NASA Contractor Report 159117

NASA-CR-159117
19790020780

THE ALTERATION OF PROFILE ANALYSIS TO
ACCOMMODATE TESTING FUNCTIONS

Raymond H. Myers

RAYMOND H. MYERS
Blacksburg, Virginia 24060

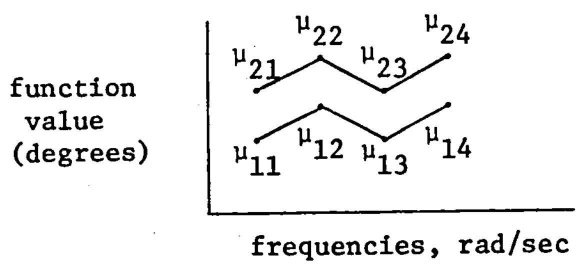
NASA Purchase Order L-72497A
July 1979

NASA
National Aeronautics and
Space Administration
Langley Research Center
Hampton, Virginia 23665

THE ALTERATION OF PROFILE ANALYSIS TO
ACCOMMODATE TESTING FUNCTIONS

The purpose of this task is to develop a methodology for testing for differences between several pilot describing functions, where the data points represent averages at various frequencies. Typically a functions represents a treatment or treatment combinations which are to be compared. Data are taken in replicates and an output is measured for each frequency. It is not unusual to experience 4-8 functions representing a like number of experimental conditions in a real-time piloted aircraft simulation. Typically, as many as 8-16 frequencies may be involved.

Consider the following plot for the special case of two functions.



Here, the μ_{ij} represents the population mean for the j^{th} frequency under the i^{th} function. The problem is to determine whether or not, simultaneously, the means for function 1 are equal to the means for function 2, across frequencies. The extension to more than two functions is also of considerable interest.

Specifically the goals of the task are as follows:

(i) Develop and describe the methodology for testing for differences between functions.

(ii) Determine how to approach the problem of assessing the power of the test. Supply charts or tables for power.

(iii) Use the power results to recommend how to design such experiments; for example determine an effective number of frequencies; in addition, what is the maximum number of functions that will allow a reasonably sensitive test. Also what is a reasonable number of replicates?

(iv) Discuss software considerations.

Basic Assumptions for the Experiment

In this section we shall discuss the distributional and data structure proposed for the test on k functions ($k \geq 2$). Consider initially two functions, in the context of Figure I. Suppose that we have two random vectors

$$\underline{x}_1 = \begin{bmatrix} x_{11} \\ x_{12} \\ \vdots \\ x_{1t} \end{bmatrix} \quad \text{and } \underline{x}_2 = \begin{bmatrix} x_{21} \\ x_{22} \\ \vdots \\ x_{2t} \end{bmatrix}$$

where x_{ip} is the scalar output for the i^{th} function and the p^{th} frequency. These are basic output measurements. It is assumed that each of the vectors follows a multivariate normal distribution with common variance-covariance matrix Σ, the latter a $t \times t$ matrix. The practical implication here is that within each function the observations are correlated and the correlation structure is the same for each of the two functions. For the first function n_1 independent vectors are taken and for the second function n_2 vectors are observed.

The extension to more than two functions is obvious. It is assumed that there are k-multidimensional vectors $\underline{x}_1, \underline{x}_2, \ldots, \underline{x}_k$,

$$\begin{bmatrix} x_{11} \\ x_{12} \\ \vdots \\ x_{1t} \end{bmatrix}, \quad \begin{bmatrix} x_{21} \\ x_{22} \\ \vdots \\ x_{2t} \end{bmatrix} \quad \cdots \quad \begin{bmatrix} x_{k1} \\ x_{k2} \\ \vdots \\ x_{kt} \end{bmatrix}$$

The assumption is made that $\underline{x}_i$ is multivariate normal with mean $\underline{\mu}_i$ $(i=1,2,\ldots,k)$ and variance-covariance matrix Σ.

Hypothesis

Consider again the case of two functions illustrated in Figure I. A reasonable hypothesis that accomplishes the goals outlined here, is given by

$$H_0: \begin{bmatrix} \mu_{11} = \mu_{21} \\ \mu_{12} = \mu_{22} \\ \mu_{13} = \mu_{23} \\ \mu_{14} = \mu_{24} \end{bmatrix} \tag{1}$$

or, in terms of vectors,

$$H_0: \underline{\mu}_1 = \underline{\mu}_2 ,$$

where $\underline{\mu}_i$ is the mean vector for the i^{th} function. One can easily see that this is the hypothesis that the experimenter needs to test. Namely, the null hypothesis states that the mean value of the two functions are equivalent at each frequency. In the general case of k functions and t frequencies the hypothesis is

$$H_0: \underline{\mu}_1 = \underline{\mu}_2 = \underline{\mu}_3 = \dots = \underline{\mu}_k \qquad (2)$$

where each vector is t-dimensional. We need a test procedure for testing the joint statement given in (2). In addition, we need guidelines on the experimental design that will adequately give evidence in favor of the alternate hypothesis when it is appropriate.

Profile Analysis

The problem posed here appears to be very similar to Profile Analysis [2] which is a multivariate technique in which a "repeated measures" design is used and the treatments are broken into groups. Basically, measurements on individuals (or individual items) are conducted for each of say t treatments, with the observations correlated from treatment to treatment. The purpose of the experiment is to determine if there are differences in treatment means. When the treatments are put into an additional classification variable called "groups" then the procedure becomes what is called Profile Analysis. Consider, again Figure I and allow the functions to take the role of the groups and frequencies take the role of the treatments. This would appear to be an application of Profile Analysis.

In a Profile Analysis the test on treatments is not valid unless no interaction exists between treatments and groups. Thus, initially a test for parallelism (no interaction) is usually conducted before the test on treatments is attempted. The hypothesis of the test on parallelism is given by

$$H_0: \begin{bmatrix} \mu_{11} - \mu_{12} = \mu_{21} - \mu_{22} \\ \mu_{13} - \mu_{12} = \mu_{23} - \mu_{22} \\ \mu_{14} - \mu_{13} = \mu_{24} - \mu_{23} \end{bmatrix}$$

The test involves a Hotelling's T^2 [2] and is quite easy to extend to more than two groups.

Alteration of Profile Analysis to Accommodate Testing Functions

One can easily see (easier to visualize in the case of two functions or groups) that if groups and interactions between groups and frequencies (functions and frequencies) are not statistically significant, one can interpret this as implying that functions are not significantly different <u>point by point</u> or <u>frequency by frequency</u>. One mode of verfication of this is to note that if two functions are parallel between frequencies the only way that the two function averages can be the same is for the two functions to coincide. On the other hand, function averages might be very close together but because of non-parallelism (interaction) the two functions may be far from coincidental. Thus neither a significance test on groups (function) averages or interaction will suffice for our purposes. Rather it is the joint hypothesis of the two that is relevant.

Searching into the structure of the problem more closely suggests that the condition of coincidental functions we have described here is identical to the hypothesis in (1) or (2) for the general case. Thus it can be said that we have a profile analysis in which we are attempting to simultaneously detect parallelism and group (function) effects but we are doing it with a single hypothesis and thus a single test and not with two tests.

Test Statistic and Procedure

For the hypothesis in (1), we are simply testing the equality of two mean vectors. For the j^{th} function, n_j independent vectors (replicates) are obtained and we have two averages $\bar{\underline{x}}_1$, $\bar{\underline{x}}_2$ which are joint estimates of the mean vectors $\underline{\mu}_1$ and $\underline{\mu}_2$ (t-dimensional). The sample data is used to compute an estimate of the variance-covariance Σ. Thus we have for the first function, n_1 independent vectors, and n_2 independent vectors for the second function. The estimate of Σ is found by obtaining sample variances and covariances for each function and pooling over functions. Thus the estimates have $n_1 + n_2 - 2$ degrees of freedom. The test statistic is given by [2]

$$T^2 = (\bar{\underline{x}}_1 - \bar{\underline{x}}_2) S^{-1} (\bar{\underline{x}}_1 - \bar{\underline{x}}_2) \qquad (3)$$

where

$$S = \hat{\Sigma}\left(\frac{1}{n_1} + \frac{1}{n_2}\right)$$

The statistic in equation (3) follows Hotelling's T^2 distribution with $n_1 + n_2 - 2$ degrees of freedom. If the same number of replicates are taken for each function, say n, then of course $n_1 + n_2 - 2$ is replaced by 2(n-1). The statistic

$$\frac{(n_1+n_2-t-1)}{(n_1+n_2-2)\,t} \; T^2 \qquad (4)$$

follows a F distribution with t and $n_1 + n_2 - t - 1$ degrees of freedom. This fact allows us to handle the power quite easily.

The extension to more than two functions is quite easily done. Consider the hypothesis of equation (2), which, written in expanded form is given by

$$\begin{aligned} H_0:\; & \mu_{11} = \mu_{21} = \mu_{31} = \cdots = \mu_{k1} \\ & \mu_{12} = \mu_{22} = \mu_{32} = \cdots = \mu_{k2} \\ & \vdots \\ & \mu_{1t} = \mu_{2t} = \mu_{3t} = \cdots = \mu_{kt} \end{aligned}$$

Once again, there are vectors of averages $\bar{\underline{x}}_1, \bar{\underline{x}}_2, \ldots, \bar{\underline{x}}_k$, t-dimensional. This data is used to compute a <u>pooled</u> estimate of the variance-covariance matrix $\hat{\Sigma}$. In addition a t×t matrix (hypothesis matrix) is given by

$$h_{rs} = \sum_{j=1}^{k} \frac{1}{n_j} T_{rj} T_{js} - \frac{1}{N} G_r G_s \quad .$$

Here T_{jr} is the sum of the data on the r^{th} frequency for function j. $N = \sum_{j=1}^{k} n_j$. G_r is the grand total of all observations on the r^{th} frequency. The test statistic is the <u>largest</u> eigenvalue of the matrix. HE^{-1}, where E is the error sum of squares matrix, which is given by

$$E = \hat{\Sigma} \cdot \text{d.f.}$$

where d.f. are the degrees of freedom for estimating the variance-covariance matrix (d.f. = $\sum_{i=1}^{k} n_i - k$).

Examples of the computation for both the special case and the general case will be given in a later section. Charts for the use of the largest root statistic can be found in Morrison [2]. A computer program for the computation of the test statistic will be found in the Appendix.

Power of Tests

The power associated with the case of two functions is quite easy to handle from a theoretical point of view. In a practical sense there are some difficulties that are not insurmountable. The test statistic in (3) follows F_{t,n_1+n_2-t-1} under the hypothesis of equality of means stipulated in H_0. One can immediately notice that the denominator degrees of freedom automatically imply an experimental restriction, namely that $n_1 + n_2 > (t+1)$. The test procedure, of course, involves rejection of H_0 if the calculated test statistic in (4) exceeds the upper tail percentage point of the F-distribution. If $\underline{\mu}_1 \neq \underline{\mu}_2$ the test statistic follows the non central F-distribution [1] with degrees of freedom t and $n_1 + n_2 - t - 1$, and with non centrality parameter

$$\delta^2 = \frac{n_1 n_2}{n_1+n_2} (\underline{\mu}_1-\underline{\mu}_2)'\Sigma^{-1}(\underline{\mu}_1-\underline{\mu}_2) \quad .$$

Thus for a specific difference $\underline{\mu}_1 - \underline{\mu}_2$ between the mean vectors and a specific variance covariance matrix , the power is given by the probability

$$Pr[F'_{t,n_1+n_2-t-1,\delta^2} > F_{\alpha,t,n_1+n_2-t-1}]$$

where $F'_{t,n_1+n_2-t-1,\delta^2}$ is a non central F variate and F_{α,t,n_1+n_2-t-1} is the upper α^{th} percentage point of the F distribution. Thus the probability distribution of the non central F is needed. This distribution is discussed in detail in Graybill [1]. Power charts are offered in Morrison [2] and are quite easy to use. An explanation of their use is given in the text.

In order to compute the power one must know Σ, or at least have an estimate of it. For the study described here, hypothetical values for the variance-covariance matrices are used, these values generally taken from previous studies at NASA Langley. A power study is described in the following section. This study was done to arrive at appropriate values of sample sizes (number of replicates) and number of frequencies that provide good power for specific values (in terms of order of magnitude) of the differences $(\underline{\mu}_1-\underline{\mu}_2)$ that are being detected between the functions.

The power computation for the test for more than two functions is extremely difficult and is wrought with practical difficulties that actually prohibit its use. Approximations are available [3]. However, the approximations involve many quantities that must be estimated from the data. The non central F probabilities are monotonic increasing functions of the non centrality parameter which, if one investigates the structure of δ^2, is very pleasing. That is, δ^2 represents "how false" the null hypothesis is before we would wish to reject H_0 with a high probability, and as δ^2 grows large we would expect the power to increase monotonically. However, the structure of the approximations of the power of the "largest root test" is such that several non centrality parameters appear and the power is not a monotonic function of them which suggests that for at least some cases the approximation is not particularly good.

It would seem reasonable then that sample sizes and other recommendations should be done for testing _two functions at a time_. In addition, the user can assess his own power in a particular case by computing the power through the use of tables of the non central F. This will be reasonable as long as one _does not study too many functions simultaneously_. This does not mean that one must restrict the number of functions in the test, but rather in a power assessment it would only be accurate if _the number of functions were kept low, say four or less_. It should be emphazised here that the value, t, the number of frequencies, n_i, the number of replicates are more crucial than k, the number of functions.

Example of Test Computation

In this section we present some hypothetical data and numerical examples of the computation for cases of two functions and three functions. The output here is not what NASA Langley personnel has at its disposal but it does represent a representative type format for the output. For the case of two functions there are $t = 5$ frequencies and $n_1 = n_2 = 4$.

OBS	GROUP	Y1	Y2	Y3	Y4	Y5
1	1	12.4	9.5	15.4	13.8	15.2
2	1	13.3	10.2	14.7	13.9	15.4
3	1	13.1	10.4	14.9	14.2	14.4
4	1	12.8	10.1	15.2	14.0	14.9
5	2	10.6	9.3	12.3	12.3	11.6
6	2	9.3	8.7	12.1	12.1	12.4
7	2	9.9	9.0	11.6	11.8	12.0
8	2	10.1	8.8	11.9	11.9	11.9

ANALYSIS OF VARIANCE PROCEDURE

CLASS LEVEL INFORMATION

CLASS	LEVELS	VALUES
GROUP	2	1 2

NUMBER OF OBSERVATIONS IN DATA SET = 8

12

ANALYSIS OF VARIANCE PROCEDURE

EPENDENT VARIABLE: Y1

OURCE	DF	SUM OF SQUARES	MEAN SQUARE	F VALUE	PR > F	R-SQUARE	C.V.
ODEL	1	17.11125000	17.11125000	77.34	0.0001	0.928005	4.1125
RROR	6	1.32750000	0.22125000		STD DEV		Y1 MEAN
ORRECTED TOTAL	7	18.43875000			0.47037219		11.43750000

OURCE	DF	ANOVA SS	F VALUE	PR > F
ROUP	1	17.11125000	77.34	0.0001

ANALYSIS OF VARIANCE PROCEDURE

DEPENDENT VARIABLE: Y2

SOURCE	DF	SUM OF SQUARES	MEAN SQUARE	F VALUE	PR > F	R-SQUARE	C.V.
MODEL	1	2.42000000	2.42000000	22.00	0.0034	0.785714	3.4912
ERROR	6	0.66000000	0.11000000		STD DEV		Y2 MEAN
CORRECTED TOTAL	7	3.08000000			0.33166248		9.50000000

SOURCE	DF	ANOVA SS	F VALUE	PR > F
GROUP	1	2.42000000	22.00	0.0034

ANALYSIS OF VARIANCE PROCEDURE

DEPENDENT VARIABLE: Y3

SOURCE	DF	SUM OF SQUARES	MEAN SQUARE	F VALUE	PR > F	R-SQUARE	C.V.
MODEL	1	18.91125000	18.91125000	203.53	0.0001	0.971364	2.2559
ERROR	6	0.55750000	0.09291667		STD DEV		Y3 MEAN
CORRECTED TOTAL	7	19.46875000			0.30482235		13.51250000

SOURCE	DF	ANOVA SS	F VALUE	PR > F
GROUP	1	18.91125000	203.53	0.0001

ANALYSIS OF VARIANCE PROCEDURE

DEPENDENT VARIABLE: Y4

SOURCE	DF	SUM OF SQUARES	MEAN SQUARE	F VALUE	PR > F	R-SQUARE	C.V.
MODEL	1	7.60500000	7.60500000	194.17	0.0001	0.970026	1.5224
ERROR	6	0.23500000	0.03916667		STD DEV		Y4 MEAN
CORRECTED TOTAL	7	7.84000000			0.19790570		13.00000000

SOURCE	DF	ANOVA SS	F VALUE	PR > F
GROUP	1	7.60500000	194.17	0.0001

ANALYSIS OF VARIANCE PROCEDURE

DEPENDENT VARIABLE: Y5

SOURCE	DF	SUM OF SQUARES	MEAN SQUARE	F VALUE	PR > F	R-SQUARE	C.V.
MODEL	1	18.00000000	18.00000000	120.67	0.0001	0.952633	2.8662
ERROR	6	0.89500000	0.14916667		STD DEV		Y5 MEAN
CORRECTED TOTAL	7	18.89500000			0.38622101		13.47500000

SOURCE	DF	ANOVA SS	F VALUE	PR > F
GROUP	1	18.00000000	120.67	0.0001

ANALYSIS OF VARIANCE PROCEDURE

E = ERROR SS&CP MATRIX

DF=6	Y1	Y2	Y3	Y4	Y5
Y1	1.32750000	0.76500000	-0.22250000	0.22250000	-0.58250000
Y2	0.76500000	0.66000000	-0.21500000	0.25000000	-0.49000000
Y3	-0.22250000	-0.21500000	0.55750000	0.12750000	-0.06750000
Y4	0.22250000	0.25000000	0.12750000	0.23500000	-0.27000000
Y5	-0.58250000	-0.49000000	-0.06750000	-0.27000000	0.89500000

PARTIAL CORRELATION COEFFICIENTS FROM THE ERROR SS&CP MATRIX / PROB > !R!

DF=5	Y1	Y2	Y3	Y4	Y5
Y1	1.000000 0.0000	0.817283 0.0248	-0.258637 0.5755	0.398363 0.3761	-0.534401 0.2166
Y2	0.817283 0.0248	1.000000 0.0000	-0.354441 0.4354	0.634796 0.1256	-0.637548 0.1235
Y3	-0.258637 0.5755	-0.354441 0.4354	1.000000 0.0000	0.352252 0.4384	-0.095559 0.8385
Y4	0.398363 0.3761	0.634796 0.1256	0.352252 0.4384	1.000000 0.0000	-0.588733 0.1643
Y5	-0.534401 0.2166	-0.637548 0.1235	-0.095559 0.8385	-0.588733 0.1643	1.000000 0.0000

ANALYSIS OF VARIANCE PROCEDURE

H = ANOVA SS&CP MATRIX FOR: GROUP

DF=1	Y1	Y2	Y3	Y4	Y5
Y1	17.11125000	6.43500000	17.98875000	11.40750000	17.55000000
Y2	6.43500000	2.42000000	6.76500000	4.29000000	6.60000000
Y3	17.98875000	6.76500000	18.91125000	11.99250000	18.45000000
Y4	11.40750000	4.29000000	11.99250000	7.60500000	11.70000000
Y5	17.55000000	6.60000000	18.45000000	11.70000000	18.00000000

CHARACTERISTIC ROOTS AND VECTORS OF: E INVERSE * H, WHERE H = ANOVA SS&CP MATRIX FOR: GROUP   E = ERROR SS&CP MATRIX

CHARACTERISTIC ROOT	PERCENT	CHARACTERISTIC VECTOR V'EV=1 Y1	Y2	Y3	Y4	Y5
201.51859391	100.00	0.41787631	0.79418981	1.07215163	0.44848061	1.25686420
0.00000000	0.00	0.90392450	-4.05018675	-2.51037663	5.22322795	-0.21822003
0.00000000	0.00	0.85487543	-0.51987437	-0.28125967	0.15670677	-0.45645118
0.00000000	0.00	-0.16047073	-0.32073506	0.81093019	0.24622078	-0.71718513
0.00000000	0.00	-1.18817472	2.25265548	0.13465891	0.05632777	0.15785824

```
                         ANALYSIS OF VARIANCE PROCEDURE
            MANOVA TEST CRITERIA FOR THE HYPOTHESIS OF NO OVERALL GROUP EFFECT
                         H = ANOVA SS&CP MATRIX FOR: GROUP
                         E = ERROR SS&CP MATRIX
                         P = DEP. VARIABLES =          5
                         Q = RANK OF H      =          1
                         NE= DF OF E        =          6
                         S = MIN(P,Q)       =          1
                         M = .5(ABS(P-Q)-1) =        1.5
                         N = .5(NE-P-1)     =        0.0
-----------------------------------------------------------------------------------------------
HOTELLING-LAWLEY TRACE = TR(E**-1*H) =          201.51859391     (SEE PILLAI'S TABLE #3)
        F APPROXIMATION = 2(S*N+1)*TR(E**-1*H)/(S*S*(2M+S+1))    WITH S(2M+S+1) AND 2(S*N+1) DF
                 F(5,2) =      80.61     PROB > F = 0.0123
-----------------------------------------------------------------------------------------------
PILLAI'S TRACE         V = TR(H*INV(H+E)) =        0.99506218     (SEE PILLAI'S TABLE #2)
        F APPROXIMATION = (2N+S+1)/(2M+S+1) * V/(S-V)            WITH S(2M+S+1) AND S(2N+S+1) DF
                 F(5,2) =      80.61     PROB > F = 0.0123
-----------------------------------------------------------------------------------------------
WILKS' CRITERION       L = DET(E)/DET(H+E) =       0.00493782     (SEE BKA VOL 53 P 347)
        W = -(NE-.5(P-Q+1))*LN(L)         =             18.5879
-----------------------------------------------------------------------------------------------
ROY'S MAXIMUM ROOT CRITERION =                  201.51859391     (SEE AMS VOL 31 P 625)
        FIRST CANONICAL VARIABLE YIELDS AN F UPPER BOUND
                 F(1,6) =    1209.11      (UPPER BOUND)
-----------------------------------------------------------------------------------------------
```

20

```
//QUICKE SMART,SAS78,SHARON                                                             08.53.27 03/07/78

1                              S T A T I S T I C A L   A N A L Y S I S   S Y S T E M            8:53 TUESDAY, MARCH 7, 1978
NOTE: THE JOB SSSSSD6 HAS BEEN RUN UNDER RELEASE 76.5 OF SAS AT VPI & SU.

1          DATA ONE; INPUT GROUP Y1-Y5;
2          CARDS;

NOTE: DATA SET WORK.ONE HAS 12 OBSERVATIONS AND 6 VARIABLES.
NOTE: THE DATA STATEMENT USED 0.43 SECONDS AND 128K.

15         PROC PRINT;

NOTE: THE PROCEDURE PRINT USED 0.81 SECONDS AND 128K AND PRINTED PAGE 1.

15                   PROC ANOVA; CLASSES GROUP;
16         MODEL Y1-Y5=GROUP;
17         MANOVA H=GROUP/PRINTH PRINTE;

NOTE: THE PROCEDURE ANOVA USED 1.53 SECONDS AND 170K AND PRINTED PAGES 2 TO 10.
NOTE: SAS USED 170K MEMORY.

NOTE: BARR, GOODNIGHT, SALL AND HELWIG
      SAS INSTITUTE INC.
      P.O. BOX 10066
      RALEIGH, N.C. 27605

PROCESSOR ENDED, COMPLETION CODE = S000, U0000
             CARDS     LINES   CPU SEC      EXCPS       KBS    PRIORITY      TOTAL
NUMBER OF      17        24        3          97        780     PRIORITY           KBS=1024 BYTE-SEC
CHARGES       .02       .02      .07         .05        .02       .14        .32
```

