or linear component exceeds the t value. Since

$$\frac{\psi_{p_i}}{\Sigma c_i^2} = -20.2$$

is negative, this indicates that the curve has a downward linear
trend from left to right.

For this example, the orthogonal polynomials provide at least
as good if not simpler characterization of the curve as that pro-
vided by the Haar functions. This will be the case when the curve
can be closely fitted by a simple combination of polynomials, such
as by a linear function, a quadratic function or a combination of
the two.

TABLE XI

HAAR COMPONENTS OF SYN. Δ^9-THC
PERFORMANCE CURVE

i	ψ_{h_i}	$\underline{h}'_i \Sigma' \underline{h}_i$	$\dfrac{(\psi_{h_i})^2}{s^2/n\ \underline{h}'_i \Sigma' \underline{h}_i}$	comparison
1	200.6	6	1188.5 *	level
2	38.0	6	42.68*	1st ½ - 2nd ½
3	-3.3	2	.96	1st ¼ - 2nd ¼
4	7.5	4	2.49	3rd ¼ - 4th ¼

TABLE XII

POLYNOMIAL COMPONENTS OF SYN. Δ^9-THC
PERFORMANCE CURVE

i	ψ_{p_i}	$\underline{p}'_i \Sigma' \underline{p}_i$	$\dfrac{(\psi_{p_i})^2}{s^2/n\ \underline{p}'_i \Sigma' \underline{p}_i}$	trend
1	-80.2	30	37.82*	linear
2	-10.8	6	-1.89	quadratic
3	29.6	30	2.27	cubic

*significant at .05 level
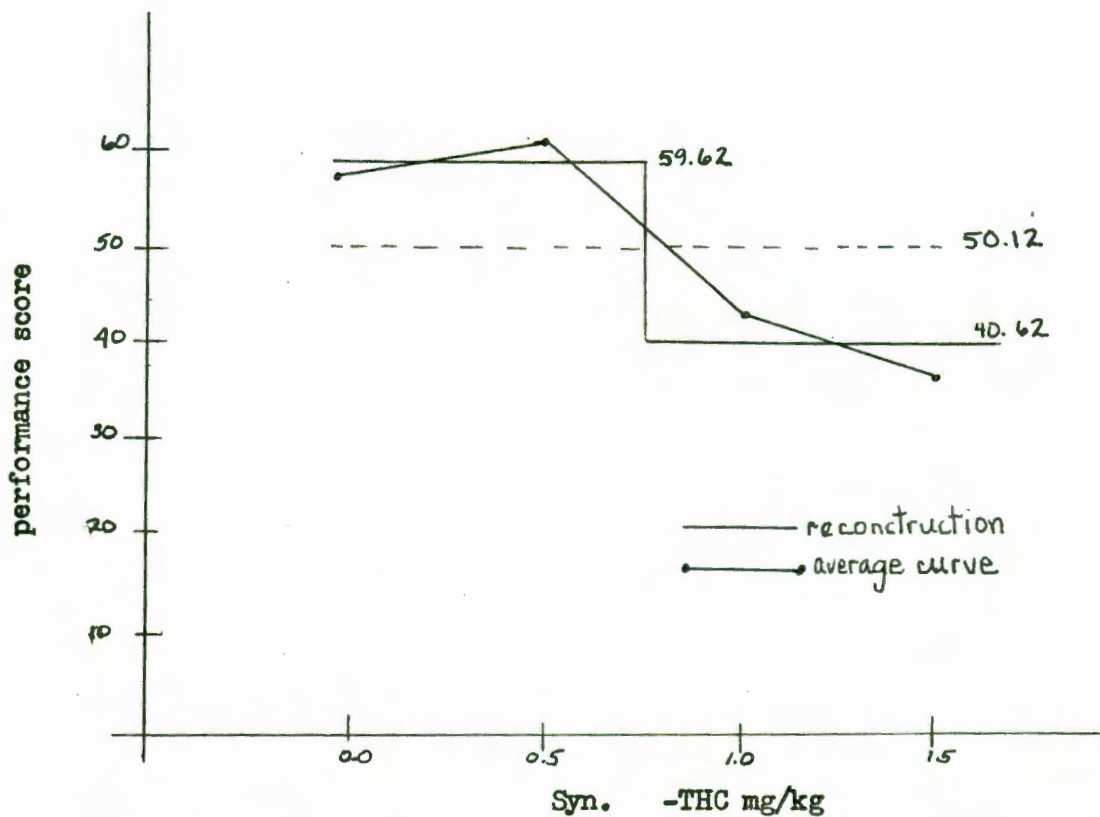
Figure 14. Curve of average performance scores under syn. Δ^9-THC and reconstruction of the curve from the significant Haar components.

