DOCUMENT RESUME

ED 218 344 TM 820 385

AUTHOR Fetler, Mark
TITLE Unidimensionality in the California Assessment Program's Third Grade Test.
PUB DATE Mar 82
NOTE 41p.; Paper presented at the Annual Meeting of the American Educational Research Association (66th, New York, NY, March 19-23, 1982); Tables 3-4 and Figures marginally legible due to small print.

EDRS PRICE MF01/PC02 Plus Postage.
DESCRIPTORS *Achievement Tests; Educational Assessment; Elementary Education; Grade 3; *Latent Trait Theory; *Models; Test Construction
IDENTIFIERS *Assumptions (Testing); California; California Assessment Program; *Unidimensional Scaling

ABSTRACT

The unidimensionality of the reading and mathematics tests of the California Assessment Program was investigated. During April of 1980, all third grade students in California took a test consisting of reading, mathematics, and written language items. Item intercorrelation matrices were constructed for each skill, and the latent roots extracted. Methods of outlier analysis were applied to look for large drops in the size of the latent roots from each matrix. Mathematics skills tended to have smaller initial roots and a second root closer to the first than the reading skills. The evidence does not support the hypothesis of unidimensionality for the money and fractions skills, but does not contradict the hypothesis for the mathematical applications skill or for reading skills. (Author/BW)

ED218344

Unidimensionality in the California Assessment Program's Third Grade Test

by

Mark Fetler
California Assessment Program
California Department of Education
Sacramento, California, 95814

Running head: Unidimensionality

AUTHOR'S NOTE

This article was originally written for presentation at the annual meeting of the American Educational Research Association, New York, March, 1982. Special thanks are owed to Dale Carlson of the California Department of Education. Dr. Carlson's encouragement, criticism, and generosity are largely responsible for any success of this study. I am especially indebted to Dave Wiley and Annegret Harnischfeger for their exceptionally lucid discussion of the problem of undimensionality of CAP skills and their penetrating analyses of the phenomenon. I, of course, retain sole responsible for any remaining errors of logic or execution. The views expressed here are not necessarily those of the California Department of Education.

ABSTRACT

This paper reports the results of an investigation of the unidimensionality of the reading and mathematics tests of the California Assessment Program. Item intercorrelation matrices were constructed for each skill, and the latent roots extracted. Methods of outlier analysis are applied to look for large drops in the size of the latent roots from each matrix. Results of factor analyses are examined, and item content is examined to check the outcomes of the statistical procedures.

INTRODUCTION

In 1980 the California Assessment Program (CAP) introduced a new test for the third grade, which is administered each spring to more than 280,000 students. Scaled scores are reported for 60 different skills in reading, written language, and mathematics. The model used to estimate the scores is based on item response curve (IRC) theory, and includes parameters for the difficulty and reliability of items (Mislevy and Bock, 1980).

The mathematical model used to measure performance on a CAP skill, assumes that the probability of success on an item depends only on the item parameters and on examinee ability. This is the assumption of unidimensionality. Lord (1980, pp. 20-21) suggests that if the items measure one dimension, and if the IRC model holds, then the matrix of tetrachoric intercorrelations will be of unit rank. To examine this possibility Lord suggests plotting the latent roots of the correlation matrix. "If (1) the first root is large compared to the second and (2) the second root is not much larger than any of the others, then the items are approximately unidimensional." A direct application of this procedure to CAP's third grade test is not possible for several reasons. The CAP test has 30 forms and is administered according to a multiple matrix sampling design. Although skills contain from ten to twenty items, each student sees at most one of these. Furthermore, the school is the basic unit for reporting and analysis. Mislevy and Bock (1980) describe the application of IRC theory to this situation. They note that "under the CAP multiple matrix sampling design, each of a school's responses to items from a given skill has been obtained from a different randomly selected pupil. For the school level analysis, then, the standard assumption in IRC theory of 'local' or 'conditional' independence is satisfied perfectly." Mislevey and Bock's paper contains a detailed account of the derivation of scores.

An analysis of unidimensionality of CAP skills was undertaken by Wiley and Harnischfeger (1981) using data and analyses supplied by the author. Their procedure involved examination of the distribution of successive differences of logarithms of roots to locate differences in spacing of roots which might indicate discrepancies in root size.

Method

Materials and procedure. The third grade test was administered during April of 1980 to all third grade students according to standardized procedures by school personnel. Each student was asked to complete one of thirty different forms, spiraled to insure even distribution of all forms, each form containing nine reading items, twelve mathematics, and thirteen written language items. Allocation of reading and mathematics items to skills is shown in Tables 1 and 2. Ample time was allowed for completion. After test administration the documents were collected and processed off-site.

Insert tables 1 and 2 about here.

Procedures for development of the items were designed to assure content validity. Specifications for each skill were carefully and narrowly defined. Items were written in accordance with the specifications and underwent several rounds of content review and field testing. Given the relative fineness of definition of specifications, these procedures tended to support the hypothesis of undiminsionality of item content in each skill. This is a matter of considered logical, and not statistical judgement. While statistical procedures discussed here provide a useful adjunct for making a judgement of unidimensionality, they are not proposed as a definitive criterion.

Analysis. For each item in a skill the number of attempts and the number of correct responses were tabulated and used to calculate a logit item score for each school. Only larger schools, having at least two responses per item were included in the analysis. Latent roots were extracted from the intercorrelation matrix of the logit item scores. The matrix was factor analyzed by the principal axis method, factors with eigenvalues greater than 1.0 were retained, and varimax rotated. This procedure was repeated for each of the 17 reading skills and the 20 mathematics skills.

Statistical procedures described by Barnett and Lewis (1978) used to examine distributions for outliers can be applied to examine the hypothesis of unidimensionality. Lord's procedure asks whether the second root, in addition to the first, is large with respect to remaining roots. This is like asking whether the second root, in addition to the first, is an upper outlier with respect to remaining roots. If the answer is yes, this can be interpreted as lack of support for the hypothesis of unidimensionality.

A class of statistics discussed by Barnett and Lewis involve taking ratios of difference scores. The numerator of the ratio is the difference between the outlier and some function of the remaining observations. The denominator is a measure of spread, such as a range or standard deviation. The search for more than one upper outlier may involve consecutive or sequential testing. Each value is compared with its next lower neighbor or with a linear combination of lower neighbors.

Results

Table 3 displays for each skill the first two roots, the mean of the remaining roots and the ratio (root2-mean)/(root1-root2). Relatively larger ratios can be interpreted in terms of a second root that is closer to the first root than to the mass of remaining roots. In general, math skills tend to have larger ratios and smaller initial roots than reading skills. The largest ratio belongs to the nature of numbers-money and fractions skill.

Insert table 3 about here.