OBS	GROUP	Y1	Y2	Y3	Y4	Y5
1	1	12.4	9.5	15.4	13.8	15.2
2	1	13.3	10.2	14.7	13.9	15.4
3	1	13.1	10.4	14.9	14.2	14.4
4	1	12.8	10.1	15.2	14.0	14.9
5	2	10.6	9.3	12.3	12.3	11.6
6	2	9.3	8.7	12.1	12.1	12.4
7	2	9.9	9.0	11.6	11.8	12.0
8	2	10.1	8.8	11.9	11.9	11.9
9	3	14.8	12.4	17.4	16.0	14.8
10	3	14.9	11.7	17.1	16.0	14.9
11	3	15.3	11.9	16.5	16.3	15.1
12	3	15.1	12.2	16.8	15.9	15.0

function

5 [illegible]

22

ANALYSIS OF VARIANCE PROCEDURE

CLASS LEVEL INFORMATION

CLASS	LEVELS	VALUES
GROUP	3	1 2 3

NUMBER OF OBSERVATIONS IN DATA SET = 12

ANALYSIS OF VARIANCE PROCEDURE

DEPENDENT VARIABLE: Y1

SOURCE	DF	SUM OF SQUARES	MEAN SQUARE	F VALUE	PR > F	R-SQUARE	C.V.
MODEL	2	51.43166667	25.71583333	156.91	0.0001	0.972121	3.2045
ERROR	9	1.47500000	0.16388889		STD DEV		Y1 MEAN
CORRECTED TOTAL	11	52.90666667			0.40483193		12.63333333

SOURCE	DF	ANOVA SS	F VALUE	PR > F
GROUP	2	51.43166667	156.91	0.0001

ANALYSIS OF VARIANCE PROCEDURE

DEPENDENT VARIABLE: Y2

SOURCE	DF	SUM OF SQUARES	MEAN SQUARE	F VALUE	PR > F	R-SQUARE	C.V.
MODEL	2	19.76000000	9.88000000	93.60	0.0001	0.954128	3.1391
ERROR	9	0.95000000	0.10555556		STD DEV		Y2 MEAN
CORRECTED TOTAL	11	20.71000000			0.32489314		10.35000000

SOURCE	DF	ANOVA SS	F VALUE	PR > F
GROUP	2	19.76000000	93.60	0.0001

ANALYSIS OF VARIANCE PROCEDURE

EPENDENT VARIABLE: Y3

OURCE	DF	SUM OF SQUARES	MEAN SQUARE	F VALUE	PR > F	R-SQUARE	C.V.
ODEL	2	50.42166667	25.21083333	225.21	0.0001	0.980410	2.2825
RROR	9	1.00750000	0.11194444		STD DEV		Y3 MEAN
ORRECTED TOTAL	11	51.42916667			0.33458100		14.65833333

OURCE	DF	ANOVA SS	F VALUE	PR > F
ROUP	2	50.42166667	225.21	0.0001

ANALYSIS OF VARIANCE PROCEDURE

DEPENDENT VARIABLE: Y4

SOURCE	DF	SUM OF SQUARES	MEAN SQUARE	F VALUE	PR > F	R-SQUARE	C.V.
MODEL	2	32.41166667	16.20583333	448.78	0.0001	0.990072	1.3557
ERROR	9	0.32500000	0.03611111		STD DEV		Y4 MEAN
CORRECTED TOTAL	11	32.73666667			0.19002924		14.01666667

SOURCE	DF	ANOVA SS	F VALUE	PR > F
GROUP	2	32.41166667	448.78	0.0001

ANALYSIS OF VARIANCE PROCEDURE

DEPENDENT VARIABLE: YS

SOURCE	DF	SUM OF SQUARES	MEAN SQUARE	F VALUE	PR > F	R-SQUARE	C.V.
MODEL	2	23.80166667	11.90083333	113.34	0.0001	0.961813	2.3201
ERROR	9	0.94500000	0.10500000		STD DEV		YS MEAN
CORRECTED TOTAL	11	24.74666667			0.32403703		13.96666667

SOURCE	DF	ANOVA SS	F VALUE	PR > F
GROUP	2	23.80166667	113.34	0.0001

ANALYSIS OF VARIANCE PROCEDURE

E = ERROR SS&CP MATRIX

DF=9	Y1	Y2	Y3	Y4	Y5
Y1	1.47500000	0.70000000	-0.47750000	0.29750000	-0.49750000
Y2	0.70000000	0.95000000	-0.06500000	0.19000000	-0.54000000
Y3	-0.47750000	-0.06500000	1.00750000	0.00750000	-0.21750000
Y4	0.29750000	0.19000000	0.00750000	0.32500000	-0.23000000
Y5	-0.49750000	-0.54000000	-0.21750000	-0.23000000	0.94500000

PARTIAL CORRELATION COEFFICIENTS FROM THE ERROR SS&CP MATRIX / PROB > !R!

DF=8	Y1	Y2	Y3	Y4	Y5
Y1	1.000000	0.591344	-0.391701	0.429684	-0.421387
	0.0000	0.0718	0.2630	0.2152	0.2252
Y2	0.591344	1.000000	-0.066440	0.341940	-0.569923
	0.0718	0.0000	0.8553	0.3335	0.0854
Y3	-0.391701	-0.066440	1.000000	0.013107	-0.222906
	0.2630	0.8553	0.0000	0.9713	0.5359
Y4	0.429684	0.341940	0.013107	1.000000	-0.415021
	0.2152	0.3335	0.9713	0.0000	0.2330
Y5	-0.421387	-0.569923	-0.222906	-0.415021	1.000000
	0.2252	0.0854	0.5359	0.2330	0.0000

ANALYSIS OF VARIANCE PROCEDURE

H = ANOVA SS&CP MATRIX FOR: GROUP

DF=2	Y1	Y2	Y3	Y4	Y5
Y1	51.43166667	30.83000000	50.87416667	40.58583333	31.66083333
Y2	30.83000000	19.76000000	30.14000000	25.03000000	16.63000000
Y3	50.87416667	30.14000000	50.42166667	39.95083333	31.97083333
Y4	40.58583333	25.03000000	39.95083333	32.41166667	23.69666667
Y5	31.66083333	16.63000000	31.97083333	23.69666667	23.80166667

CHARACTERISTIC ROOTS AND VECTORS OF: E INVERSE * H, WHERE H = ANOVA SS&CP MATRIX FOR: GROUP   E = ERROR SS&CP MATRIX

CHARACTERISTIC ROOT	PERCENT	CHARACTERISTIC VECTOR V'EV=1 Y1	Y2	Y3	Y4	Y5
450.21043656	98.63	0.54500665	0.31402293	0.84238821	0.89104337	1.10467913
6.25629712	1.37	-0.59176348	0.48946933	-0.27940395	1.18298042	-0.53870690
0.00000000	0.00	-0.07407905	1.07117428	0.28708306	-1.21733547	0.17646957
0.00000000	0.00	0.26891350	-0.61165940	0.80057146	-0.42590350	-0.58166402
0.00000000	0.00	0.91375709	-0.38162881	-0.03212027	-0.54292847	-0.36515582

```
                              ANALYSIS OF VARIANCE PROCEDURE
              MANOVA TEST CRITERIA FOR THE HYPOTHESIS OF NO OVERALL GROUP EFFECT
                              H = ANOVA SS&CP MATRIX FOR: GROUP
                              E = ERROR SS&CP MATRIX
                              P = DEP. VARIABLES =        5
                              Q = RANK OF H      =        2
                              NE= DF OF E        =        9
                              S = MIN(P,Q)       =        2
                              M = .5(ABS(P-Q)-1) =      1.0
                              N = .5(NE-P-1)     =      1.5
---------------------------------------------------------------------------------------------
HOTELLING-LAWLEY TRACE = TR(E**-1*H) =          456.46673368     (SEE PILLAI'S TABLE #3)
      F APPROXIMATION = 2(S*N+1)*TR(E**-1*H)/(S*S*(2M+S+1))     WITH S(2M+S+1) AND 2(S*N+1) DF
            F(10,6) =    182.59      PROB > F = 0.0001
---------------------------------------------------------------------------------------------
PILLAI'S TRACE            V = TR(H*INV(H+E)) =    1.85997240     (SEE PILLAI'S TABLE #2)
      F APPROXIMATION = (2N+S+1)/(2M+S+1) * V/(S-V)             WITH S(2M+S+1) AND S(2N+S+1) DF
            F(10,12) =    15.94      PROB > F = 0.0001
---------------------------------------------------------------------------------------------
WILKS' CRITERION          L = DET(E)/DET(H+E) =   0.00030543     (SEE BKA VOL 53 P 347)
      X = -(NE-.5(P-Q+1))*LN(L)          =       56.6566
---------------------------------------------------------------------------------------------
ROY'S MAXIMUM ROOT CRITERION =                   450.21043656     (SEE AMS VOL 31 P 625)
      FIRST CANONICAL VARIABLE YIELDS AN F UPPER BOUND
            F(2,9) =   2025.95      (UPPER BOUND)
---------------------------------------------------------------------------------------------
```

See Charts in Morrison

Significant at 0.01 level

31

First, all univariate F-statistics are given comparing function values for separate frequencies. In addition, the $\hat{\Sigma}$ and H matrices are given. The Hotelling's T^2 value is given for the case of two functions and the largest eigenvalue of $E^{-1}H$ is given for the case of three functions. In both cases, the test statistic is significant. These examples are shown in the Appendix.

Recommendation of Experimental Design Through Power Study

An extensive Power Study was made in which the power of the test was computed for varying number of frequencies, number of replicates, correlation structure, and assigned differences between the function means. A set of plots are provided in order to give the reader an indication of what magnitude of power can be expected for various combinations of the parameters and to devise recommendations of what are reasonable values for the number of replicates and frequencies, given a particular correlation structure. It should be emphasized that regardless of power considerations there is a certain restriction that must hold. For the case of two functions, if we call n the number of replicates for each frequency-function combination, the denominator degrees of freedom for the F statistic will be non-positive unless

$$2n > t - 1 \qquad (5)$$

Thus equation (5) represents a necessary restriction in the case of two functions. In the general case it would be adviseable to keep the same restriction, i.e., that given by equation (5) because it is unlikely that the quality of the test (i.e., power) will be good at all unless n satisfies equation (5). In addition since is is recommended to compute the power on the basis of Hotelling's T^2 (tests on two functions) one cannot assess the power unless equation (5) holds.

We attempt here to give some rationale regarding our power study. The most difficult variable to cope with was the covariance structure. Keep in mind that the correlations represent association between observations from one frequency to another within functions. The following covariance or correlation structures were studied:

(a) Low correlation, correlation constant (0.1, 0.2)

(b) High correlation, (0.6-0.9)

(c) Mixed correlation between high and low (0.1-0.95) .

In addition, the power (through the noncentrality parameter) is a function of $\underline{\mu}_1 - \underline{\mu}_2$, the difference between the means of the two functions. Here, we are assuming that there is a <u>typical difference</u> which is a <u>proportion of σ</u>, where σ is the same at each frequency. Proportions of 0.3, 0.5, 0.7, and 1.0 are used in this study. We feel that it is reasonable to assume that 0.7σ - 1.0σ is a reasonable range in the difference between the two functions which one would like to detect. It is clear from the plots that for any value less than 0.7σ, experimentation is much too demanding if one wishes to obtain high power. The significance level of the test was fixed at $\alpha = 0.10$. Plots of power against frequency are shown in the Appendix for various correlation structures. Curves are shown for various proportions of σ which one is attempting to detect.

The purpose of the plots is to attempt to show the role of t, the number of frequencies and n, the number of replicates, on the power. The plots indicate the following:

(a) All correlations equal - low correlations (0.1, 0.2). The role of t, the number of frequencies is displayed dramatically in the plots shown. As t grows large, the power is reduced, and for $t > 10$ the only condition under which power is moderately good is when $n > 10$. Thus one can assume that the test will be ineffective unless $t < 10$; but if t must exceed 10, the number of replicates should be 10-12.

(b) High correlation (0.7-0.9). Here it is very clear that despite the number of frequencies, in order to obtain any effectiveness in the test, at least 12 replicates should be used. Of course, it is probably an unusual experimental situation for which all correlations would be high.

(c) Correlations mixed between high and low. This condition likely represents the most typical experimental condition that one confronts. Again, however, it appears as if small numbers of replicates will not give power values that are acceptable if one desires a quality test. For 8-10 frequencies, $n = 10$ is barely acceptable whereas for $t > 10$, a sample size of at least 12 is necessary.

Conclusions

A test procedure is given for testing equality of k different pilot describing functions. The method for handling the power for the case of two functions is described. It is felt that this is not an unreasonable assessment of power as long as the number of functions is four or less. For rather typical types of correlation structures a power study is made in order to determine reasonable number of replicates and frequencies. Two other recommendations come to light, in addition to those involving the experimental design. These are given in the following two paragraphs.

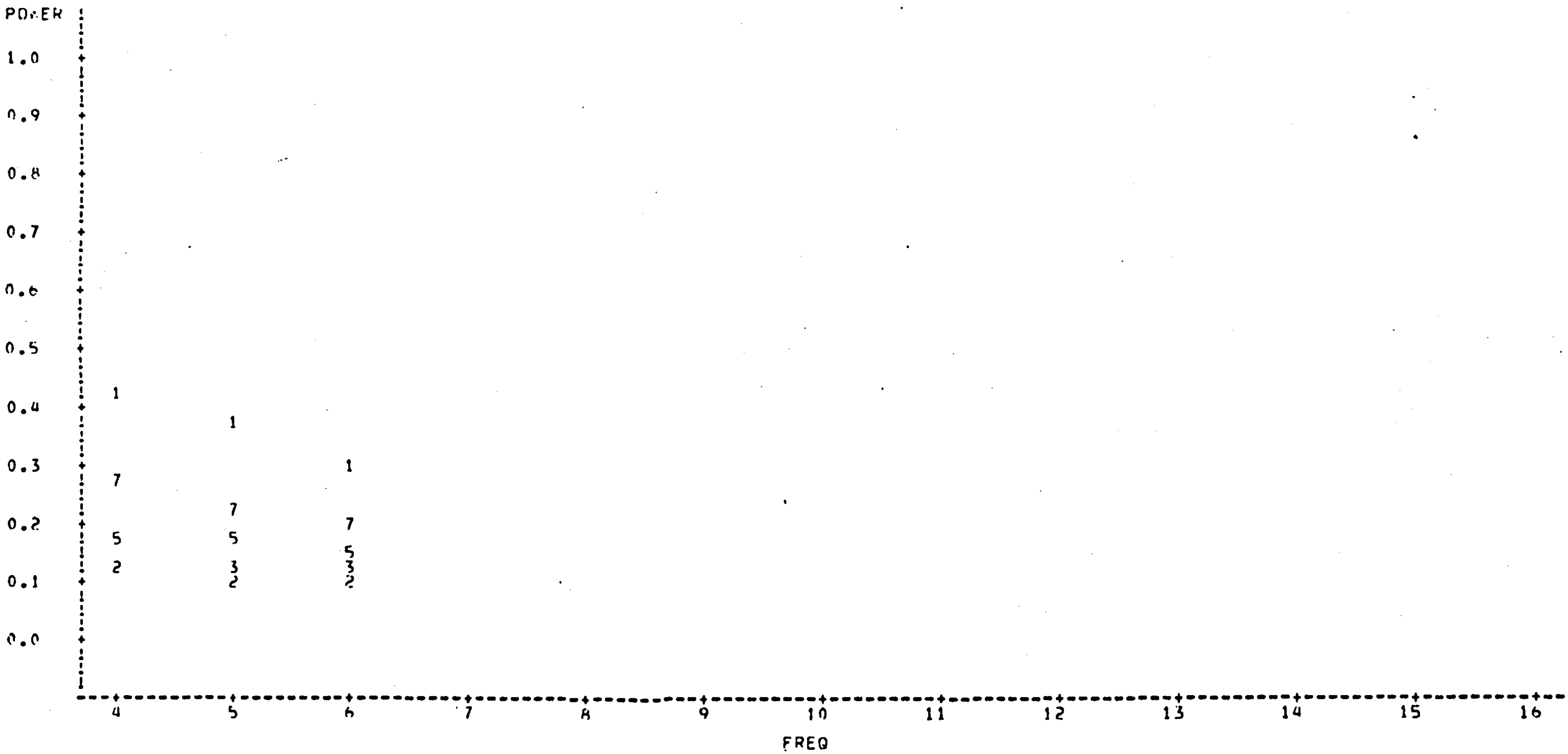
In using the techniques described here the NASA Langley personnel should be very careful to sample to determine what correlation category prevails. The power is very much dependent on the correlation. In fact, if possible it might be adviseable to take data initially to determine the approximate correlation structure, then supplement with the required data according to the experimental design recommendations.

The user should not be attached to a particular significance level. We used $\alpha = 0.10$ for the power study. However, because of the inherent lack of power in tests such as these one should be prepared to consider that a test which is significant at the 0.15 or even 0.20 level is evidence in favor of differences between functions. This is particularly true if one is forced to use a large number of frequencies.

CORRELATION = .1

N1=5

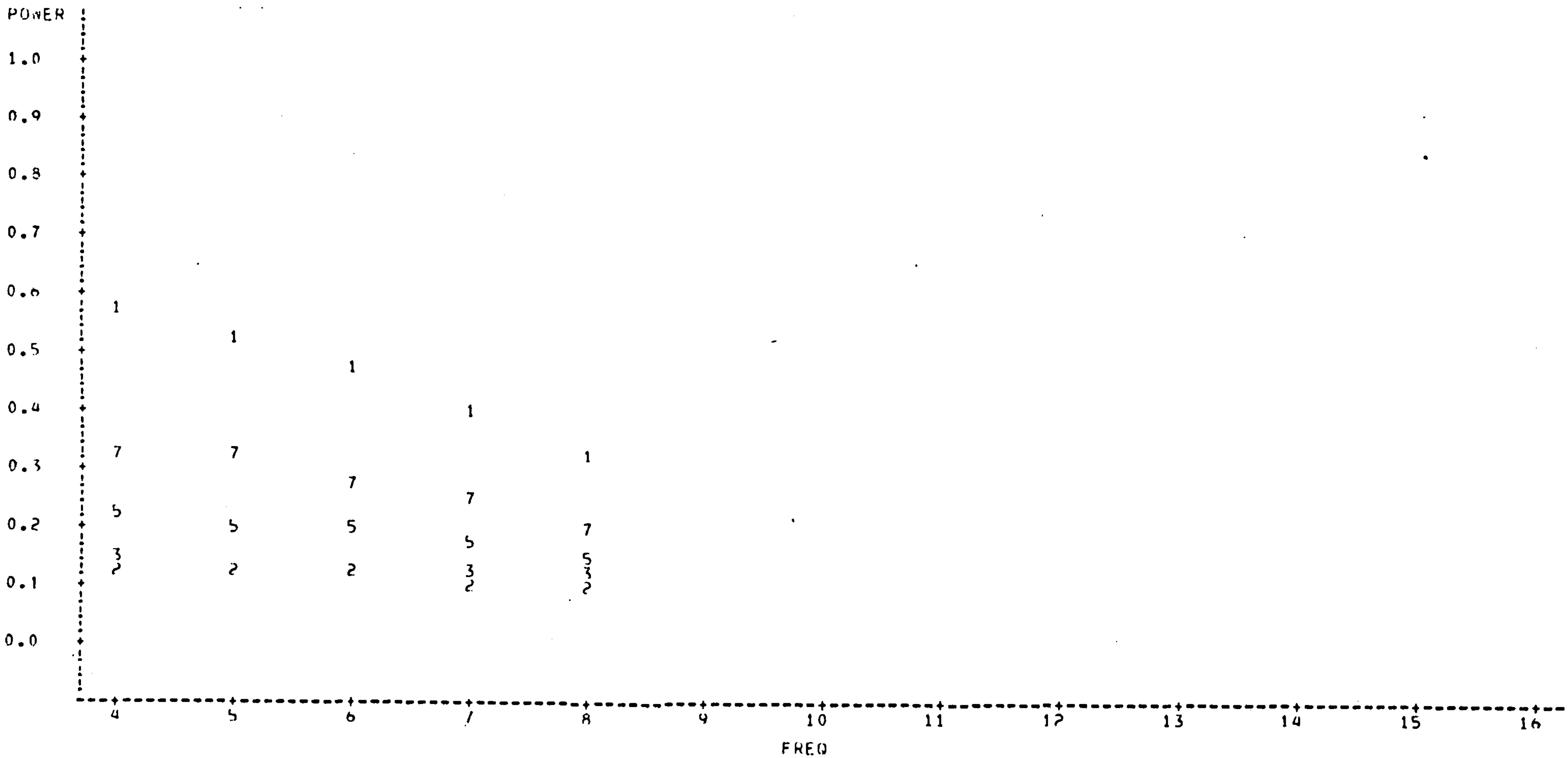
PLOT OF POWER*FREQ SYMBOL IS VALUE OF PC



NOTE: 1 OBS HIDDEN

N1=6

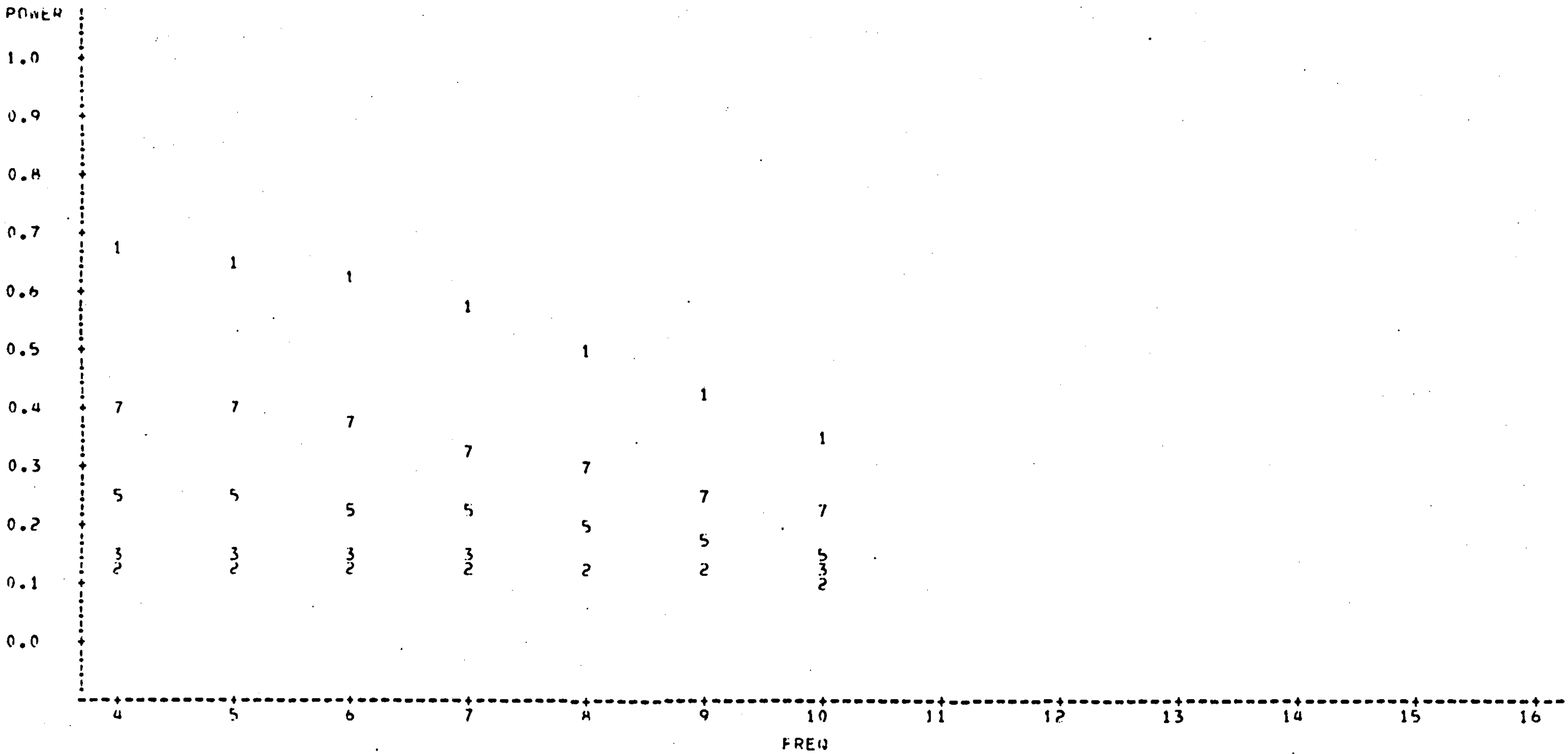
PLOT OF POWER*FREQ   SYMBOL IS VALUE OF PC