IV. COMPONENT BY COMPONENT COMPARISON OF TWO CURVES

A natural extension of the determination of the components of
one curve is the comparison of the components of two or more curves.
This is a useful procedure when one wishes to compare the shape of
two curves. It may be, for example, that two curves differ greatly
in their overall level but have similar shapes as in figure 15. A
confidence band for the difference between the two curves would indi-
cate that there is a significant difference between the two curves at
each point, but would not indicate that the two curves were of
similar shape. Or, perhaps one would like to say that the first
half of two curves are the same shape, but after that diverge as
in figure 16.

The method presently used for making this kind of comparison
would be comparing the polynomial components of the curves. How-
ever, the difficulties presented in interpreting the polynomial
components of one curve are only compounded when comparing the com-
ponents from two or more curves. If one or more of the curves had
components beyond the first and second which are significant, compar-
ing the Haar components rather than polynomial components may pro-
vide a more meaningful analysis.

Whether using Haar components or orthogonal polynomial com-
ponents, making a component by component comparison involves taking
the linear combination of co-ordinate means of one curve and sub-
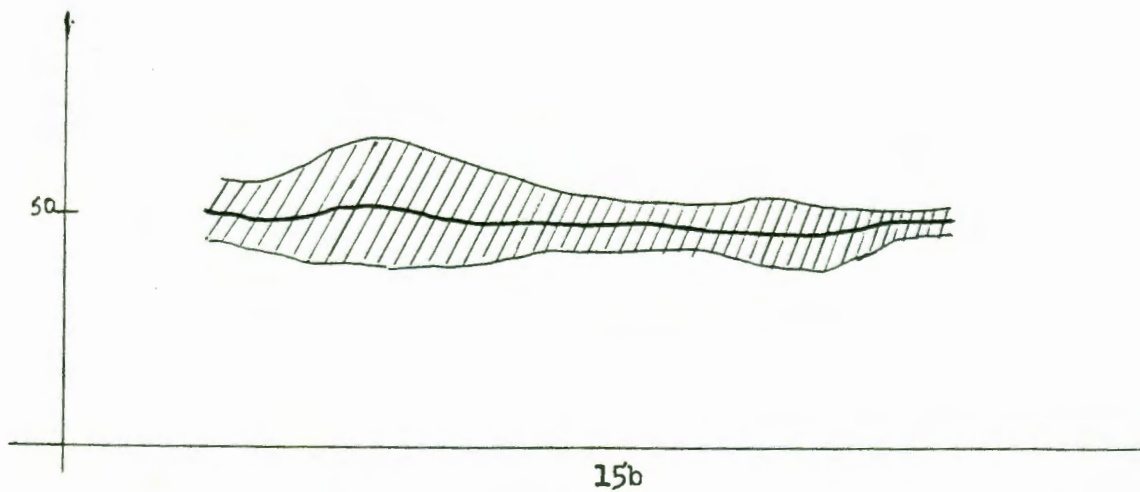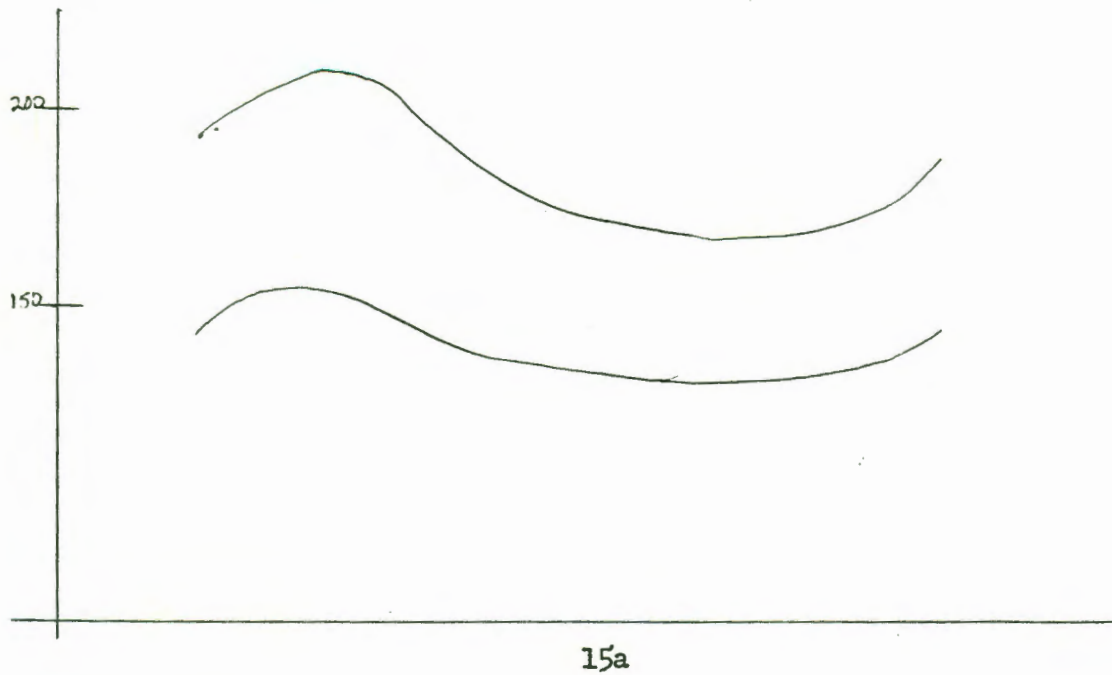tracting it from the same linear combination of the co-ordinate