Figure 1 is a scatterplot with (root1-root2)/range and (root2-mean)/range on the horizontal axis. Reading skills are denoted by "R", and mathematics skills by "M". Using the range as a divisor tended to produce a straight line plot, as shown, while similar plots using no divisor or the standard deviation were more amorphous, e.g. figure 2. The reading skills are grouped in the upper left corner of the plot. This indicates that in reading there was a larger drop from the first to the second root than in math, and that the second root in reading tends to be closer to the mean of remaining roots. The two extreme points in the lower right of the plot represent money and fractions and basic skills operations.

Insert figures 1, 2, 3 and 4 about here.

Figure 3 is a plot of the roots of the money and fractions skill. For comparison purposes figure 4 is a plot of the roots of the nature of numbers applications skill. The second plot corresponds more nearly to the ideal for a unidimensional skill. Plots of the logs of the eigenvalues for these two skills are shown in figures 5 and 6. The first root is off the scale on both plots. The roots for the applications skill appear to form an unbroken curve, while the roots for money and fractions do not, with the second and possibly the third root seeming to deviate.

Insert figures 5, 6, 7 and 8 about here.

A plot of the linear regression of the logs of the roots on serial position, excluding the first root, is shown in Figure 5. The equation of the line is y(predicted) = -.054*x + .347 . With the exception of the second root, most of the points lie close to the line. Given that its deviation is .137, and the the standard error of estimate is .055, this point is 2.5 standard errors from the line.

Figures 7 and 8 display consecutive differences of logs of roots. That is, the first value is (log1-log2), the second is (-log2 - log3), and so on. Again, for money and fractions the second root does not appear to belong to the remaining mass. By contrast, the applications skill has one large initial difference, followed by a sequence of much smaller differences having no apparent upward or downward trend.

For reference with the factor analysis, the items in the money and fractions skill are displayed in the appendix. Based on logical analysis of content, there appear to be two distinct types: Those involving the shading of a geometric figure; and those relating to money. There are several major variations in format. Seven money items involve translating a picture of coins into an amount of money, and two involve translating numeric to longhand written amounts. Three fractions items involve translating a numeric fraction into a shaded geometric shape and the remaining involve translating a shaded shape into a particular numeric fraction.

Results of the factor analysis are shown in Table 4. Although not totally unambiguous, some patterns may be discerned. Items 2, 3, 4 and 6 (fractions) load positively on factor one. Items 8 and 10, and to a lesser extent 12 and 14 (money), loaded positively on factor two. Items 5, 11 and 13 (money) loaded negatively on factor three. The situation is not as clear for factors four and five. Items 4 (fractions) and 12 and 13 (money) load on factor four, and items 1 and 15 (fractions) and 7 and 11 (money) load positively on factor five.

Insert table 4 about here.

A question remains as to the extent item formats might have influenced the factor loadings. One of the fraction item formats includes a fraction in the stem and options are geometrical figures (items 1, 3, and 15). The other fraction item format includes a geometrical figure in the stem, and the options are fractions (items 2, 4 and 6). In fact, items 2, 4 and 6 load on factor one, and items 1 and 15 load on factor 5. Unexpectedly, item 3 loads on factor one. While the evidence for a format effect is suggestive, it certainly is not conclusive. A similar result obtained with money item formats. Items 9 and 11, involving translation of of numerical and written amounts, did load together on factor three. Unexpectedly, they are joined by item 5, which involved recognition of a coin denomination. Additionally, item 11 loaded rather heavily on factor five, which seems to contain a little of everything.

Discussion

One of the striking features of Table 3 is the difference between mathematics and reading skills. Reading skills tend to have larger initial roots, smaller ratios, and have a smaller range of ratios. By contrast the mathematics ratios have a much greater range, with basic facts operations and money and fractions as two upper values. If the ratio by itself could be considered an index of unidimensionality, it seems that reading skills are more so than mathematics skills. Even so, it is true that great care was lavished on skills to assure narrowly defined content. Possibly variations in mathematics item formats, which have no corresponding counterparts in reading, are partly responsible for the difference.

The plots of difference statistics, figures 1 and 2, show a clumping and segregation of reading skills, confirming a similar trend in Table 3 of roots and ratios. Again, the two math skills which stand out from the rest are basic facts operations and money and fractions.

Figure 1, with its nearly linear plot stands in contrast to the more amorphous shape of figure 2. The effect of dividing by the range increases the value of of the variable on the ordinate, (root1-root2)/range, relative to the value on the abscissa, (root2-mean)/range. For skills with relatively large second roots the values on the ordinate are small relative to those on the abscissa. As a result, the plot displays such outlying skills on the lower right.

Figure 3 is a plot of the latent roots for money and fractions, and figure 4 displays the roots for the applications skill. The relatively larger second root of money and fractions is the main difference between the two plots. The remaining figures attempt to make the difference more evident. If the logs of the roots are plotted, e.g. in figures 4 and 5, the descrease in value of the roots with increasing serial position is more evident. Tentatively, one can compute a regression line. deviations from the line. On this basis the second root of money and fractions was shown to be an outlier. Visual inspection of the applications roots indicates no such outliers. Certain limitations of this approach must be acknowledged. It is assumed that the distributions of latent roots satisfy those needed for regression. Even if these assumptions do not strictly hold, it may be that the analysis yields an acceptable rule of thumb, and in this case it appears to be so.

Figures 7 and 8 are plots of differences of successive pairs of eigenvalues for the two skills under consideration. Large values indicate a large drop in root size. The results are generally consistent with earlier figures. The applications skill shows a large initial drop, followed by much smaller diffences. The money and fractions skill shows a much smaller initial drop, followed by relatively large second and third differeces, with the remaining differences similar to those in applications.

Comparison of the factor loadings, displayed in Table 4, with the items in the appendix seems to confirm that factor one respresents money items and that factors two and three represent fractions. Factors four anf fixe were mixed. There was some evidence for format factors. However, this hypothesis was clouded somewhat by the potpourri of items loading on factors four and five.

Several techniques have been presented for examining the hypothesis of unidimensionality. The evidence does not support this hypothesis for the money and fractions skill, but did not contradict the hypothesis for the applications skill. The result confirmed a feature that had been built into the test. The money and fractions skill was designed to have two types of items. This was the legitimate and considered decision of the committee responsible for design of the test, and related to a need for a certain type of score. Making statistical judgements about unidimensionality was aided considerably by the presence of many other skills for comparison purposes. This helped provide a sense of what was genuinely unusual and what was a mild aberration. In addition, there were definite differences across content areas. Comparison of an isolated mathematics skill with an isolated reading skill would have been misleading. In hindsight, it seems important to be aware of such differences.

Examination of the figures is probably a reasonable way to make judgements about unidimensionality. Still, there is often a felt need for a statistic and a significance test. Keeping in mind caveats about the assumptions of regression analysis, it seems that examination of residuals would be one way to procede. In final analysis one must take a close look at item content. Here factor analysis proved to be a useful adjunct.

BIBLIOGRAPHY

Barnett, V. & Lewis, T. Outliers in Statistical Data. New York: Wiley, 1978.

Lord, F. & Novick, M. Statistical theories of mental test scores. Menlo Park, California; Addison Wesley, 1968.