NOTE: 2 OBS HIDDEN

N1=7

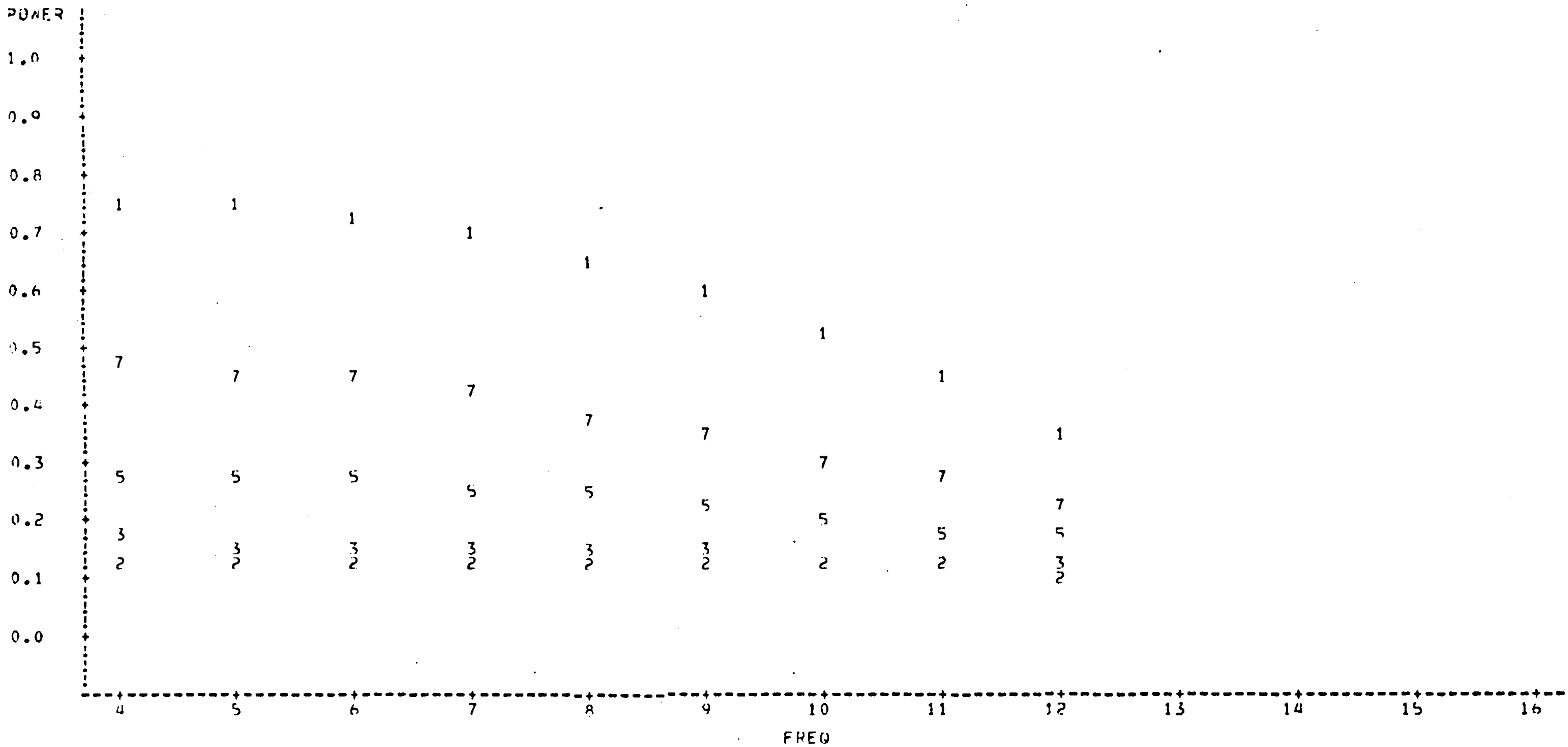
PLOT OF POWER*FREQ SYMBOL IS VALUE OF PC



NOTE: 2 OBS HIDDEN

N1=8

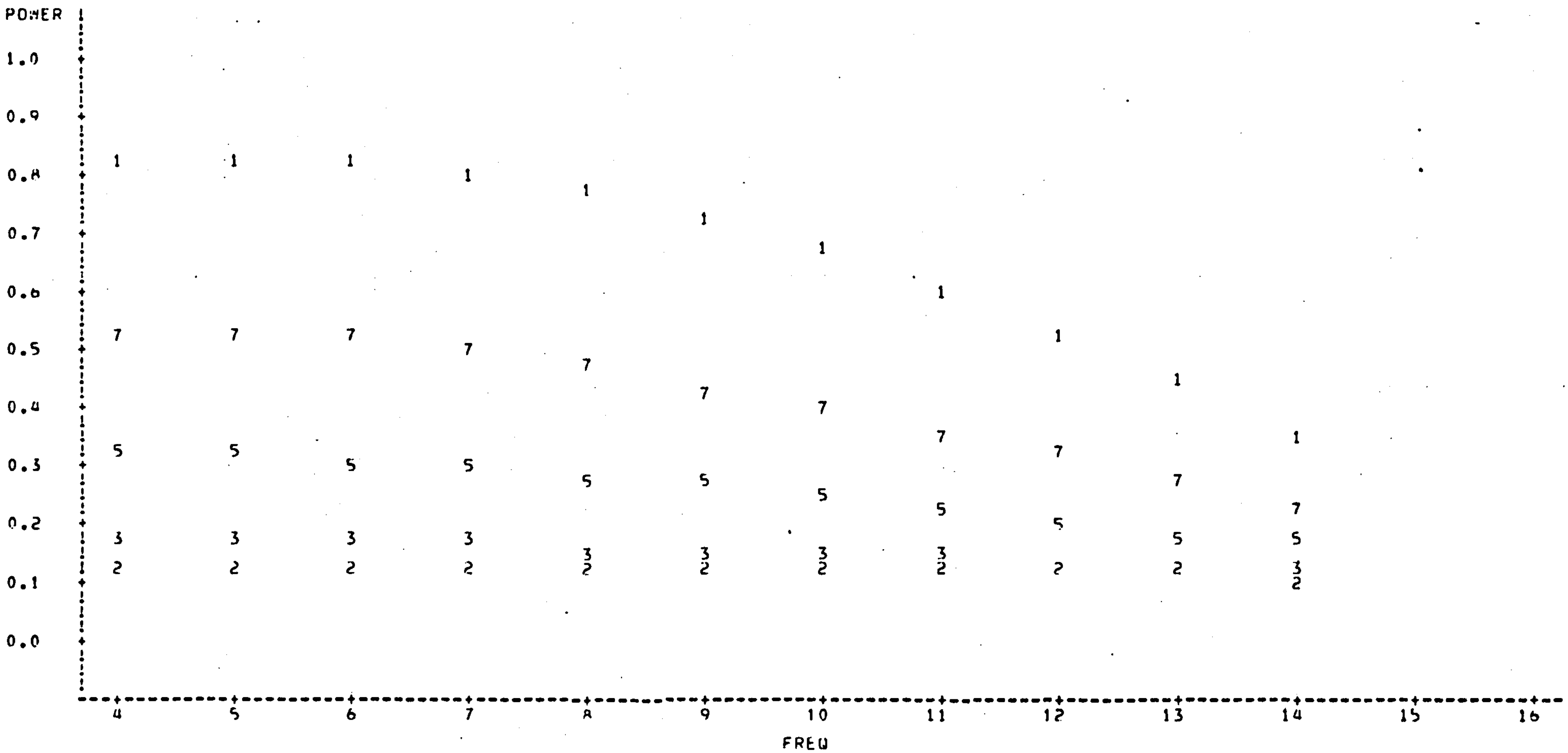
PLOT OF POWER*FREQ   SYMBOL IS VALUE OF PC



NOTE: 2 OBS HIDDEN

NJ=9

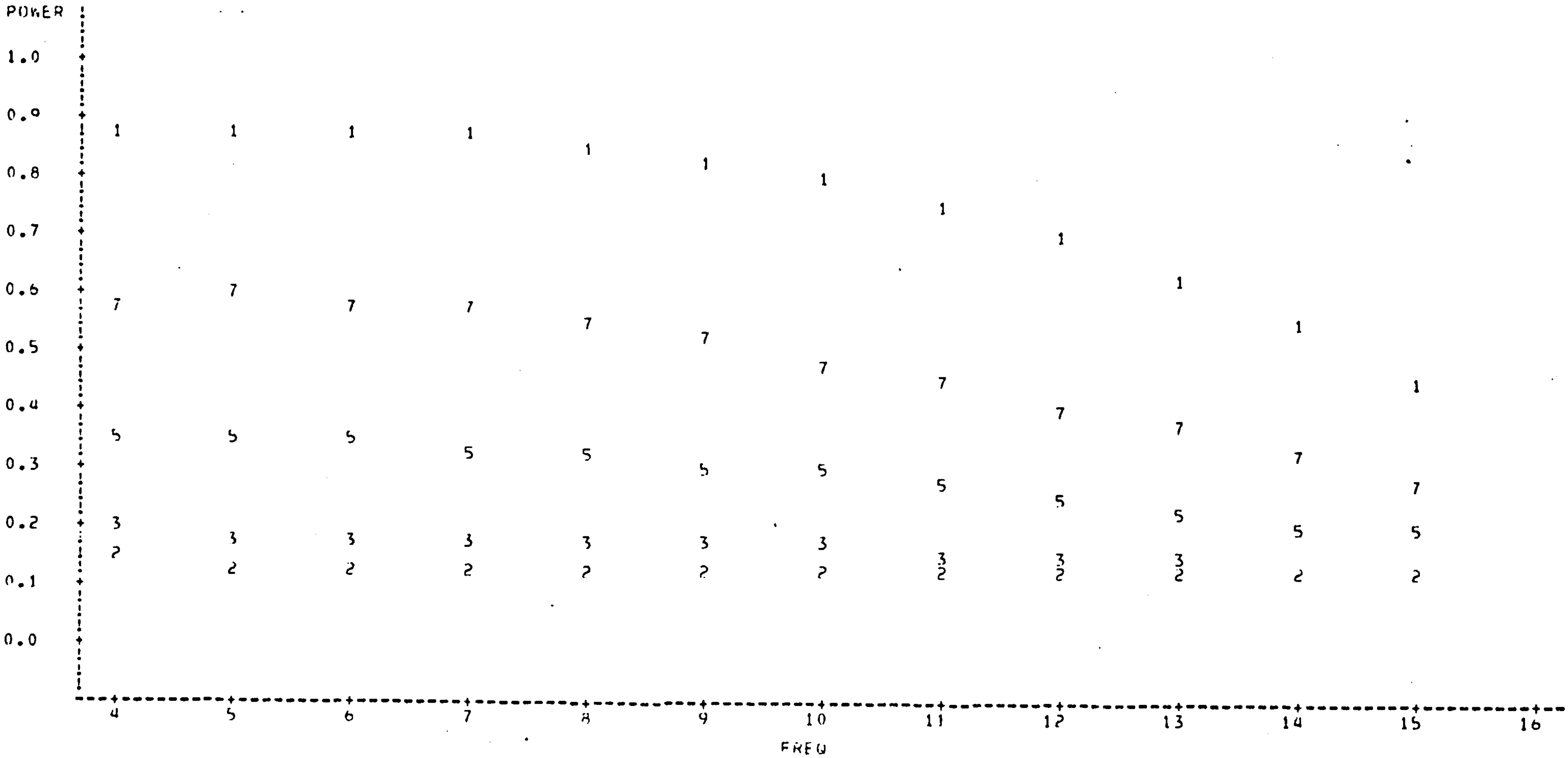
PLOT OF POWER*FREQ   SYMBOL IS VALUE OF PC



NOTE: 2 OBS HIDDEN

N1=10

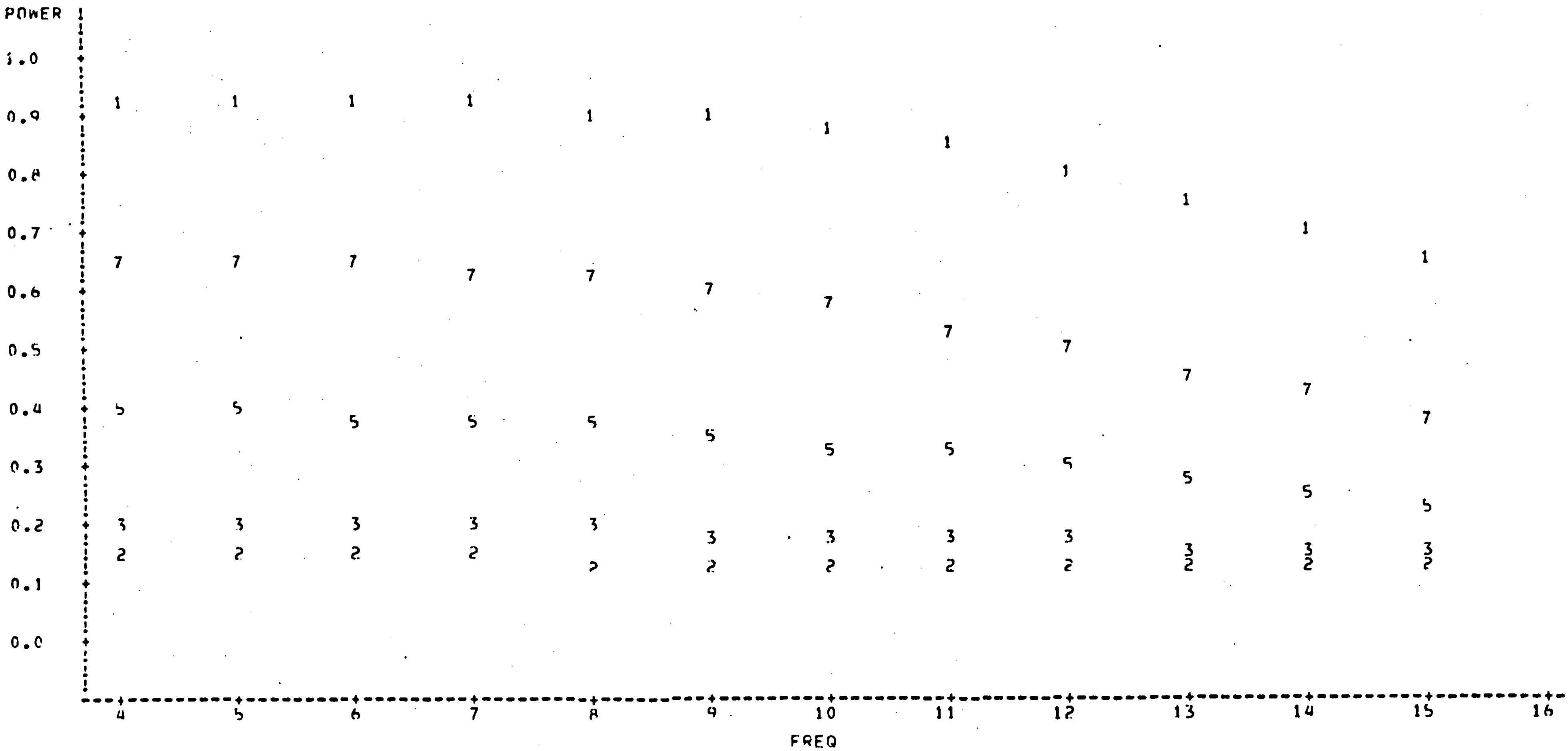
PLOT OF POWER*FREQ   SYMBOL IS VALUE OF PC



NOTE: 2 OBS HIDDEN

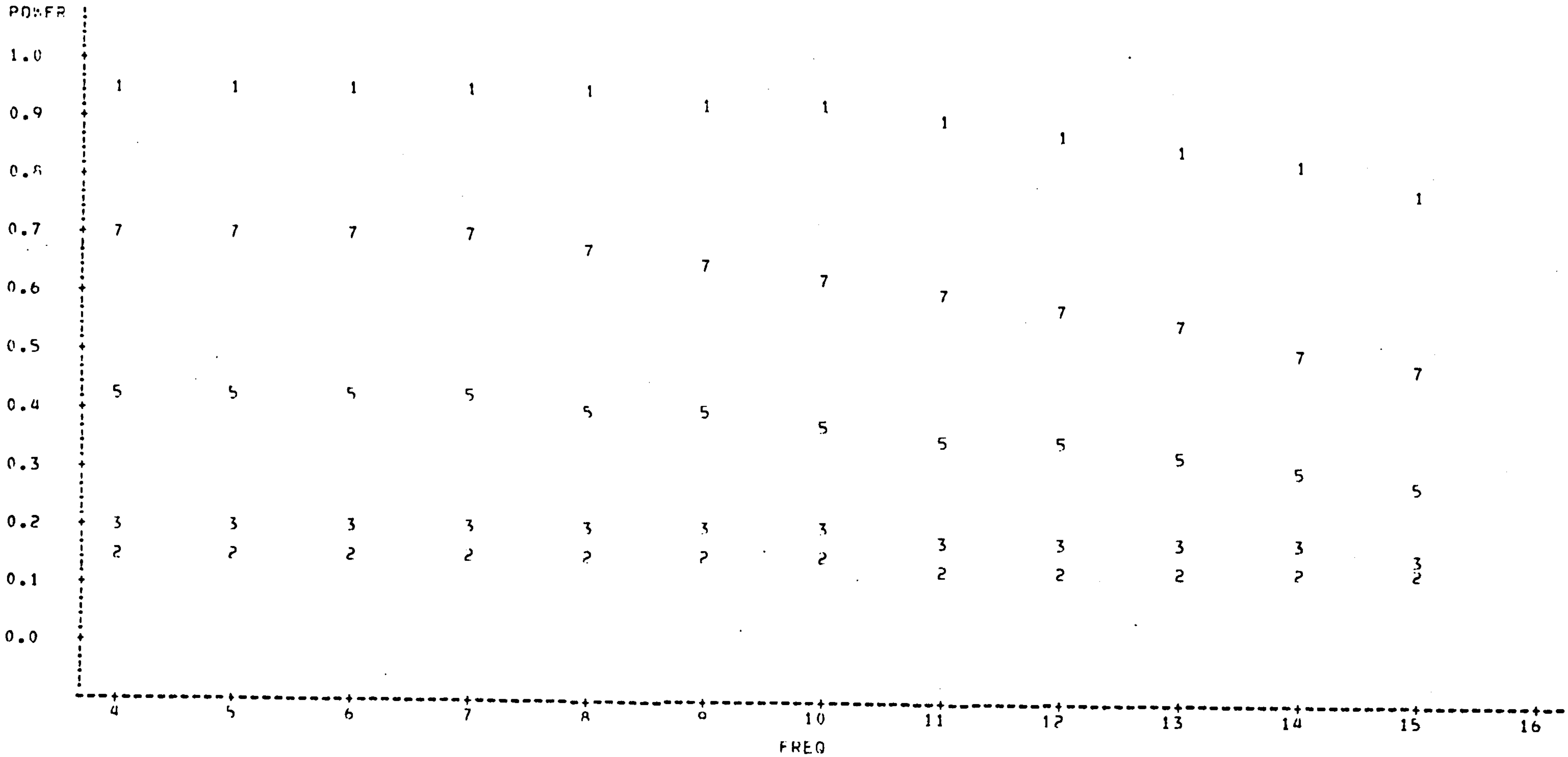
N1=11

PLOT OF POWER*FREQ   SYMBOL IS VALUE OF PC



N1=12

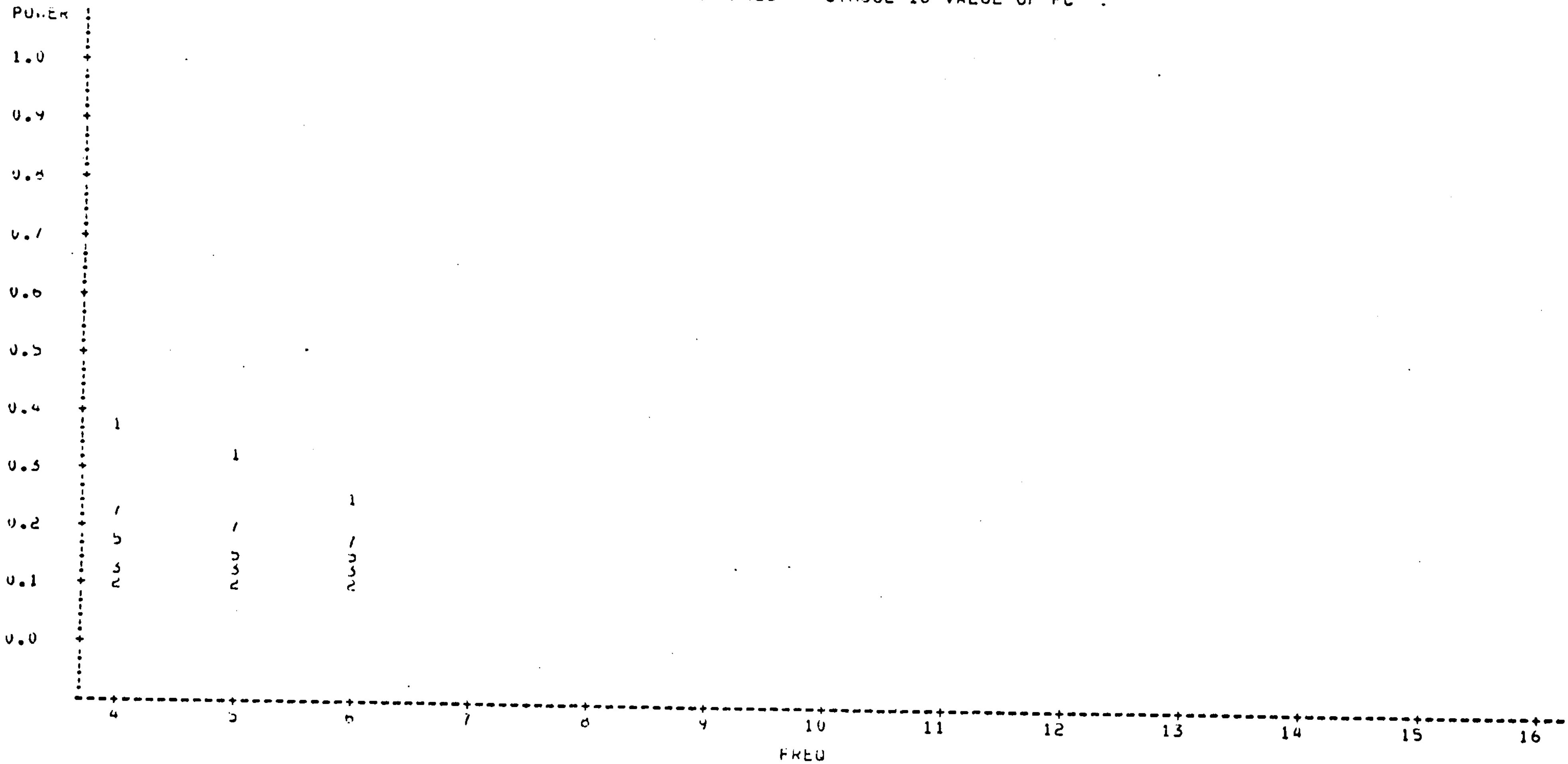
PLOT OF POWER*FREQ SYMBOL IS VALUE OF PCT