Figure 15. The curves in 15a are of similar shape but different level. 15b shows the confidence band for the difference between the curves in 15a with a significant difference at each point.
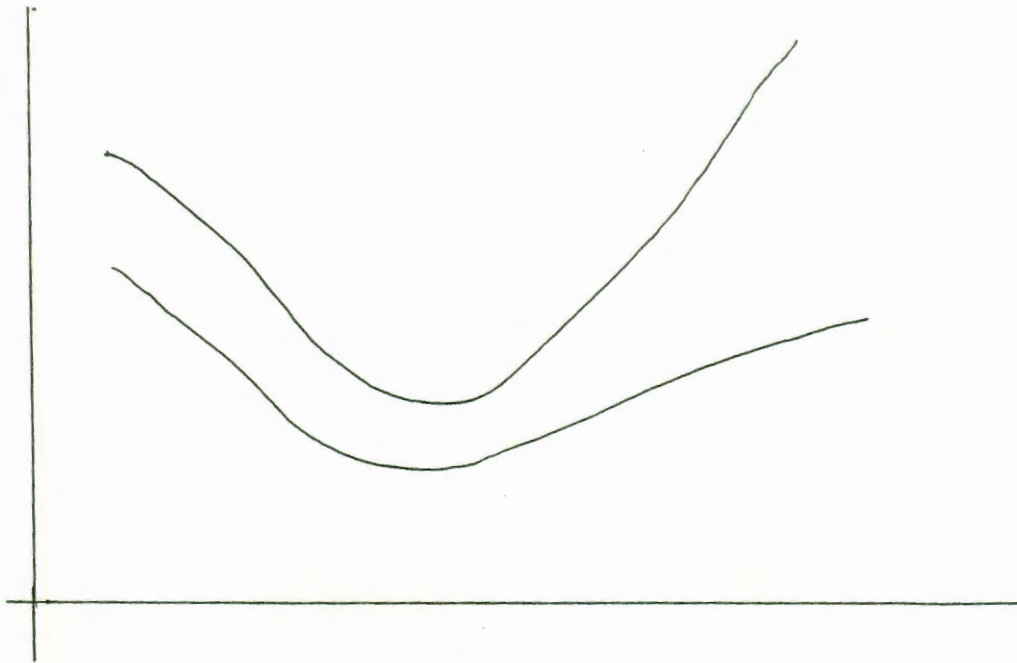
Figure 16. Curves with the same shape in the first half but with divergent shapes in the second half.

means of the second curve. A statistical test is then made of this difference by dividing it by a factor of the covariance matrix and comparing this to a test statistic.

CASE I: COVARIANCE MATRIX UNKNOWN

For this case with   unknown, the following expression holds:

$$(22) \qquad \rho \left(\left| \frac{(\underline{1}'\mu_1 - \underline{1}'\mu_2) - (\underline{1}'\overline{Y}_1 - \underline{1}'\overline{Y}_2)}{\left(\dfrac{2 \ \underline{1}'S\,\underline{1}}{n} \right)^{\frac{1}{2}}} \right| \leq t^{\alpha/2K}_{2(n-1)} \right) = 1-\alpha,$$

where S is the pooled estimate of the covariance matrix and where k is the number of components being compared. One would reject the null hypothesis that $\underline{1}'\overline{Y}_1 - \underline{1}'\overline{Y}_2 = 0$ if

$$(23) \qquad \frac{\underline{1}'\overline{Y}_1 - \underline{1}'\overline{Y}_2}{\left(\dfrac{2 \ \underline{1}'S\,\underline{1}}{n} \right)^{\frac{1}{2}}} \ > \ t^{\alpha/2K}_{2(n-1)}$$

For example, consider the case of the two p-waves shown in figure 2 (p. 20). The difference between respective Haar components was calcaulated. Table XIII gives values of $\underline{h}'_i\overline{Y}_1$, $\underline{h}'_i\overline{Y}_2$, $(\underline{h}'_i\overline{Y}_1 - \underline{h}'_i\overline{Y}_2)$, and $\dfrac{\underline{h}'_i\overline{Y}_1 - \underline{h}'_i\overline{Y}_2}{\dfrac{2 \ \underline{h}'_i S\,\underline{h}_i}{n}}$. Using expression (23), the significant differences were determined and are shown with an asterisk in table XIII.