Lord, F. Applications of Item Response Theory to Practical Testing Problems Hillsdale, New Jersey; Earlbaum, 1980.

Mislevy, R. and Bock, D. Estimation of CAP scaled scores. International Educational Services Technical Report # 100. Chicago; International Educational Services, (undated).

Wiley, D. & Harnischfeger, A. Assessing and Reporting CAP Results. Kenilworth, Illinois: The Beacon Institute, Inc., 1981.

1045-0403-7 r0501-4 es346/lca/B2

Table 1

Survey of Basic Skills: Grade 3
Reading Items

Skill area	Number of items
Word identification	60
Phonics	30
Vowels	15
Consonants	15
Structural analysis	30
Prefixes, suffixes, and roots	18
Contractions and compound words	12
Vocabulary	30
Recognizing word meanings	16
Using context	14
Comprehension	150
Literal	74
Details	37
- From a single sentence	20
- From two or three sentences	17
Pronoun references	18
Sequence	19
Inferential	76
Main ideas	19
Cause and effect	20
Drawing conclusions	37
- About characters	15
- From details	12
- From overall	10
Study locational	30
Alphabetizing	15
Table of contents	15

Table 2

Survey of Basic Skills: Grade 3
Mathematic Items

Skill area	Number of items
Counting and place value	45
Skills	30
Applications	15
Operations	155
Basic facts	25
Addition	30
Subtraction	30
Multiplication	30
Applications	40
Basic facts	13
Addition/subtraction	15
Multiplication	12
Nature of numbers and properties	45
Properties and relationships	15
Money and fractions	15
Applications	15
Geometry	30
Skills	20
Applications	10
Measurement	40
Linear measures	15
Other measures	15
Applications	10
Patterns and graphs	30
Skills	15
Applications	15
Analysis and models	15

Table 3

STATISTICAL ANALYSIS SYSTEM

NAME	E1	E2	D1_D2	D2_MEAN	RATIO
VOWELS	3.79334	1.00613	2.788	0.222	0.079
CONSONANTS	3.81205	0.98967	2.822	0.203	0.073
PREFIXES SUFFIXES AND ROOTS	4.40442	1.04520	3.359	0.261	0.078
CONTRACTIONS	3.26113	0.96769	2.295	0.188	0.082
WORD MEANING	4.15046	0.99105	3.159	0.220	0.070
CONTEXT	3.47963	0.99375	2.506	0.202	0.080
SINGLE SENTENCES	4.20912	1.08879	3.120	0.272	0.087
TWO OR THREE SENTENCES	3.46451	1.05307	2.411	0.224	0.092
PRONOUNS	4.15753	1.06792	3.090	0.270	0.087
SEQUENCES	3.73592	1.11265	2.624	0.280	0.107
MAIN IDEA	4.34277	1.01780	3.325	0.215	0.065
CAUSE AND EFFECT	4.72079	1.06765	3.653	0.278	0.076
ABOUT CHARACTER	3.84254	1.01641	2.826	0.236	0.084
FROM DETAILS	2.79704	1.02659	1.770	0.211	0.119
OVERALL MEANING	2.77484	0.93722	1.838	0.151	0.082
ALPHABETIZING	3.23627	1.03920	2.197	0.214	0.093
TABLE OF CONTENTS	4.11222	0.97017	3.143	0.207	0.066
COUNTING AND PLACE VALUE SKILL	4.48298	1.35243	3.131	0.489	0.156
COUNTING AND PLACE VALUE APPLI	2.89719	1.15444	1.743	0.312	0.179
BASIC FACTS OPERATIONS	3.20732	1.65462	1.553	0.779	0.502
ADDITION OPERATIONS	3.42149	1.47795	1.944	0.582	0.299
SUBTRACTION OPERATIONS	3.95433	1.67908	2.276	0.809	0.355
MULTIPLICATION OPERATIONS	4.37074	1.38740	2.984	0.522	0.175
BASIC FACTS APPLICATIONS	2.33419	1.22082	1.113	0.362	0.325
ADD AND SUBSTRACT APPLICATIONS	2.17420	1.20009	0.974	0.306	0.314
MULTIPLICATION APPLICATIONS	2.41209	1.07589	1.336	0.225	0.168
NATURE OF NUMBER PROPERTIES	2.59328	1.18024	1.413	0.317	0.224
NATURE OF NUMBERS MONEY	2.25619	1.45538	0.801	0.587	0.733
NATURE OF NUMBER APPLICATIONS	2.45961	1.13259	1.327	0.255	0.192
GEOMETRIC SKILLS	2.90831	1.34872	1.560	0.474	0.304
GEOMETRY APPLICATIONS	2.12554	1.10355	1.022	0.257	0.252
LINEAR MEASURES	2.52374	1.18732	1.336	0.319	0.239
OTHER MEASURES	2.56191	1.23282	1.329	0.371	0.279
MEASUREMENT APPLICATIONS	2.07875	1.08181	0.997	0.227	0.228
PATTERNS AND GRAPHS	2.77336	1.14513	1.628	0.293	0.180
PATTERNS AND GRAPHS APPLICATIO	3.06108	1.21085	1.850	0.336	0.203
ANALYSIS AND MODELS	2.17009	1.17168	0.998	0.275	0.275