CORRELATION = .2

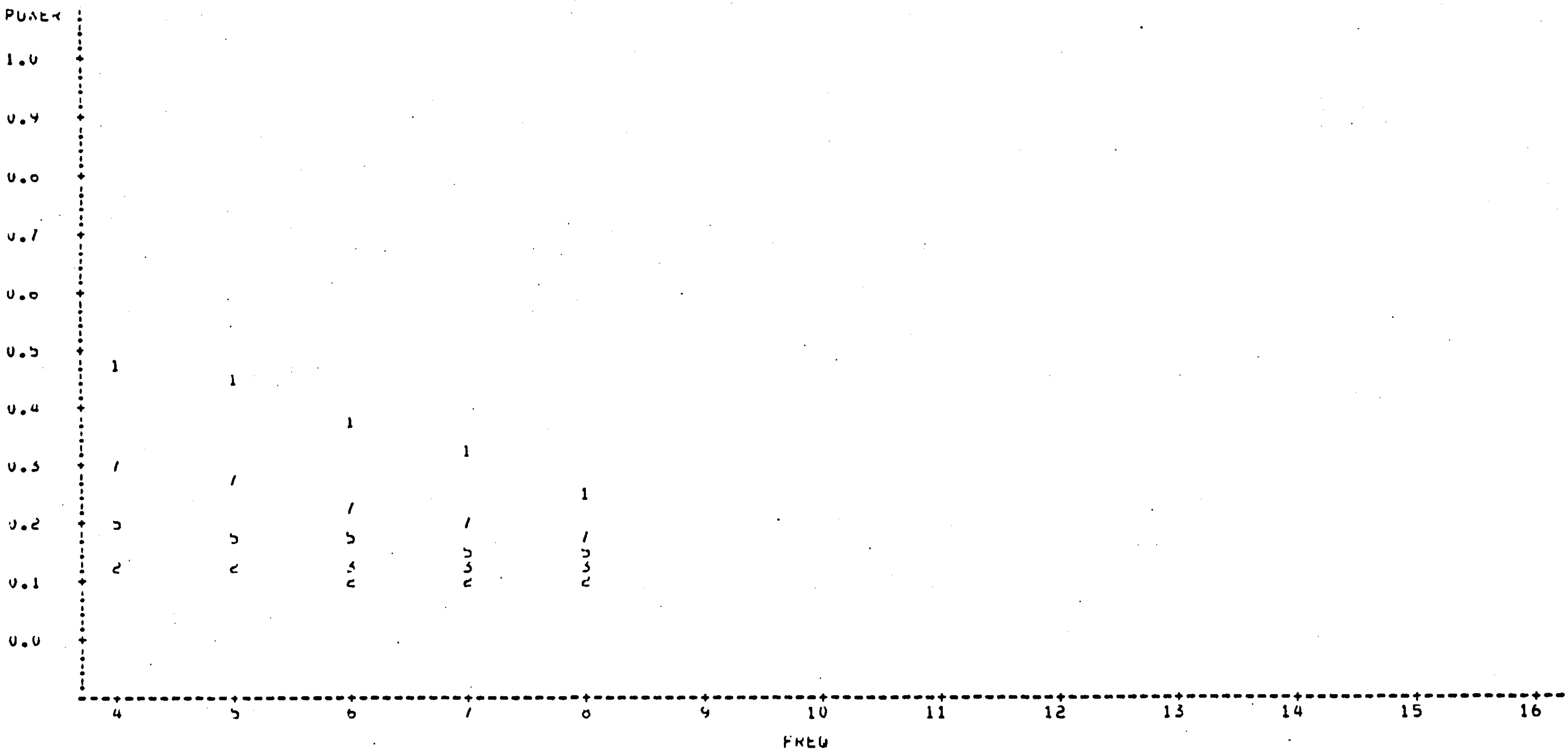
N1=5

PLOT OF POWER*FREQ SYMBOL IS VALUE OF PC



N1=6

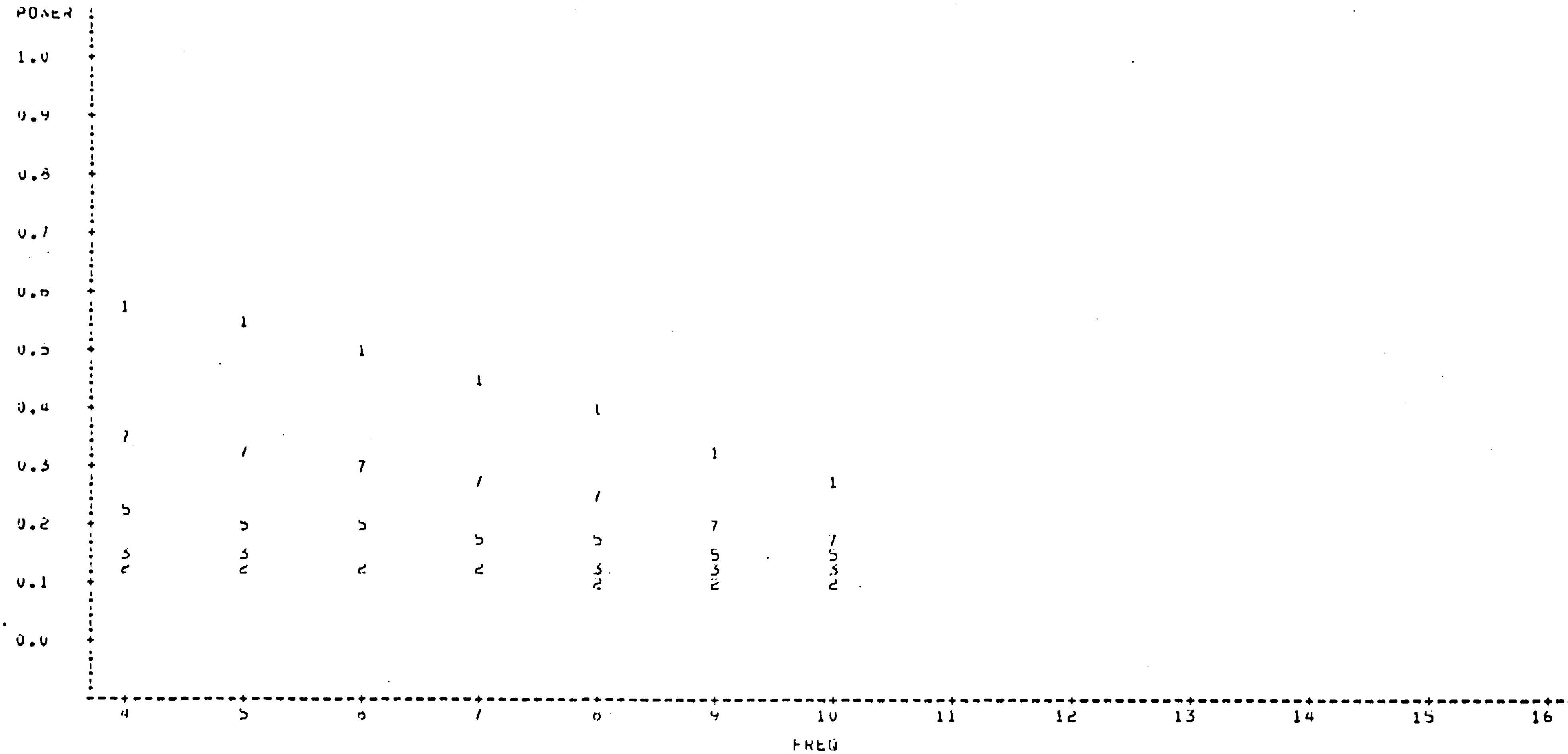
PLOT OF POWER*FREQ   SYMBOL IS VALUE OF PC



NOTE: 2 OBS HIDDEN

N1=7

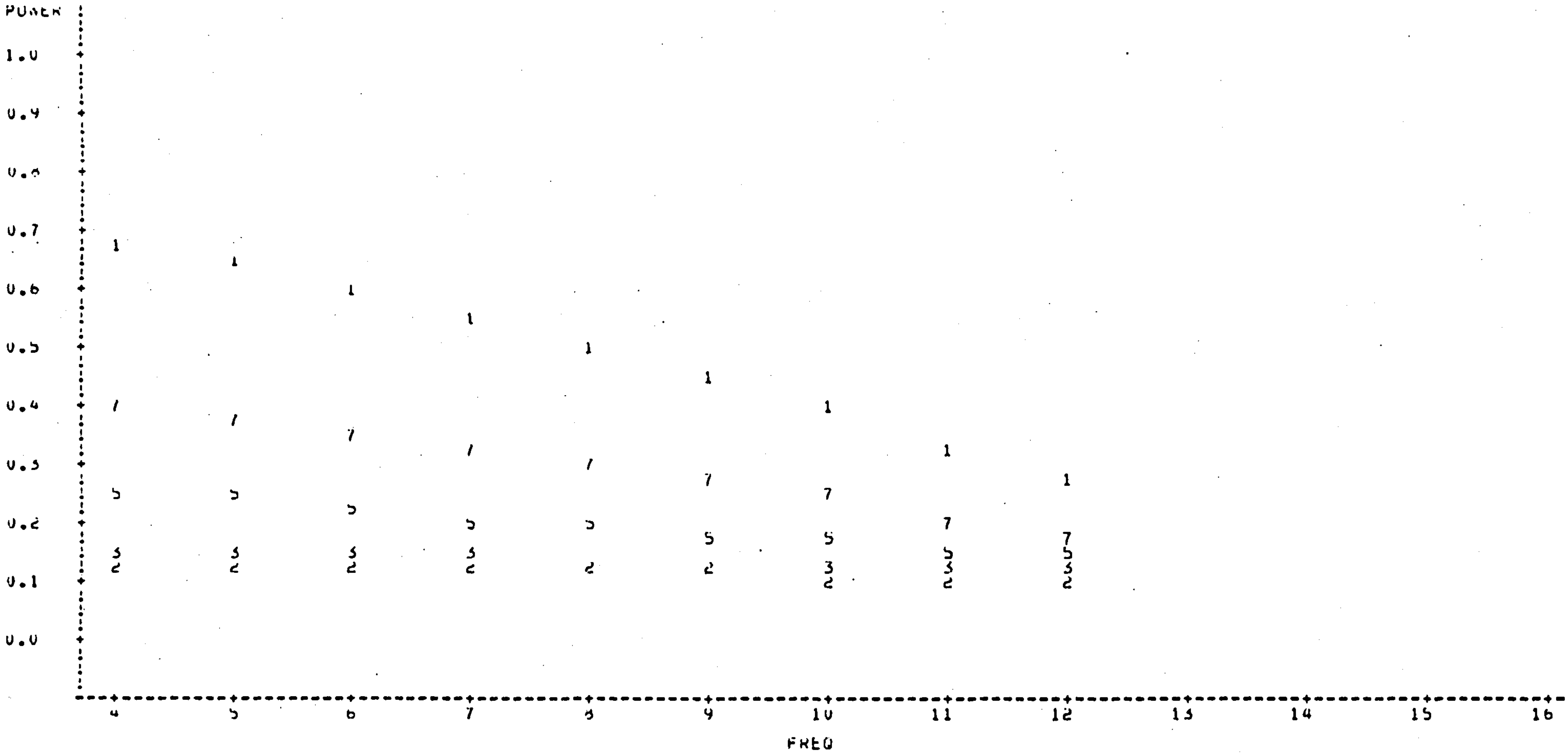
PLOT OF POWER*FREQ SYMBOL IS VALUE OF PC



NOTE: 2 OBS HIDDEN

N1=8

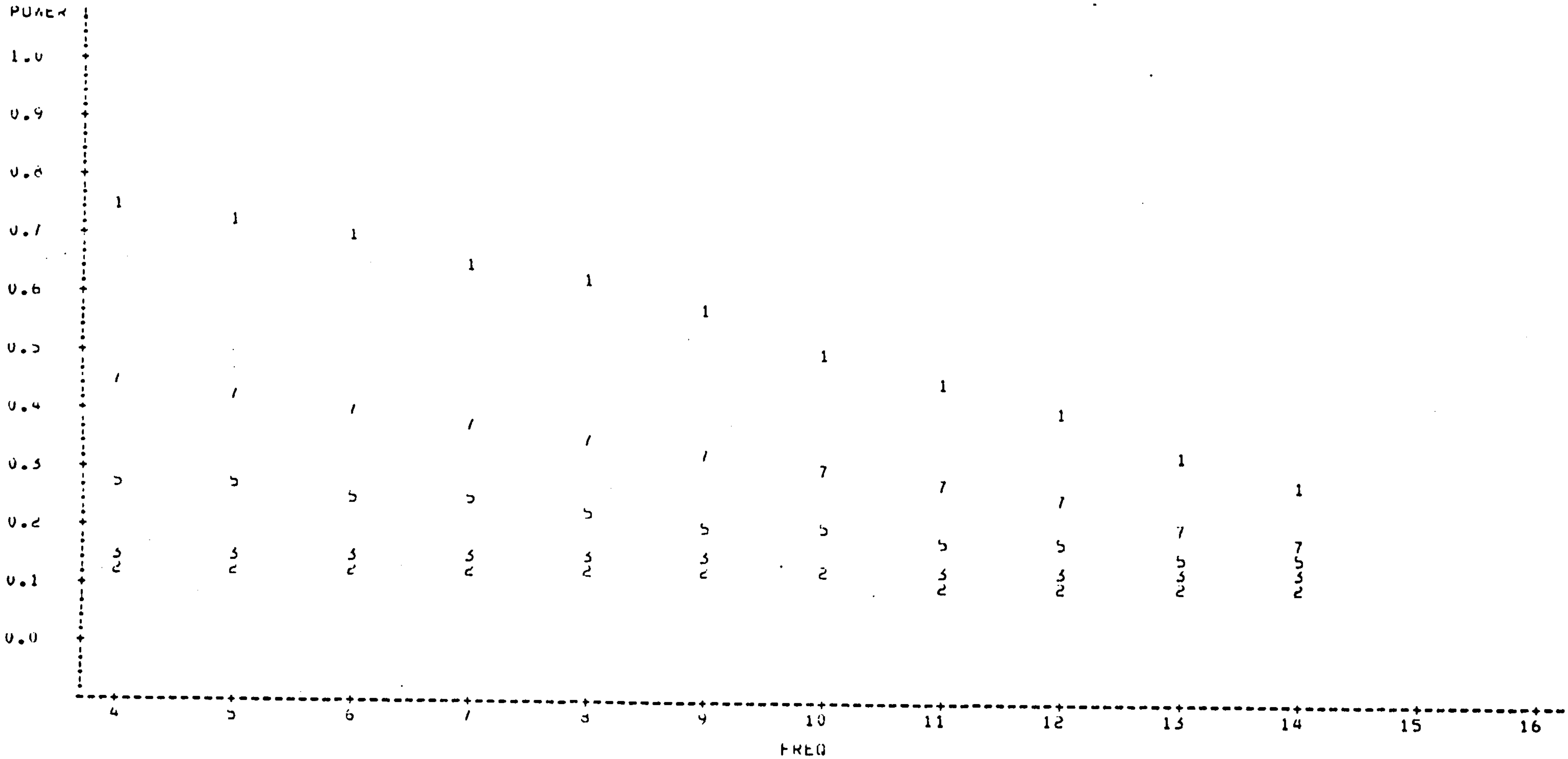
PLOT OF POWER*FREQ   SYMBOL IS VALUE OF PC



NOTE: 2 OBS HIDDEN

N1=9

PLOT OF POWER*FREQ   SYMBOL IS VALUE OF PC

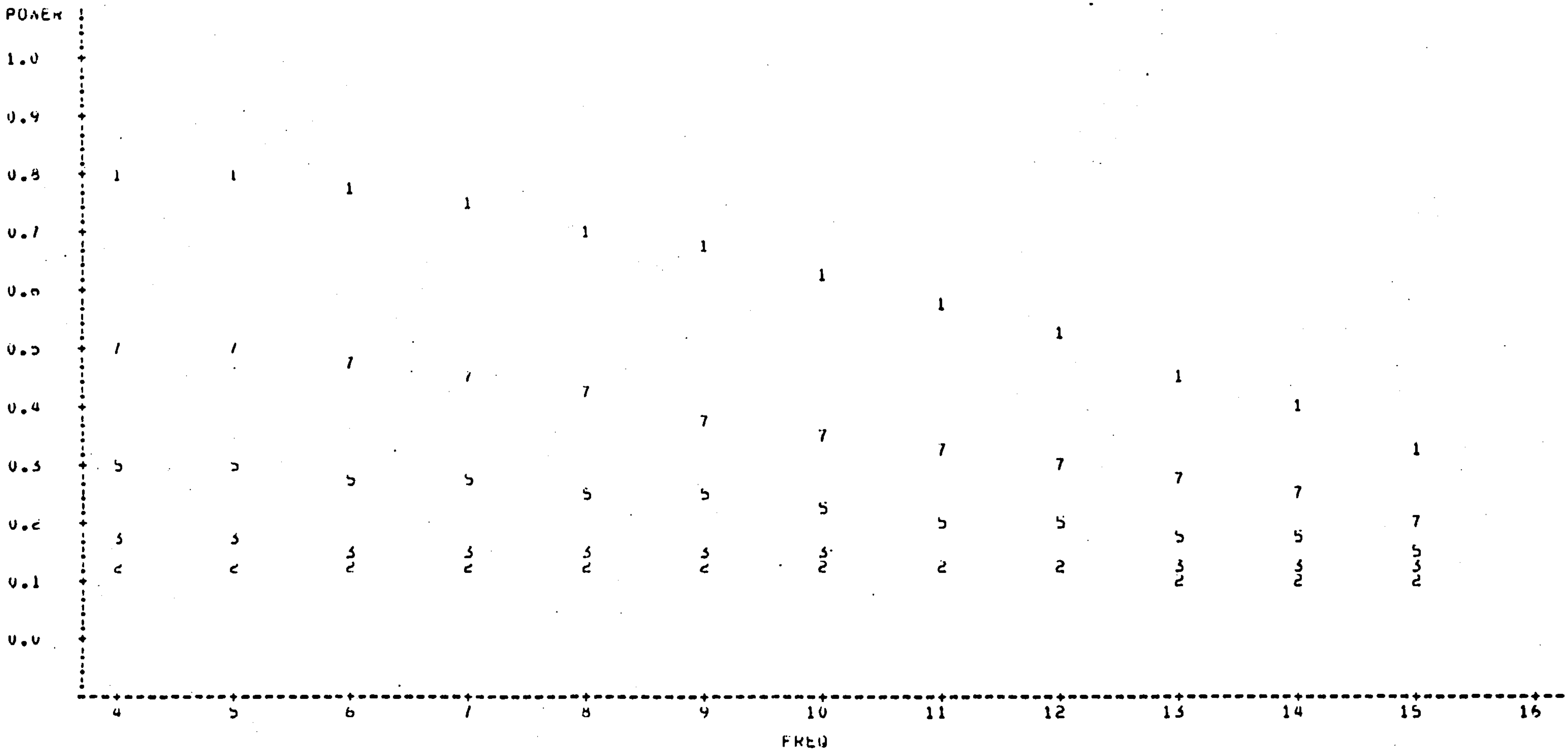


NOTE: 1 OBS HIDDEN

S T A T I S T I C A L   A N A L Y S I S   S Y S T E M   14:22 MONDAY, APRIL 3, 1978   22

N1=10

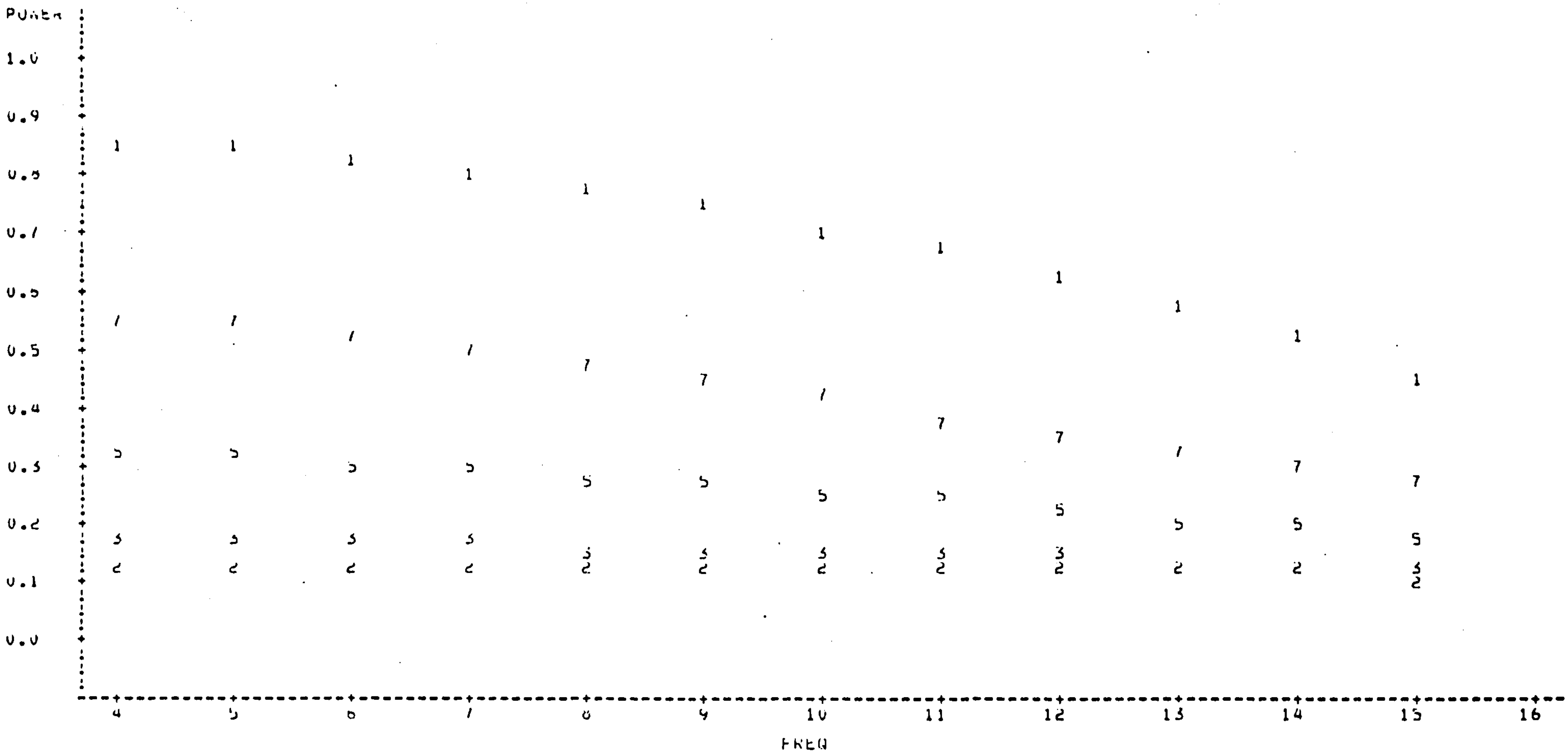
PLOT OF POWER*FREQ   SYMBOL IS VALUE OF PC