Since the difference between the first Haar component in each curve was significant, the two curves differed in overall level. Since the difference was positive, the first curve was higher on the average than the second. The difference between the second Haar component of each curve was also significant, with the first curve

TABLE XIII

COMPARISON OF HAAR COMPONENTS OF P-WAVES
BEFORE AND AFTER ADMINISTRATION
OF Δ^9-THC

i	$\underline{h}_i'\bar{y}_1$	$\underline{h}_i'\bar{y}_2$	$\underline{h}_i'\bar{y}_1 - \underline{h}_i'\bar{y}_2$	$\dfrac{h_i'\bar{y}_1 - h_i'\bar{y}_2}{\left(\dfrac{2h_i'sh_i}{n}\right)^{1/2}}$
1	5531.44	4991.08	540.36	49.8*
2	173.28	-38.24	211.52	12.0*
3	-94.24	73.14	-167.38	-23.35*
4	191.80	-151.82	343.62	42.19*
5	-22.94	1.98	-24.92	-12.60*
6	-20.46	37.28	-66.74	-16.29*
7	91.24	-37.30	128.54	32.99*
8	-2.04	-24.20	22.16	7.30*

*significant at .01 level

showing a greater difference in level between first half and second half than is evident in the second curve. In fact, as can be seen from the values of $\underline{h}'_1\bar{y}_1$ and $\underline{h}'_2\bar{y}_2$, the difference is one of direction as well as size, with the first curve having a higher first half and the second curve a higher second half.

The difference between the third Haar components is also significant, the first curve having a negative value for $\underline{h}'_3\bar{y}_1$, while the second curve has a positive value. Hence, the first curve rises from the first quarter to the second, while the second curve decreases. Looking at the difference between the fourth Haar components, one sees that there is also a significant difference here with the first curve showing a decrease from the third quarter to the fourth and the second curve showing an increase from third quarter to fourth.

From the analysis so far, it can be seen that these two curves differ significantly in level as well as shape. Curve 2 seems to be a rough reversal of curve 1 with curve 1, however, showing a greater difference in level between first half and second half than is evident in curve 2.

Analysis of the remaining differences in components will add more detail to the difference in shape. One notes that curve 1 shows a greater change in level from the first eighth to the second than is shown in curve 2. Curve 2 remains rather flat in this region ($\underline{h}'_5\bar{y}_2$ = 1.98). Differences in the sixth and seventh components reflect the reversed trends in the two curves mentioned earlier. Differences in the eighth components of the two curves indicate that the second curve shows a much greater change in

level from the seventh eighth to the eighth eighth than that shown in curve 2.

These two curves differed markedly in shape with all of the eight components which were compared showing significant differences. The confidence band for the difference between the two curves (see figure 3) showed a difference at all but two points. Doing the component by component analysis of these two curves allows one to say more about how the two curves differ. For this experiment, this analysis of the components allows one to say that the administration of Δ^9-THC lowered the overall level of the p-wave and actually effected a reversal in the shape of the curve.

CASE II: COVARIANCE MATRIX KNOWN EXCEPT FOR A SCALAR

For this case, with $\Sigma = \sigma^2 \Sigma'$, the following expression is true:

$$(24) \quad P\left(\left| \frac{(\underline{1}'\mu_1 - \underline{1}'\mu_2) - (\underline{1}'\underline{y}_1 - \underline{1}'\underline{y}_2)}{\left(\dfrac{2s^2 \underline{1}'\Sigma'\underline{1}}{n} \right)^{1/2}} \right| \leq t^{\alpha/2K}_{2(n-1)} \right) = 1 - \alpha$$

where s^2 is the pooled estimate of σ^2. One would reject the null hypothesis that $(\underline{1}'\mu_1 - \underline{1}'\mu_2) = 0$ if

$$(25) \quad \frac{\underline{1}'\underline{y}_1 - \underline{1}'\underline{y}_2}{\left(\dfrac{2s^2 \underline{1}'\Sigma'\underline{1}}{n} \right)^{1/2}} \; > \; t^{\alpha/2K}_{2(n-1)}$$