Table 4

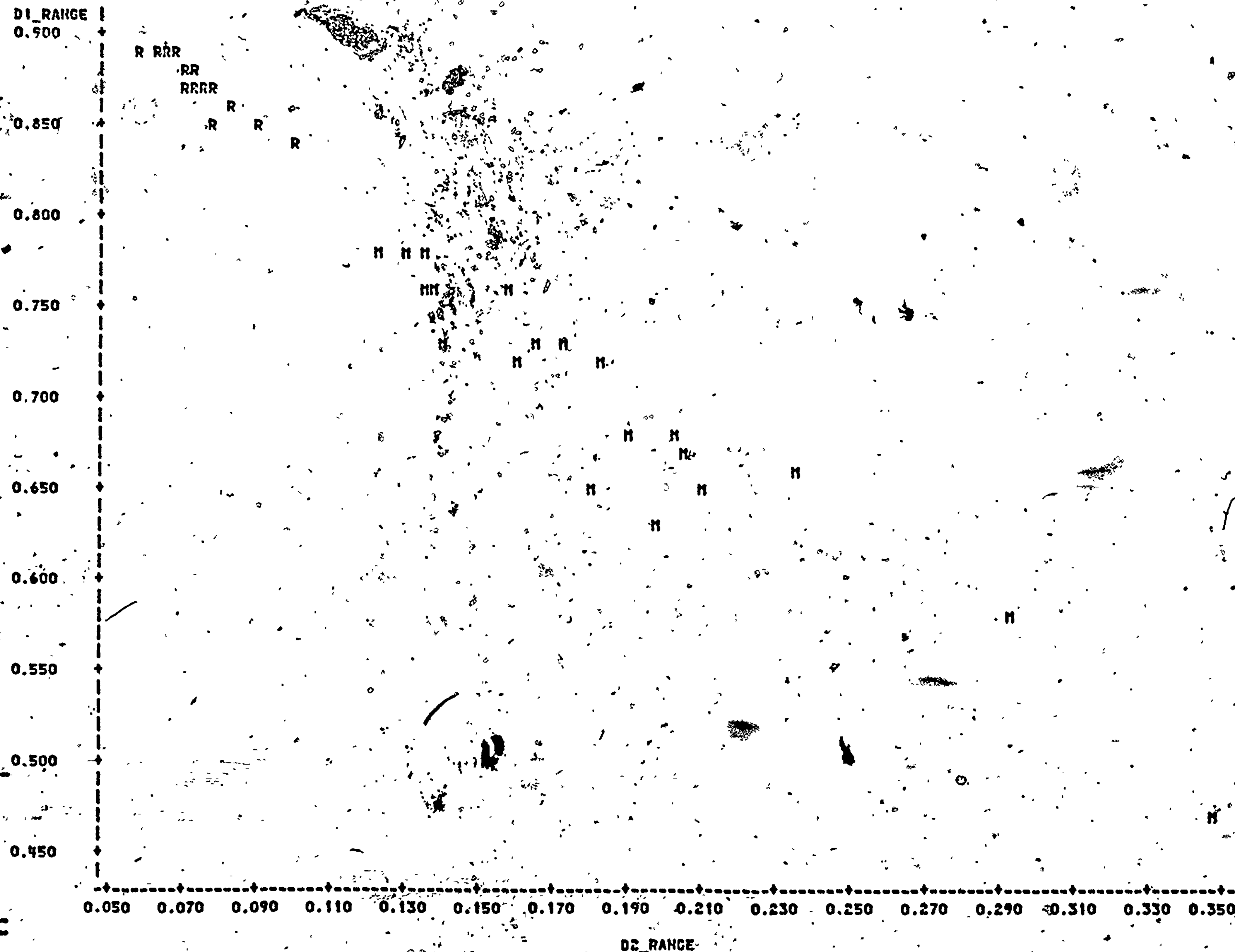
VARIMAX ROTATED FACTOR MATRIX

	FACTOR1	FACTOR2	FACTOR3	FACTOR4	FACTOR5
Z216-1	0.09046	0.03195	0.23224	-0.10898	0.69658
Z217-2	0.69334	0.03022	-0.12663	0.01893	0.02761
Z218-3	0.57113	0.16890	-0.03853	0.06809	0.25029
Z219-4	0.38938	-0.05232	0.14114	-0.43360	0.31083
Z220-5	0.08596	0.06332	-0.76075	0.02343	-0.03806
Z221-6	0.68370	-0.16599	0.00185	-0.29597	0.05600
Z222-7	-0.25979	0.12374	-0.18183	-0.37703	0.40907
Z223-8	-0.10430	0.67792	-0.01791	0.01535	0.03725
Z224-9	0.04499	-0.15221	-0.48093	-0.38966	0.05449
Z225-10	0.13151	0.70842	0.02457	-0.03269	0.07523
Z226-11	0.10731	-0.13670	-0.44683	0.19633	0.49213
Z227-12	0.07687	0.32763	0.04338	-0.45574	-0.11599
Z228-13	0.04496	-0.01896	-0.10358	-0.69313	0.06534
Z229-14	-0.02069	0.37002	-0.40731	-0.26628	-0.04421
Z230-15	0.21133	0.06319	-0.09681	-0.00784	0.54099

Figure 1.