NOTE: 2 OBS HIDDEN

N1=11

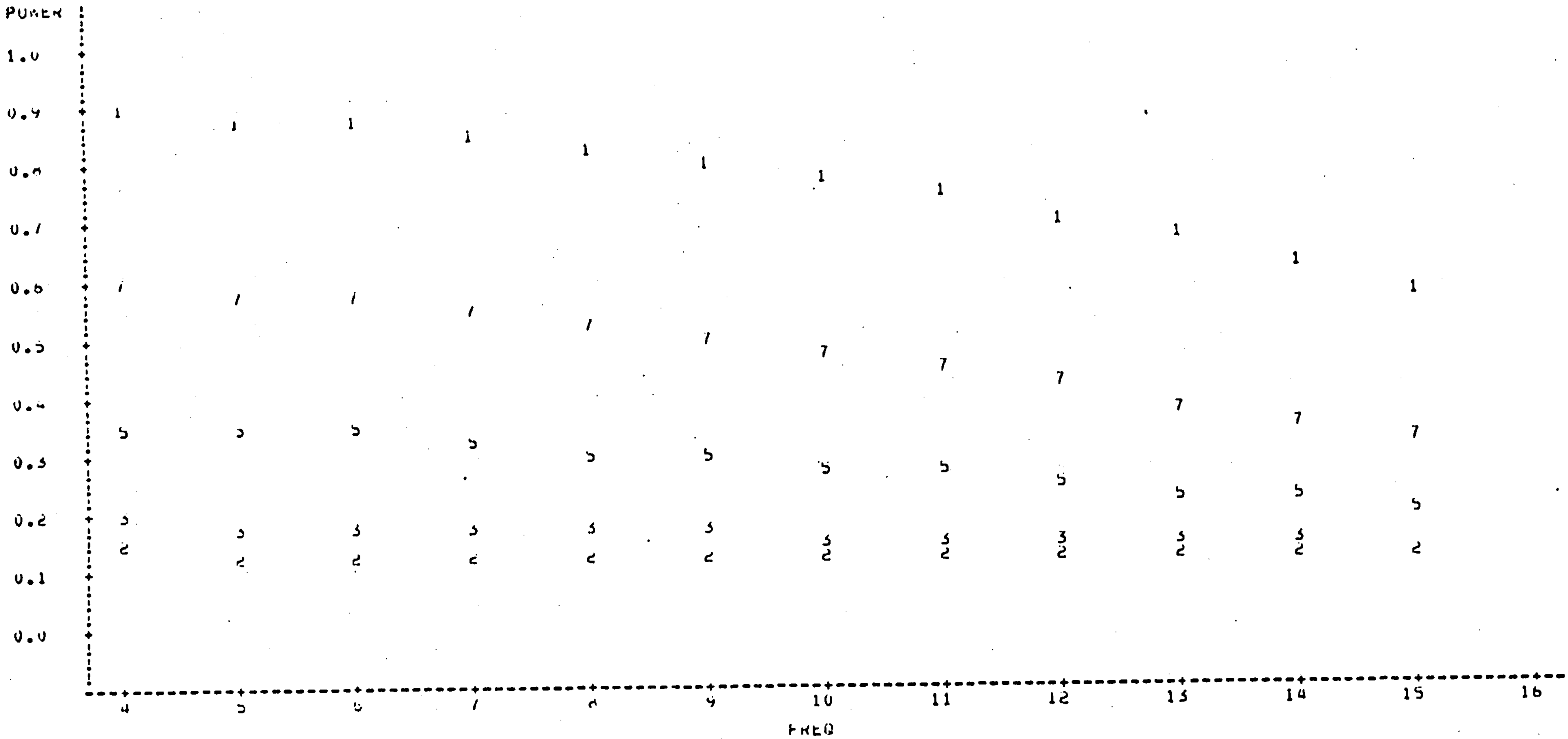
PLOT OF POWER*FREQ SYMBOL IS VALUE OF PC



NOTE: 2 OBS HIDDEN

N1=12

PLOT OF POWER*FREQ   SYMBOL IS VALUE OF PC



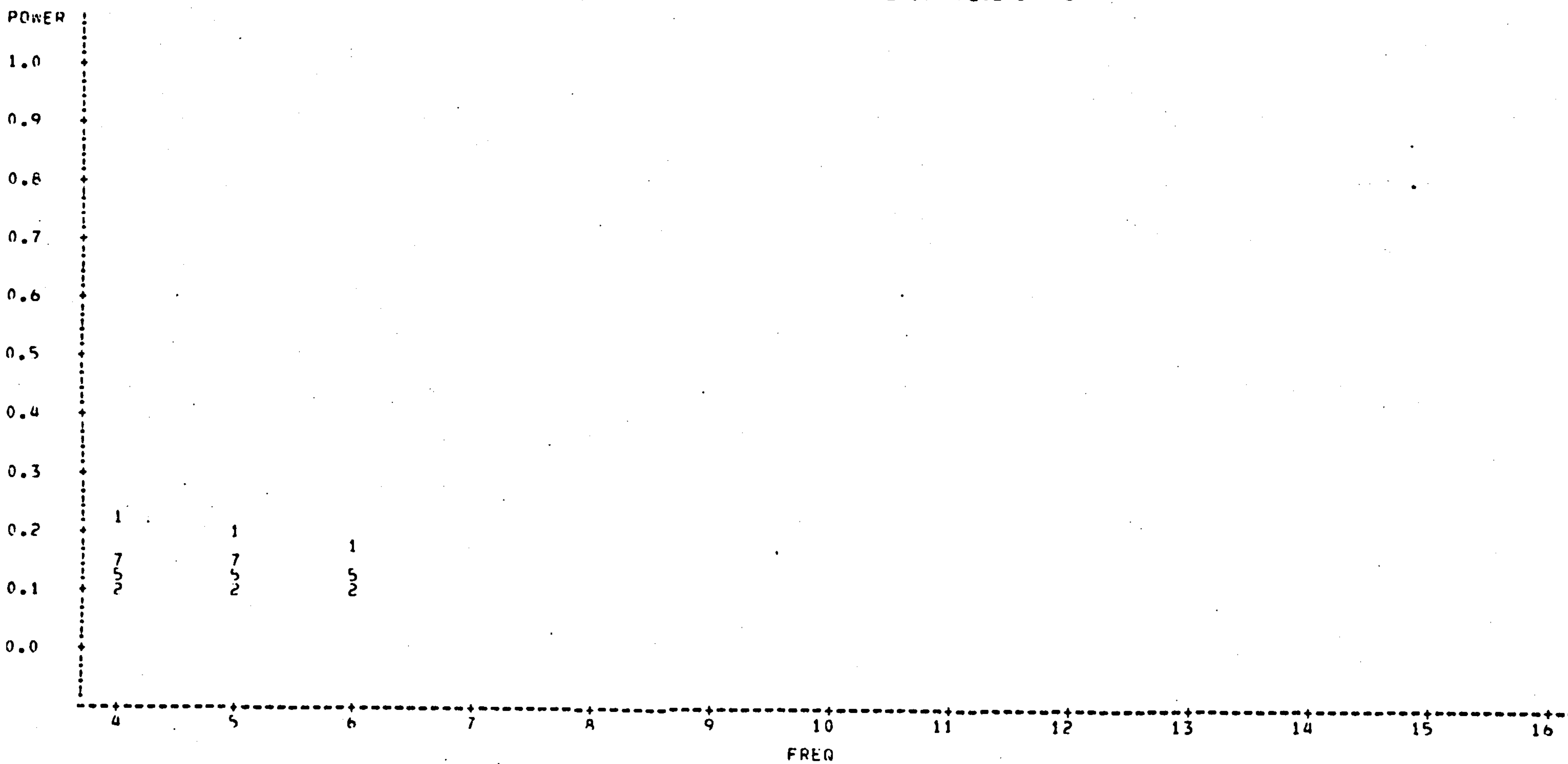
NOTE: 1 OBS HIDDEN

53

HIGH CORRELATION

N1=5

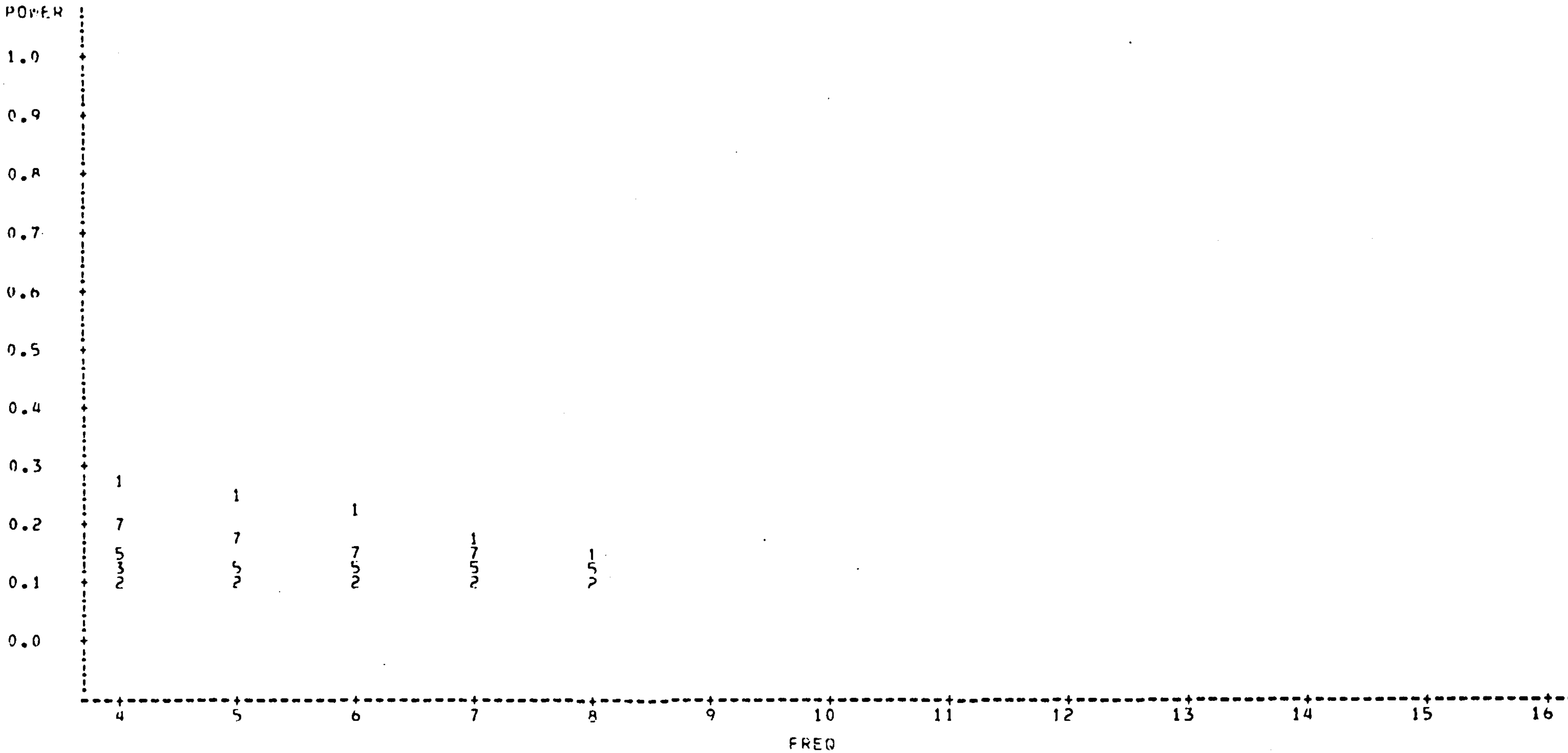
PLOT OF POWER*FREQ   SYMBOL IS VALUE OF PC



NOTE: 4 OBS HIDDEN

N1=6

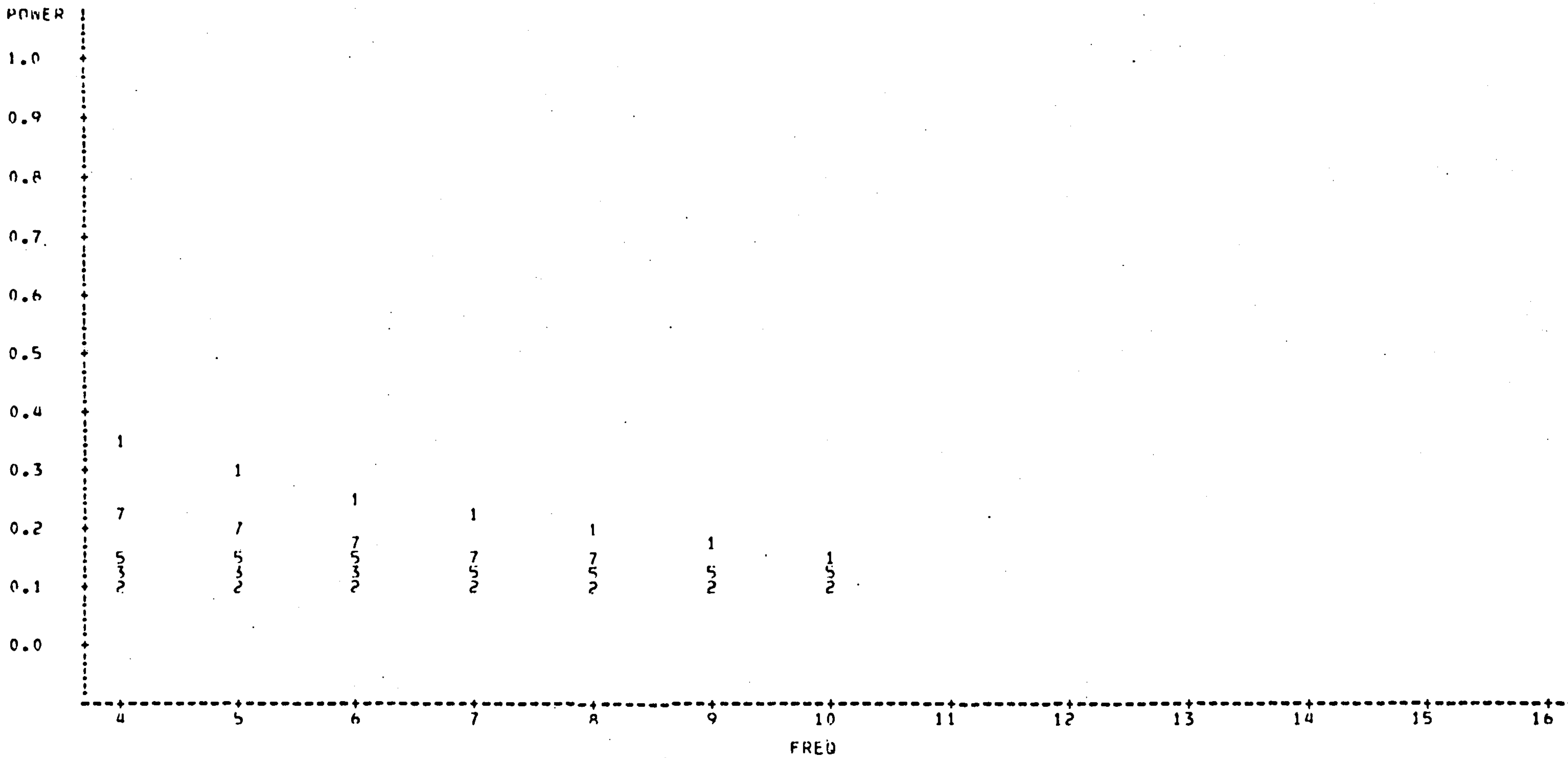
PLOT OF POWER*FREQ   SYMBOL IS VALUE OF PC



NOTE: 5 OBS HIDDEN

N1=7

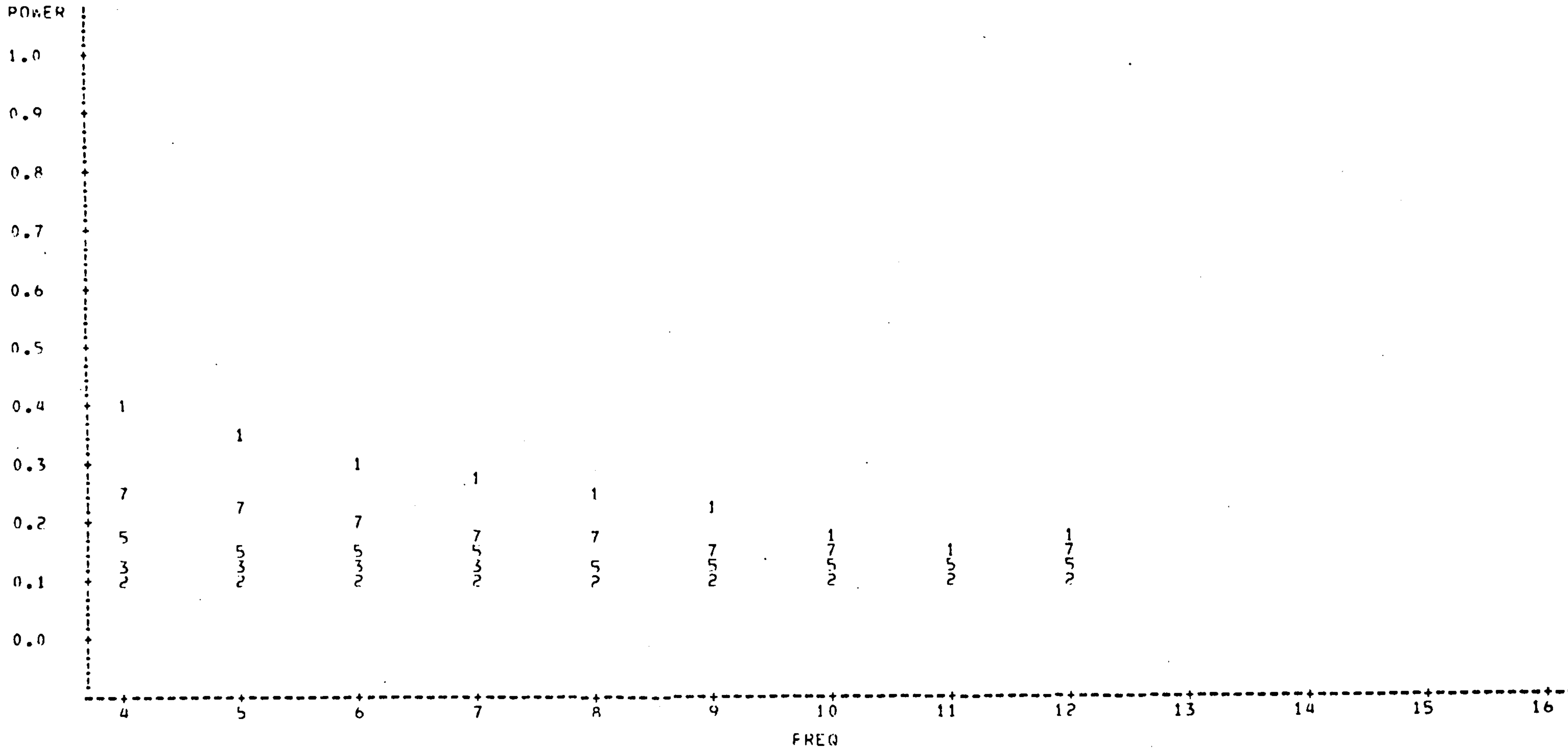
PLOT OF POWER*FREQ   SYMBOL IS VALUE OF PC.



NOTE: 6 OBS HIDDEN

N1=8

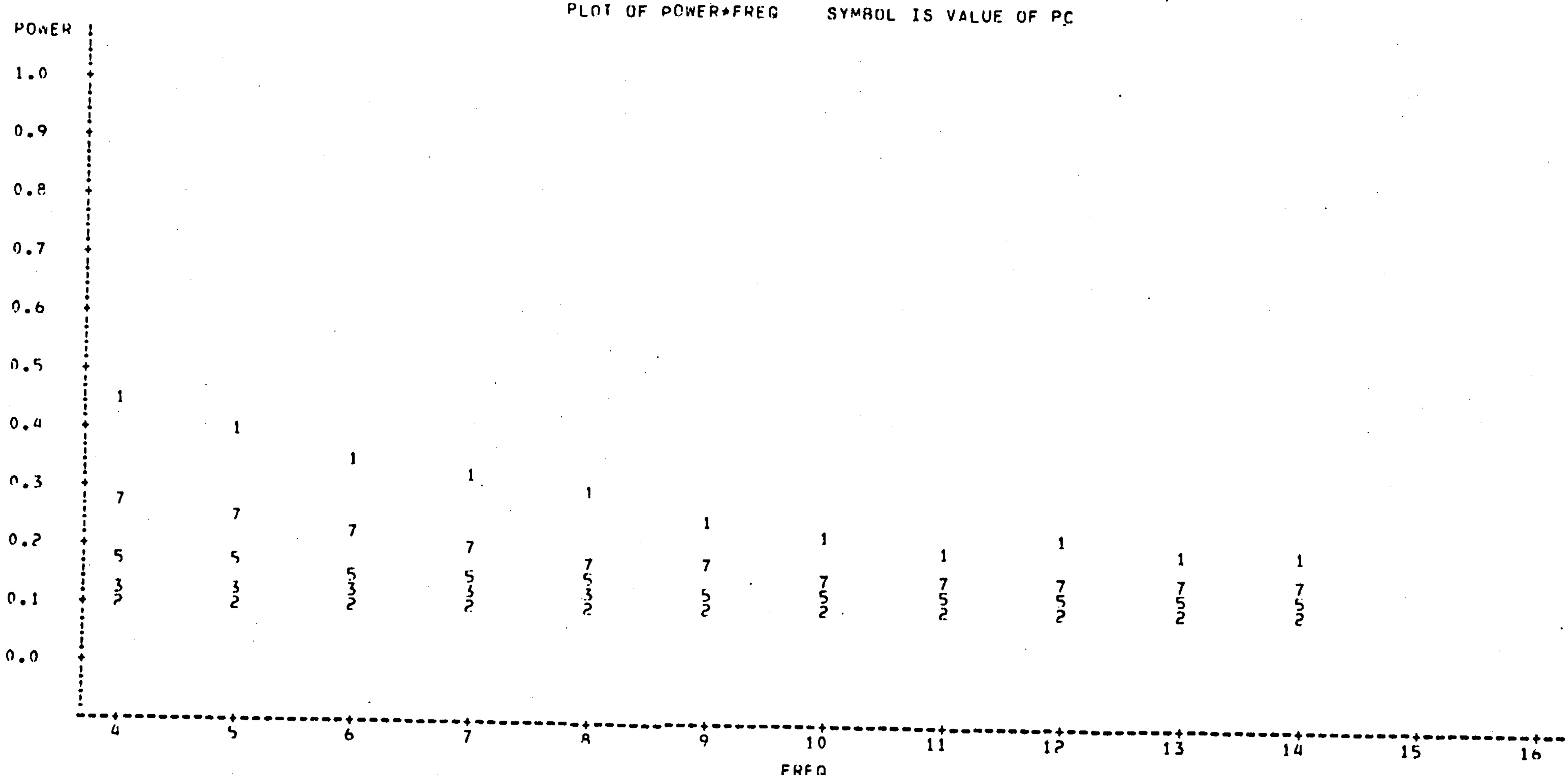
PLOT OF POWER*FREQ   SYMBOL IS VALUE OF PC



NOTE: 6 OBS HIDDEN

N1=9

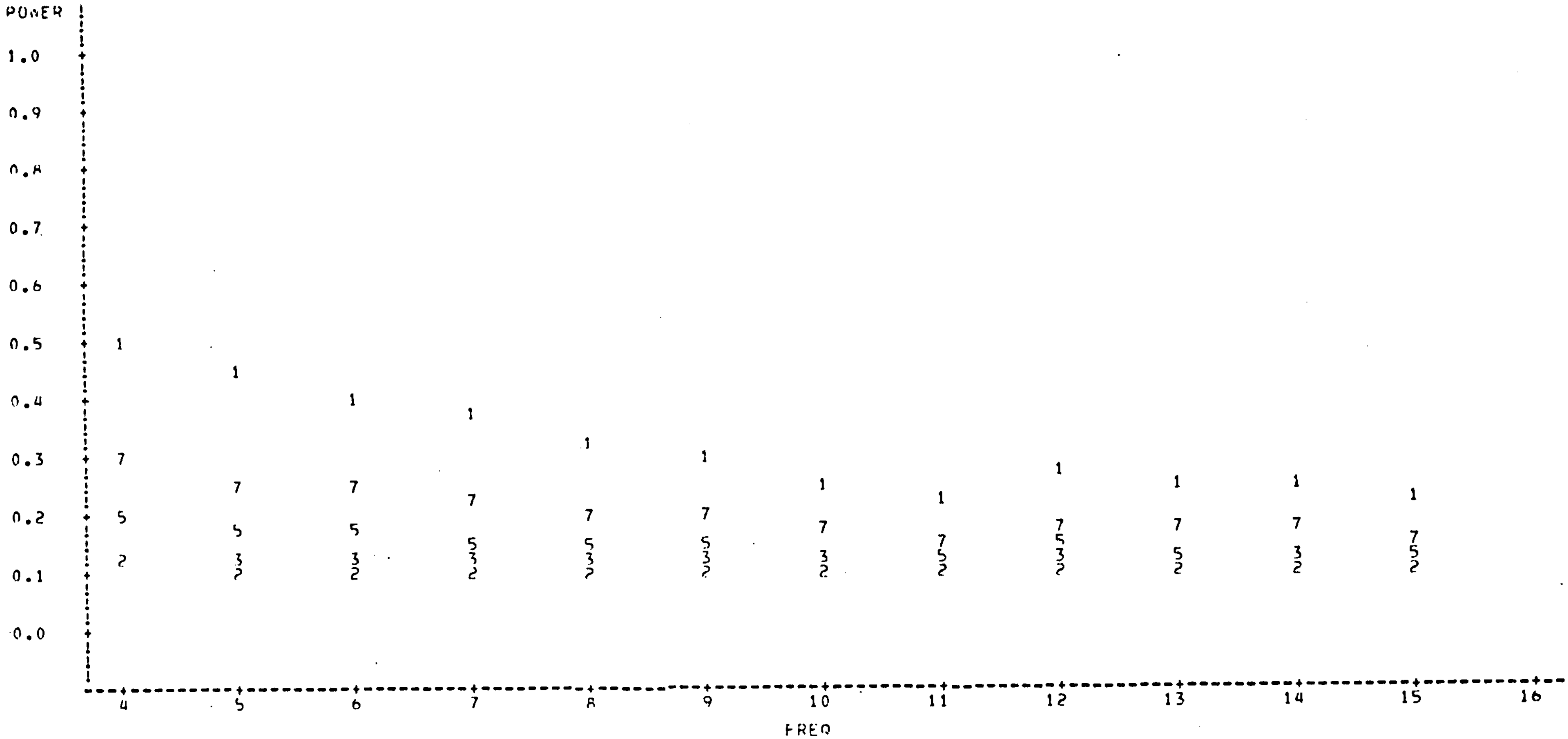
PLOT OF POWER*FREQ   SYMBOL IS VALUE OF PC