For an example, we will again use the data from the experiment on the dosage level effects on performance of a task (p. 22). The mean curve for the syn. Δ^9-THC was found to be $\underline{y} = (58.0, 61.3, 44.4, 36.9)$ and for the Δ^9-THC was $\underline{y} = (63.3, 59.6, 33.2, 21.8)$. The pooled estimate of s^2 would be 53.61. Table XIV shows the value of $\underline{h}'_i\underline{y}_1$, $\underline{h}'_i\underline{y}_2$, $\underline{h}'_i\underline{y}_1 - \underline{h}'_i\underline{y}_2$ and $\dfrac{\underline{h}'_i\underline{y}_1 - \underline{h}'_i\underline{y}_2}{\left(\dfrac{2s\,\underline{h}'_i \Sigma'\underline{h}_i}{n} \right)^{1/2}}$.

For $\alpha = .05$, $t^{\alpha/8}_{2(9)} = 2.81$. Thus the difference between the first Haar components and between the second Haar components would be significant. Since,

$$\frac{\psi_{11} - \psi_{21}}{\Sigma c_i^2} = \frac{22.7}{4} = 5.67 \qquad \frac{\psi_{12} - \psi_{22}}{\Sigma c_i^2} = \frac{-29.9}{4} = -7.47$$

the overall level of the first curve is 5.67 units higher than the

TABLE XIV

COMPARISON OF HAAR COMPONENTS OF DOSE-RESPONSE CURVES OF SYN. Δ^2-THC AND Δ^9-THC

i	$\underline{h}_i'\bar{Y}_1$	$\underline{h}_i'\bar{Y}_2$	$\underline{h}_i'\bar{Y}_1 - \underline{h}_i'\bar{Y}_2$	$\dfrac{b_i'\bar{Y}_1 - b_i'\bar{Y}_2}{(^{2s^2}\!/_n\, b_i'\Sigma'b_i)^{1/2}}$
1	200.6	177.9	22.7	2.83*
2	38.0	67.9	-29.9	-3.73*
3	-3.3	3.7	-7.0	-1.51
4	7.5	11.4	-3.9	-.59

*significant at .05 level

second, and the second curve showed a greater difference between the first half and the second that the first curve showed.

In an interpretation of these results with reference to this particular experiment, it could be said that the performance scores were on the average higher with the syn. Δ^9-THC than with Δ^9-THC, and that the two higher dosage levels of the second drug effected a greater change in performance than the two higher levels of drug 1.

VI. CONCLUSION

The interpretaion of data from psychological experiments often involves the analysis and comparison of curves. This paper has summarized methods presented in texts on psychological statistics for the comparison of curves and has presented some alternative methods.

The first kind of analysis discussed was that of finding a confidence band around one curve. There are no methods commonly in use for finding this confidence band. It was mentioned that one might be tempted to find a confidence interval around each point using an α level of significance for each interval. Doing so does not take into account the degree of correlation between the coordinate means and will result in an overall significance level which would be much too low. The alternative discussed allowed one to determine simultaneous intervals such that all of the intervals would cover their respective mean with $1-\alpha$ probability.

The second kind of analysis discussed was finding a confidence band around the difference between two curves. Current methods presented were visual inspection, analysis of variance, and Hotelling's T^2. Visual inspection of curves presents the same difficulty as visual inspection of single values, in that variance and covariances may be so large that observed differences would not be statistically significant. Visual analysis can however point out obvious differences in different portions of the curve which would be obscurred by the global analysis of ANOVA or Hotelling's T^2.

Also, in the case of repeated measure, the use of ANOVA requires a prior test on the shape of the covariance matrix which involves difficult and computer time-consuming computations.

The alternative method discussed involved finding simultaneous confidence intervals about the difference between co-ordinate means. This method allows one to determine rather easily at which points two curves differ significantly and at which points the difference is not significant. This method eliminates the need for finding the determinant of the covariance matrix which is necessary for ANOVA with repeated measures.

Next discussed was finding the significant components of a curve. The best method presently discussed in texts on psychological statistics is the method of orthogonal polynomials. The problem with this is the difficulty in interpreting the meaning of significant trends beyond quadratic. When trends such as cubic and quartic and beyond are significant, it becomes difficult to visualize the shape of the curve from knowledge of the significant trends. The alternative presented was the use of Haar functions. All of the significant Haar functions may be interpreted and the shape of the curve determined no matter how many components are significant. In some cases, however, using Haar components may provide less easily interpreted results if the curve can be characterized by a simple combination of low degree polynomials.

The last type of analysis discussed was that of comparing the components of two or more curves. The comparison of orthogonal polynomial components was discussed as the method presently in use. The results of this method are difficult to interpret if one or more

of the curves has several significant components. The comparison of Haar components may in such cases provide a more easily understood summarization of the differences in shape of two curves.