EIGENVALUE DIFFERENCE STATISTICS

PLOT OF D1_RANGE*D2_RANGE   SYMBOL IS VALUE OF AREA



NOTE: 3 OBS HIDDEN

Figure 2

EIGENVALUE DIFFERENCE STATISTICS

PLOT OF D1_D2*D2_MEAN   SYMBOL IS VALUE OF AREA

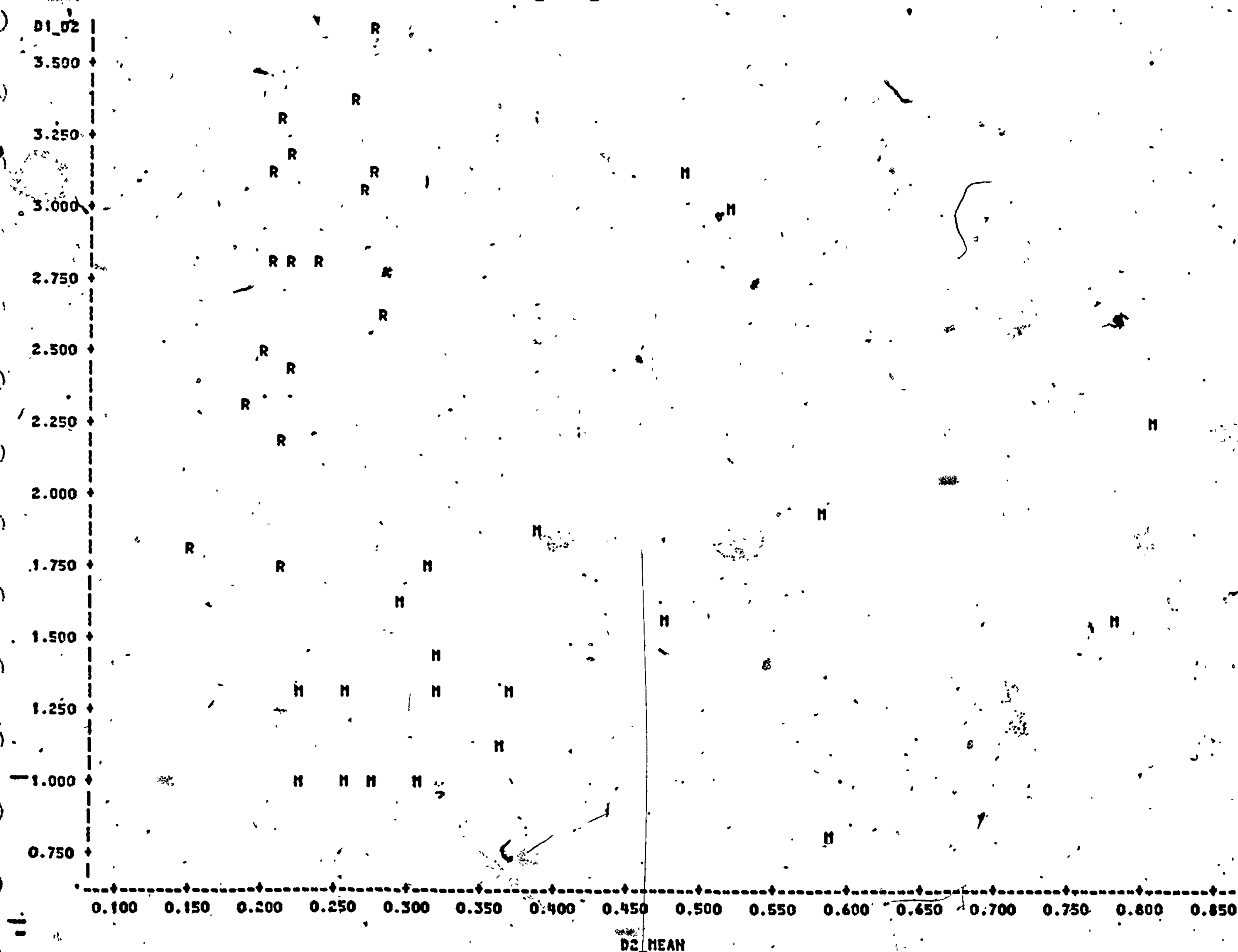


Figure 3

ANALYSIS OF UNTRANSFORMED SCORES