NOTE: 6 OBS HIDDEN

N1=10

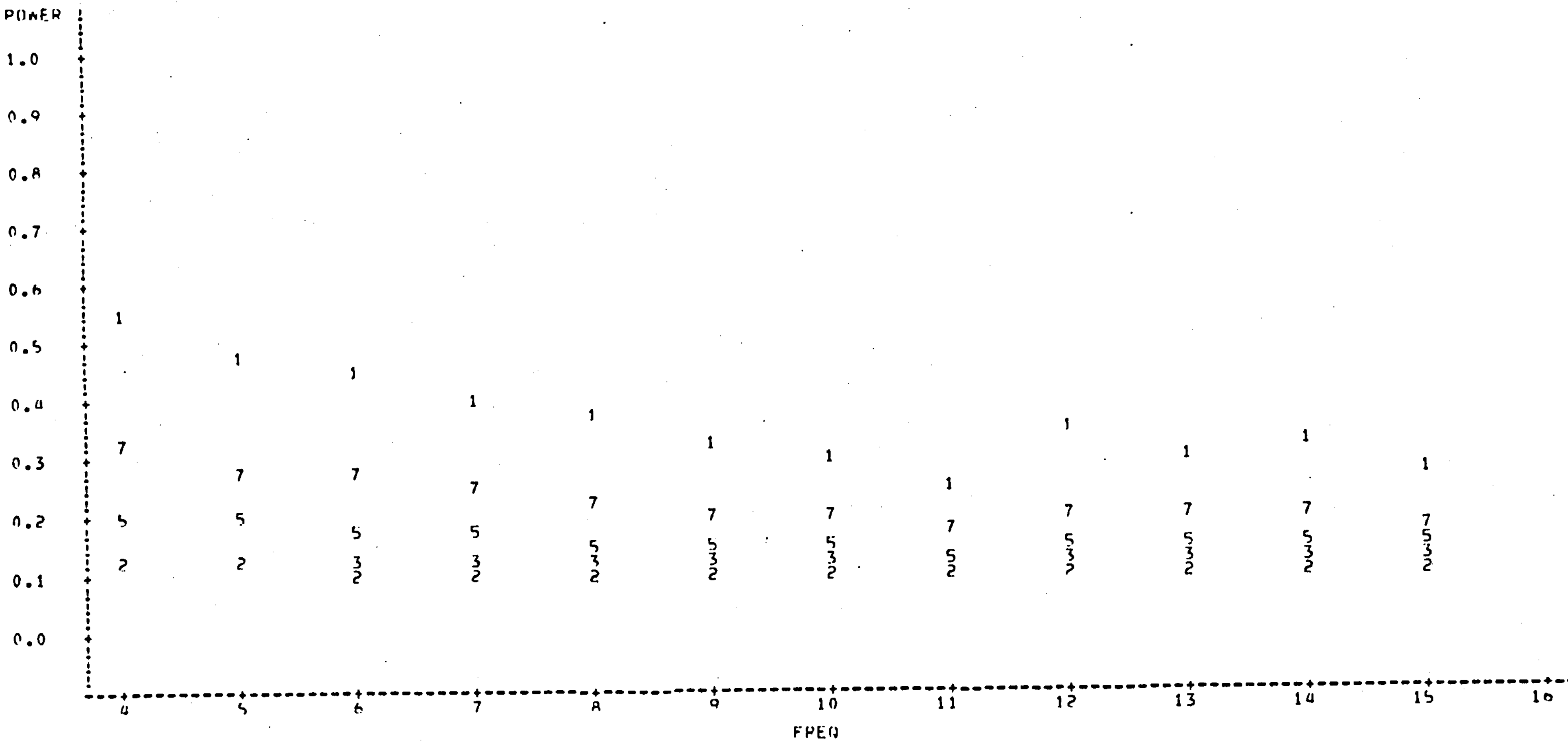
PLOT OF POWER*FREQ   SYMBOL IS VALUE OF PC



NOTE: 6 OBS HIDDEN

N1=11

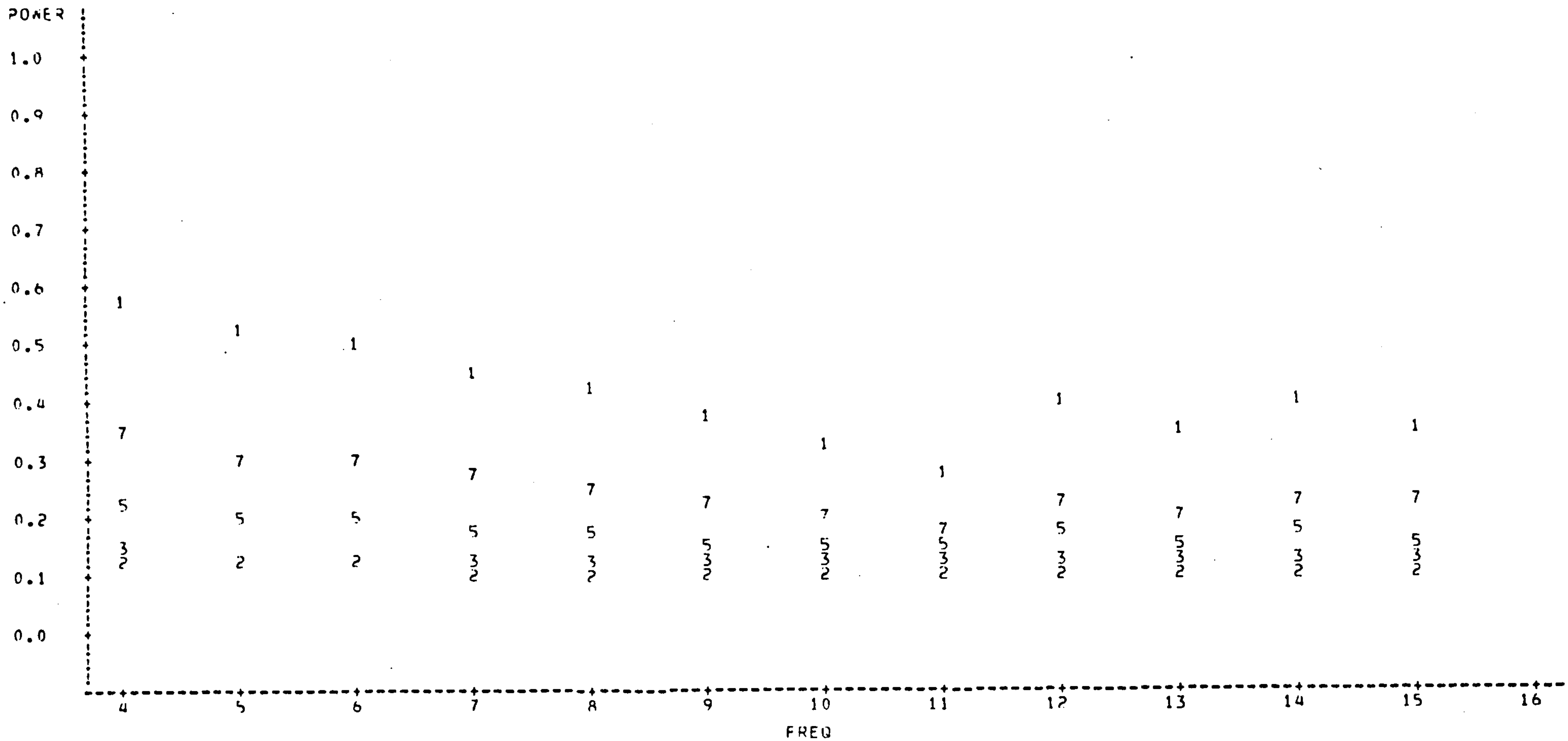
PLOT OF POWER*FREQ   SYMBOL IS VALUE OF PC



NOTE: 3 OBS HIDDEN

N1=12

PLOT OF POWER*FREQ   SYMBOL IS VALUE OF PC

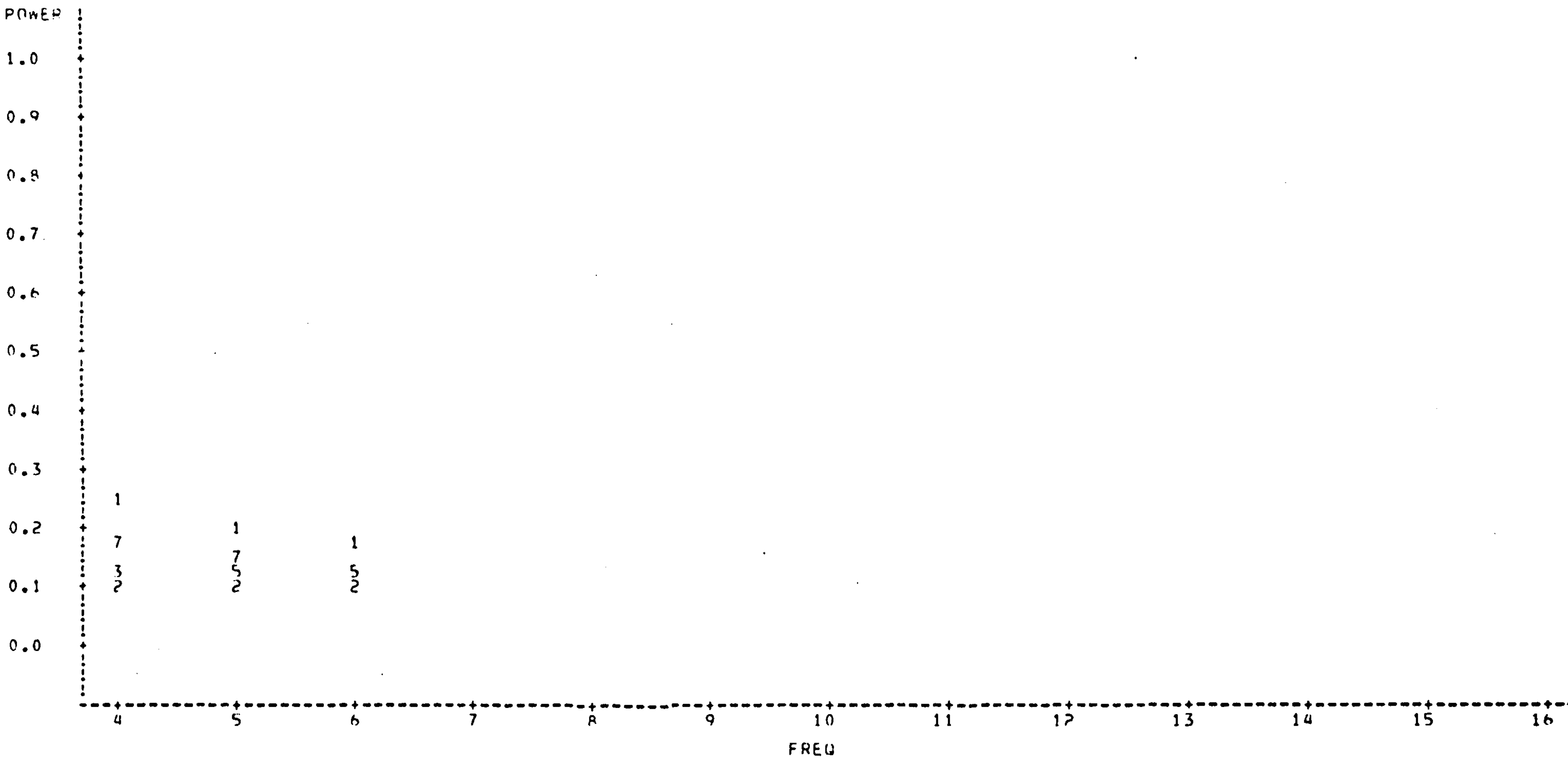


NOTE: 2 OBS HIDDEN

MIXED CORRELATION

N1=5

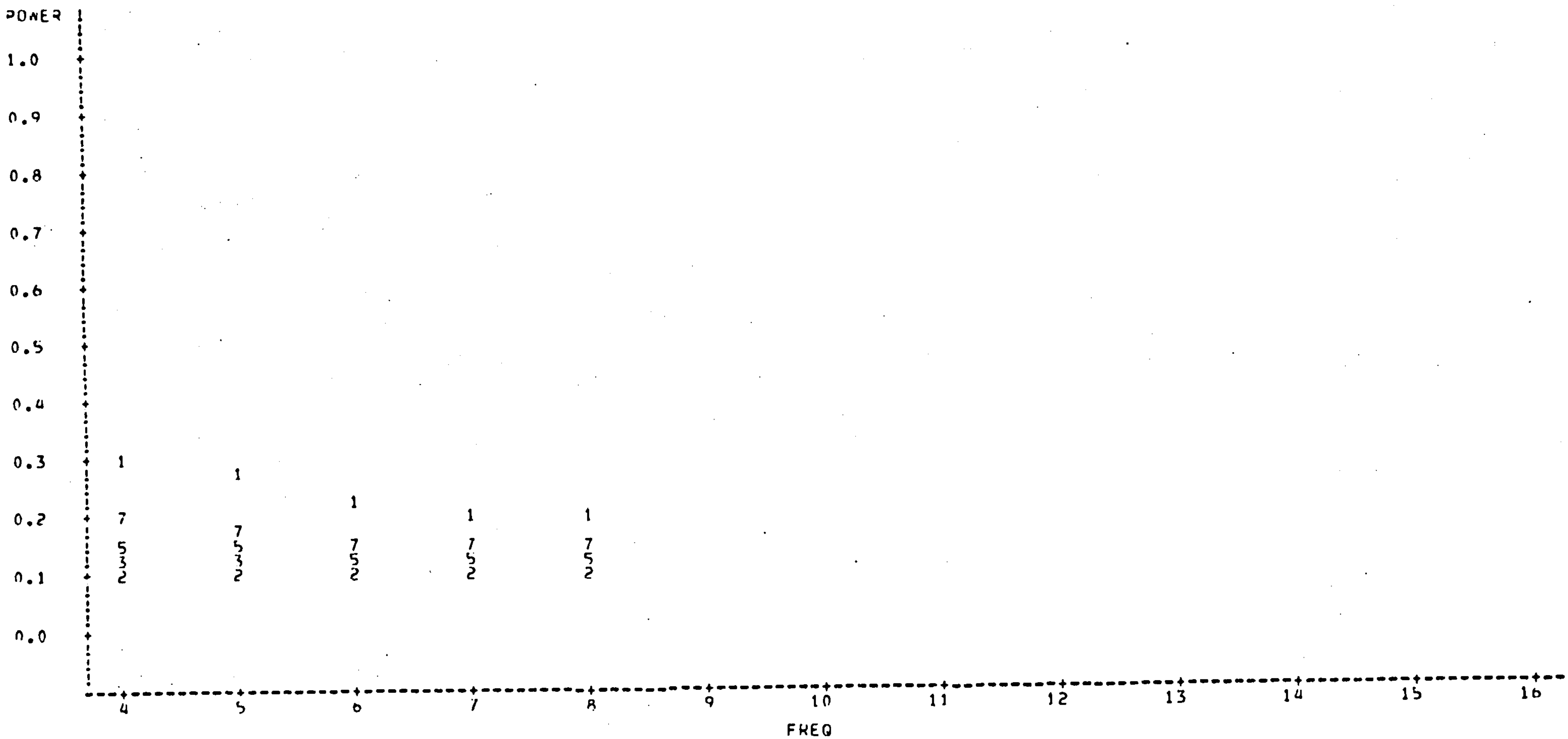
PLOT OF POWER*FREQ   SYMBOL IS VALUE OF PC.



NOTE: 4 OBS HIDDEN

N1=6

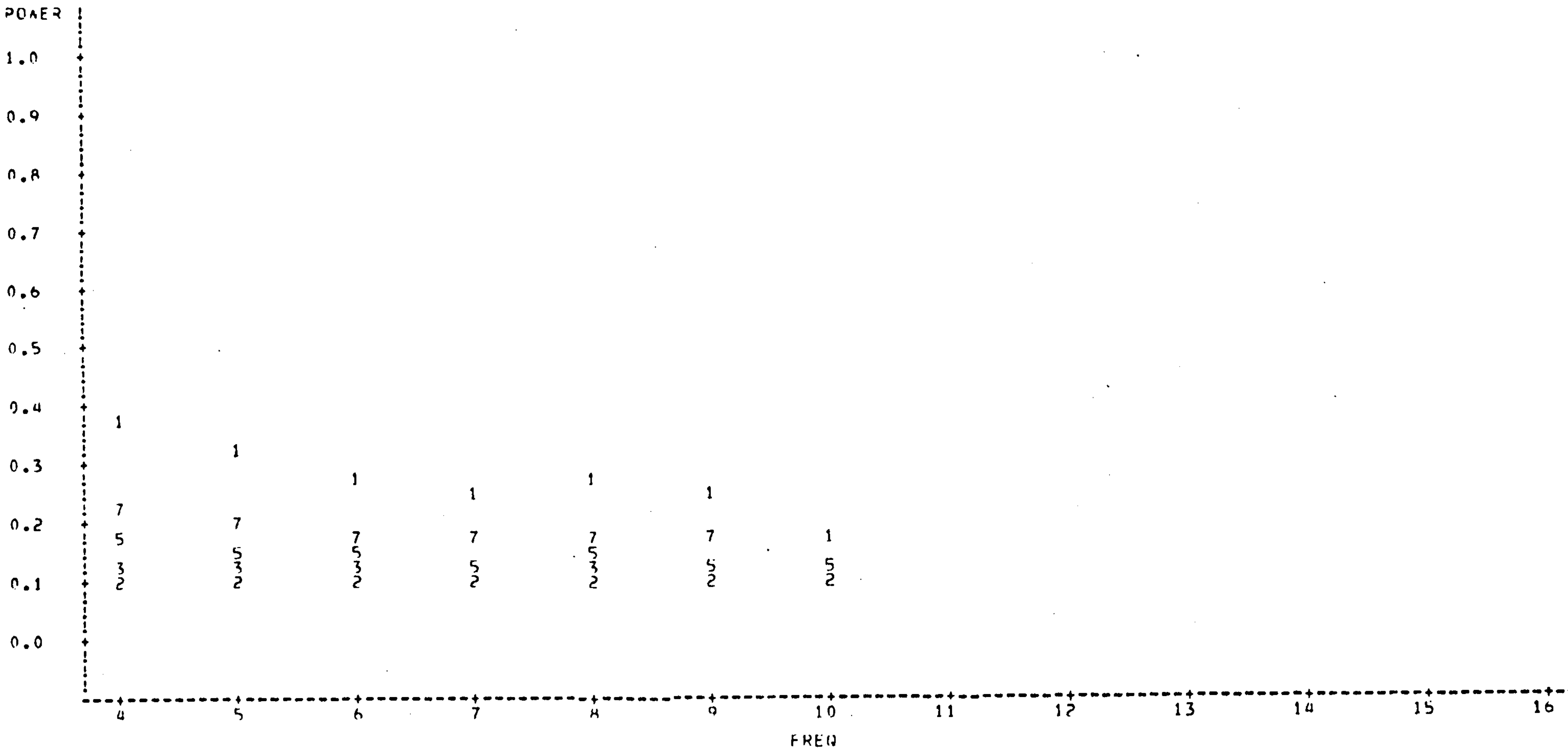
PLOT OF POWER*FREQ   SYMBOL IS VALUE OF PC



NOTE: 3 OBS HIDDEN

N1=7

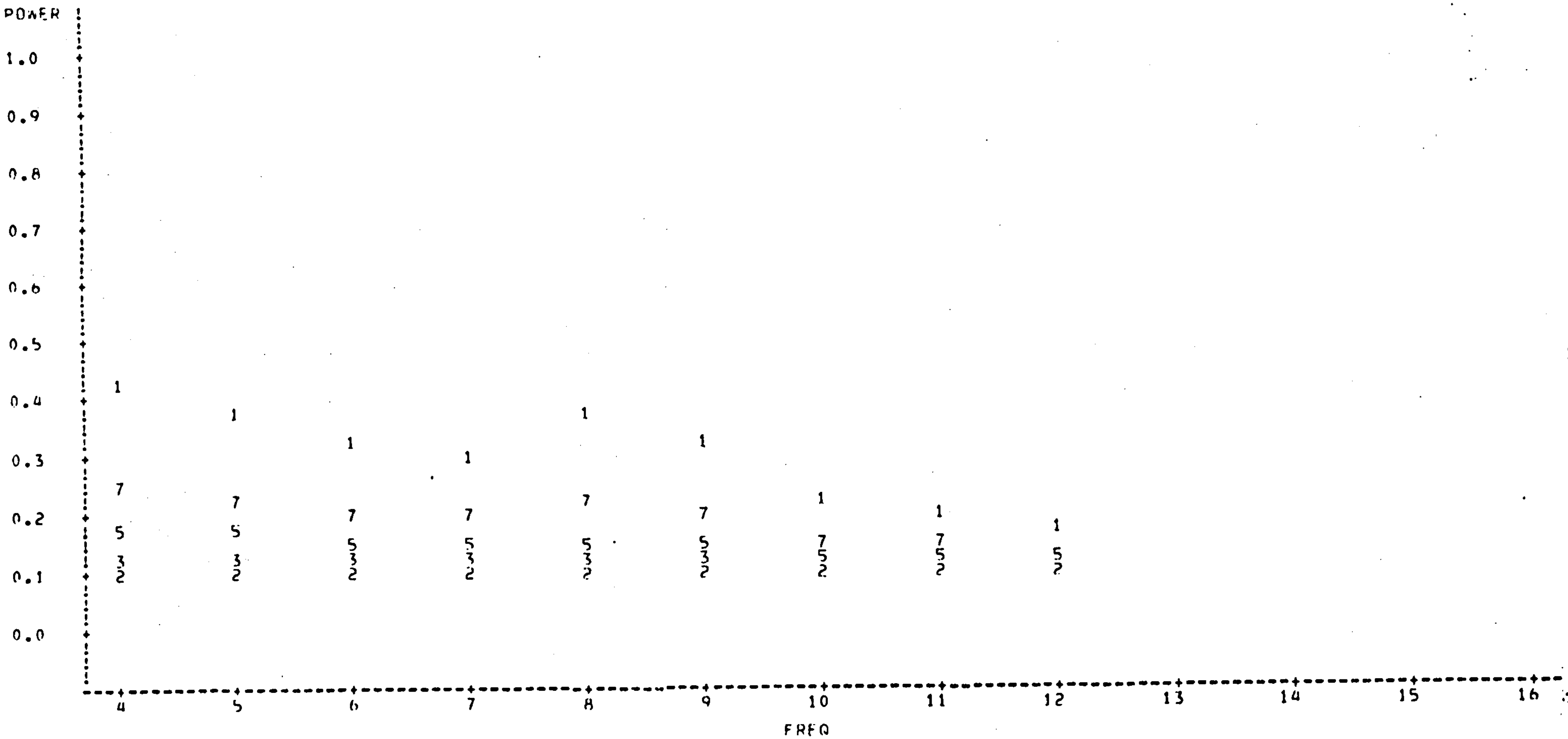
PLOT OF POWER*FREQ   SYMBOL IS VALUE OF PC .