As discussed above, the new methods can in some cases provide advantages over existing techniques. It is hoped that these new techniques will be incorporated into the analysis of psychological data.

REFERENCES

Box, G. E. D.: Some theorems on quadratic forms applied in the study of analysis of variance problems, Ann. Math. Statist., 1954, 25, 290-302, 484-498.

Harmuth, H.: Transmission of Information by Orthogonal Functions, New York: Springer - Vertag, 1972.

Hays, W. C.: Statistics for the Social Sciences, New York: Holt, Rinehart, and Winston, 1973.

Miller, Simultaneous Statistical Inference, New York: McGraw-Hill, 1966.

Morrison, D. F.: Multivariate Statistical Methods, New York: McGraw-Hill, 1967.

Roy, S. N. and R. C. Bose: Simultaneous confidence interval estimation, Ann. Math. Statist., 1953, 24, 513-536.

Winer, B. J., Statistical Principles in Experimental Design, New York: McGraw-Hill, 1971.

APPENDIX

The following is an explanation of the vector and matrix
multiplication which is used throughout the text to simplify not-
ation and calculation.

Let $\underline{x}$ and $\underline{y}$ be vectors. A vector as it is normally written
($\underline{x}$) is usually considered to be a column vector, that is, with its
elements arranged in a column:

$$\underline{x} = \begin{pmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \end{pmatrix}.$$

A vector written with a prime ($\underline{x}'$) is a row vector:

$$\underline{x}' = (x_1, x_2, x_3, x_4).$$

By definition, a column vector can be multiplied on the left by
a row vector. The product is found by multiplying corresponding
elements and summing these products:

$$\underline{x}'\underline{y} = (x_1, x_2, x_3, x_4) \begin{pmatrix} y_1 \\ y_2 \\ y_3 \\ y_4 \end{pmatrix} = x_1 y_1 + x_2 y_2 + x_3 y_3 + x_4 y_4$$

For example if $\underline{x} = \begin{pmatrix} 1 \\ 1 \\ -1 \\ -1 \end{pmatrix}$ and $\underline{y} = \begin{pmatrix} 36 \\ 39 \\ 43 \\ 35 \end{pmatrix}$,

$$\underline{x}'\underline{y} = (1, 1, -1, -1) \begin{pmatrix} 36 \\ 49 \\ 43 \\ 35 \end{pmatrix} = 36 + 39 - 43 - 35 = -3$$

The multiplication of a vector times a matrix is similar. A
matrix can be multiplied on the left by a row vector and on the right

by a column vector:

$$\underline{x}' \cdot S \quad \text{or} \quad S \cdot \underline{x}$$

Let S be a matrix such that

$$S = \begin{pmatrix} s_{11} & s_{12} & s_{13} & s_{14} \\ s_{21} & s_{22} & s_{23} & s_{24} \\ s_{31} & s_{32} & s_{33} & s_{34} \\ s_{41} & s_{42} & s_{43} & s_{44} \end{pmatrix}$$

S could be considered to be made up of four column vectors of four elements each, or four row vectors of four elements each:

$$S = \begin{pmatrix} s_{11} \\ s_{21} \\ s_{31} \\ s_{41} \end{pmatrix} \begin{pmatrix} s_{12} \\ s_{22} \\ s_{32} \\ s_{42} \end{pmatrix} \begin{pmatrix} s_{13} \\ s_{23} \\ s_{33} \\ s_{43} \end{pmatrix} \begin{pmatrix} s_{14} \\ s_{24} \\ s_{34} \\ s_{44} \end{pmatrix}$$

$$\text{or} \quad S = \begin{pmatrix} (s_{11} & s_{12} & s_{13} & s_{14}) \\ (s_{21} & s_{22} & s_{23} & s_{24}) \\ (s_{31} & s_{32} & s_{33} & s_{34}) \\ (s_{41} & s_{42} & s_{43} & s_{44}) \end{pmatrix}$$

Considering S to be composed of 4 column vectors, left multiplication by a row vector consists of multiplying the row vector times each of the columns. The result is a row vector with 4 elements:

$$\underline{x}' \cdot S = (x_1 \ x_2 \ x_3 \ x_4) \cdot \left(\begin{pmatrix} s_{11} \\ s_{21} \\ s_{31} \\ s_{41} \end{pmatrix} \begin{pmatrix} s_{12} \\ s_{22} \\ s_{32} \\ s_{42} \end{pmatrix} \begin{pmatrix} s_{13} \\ s_{23} \\ s_{33} \\ s_{43} \end{pmatrix} \begin{pmatrix} s_{14} \\ s_{24} \\ s_{34} \\ s_{44} \end{pmatrix} \right)$$