NAME=NATURE OF NUMBERS MONEY

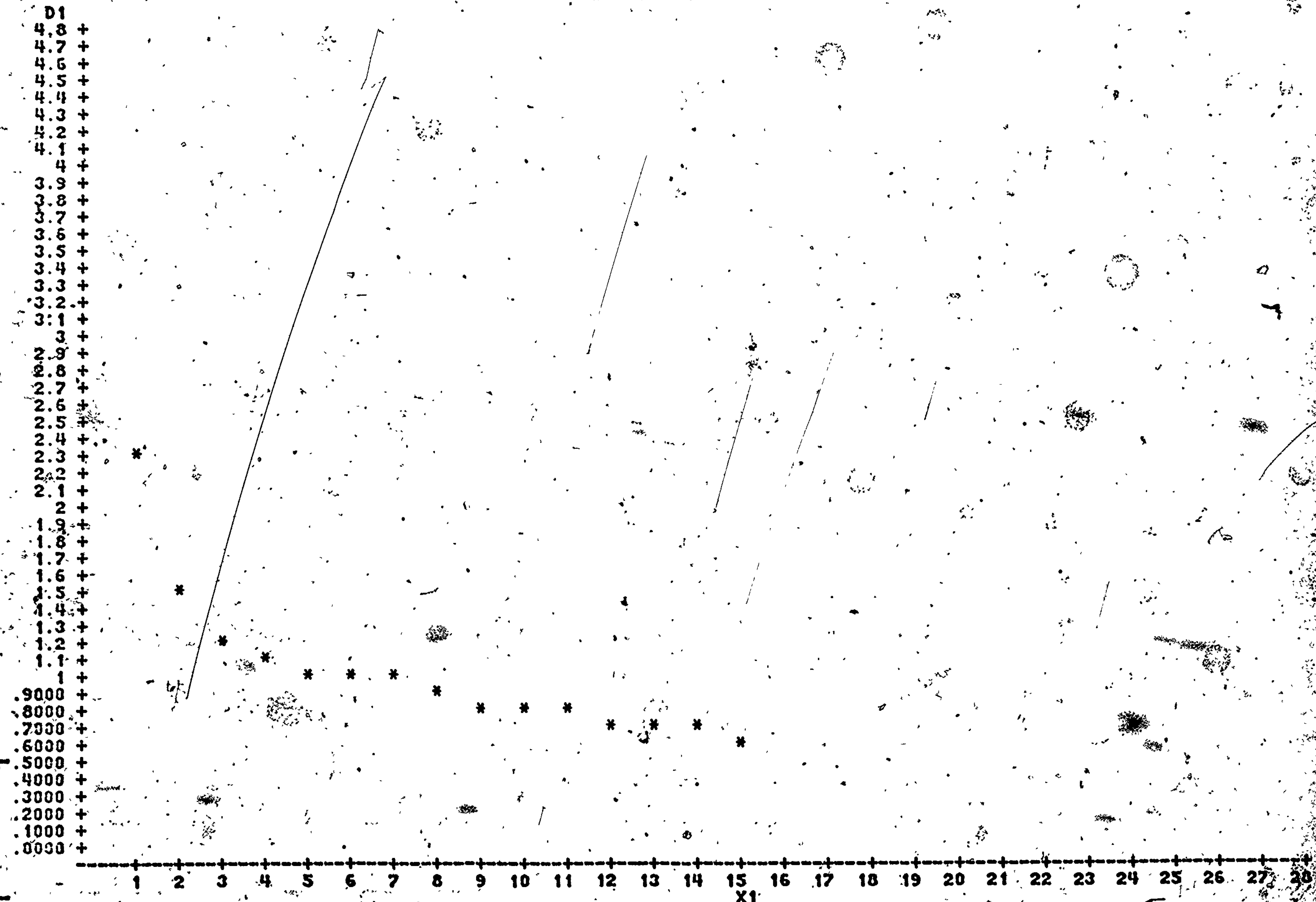


Figure 4

ANALYSIS OF UNTRANSFORMED SCORES
NAME=NATURE OF NUMBER APPLICATIONS

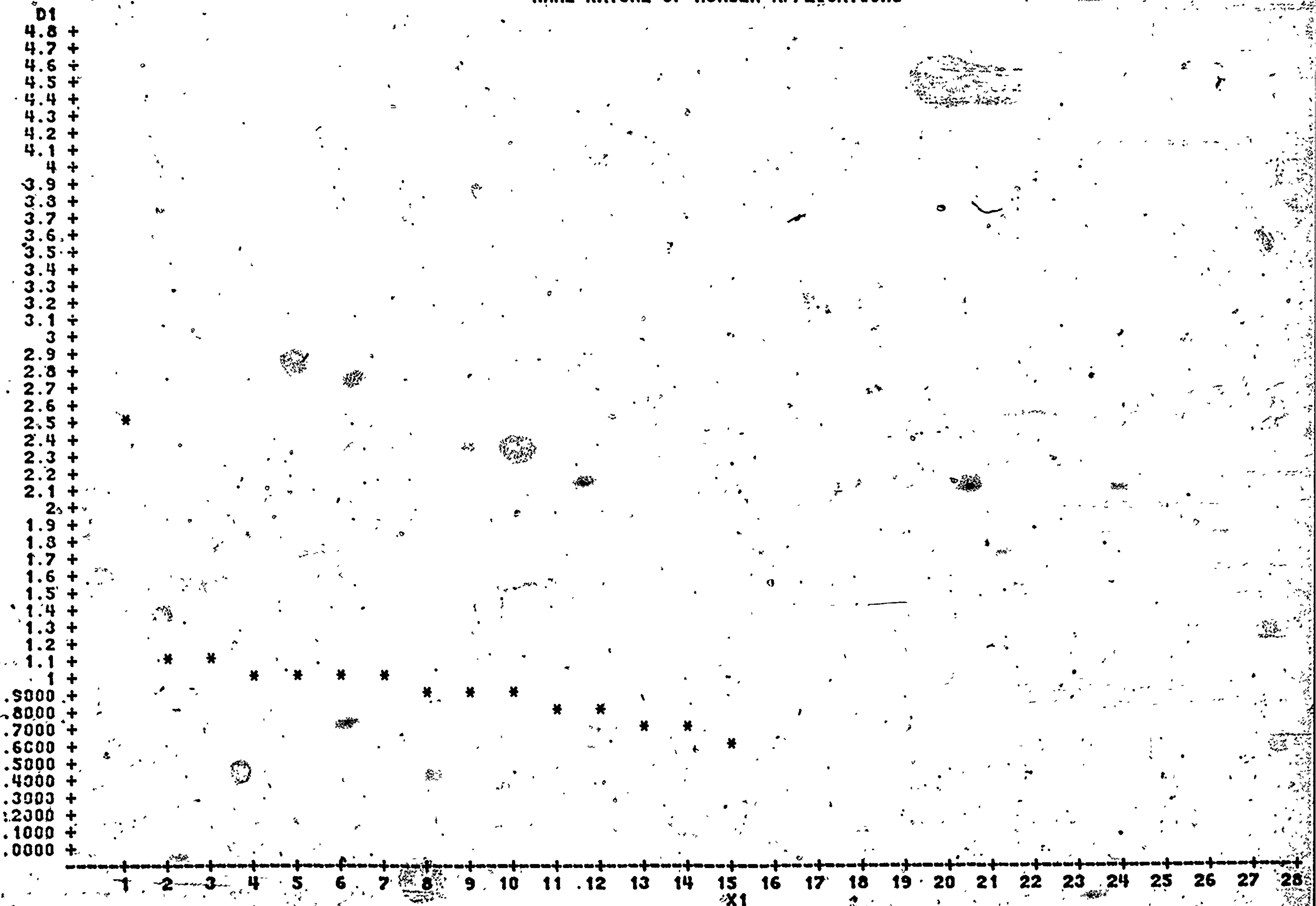


Figure 5

LOG OF EIGENVALUES

NAME=NATURE OF NUMBERS MONEY

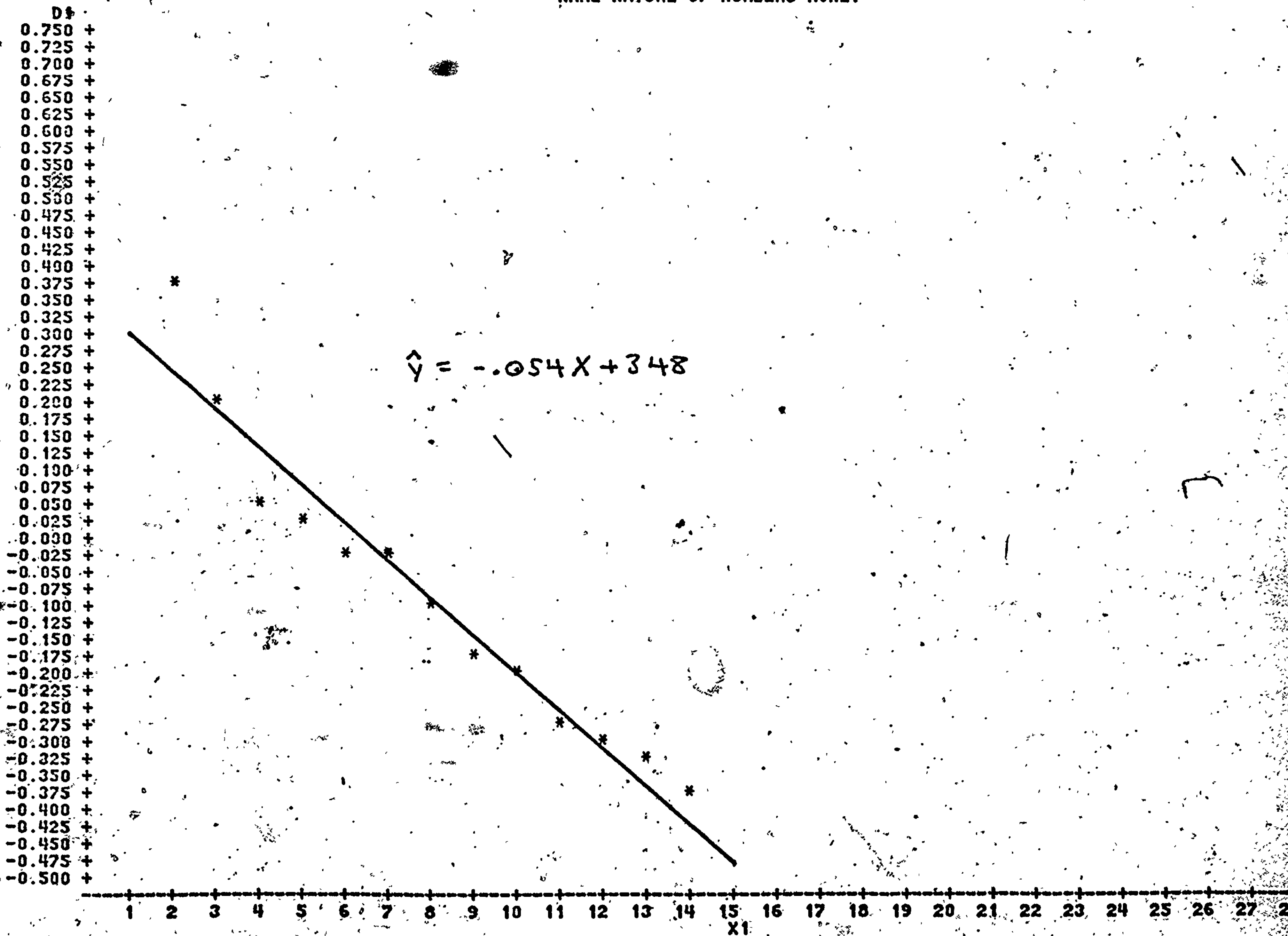