NOTE: 4 OBS HIDDEN

N1=8

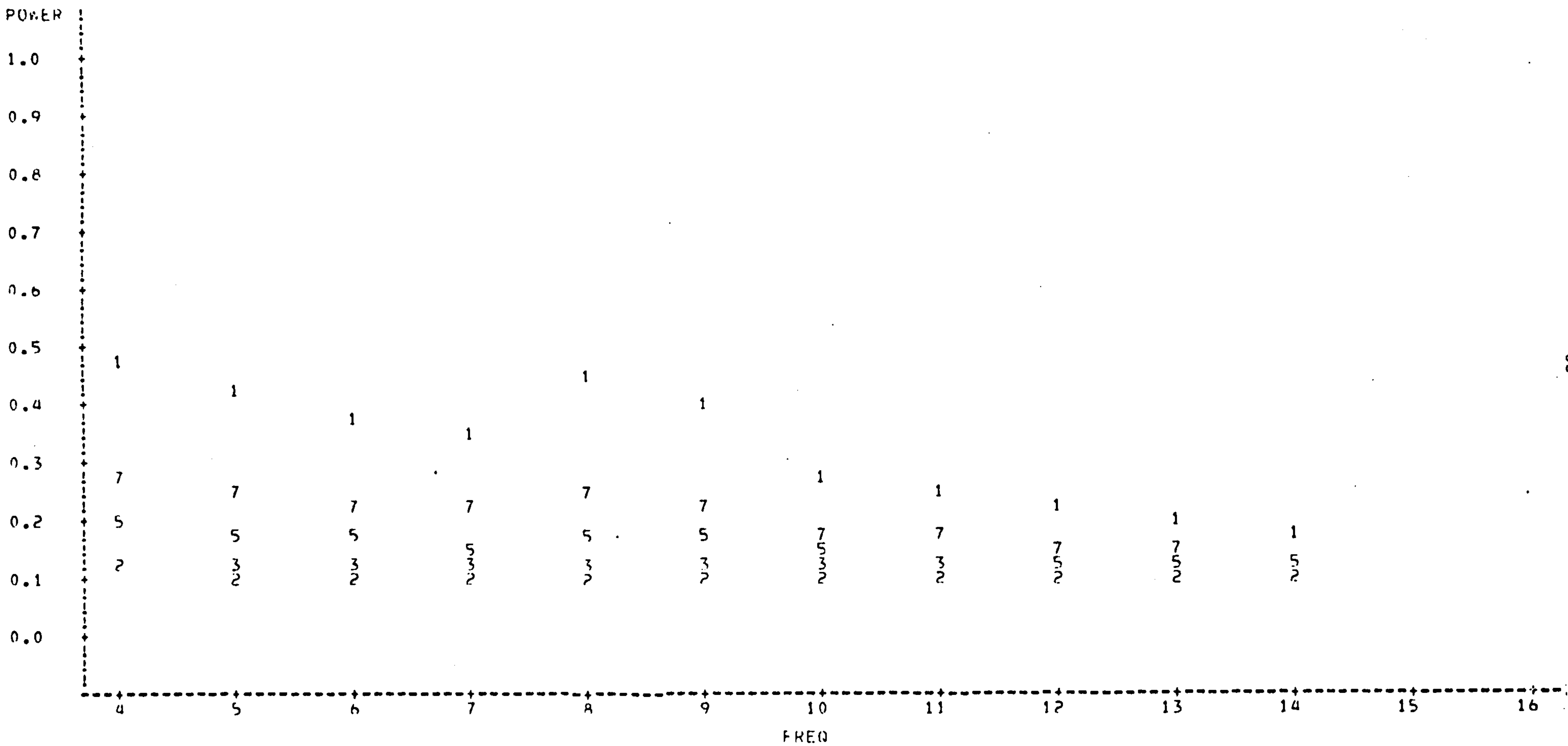
PLOT OF POWER*FREQ SYMBOL IS VALUE OF PC



NOTE: 4 OBS HIDDEN

N1=9

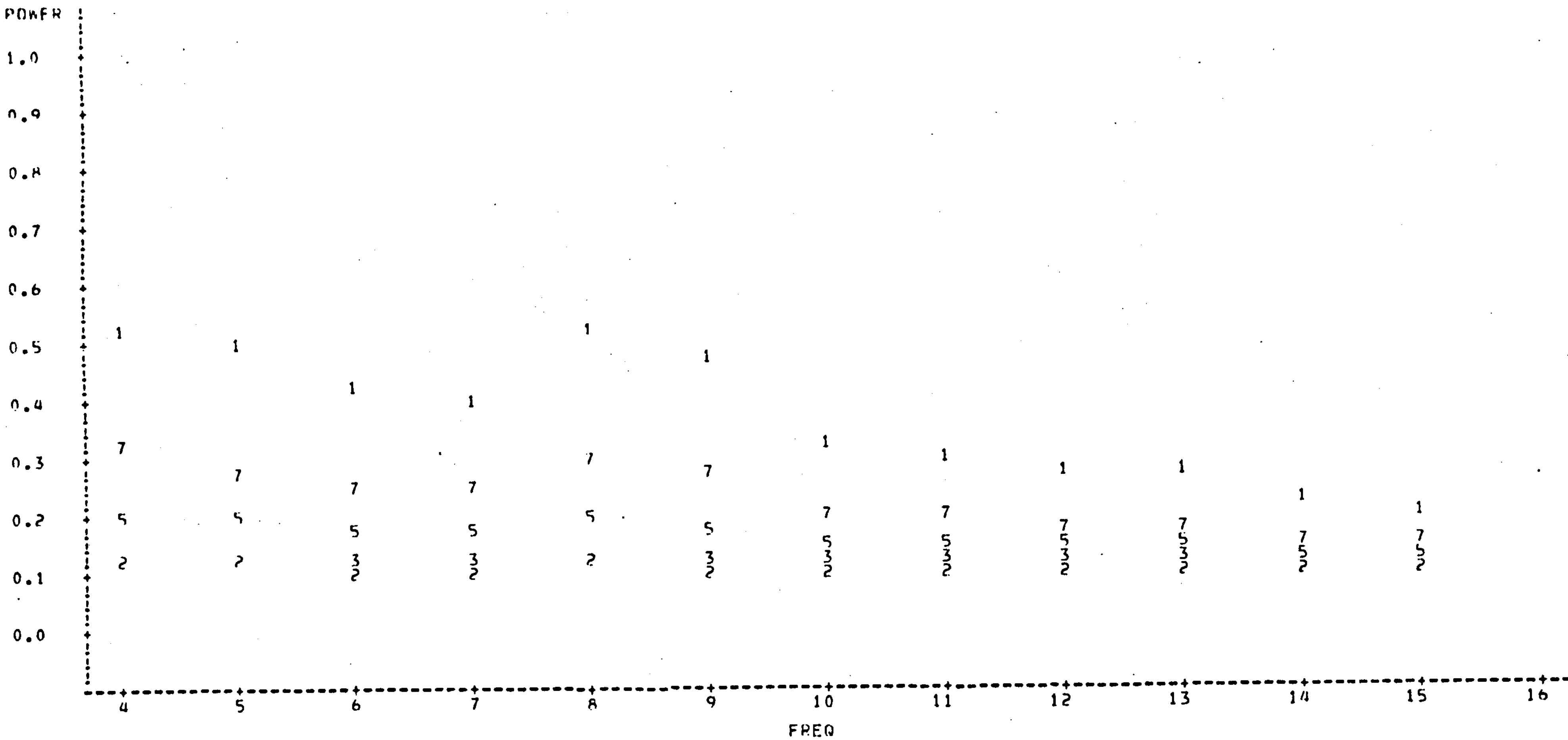
PLOT OF POWER*FREQ SYMBOL IS VALUE OF PC



NOTE: 6 OBS HIDDEN

N1=10

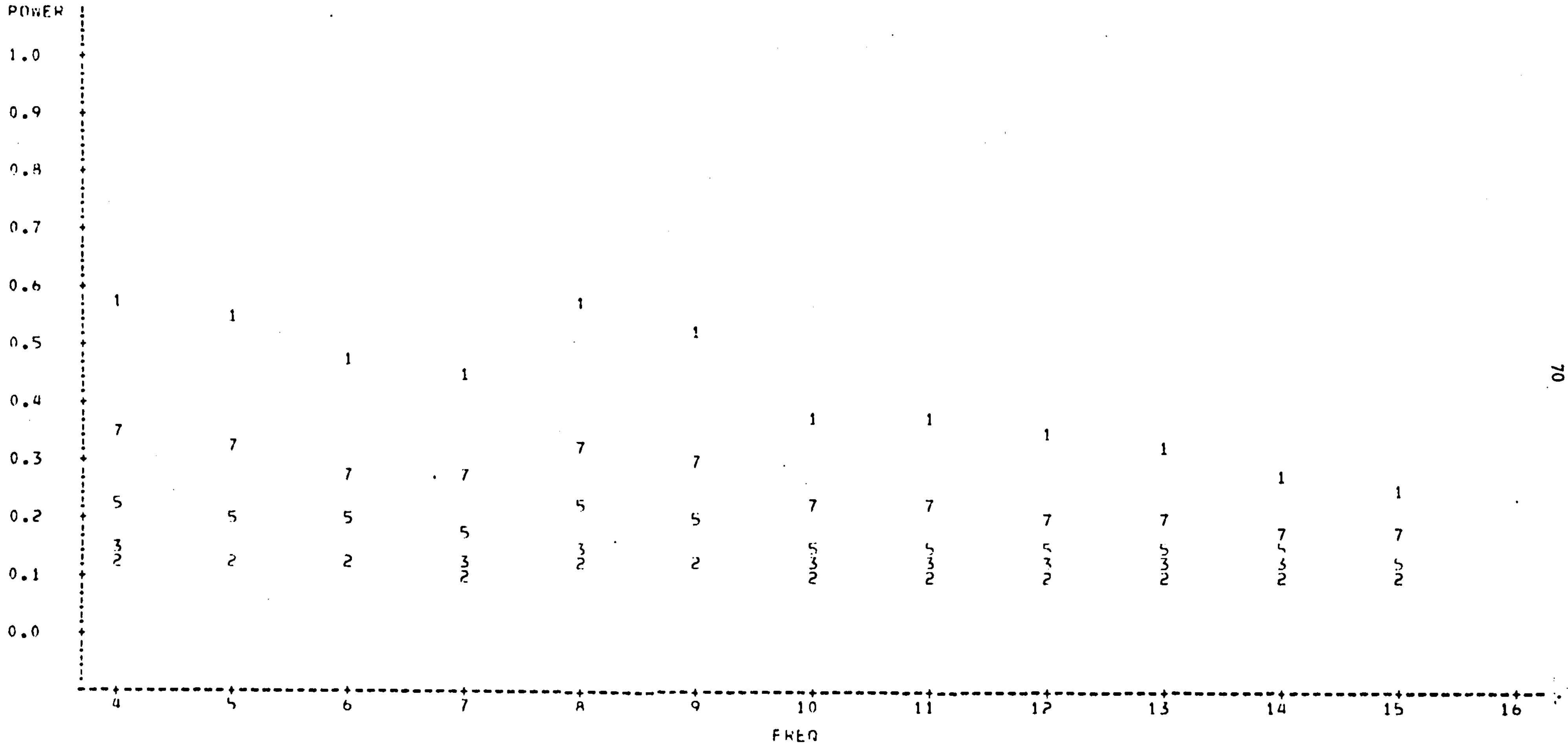
PLOT OF POWER*FREQ   SYMBOL IS VALUE OF PC



NOTE: 5 OBS HIDDEN

N1=11

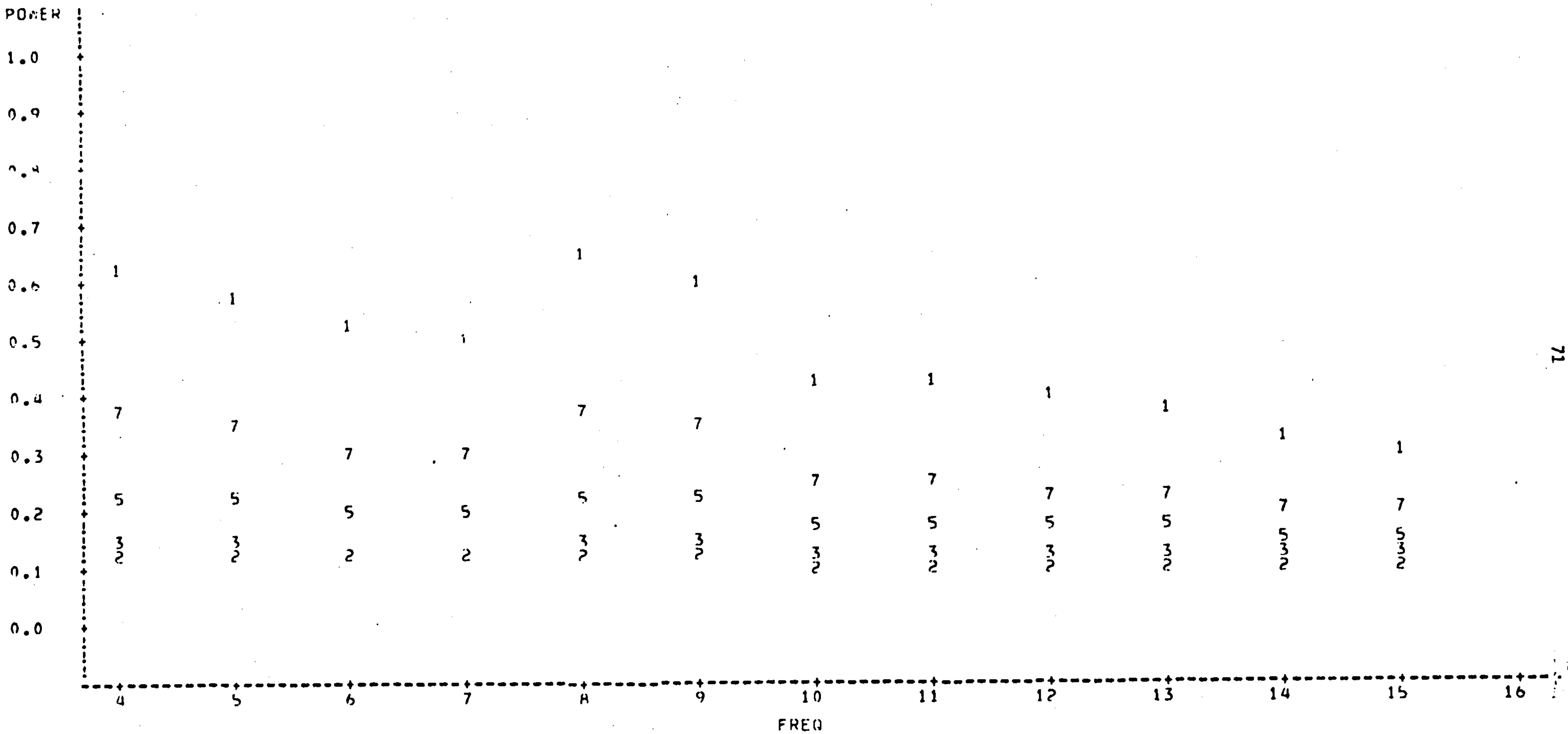
PLOT OF POWER*FREQ   SYMBOL IS VALUE OF PC



NOTE: 4 OBS HIDDEN

N1=12

PLOT OF POWER*FREQ   SYMBOL IS VALUE OF PC



NOTE: 2 OBS HIDDEN

Bibliography

1. Graybill, Franklin A., *Introduction to Linear Statistical Models*, Duxbury Press, 1976.

2. Morrison, Donald, *Multivariate Statistical Methods*, Second Edition McGraw-Hill, 1977.

3. Posten, Harry, "Power of the Likelihood-Ratio Test of the General Linear Hypothesis in Multivariate Analysis," Ph.D. dissertation, VPI&SU, 1960.

APPENDIX

Computer Program and Writeup

There are three subroutines supplied in the appendix for calculating various multivariate statistics for many hypothesis testing situations.

Some notation:

H - The hypothesis sum of squares matrix.

E - The error sum of squares matrix.

NP - The dimension of H and E.

NUH - The hypothesis degrees of freedom.

NUE - The error degrees of freedom.

The first subroutine ROY requires as input:

H, E - Square matrices of order NP stored columnwise in the vectors H and E respectively.

NUH, NUE - Degrees of freedom are also supplied by the calling program.

RTS - A work vector of length NP.

N - A work vector of length NP*NP.

Output:

RTS - Contains the ordered eigenvalues of HE^{-1}.

theta = $(\frac{\lambda max}{1+\lambda max})$ where λmax is the largest eigenvalue of HE^{-1}.

LS, RM, RN are S, M, N respectively for use in Heck's charts.

The second subroutine NILKS needs as input H, E, NP, NUH, and NUE as previously described, where H and E are square NP*NP matrices stored columnwise in a vector of the same name.

Output:

$W1 = \frac{|E|}{|E+H|}$ for comparison with Wall's Tables.

W2 = -(NUE-(NP-NUH+1)/2)*LN(W1) for comparison with chi-square tables, in large samples where W2 has NP*NUH degrees of freedom.

W3 = F-ratio having INUM numerator and IDENOM denominator degrees of freedom when and exact F-test is available otherwise W3=0 and INUM=IDENOM=0.

The third subroutine HOTEL can calculate M T^2-statistics. The necessary input includes

X - a vector of length M*NP containing M sets of mean vectors of length NP stored columnwise.

S - Sigma - where X(J), (J=1,M) has covariance-matrix (1/A(J))*Sigma.

A(J) - A vector of length n containing $\frac{n_1 n_2}{n_1+n_2}$ (where the n_i are the number of replicates in the i^{th} function), stored columnwise.

T2 - is a work vector of length M.

Output:

S: S-inverse

T2 - contains the M values of Hotelling's T^2 for direct comparison with tables

IEH - comes from matrix inversion routine

= 0 No error occurred during inversion

= -1 No result because of wrong input parameter n or the matrix is not positive definite