$$= \left((x_1 \; x_2 \; x_3 \; x_4) \begin{pmatrix} s_{11} \\ s_{21} \\ s_{31} \\ s_{41} \end{pmatrix}, \; (x_1 \; x_2 \; x_3 \; x_4) \begin{pmatrix} s_{12} \\ s_{22} \\ s_{32} \\ s_{42} \end{pmatrix}, \right.$$

$$\left. (x_1 \; x_2 \; x_3 \; x_4) \begin{pmatrix} s_{13} \\ s_{23} \\ s_{33} \\ s_{34} \end{pmatrix}, \; (x_1 \; x_2 \; x_3 \; x_4) \begin{pmatrix} s_{14} \\ s_{24} \\ s_{34} \\ s_{44} \end{pmatrix} \right)$$

$$= (x_1 s_{11} + x_2 s_{21} + x_3 s_{31} + x_4 s_{41}, \; x_1 s_{12} + x_2 s_{22} + x_3 s_{32} + x_4 s_{42},$$
$$x_1 s_{13} + x_2 s_{23} + x_3 s_{33} + x_4 s_{43}, \; x_1 s_{14} + x_2 s_{24} + x_3 s_{34} + x_4 s_{44})$$

For example:

$$(1, 1, -1, -1) \begin{pmatrix} 6 & 8 & 7 & 5 \\ 9 & 3 & 2 & 4 \\ 8 & 6 & 1 & 5 \\ 7 & 6 & 2 & 3 \end{pmatrix} = \left((1, 1, -1, -1) \begin{pmatrix} 6 \\ 9 \\ 8 \\ 7 \end{pmatrix}, \; (1, 1, -1, -1) \begin{pmatrix} 8 \\ 3 \\ 6 \\ 6 \end{pmatrix}, \right.$$

$$\left. (1, 1, -1, -1) \begin{pmatrix} 7 \\ 2 \\ 1 \\ 2 \end{pmatrix}, \; (1, 1, -1, -1) \begin{pmatrix} 5 \\ 4 \\ 5 \\ 3 \end{pmatrix} \right)$$

$$= \left((6 + 9 - 8 - 7), \; (8 + 3 - 6 - 6), \; (7 + 2 - 1 - 2), \; (5 + 4 - 5 - 3) \right)$$

$$= (2, -1, 6, 1)$$

A matrix can be right multiplied by a column vector. In this

case, consider the matrix to be composed of rows. Each row is multiplied separately times the column vector, with the result being a column vector:

$$S \cdot \underline{x} = \begin{pmatrix} (s_{11}, & s_{12}, & s_{13}, & s_{14}) \\ (s_{21}, & s_{72}, & s_{23}, & s_{24}) \\ (s_{31}, & s_{32}, & s_{33}, & s_{34}) \\ (s_{41}, & s_{42}, & s_{43}, & s_{44}) \end{pmatrix} \cdot \begin{pmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \end{pmatrix} = \begin{pmatrix} s_{11}x_1 + s_{12}x_2 + s_{13}x_3 + s_{14}x_4 \\ s_{21}x_1 + s_{22}x_2 + s_{23}x_3 + s_{24}x_4 \\ s_{31}x_1 + s_{32}x_2 + s_{33}x_3 + s_{34}x_4 \\ s_{41}x_1 + s_{42}x_2 + s_{43}x_3 + s_{44}x_4 \end{pmatrix}$$

For example:

$$\begin{pmatrix} 6 & 8 & 7 & 5 \\ 9 & 3 & 2 & 4 \\ 8 & 6 & 1 & 5 \\ 7 & 6 & 2 & 3 \end{pmatrix} \cdot \begin{pmatrix} 1 \\ 1 \\ -1 \\ -1 \end{pmatrix} = \begin{pmatrix} 6 + 8 - 7 - 5 \\ 9 + 3 - 2 - 4 \\ 8 + 6 - 1 - 5 \\ 7 + 6 - 2 - 3 \end{pmatrix} = \begin{pmatrix} 2 \\ 6 \\ 9 \\ 8 \end{pmatrix}$$

It should be noted that $\underline{x}' \cdot S$ does not necessarily equal $S \cdot \underline{x}$.