Figure 6

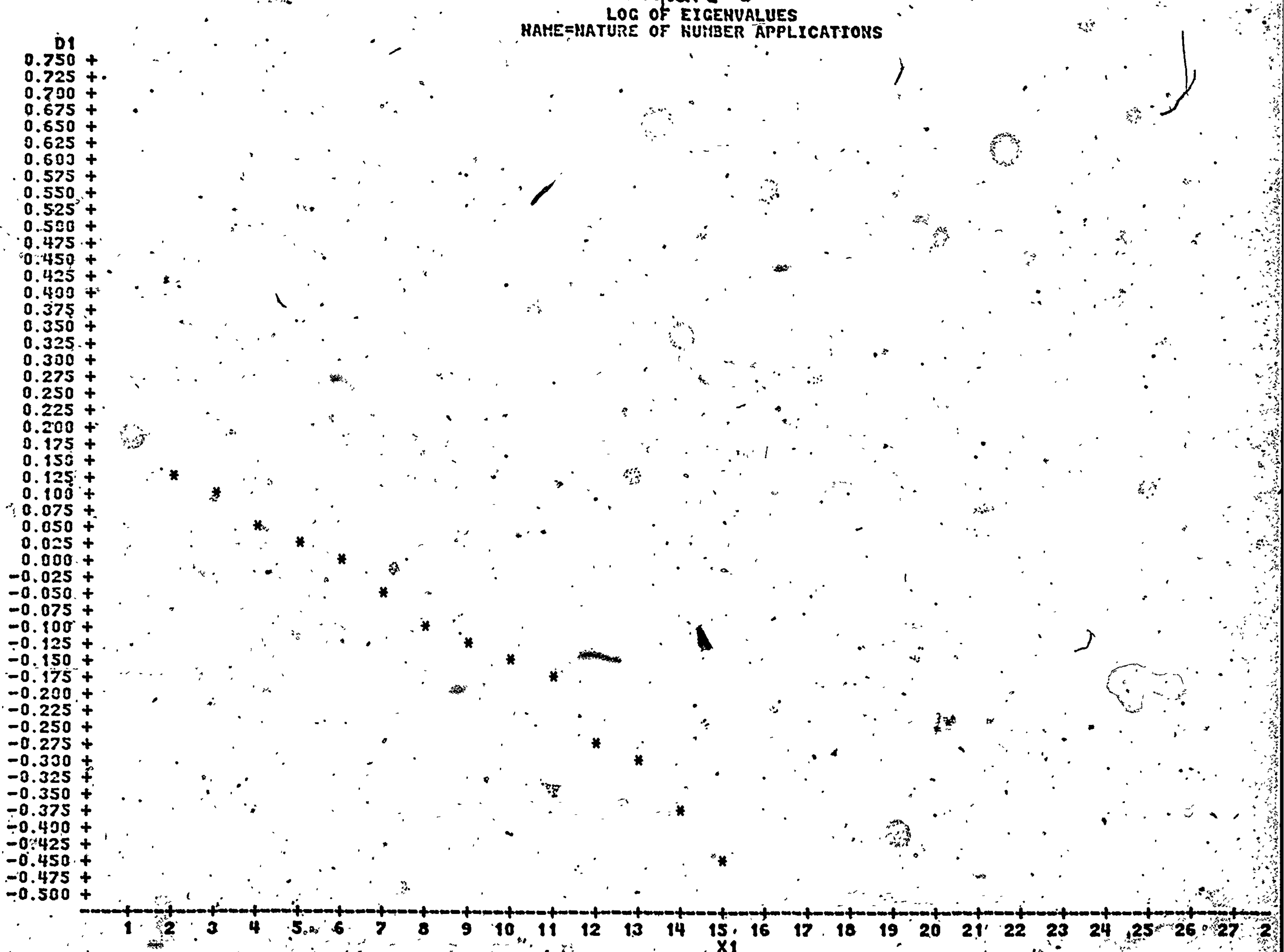

LOG OF EIGENVALUES
NAME=NATURE OF NUMBER APPLICATIONS
D1
X1

Figure 7

DIFFERENCE OF LOGS OF EIGENVALUES

NAME=NATURE OF NUMBERS MONEY

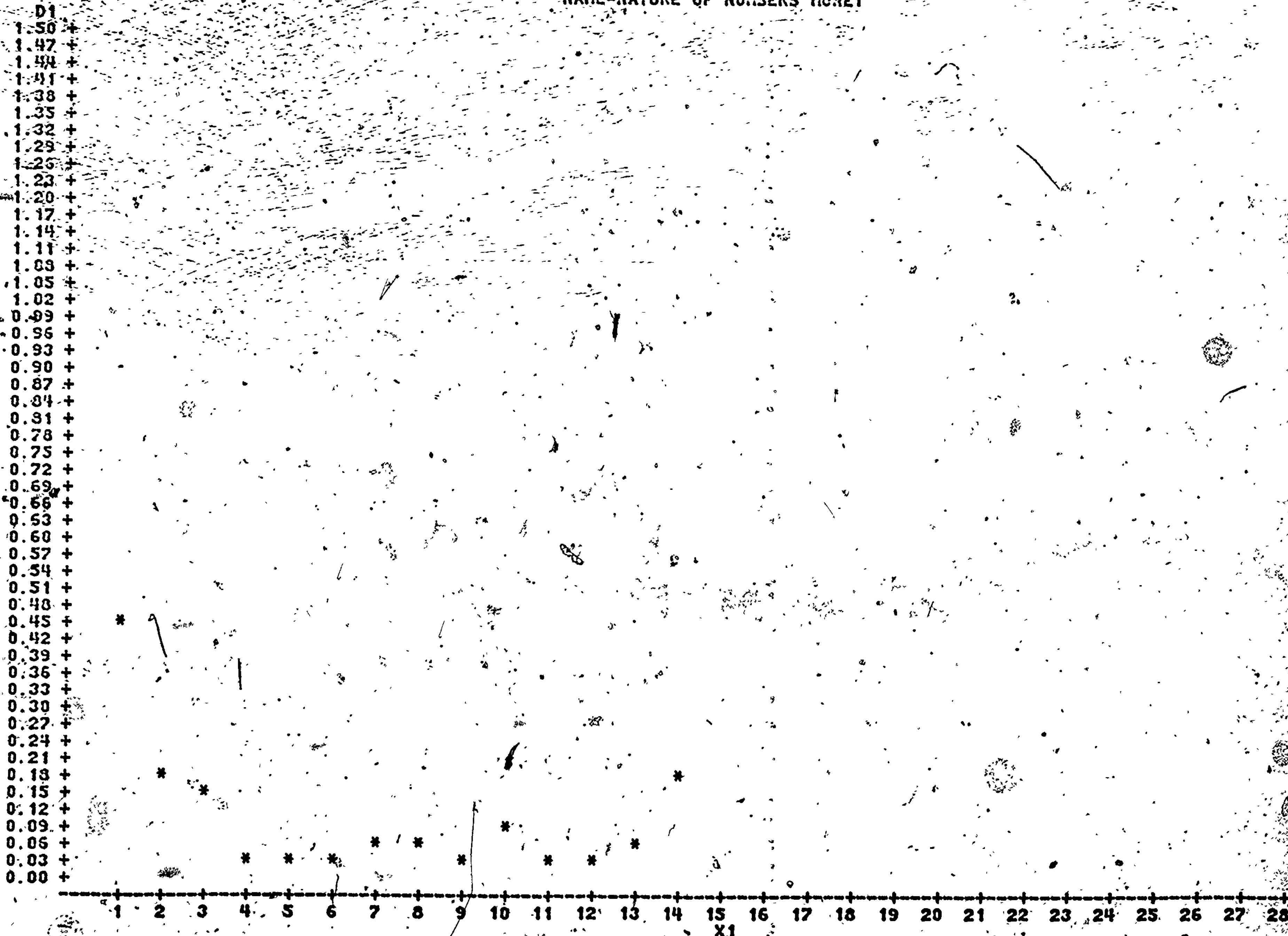
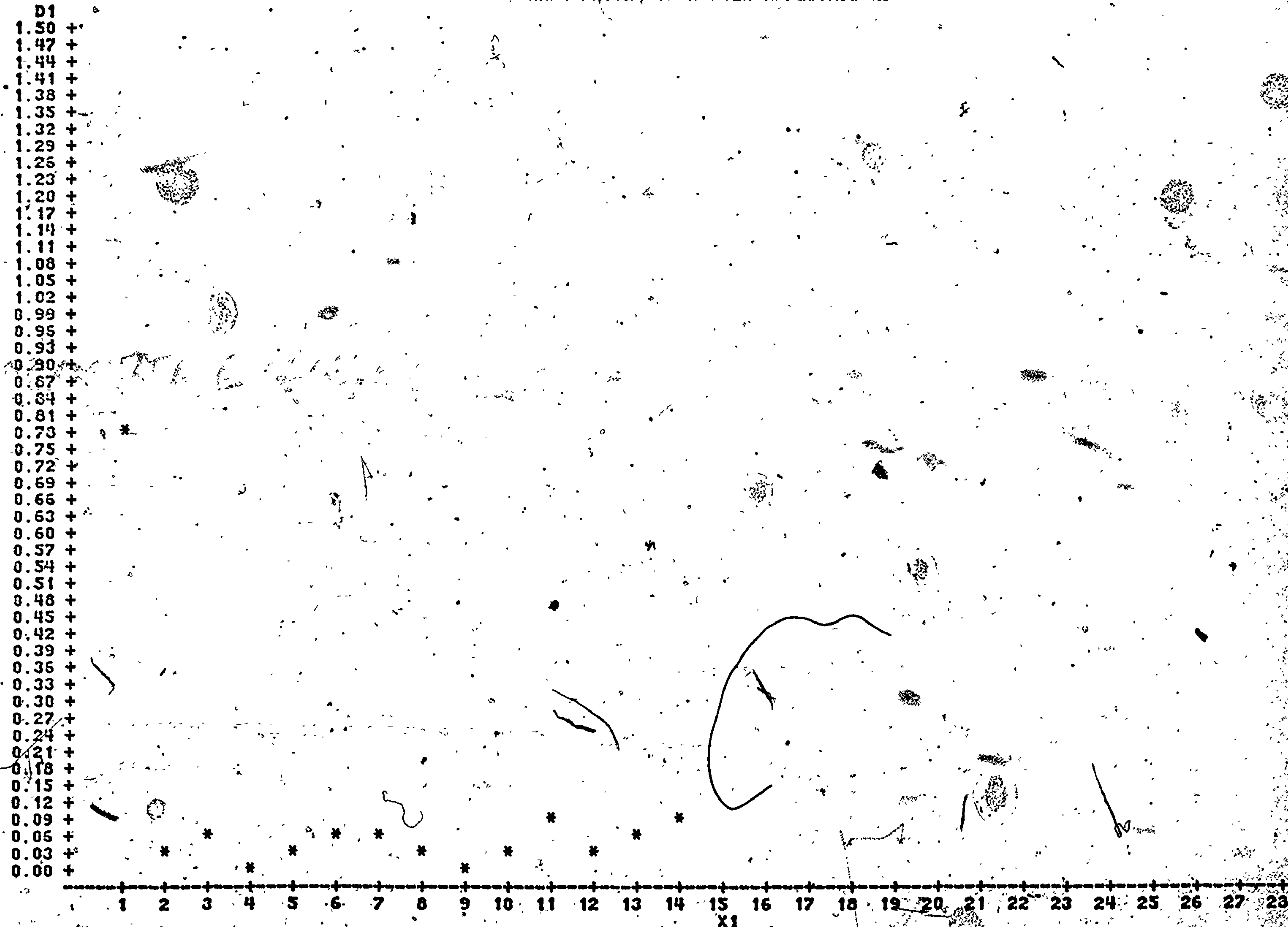