= k if there is a loss of significance.

```
      SUBROUTINE ROY(H,E,W,NP,NUH,NUE,RTS,THETA,LS,RM,RN)
C
C     SUBROUTINES REQUIRED: DNROOT AND DEIGEN SUPPLIED BY USER.
C
C     MATRICES H DUE TO HYPOTHESIS AND E DUE TO ERROR, DIMENSION NP, AND
C     DEGREES OF FREEDOM, NUH FOR HYPOTHESIS AND NUE FOR ERROR, ARE SUPPLIED IN
C     THE CALLING PROGRAM.
C     H AND E ARE SQUARE MATRICES OF ORDER NP STORED COLUMNWISE.
C     RTS AND W ARE WORK VECTORS OF DIMENSIONS NP AND NP*NP, RESPECTIVELY.
C     VALUES RETURNED ARE THE VECTOR RTS OF ORDERED LATENT ROOTS OF H*E-INVERSE,
C     THETA=(LARGEST ROOT)/(1.+LARGEST ROOT), AND PARAMETERS LS, RM, AND RN FOR
C     REFERENCE TO HECK'S CHARTS (DROP THE FIRST LETTER OF EACH).
C
      DIMENSION H(1),E(1),W(1),RTS(1)
      DOUBLE PRECISION H,E,RTS,THETA,X,W
      IF(NUH-NP)15,15,20
   15 LS=NUH
      GO TO 25
   20 LS=NP
   25 RM=(FLOAT(IABS(NUH-NP)-1))/2.
      RN=(FLOAT(NUE-NP-1))/2.
      CALL DNROOT(NP,H,E,RTS,W)
      THETA=RTS(1)/(1.D0+RTS(1))
      RETURN
      END

      SUBROUTINE WILKS(H,E,NP,NUH,NUE,W1,W2,W3,INUM,IDENOM,IEW)
C
C     SUBROUTINES REQUIRED: DMFSD FROM SSP LIBRARY.
C
C     MATRICES H DUE TO HYPOTHESIS AND E DUE TO ERROR, DIMENSION NP, AND
C     DEGREES OF FREEDOM, NUH FOR HYPOTHESIS AND NUE FOR ERROR, ARE PROVIDED
C     BY THE CALLING PROGRAM.
C     H AND E ARE STORED COLUMN-WISE AS UPPER TRIANGULAR MATRICES.
C
C     OUTPUT INCLUDES:
C        W1=DET(E)/DET(E+H) FOR COMPARISON WITH WALL'S TABLES,
C        W2=-(NUE-(NP-NUH+1)/2)*LN(W1) FOR COMPARISON WITH CHI-SQUARE HAVING
C           NP*NUH DEGREES OF FREEDOM IN LARGE SAMPLES (LN(.) IS NATURAL LOG).
C        W3=F-RATIO HAVING INUM NUMERATOR AND IDENOM DENOMINATOR DEGREES
C           OF FREEDOM WHEN EXACT F-TESTS ARE AVAILABLE; OTHERWISE W3=0.0,
C           INUM=0, AND IDENOM=0.
C
      DIMENSION H(1),E(1)
      DOUBLE PRECISION H,E,T,W1,W2,W3,WM,DETE,DETT,PROD
      TOL=.000001
      L=NP*(NP+1)/2
      DO 20 I=1,L
      T=H(I)+E(I)
   20 H(I)=T
      CALL DMFSD(E,NP,TOL,IEE)
```

```
      CALL DMFSD(H,NP,TOL,IEH)
      PROD=1.D0
      L1=1
      L2=2
      DO 25 I=1,NP
      PROD=E(L1)*PROD
      L1=L1+L2
   25 L2=L2+1
      DETE=PROD**2
      PROD=1.D0
      L1=1
      L2=2
      DO 30 I=1,NP
      PROD=H(L1)*PROD
      L1=L1+L2
   30 L2=L2+1
      DETT=PROD**2
      W1=DETE/DETT
      AM=DFLOAT(NUE)-(DFLOAT(NP-NUH+1))/2.D0
      W2=-AM*DLOG(W1)
      IF(NUH.EQ.1) GO TO 50
      IF(NUH.EQ.2) GO TO 60
      IF(NP.EQ.1) GO TO 70
      IF(NP.EQ.2) GO TO 80
      GO TO 90
   50 INUM=NP
      IDENOM=NUE+NUH-NP
      W3=(1.D0-W1)*DFLOAT(IDENOM)/(W1*DFLOAT(INUM))
      GO TO 130
   60 INUM=2*NP
      IDENOM=2*(NUE+NUH-NP-1)
      W3=(1.D0-DSQRT(W1))*DFLOAT(IDENOM)/(DSQRT(W1)*DFLOAT(INUM))
      GO TO 130
   70 INUM=NUH
      IDENOM=NUE
      W3=(1.D0-W1)*DFLOAT(IDENOM)/(W1*DFLOAT(INUM))
      GO TO 130
   80 INUM=2*NUH
      IDENOM=2*(NUE-1)
      W3=(1.D0-DSQRT(W1))*DFLOAT(IDENOM)/(DSQRT(W1)*DFLOAT(INUM))
      GO TO 130
   90 W3=0.D0
      INUM=0
      IDENOM=0
  130 IF(IEE.EQ.0) GO TO 135
      GO TO 145
  135 IF(IEH.EQ.0) GO TO 140
      GO TO 145
  140 IER=0
      GO TO 150
  145 IER=1
```

```
  150 RETURN
      END

           SUBROUTINE HOTEL(S,W,NP,X,A,T2,M,IER)
C
C     SUBROUTINES REQUIRED: DSINV IN SSP LIBRARY
C
C     S IS A SYMMETRIC MATRIX OF ORDER NP STORED COLUMN-WISE IN UPPER
C           TRIANGULAR FORM.
C     A AND T2 ARE VECTORS OF LENGTH M.
C     W IS A WORK VECTOR OF LENGTH NP.
C     ASSUME VECTORS X(J) TO HAVE COVARIANCE MATRICES (1./A(J))*SIGMA, FOR
C     J=1,M. THE MATRIX S AND VECTORS X(J) ARE SUPPLIED BY THE CALLING PROGRAM
C     AS WELL AS THE DIMENSION NP AND M. THE VECTORS X(J) ARE STORED COLUMN-
C           WISE IN THE VECTOR X OF LENGTH NP*M.
C     ON RETURN S IS REPLACED BY SINVERSE, EACH X(J) IS REPLACED BY THE
C     CORRESPONDING COEFFICIENTS OF A LINEAR DISCRIMINANT FUNCTION, AND
C     THE VECTOR T2 CONTAINS VALUES OF THE HOTELLING'S T-SQUARE STATISTICS
C     FOR DIRECT COMPARISON WITH SPECIAL AID TABLES (SEE JENSEN AND HOWE, 1967)
C
      DIMENSION S(1),X(1),A(1),T2(1),W(1)
      DOUBLE PRECISION S,X,A,T2,W,SUM,SUM1
      TOL=.000001
      CALL DSINV(S,NP,TOL,IER)
      L2=1
      L3=NP
      DO 30 I=1,M
      SUM=0.D0
      L1=1
      DO 25 J=L2,L3
      DO 25 K=L2,J
      SUM1=X(K)*S(L1)*X(J)
      IF(J-K) 15,20,15
   15 SUM1=SUM1*2.D0
   20 SUM=SUM+SUM1
   25 L1=L1+1
      T2(I)=SUM*A(I)
      L2=L2+NP
   30 L3=L3+NP
      L=0
      L1=1
      L2=NP
      DO 50 I=1,M
      L3=L*NP
      DO 40 J=L1,L2
      SUM=0.D0
      DO 35 K=L1,L2
      LJ=J-L3
      LK=K-L3
      CALL LOC(LJ,LK,JK,NP,NP,1)
      SUM=SUM+S(JK)*X(K)
```

```
   35 X(LJ)=SUM
   40 CONTINUE
      DO 45 JJ=L1,L2
      JL=JJ-L3
   45 X(JJ)= (JL)
      L=L+1
      L1=L1+NP
   50 L2=L2+NP
      RETURN
      END
            SUBROUTINE DNROOT(M,A,B,XL,X)                       NR00 370
      DIMENSION A(1),B(1),XL(1),X(1)                            NR00 380
      DOUBLE PRECISION A,B,XL,X,SUMV                            NR00 460
      K=1                                                       NR00 610
      DO 100 J=2,M                                              NR00 620
      L=M*(J-1)                                                 NR00 630
      DO 100 I=1,J                                              NR00 640
      L=L+1                                                     NR00 650
      K=K+1                                                     NR00 660
  100 B(K)=B(L)                                                 NR00 670
      MV=0                                                      NR00 710
      CALL DEIGEN(B,X,M,MV)                                     NR00 720
      L=0                                                       NR00 770
      DO 110 J=1,M                                              NR00 780
      L=L+J                                                     NR00 790
  110 XL(J)=1.D/DSQRT(DABS(B(L)))                               NR00 800
      K=0                                                       NR00 810
      DO 115 J=1,M                                              NR00 820
      DO 115 I=1,M                                              NR00 830
      K=K+1                                                     NR00 840
  115 B(K)=X(K)*XL(J)                                           NR00 850
      DO 120 I=1,M                                              NR00 890
      N2=0                                                      NR00 900
      DO 120 J=1,M                                              NR00 910
      N1=M*(I-1)                                                NR00 920
      L=M*(J-1)+I                                               NR00 930
      X(L)=0.D0                                                 NR00 940
      DO 120 K=1,M                                              NR00 950
      N1=N1+1                                                   NR00 960
      N2=N2+1                                                   NR00 970
  120 X(L)=X(L)+B(N1)*A(N2)                                     NR00 980
      L=0                                                       NR00 990
      DO 130 J=1,M                                              NR001000
      DO 130 I=1,J                                              NR001010
      N1=I-M                                                    NR001020
      N2=M*(J-1)                                                NR001030
      L=L+1                                                     NR001040
      A(L)=0.D0                                                 NR001050
      DO 130 K=1,M                                              NR001060
      N1=N1+M                                                   NR001070
```

```
      N2=N2+1                                                          NR001080
  130 A(L)=A(L)+X(N1)*B(N2)                                            NR001090
      CALL DEIGEN(A,X,M,MV)                                            NR001130
      L=0                                                              NR001140
      DO 140 I=1,M                                                     NR001150
      L=L+I                                                            NR001160
  140 XL(I)=A(L)                                                       NR001170
      DO 150 I=1,M                                                     NR001210
      N2=0                                                             NR001220
      DO 150 J=1,M                                                     NR001230
      N1=I-M                                                           NR001240
      L=M*(J-1)+I                                                      NR001250
      A(L)=0.D0                                                        NR001260
      DO 150 K=1,M                                                     NR001270
      N1=N1+M                                                          NR001280
      N2=N2+1                                                          NR001290
  150 A(L)=A(L)+B(N1)*X(N2)                                            NR001300
      L=0                                                              NR001310
      K=0                                                              NR001320
      DO 180 J=1,M                                                     NR001330
      SUMV=0.D0                                                        NR001340
      DO 170 I=1,M                                                     NR001350
      L=L+1                                                            NR001360
  170 SUMV=SUMV+A(L)*A(L)                                              NR001370
  175 SUMV=DSQRT(SUMV)                                                 NR001380
      DO 180 I=1,M                                                     NR001390
      K=K+1                                                            NR001400
  180 X(K)=A(K)/SUMV                                                   NR001410
      RETURN                                                           NR001420
      END                                                              NR001430
```

MEMBER NAME EIGEN

```
C
C     ..................................................................
C
C        SUBROUTINE EIGEN
C
C        PURPOSE
C           COMPUTE EIGENVALUES AND EIGENVECTORS OF A REAL SYMMETRIC
C           MATRIX
C
C        USAGE
C           CALL EIGEN(A,R,N,MV)
C
C        DESCRIPTION OF PARAMETERS
C           A - ORIGINAL MATRIX (SYMMETRIC), DESTROYED IN COMPUTATION.
C               RESULTANT EIGENVALUES ARE DEVELOPED IN DIAGONAL OF
C               MATRIX A IN DESCENDING ORDER.
C           R - RESULTANT MATRIX OF EIGENVECTORS (STORED COLUMNWISE,
C               IN SAME SEQUENCE AS EIGENVALUES)
C           N - ORDER OF MATRICES A AND R
C           MV- INPUT CODE
C                   0   COMPUTE EIGENVALUES AND EIGENVECTORS
C                   1   COMPUTE EIGENVALUES ONLY (R NEED NOT BE
C                       DIMENSIONED BUT MUST STILL APPEAR IN CALLING
C                       SEQUENCE)
C
C        REMARKS
C           ORIGINAL MATRIX A MUST BE REAL SYMMETRIC (STORAGE MODE=1)
C           MATRIX A CANNOT BE IN THE SAME LOCATION AS MATRIX R
C
C        SUBROUTINES AND FUNCTION SUBPROGRAMS REQUIRED
C           NONE
C
C        METHOD
C           DIAGONALIZATION METHOD ORIGINATED BY JACOBI AND ADAPTED
C           BY VON NEUMANN FOR LARGE COMPUTERS AS FOUND IN 'MATHEMATICAL
C           METHODS FOR DIGITAL COMPUTERS', EDITED BY A. RALSTON AND
C           H.S. WILF, JOHN WILEY AND SONS, NEW YORK, 1962, CHAPTER 7
C
C     ..................................................................
C
      SUBROUTINE EIGEN(A,R,N,MV)
      DIMENSION A(1),R(1)
C
C        ...............................................................
C                                  (DEIGEN)
C        IF A DOUBLE PRECISION VERSION OF THIS ROUTINE IS DESIRED, THE
C        C IN COLUMN 1 SHOULD BE REMOVED FROM THE DOUBLE PRECISION
C        STATEMENT WHICH FOLLOWS.
C
C     DOUBLE PRECISION A,R,ANORM,ANRMX,THR,X,Y,SINX,SINX2,COSX,
C    1                 COSX2,SINCS,RANGE
C
C        THE C MUST ALSO BE REMOVED FROM DOUBLE PRECISION STATEMENTS
C        APPEARING IN OTHER ROUTINES USED IN CONJUNCTION WITH THIS
C        ROUTINE.
C
C        THE DOUBLE PRECISION VERSION OF THIS SUBROUTINE MUST ALSO
C        CONTAIN DOUBLE PRECISION FORTRAN FUNCTIONS.  SQRT IN STATEMENTS
```

```
MEMBER NAME  EIGEN
C           40, 68, 75, AND 78 MUST BE CHANGED TO DSQRT.  ABS IN STATEMENT
C           62 MUST BE CHANGED TO DABS. THE CONSTANT IN STATEMENT 5 SHOULD
C           BE CHANGED TO 1.0D-12.
C
C           ..........................................................
C
C           GENERATE IDENTITY MATRIX
C
    5 RANGE=1.0E-6
      IF(MV-1) 10,25,10
   10 IQ=-N
      DO 20 J=1,N
      IQ=IQ+N
      DO 20 I=1,N
      IJ=IQ+I
      R(IJ)=0.0
      IF(I-J) 20,15,20
   15 R(IJ)=1.0
   20 CONTINUE
C
C           COMPUTE INITIAL AND FINAL NORMS (ANORM AND ANORMX)
C
   25 ANORM=0.0
      DO 35 I=1,N
      DO 35 J=I,N
      IF(I-J) 30,35,30
   30 IA=I+(J*J-J)/2
      ANORM=ANORM+A(IA)*A(IA)
   35 CONTINUE
      IF(ANORM) 165,165,40
   40 ANORM=1.414*SQRT(ANORM)
      ANRMX=ANORM*RANGE/FLOAT(N)
C
C           INITIALIZE INDICATORS AND COMPUTE THRESHOLD, THR
C
      IND=0
      THR=ANORM
   45 THR=THR/FLOAT(N)
   50 L=1
   55 M=L+1
C
C           COMPUTE SIN AND COS
C
   60 MQ=(M*M-M)/2
      LQ=(L*L-L)/2
      LM=L+MQ
   62 IF( ABS(A(LM))-THR) 130,65,65
   65 IND=1
      LL=L+LQ
      MM=M+MQ
      X=0.5*(A(LL)-A(MM))
   68 Y=-A(LM)/ SQRT(A(LM)*A(LM)+X*X)
      IF(X) 70,75,75
   70 Y=-Y
   75 SINX=Y/ SQRT(2.0*(1.0+( SQRT(1.0-Y*Y))))
      SINX2=SINX*SINX
   78 COSX= SQRT(1.0-SINX2)
      COSX2=COSX*COSX
```

```
MEMBER NAME  EIGEN
      SINCS =SINX*COSX
C
C        ROTATE L AND M COLUMNS
C
      ILQ=N*(L-1)
      IMQ=N*(M-1)
      DO 125 I=1,N
      IQ=(I*I-I)/2
      IF(I-L) 80,115,80
   80 IF(I-M) 85,115,90
   85 IM=I+MQ
      GO TO 95
   90 IM=M+IQ
   95 IF(I-L) 100,105,105
  100 IL=I+LQ
      GO TO 110
  105 IL=L+IQ
  110 X=A(IL)*COSX-A(IM)*SINX
      A(IM)=A(IL)*SINX+A(IM)*COSX
      A(IL)=X
  115 IF(MV-1) 120,125,120
  120 ILR=ILQ+I
      IMR=IMQ+I
      X=R(ILR)*COSX-R(IMR)*SINX
      R(IMR)=R(ILR)*SINX+R(IMR)*COSX
      R(ILR)=X
  125 CONTINUE
      X=2.0*A(LM)*SINCS
      Y=A(LL)*COSX2+A(MM)*SINX2-X
      X=A(LL)*SINX2+A(MM)*COSX2+X
      A(LM)=(A(LL)-A(MM))*SINCS+A(LM)*(COSX2-SINX2)
      A(LL)=Y
      A(MM)=X
C
C        TESTS FOR COMPLETION
C
C        TEST FOR M = LAST COLUMN
C
  130 IF(M-N) 135,140,135
  135 M=M+1
      GO TO 60
C
C        TEST FOR L = SECOND FROM LAST COLUMN
C
  140 IF(L-(N-1)) 145,150,145
  145 L=L+1
      GO TO 55
  150 IF(IND-1) 160,155,160
  155 IND=0
      GO TO 50
C
C        COMPARE THRESHOLD WITH FINAL NORM
C
  160 IF(THR-ANRMX) 165,165,45
C
C        SORT EIGENVALUES AND EIGENVECTORS
C
  165 IQ=-N
```

```
MEMBER NAME  EIGEN
      DO 185 I=1,N                                                      EIGE1750
      IQ=IQ+N                                                           EIGE1760
      LL=I+(I*I-I)/2                                                    EIGE1770
      JQ=N*(I-2)                                                        EIGE1780
      DO 185 J=I,N                                                      EIGE1790
      JQ=JQ+N                                                           EIGE1800
      MM=J+(J*J-J)/2                                                    EIGE1810
      IF(A(LL)-A(MM)) 170,185,185                                       EIGE1820
  170 X=A(LL)                                                           EIGE1830
      A(LL)=A(MM)                                                       EIGE1840
      A(MM)=X                                                           EIGE1850
      IF(MV-1) 175,185,175                                              EIGE1860
  175 DO 180 K=1,N                                                      EIGE1870
      ILR=IQ+K                                                          EIGE1880
      IMR=JQ+K                                                          EIGE1890
      X=R(ILR)                                                          EIGE1900
      R(ILR)=R(IMR)                                                     EIGE1910
  180 R(IMR)=X                                                          EIGE1920
  185 CONTINUE                                                          EIGE1930
      RETURN                                                            EIGE1940
      END                                                               EIGE1950
```

MEMBER NAME DMFSD

```
C
C     ..................................................................
C
C        SUBROUTINE DMFSD
C
C        PURPOSE
C           FACTOR A GIVEN SYMMETRIC POSITIVE DEFINITE MATRIX
C
C        USAGE
C           CALL DMFSD(A,N,EPS,IER)
C
C        DESCRIPTION OF PARAMETERS
C           A      - DOUBLE PRECISION UPPER TRIANGULAR PART OF GIVEN
C                    SYMMETRIC POSITIVE DEFINITE N BY N COEFFICIENT
C                    MATRIX.
C                    ON RETURN A CONTAINS THE RESULTANT UPPER
C                    TRIANGULAR MATRIX IN DOUBLE PRECISION.
C           N      - THE NUMBER OF ROWS (COLUMNS) IN GIVEN MATRIX.
C           EPS    - SINGLE PRECISION INPUT CONSTANT WHICH IS USED
C                    AS RELATIVE TOLERANCE FOR TEST ON LOSS OF
C                    SIGNIFICANCE.
C           IER    - RESULTING ERROR PARAMETER CODED AS FOLLOWS
C                    IER=0  - NO ERROR
C                    IER=-1 - NO RESULT BECAUSE OF WRONG INPUT PARAME-
C                             TER N OR BECAUSE SOME RADICAND IS NON-
C                             POSITIVE (MATRIX A IS NOT POSITIVE
C                             DEFINITE, POSSIBLY DUE TO LOSS OF SIGNI-
C                             FICANCE)
C                    IER=K  - WARNING WHICH INDICATES LOSS OF SIGNIFI-
C                             CANCE. THE RADICAND FORMED AT FACTORIZA-
C                             TION STEP K+1 WAS STILL POSITIVE BUT NO
C                             LONGER GREATER THAN ABS(EPS*A(K+1,K+1)).
C
C        REMARKS
C           THE UPPER TRIANGULAR PART OF GIVEN MATRIX IS ASSUMED TO BE
C           STORED COLUMNWISE IN N*(N+1)/2 SUCCESSIVE STORAGE LOCATIONS.
C           IN THE SAME STORAGE LOCATIONS THE RESULTING UPPER TRIANGU-
C           LAR MATRIX IS STORED COLUMNWISE TOO.
C           THE PROCEDURE GIVES RESULTS IF N IS GREATER THAN 0 AND ALL
C           CALCULATED RADICANDS ARE POSITIVE.
C           THE PRODUCT OF RETURNED DIAGONAL TERMS IS EQUAL TO THE
C           SQUARE-ROOT OF THE DETERMINANT OF THE GIVEN MATRIX.
C
C        SUBROUTINES AND FUNCTION SUBPROGRAMS REQUIRED
C           NONE
C
C        METHOD
C           SOLUTION IS DONE USING THE SQUARE-ROOT METHOD OF CHOLESKY.
C           THE GIVEN MATRIX IS REPRESENTED AS PRODUCT OF TWO TRIANGULAR
C           MATRICES, WHERE THE LEFT HAND FACTOR IS THE TRANSPOSE OF
C           THE RETURNED RIGHT HAND FACTOR.
C
C     ..................................................................
C
      SUBROUTINE DMFSD(A,N,EPS,IER)
C
C
      DIMENSION A(1)
```

```
MEMBER NAME  DMFSD
      DOUBLE PRECISION DPIV,DSUM,A
C
C        TEST ON WRONG INPUT PARAMETER N
      IF(N-1) 12,1,1
    1 IER=0
C
C        INITIALIZE DIAGONAL-LOOP
      KPIV=0
      DO 11 K=1,N
      KPIV=KPIV+K
      IND=KPIV
      LEND=K-1
C
C        CALCULATE TOLERANCE
      TOL=ABS(EPS*SNGL(A(KPIV)))
C
C        START FACTORIZATION-LOOP OVER K-TH ROW
      DO 11 I=K,N
      DSUM=0.D0
      IF(LEND) 2,4,2
C
C        START INNER LOOP
    2 DO 3 L=1,LEND
      LANF=KPIV-L
      LIND=IND-L
    3 DSUM=DSUM+A(LANF)*A(LIND)
C        END OF INNER LOOP
C
C        TRANSFORM ELEMENT A(IND)
    4 DSUM=A(IND)-DSUM
      IF(I-K) 10,5,10
C
C        TEST FOR NEGATIVE PIVOT ELEMENT AND FOR LOSS OF SIGNIFICANCE
    5 IF(SNGL(DSUM)-TOL) 6,6,9
    6 IF(DSUM) 12,12,7
    7 IF(IER) 8,8,9
    8 IER=K-1
C
C        COMPUTE PIVOT ELEMENT
    9 DPIV=DSQRT(DSUM)
      A(KPIV)=DPIV
      DPIV=1.D0/DPIV
      GO TO 11
C
C        CALCULATE TERMS IN ROW
   10 A(IND)=DSUM*DPIV
   11 IND=IND+I
C        END OF DIAGONAL-LOOP
C
      RETURN
   12 IER=-1
      RETURN
      END
```

```
MEMBER NAME  DSINV
C
C        ..................................................................
C
C           SUBROUTINE DSINV
C
C           PURPOSE
C              INVERT A GIVEN SYMMETRIC POSITIVE DEFINITE MATRIX
C
C           USAGE
C              CALL DSINV(A,N,EPS,IER)
C
C           DESCRIPTION OF PARAMETERS
C              A      - DOUBLE PRECISION UPPER TRIANGULAR PART OF GIVEN
C                       SYMMETRIC POSITIVE DEFINITE N BY N COEFFICIENT
C                       MATRIX.
C                       ON RETURN A CONTAINS THE RESULTANT UPPER
C                       TRIANGULAR MATRIX IN DOUBLE PRECISION.
C              N      - THE NUMBER OF ROWS (COLUMNS) IN GIVEN MATRIX.
C              EPS    - SINGLE PRECISION INPUT CONSTANT WHICH IS USED
C                       AS RELATIVE TOLERANCE FOR TEST ON LOSS OF
C                       SIGNIFICANCE.
C              IER    - RESULTING ERROR PARAMETER CODED AS FOLLOWS
C                       IER=0  - NO ERROR
C                       IER=-1 - NO RESULT BECAUSE OF WRONG INPUT PARAME-
C                                TER N OR BECAUSE SOME RADICAND IS NON-
C                                POSITIVE (MATRIX A IS NOT POSITIVE
C                                DEFINITE, POSSIBLY DUE TO LOSS OF SIGNI-
C                                FICANCE)
C                       IER=K  - WARNING WHICH INDICATES LOSS OF SIGNIFI-
C                                CANCE. THE RADICAND FORMED AT FACTORIZA-
C                                TION STEP K+1 WAS STILL POSITIVE BUT NO
C                                LONGER GREATER THAN ABS(EPS*A(K+1,K+1)).
C
C           REMARKS
C              THE UPPER TRIANGULAR PART OF GIVEN MATRIX IS ASSUMED TO BE
C              STORED COLUMNWISE IN N*(N+1)/2 SUCCESSIVE STORAGE LOCATIONS.
C              IN THE SAME STORAGE LOCATIONS THE RESULTING UPPER TRIANGU-
C              LAR MATRIX IS STORED COLUMNWISE TOO.
C              THE PROCEDURE GIVES RESULTS IF N IS GREATER THAN 0 AND ALL
C              CALCULATED RADICANDS ARE POSITIVE.
C
C           SUBROUTINES AND FUNCTION SUBPROGRAMS REQUIRED
C              DMFSD
C
C           METHOD
C              SOLUTION IS DONE USING FACTORIZATION BY SUBROUTINE DMFSD.
C
C        ..................................................................
C
      SUBROUTINE DSINV(A,N,EPS,IER)
C
C
      DIMENSION A(1)
      DOUBLE PRECISION A,DIN,WORK
C
C        FACTORIZE GIVEN MATRIX BY MEANS OF SUBROUTINE DMFSD
C        A = TRANSPOSE(T) * T
      CALL DMFSD(A,N,EPS,IER)
```

```
MEMBER NAME  DSINV
      IF(IER) 9,1,1
C
C        INVERT UPPER TRIANGULAR MATRIX T
C        PREPARE INVERSION-LOOP
    1 IPIV=N*(N+1)/2
      IND=IPIV
C
C        INITIALIZE INVERSION-LOOP
      DO 6 I=1,N
      DIN=1.D0/A(IPIV)
      A(IPIV)=DIN
      MIN=N
      KEND=I-1
      LANF=N-KEND
      IF(KEND) 5,5,2
    2 J=IND
C
C        INITIALIZE ROW-LOOP
      DO 4 K=1,KEND
      WORK=0.D0
      MIN=MIN-1
      LHOR=IPIV
      LVER=J
C
C        START INNER LOOP
      DO 3 L=LANF,MIN
      LVER=LVER+1
      LHOR=LHOR+L
    3 WORK=WORK+A(LVER)*A(LHOR)
C        END OF INNER LOOP
C
      A(J)=-WORK*DIN
    4 J=J-MIN
C        END OF ROW-LOOP
C
    5 IPIV=IPIV-MIN
    6 IND=IND-1
C        END OF INVERSION-LOOP
C
C        CALCULATE INVERSE(A) BY MEANS OF INVERSE(T)
C        INVERSE(A) = INVERSE(T) * TRANSPOSE(INVERSE(T))
C        INITIALIZE MULTIPLICATION-LOOP
      DO 8 I=1,N
      IPIV=IPIV+I
      J=IPIV
C
C        INITIALIZE ROW-LOOP
      DO 8 K=I,N
      WORK=0.D0
      LHOR=J
C
C        START INNER LOOP
      DO 7 L=K,N
      LVER=LHOR+K-I
      WORK=WORK+A(LHOR)*A(LVER)
    7 LHOR=LHOR+L
C        END OF INNER LOOP
C
```

```
      A(J)=WORK                                                         DSIN1170
    6 J=J+K                                                             DSIN1180
C        END OF ROW- AND MULTIPLICATION-LOOP                            DSIN1190
C                                                                       DSIN1200
    9 RETURN                                                            DSIN1210
      END                                                               DSIN1220
```