As an example of vector and matrix multiplication involving the Haar coefficients, consider a curve having an average curve of

$$\overline{\underline{y}} = (7, 12, 15, 10)$$

and covariance matrix of:

$$S = \begin{pmatrix} 12 & 5 & 8 & 2 \\ 5 & 14 & 9 & 3 \\ 8 & 9 & 15 & 4 \\ 2 & 3 & 4 & 6 \end{pmatrix}$$

The linear combinations of the Haar coefficients with the vector means

are written as $\underline{h}'_1\bar{\underline{y}}$, $\underline{h}'_2\bar{\underline{y}}$, $\underline{h}'_3\bar{\underline{y}}$, $\underline{h}'_4\bar{\underline{y}}$ and would be computed as follows:

$$\underline{h}'_1\bar{\underline{y}} = (1, 1, 1, 1)\begin{pmatrix} 7 \\ 12 \\ 15 \\ 10 \end{pmatrix} = 7 + 12 + 15 + 10 = 44$$

$$\underline{h}'_2\bar{\underline{y}} = (1, 1, -1, -1)\begin{pmatrix} 7 \\ 12 \\ 15 \\ 10 \end{pmatrix} = 7 + 12 - 15 - 10 = -6$$

$$\underline{h}'_3\bar{\underline{y}} = (1, -1, 0, 0)\begin{pmatrix} 7 \\ 12 \\ 15 \\ 10 \end{pmatrix} = 7 - 12 + 0 + 0 = -5$$

$$\underline{h}'_4\bar{\underline{y}} = (0, 0, 1, -1)\begin{pmatrix} 7 \\ 12 \\ 15 \\ 10 \end{pmatrix} = 0 + 0 + 15 - 10 = 5$$

$\underline{h}'_i \underline{Sh}_i$ would be found by

$$\underline{h}'_1\underline{Sh}_1 = (1\ 1\ 1\ 1)\begin{pmatrix} 12 & 5 & 8 & 2 \\ 5 & 14 & 9 & 3 \\ 8 & 9 & 15 & 4 \\ 2 & 3 & 4 & 6 \end{pmatrix}\begin{pmatrix} 1 \\ 1 \\ 1 \\ 1 \end{pmatrix}$$

$$= (12 + 5 + 8 + 2, \; 5 + 14 + 9 + 3, \; 8 + 9 + 15 + 4, \; 2 + 3 + 4 + 6) \begin{pmatrix} 1 \\ 1 \\ 1 \\ 1 \end{pmatrix}$$

$$= (27, 31, 36, 15) \begin{pmatrix} 1 \\ 1 \\ 1 \\ 1 \end{pmatrix} = 27 + 31 + 36 + 15 = 109$$

$$\underline{h}_2' \, \underline{Sh}_2 = (1, 1, -1, -1) \begin{pmatrix} 12 & 5 & 8 & 2 \\ 5 & 14 & 9 & 3 \\ 8 & 9 & 15 & 4 \\ 2 & 3 & 4 & 6 \end{pmatrix} \begin{pmatrix} 1 \\ 1 \\ -1 \\ -1 \end{pmatrix}$$

$$= (12 + 5 - 8 - 2, \; 5 + 14 - 9 - 3, \; 8 + 9 - 15 - 4, \; 2 + 3 - 4 - 6) \begin{pmatrix} 1 \\ 1 \\ -1 \\ -1 \end{pmatrix}$$

$$= (7, 7, -2, -5) \begin{pmatrix} 1 \\ 1 \\ -1 \\ -1 \end{pmatrix} = 7 + 7 + 2 + 5 = 21$$

$$\underline{h}_3' \, \underline{Sh}_3 = (1, -1, 0, 0) \begin{pmatrix} 12 & 5 & 8 & 2 \\ 5 & 14 & 9 & 3 \\ 8 & 9 & 15 & 4 \\ 2 & 3 & 4 & 6 \end{pmatrix} \begin{pmatrix} 1 \\ -1 \\ 0 \\ 0 \end{pmatrix}$$

$$= (12-5,\ 5-14,\ 8-9,\ 2-3) \begin{pmatrix} 1 \\ -1 \\ 0 \\ 0 \end{pmatrix} = (7,-9,-1,-1) \begin{pmatrix} 1 \\ -1 \\ 0 \\ 0 \end{pmatrix} = 7 + 9 = 16$$

$$\underline{h}_4'\ \underline{Sh}_4 = (0,\ 0,\ 1,\ -1) \begin{pmatrix} 12 & 5 & 8 & 2 \\ 5 & 14 & 9 & 3 \\ 8 & 9 & 15 & 4 \\ 2 & 3 & 4 & 6 \end{pmatrix} \begin{pmatrix} 0 \\ 0 \\ 1 \\ -1 \end{pmatrix}$$

$$= (8-2,\ 9-3,\ 15-4,\ 4-6) \begin{pmatrix} 0 \\ 0 \\ 1 \\ -1 \end{pmatrix} = (6,6,11,-2) \begin{pmatrix} 0 \\ 0 \\ 1 \\ -1 \end{pmatrix} = 11 + 2 = 13$$