Figure 8

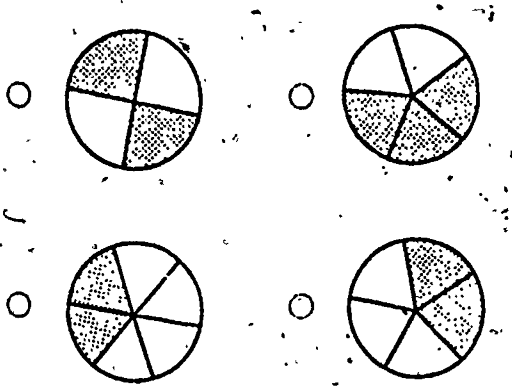
DIFFERENCE OF LOGS OF EIGENVALUES
NAME=NATURE OF NUMBER APPLICATIONS



APPENDIX

25

Which figure is $\frac{2}{5}$ shaded?



26

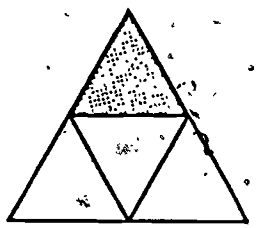
What fraction of the shape is shaded?

○ $\frac{1}{4}$

○ $\frac{4}{3}$

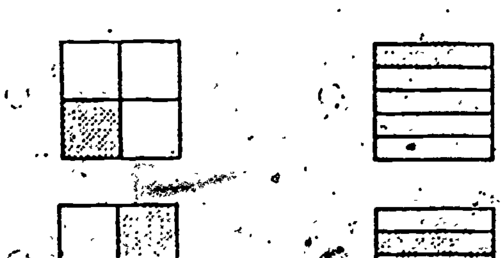
○ $\frac{3}{1}$

○ $\frac{4}{4}$



23

Which figure is $\frac{1}{4}$ shaded?



21

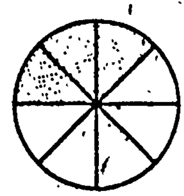
What fraction of the figure is shaded?

○ $\frac{3}{8}$

○ $\frac{3}{5}$

○ $\frac{5}{3}$

○ $\frac{3}{7}$



21

40¢ is:

21

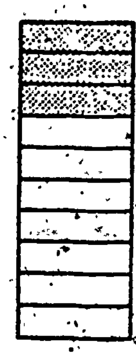
What fraction of the figure is shaded?

○ $\frac{3}{10}$

○ $\frac{4}{10}$

○ $\frac{7}{3}$

○ $\frac{3}{7}$



7

18

Which coin is this?

- ○ quarter
- ○ nickel
- ○ penny
- ○ dime

11

19

Three dollars and four cents is:

- ○ $3.40
- ○ $3.04¢
- ○ $30.4
- ○ $3.04

8

19

How much is this?

- ○ 5¢
- ○ 1¢
- ○ 10¢
- ○ 25¢

12

24

22¢ is:

○ 

○ 

○ 

○ 

9

18

One dollar and thirty-six cents is:

- $1.36
- $13.6
- $136
- $.136

10

21

How much is this?

- ○ 5¢
- ○ 25¢
- ○ 1¢
- ○ 10¢

24

How much is this?

- ○ 18¢
- ○ 82¢
- ○ $1.00
- ○ $1.20

21

How much is this?

- ○ 8¢
- ○ 28¢
- ○ 36¢
- ○ 37¢

23

Which picture is shaded one whole?