AD 271 148

A Machine Search for Moderator Variables in Massive Data

JOHN W. FRENCH

Office of Naval Research Contract Nonr-2214(00)
Project Designation NR 151-174
John W. French, Principal Investigator

Educational Testing Service

PRINCETON, NEW JERSEY

May 1961

A MACHINE SEARCH FOR MODERATOR VARIABLES IN MASSIVE DATA

John W. French

Office of Naval Research Contract Nonr-2214(00)
Project Designation NR 151-174
John W. French, Principal Investigator

Educational Testing Service
Princeton, New Jersey
May 1961

A MACHINE SEARCH FOR MODERATOR VARIABLES IN MASSIVE DATA

Abstract

The IBM 650 was programmed to compute indices which would represent validities for all possible predictors based on subjects standing low, middle, and high on all possible "moderator variables" in massive data. The indices were derived from the 3x3 contingency tables resulting from a pairing of predictors and criterion scores after both had been trichotomized. This method permitted the computational speed necessary to search a large quantity of data in a reasonable time, but in doing so it allowed distortion as a result of skewness in the distributions of the variables and of correlation between the moderator variable and either the predictor or the criterion. The program was tried out on three groups of subjects for whom 42 predictors were available. The resulting indices pointed to a small number of situations where moderator variables were operating effectively and in the same direction for two and sometimes three separate large groups of subjects. Corresponding product-moment correlations were computed. The relationship between the indices and the correlations was found to be so affected by distortion as to be unsatisfactory. For this reason the program in its present form cannot be recommended for general use. Satisfactory searching for moderator variables in this kind of data will require a larger machine or a more time-taking, more expensive

computational procedure. Nevertheless, the tryout of the present program did reveal a clear moderator variable: high visual-spatial ability interferes with the correlation between interest in English and grades in English. A possible reason for this is described.

A MACHINE SEARCH FOR MODERATOR VARIABLES IN MASSIVE DATA

Even for the same criteria some tests have been found to possess quite different validities when tried out with different groups of people and in different situations. It has sometimes been possible to find tests or other measures, to be called moderator variables, that will classify people into groups for which a test has a certain amount of validity for some criterion and other groups for which it has a very different amount of validity. A knowledge of the particular classification of this kind in which an individual belongs might be very important in the interpretation of his test scores. Classification for this purpose may be made either on the basis of some qualitative point-distribution or on the basis of ranges such as high, middle, and low on a continuous variable. Saunders (1956) has discussed the characteristics of the curved regression surface with which a criterion can be predicted by a predictor and a moderator variable. An illustration of a continuous variable acting as a moderator was found, for example, in connection with the validity of the Strong Engineering Interest score for success at an engineering college (Frederiksen and Melville 1954; Frederiksen and Gilbert 1960). It was shown that the validity of Engineering Interest was high for students having low scores on Accounting Interest, while it was low for students having high scores on Accounting Interest. This result was anticipated on the hypothesis that students with high interest in accounting are so compulsive that they work toward college success whether or not they are interested in the subject being studied.

In connection with an experimental battery designed to study comparative prediction of academic criteria, considerable effort was expended in testing hypotheses concerning the existence of moderator variables among the battery's 42 aptitude, interest, and personality measures. For example, it was thought that, for persons of moderate standing on the variable "sociability," the validity of verbal test scores for English grades might be higher than the validity for the whole group, because the disturbing effects of over-socializing or social maladjustment would be removed. This hypothesis relates to a finding of Hoyt and Norman (1954) that maladjustment as measured by the MMPI lowers the validity of aptitude tests for college grades. However, a scatter-plot of the data did not bear out the hypothesis. A number of similar hypotheses were tried for a good many combinations of the 42 predictors and the several dozen available criteria. However, a successful moderator variable was not found by this method. The method of trying out specific hypotheses is slow and expensive because of the tremendous number of possibilities that could be tried.

One possibility for obtaining a moderator variable without having to rely on successful psychological hypotheses is to develop a scale empirically. Ghiselli (1960) has shown this to be possible. He assigned a score to each of his subjects according to how well for them a certain criterion could be predicted. Questionnaire items were then selected for their relationship to this score. The scale proved to be a successful moderator variable on a new group of subjects.

Since the present data consisted of numerous, fairly homogeneous variables and since it seemed desirable to set up a method for finding existing moderator variables routinely in future data, the problem was conceived as a search for a variable with the appropriate moderator characteristics rather than the synthesis of a "predictability" scale consisting of items selected for these characteristics.

In order to effect such a search, the IBM 650 data processing machine was programmed to test the possibility that any of the 42 variables could serve as a moderator variable controlling the validity of any other variable for any of several criteria.

The data from the comparative prediction study (French 1959) which were available for use in this experiment included pure factor aptitude tests, factored personality scales, and interest inventory measures. The variables were given the following names:

1. Mathematics Reasoning
2. Associative Memory
3. Integration
4. Visualization
5. Mechanical Knowledge
6. Number Speed
7. Space and Orientation
8. Speed of Judgment
9. Fluency of Expression
10. Aiming
11. Motor Speed
12. Speed of Symbol Discrimin.
13. Carefulness
14. Meaningful Memory
15. Verbal Comprehension
16. Induction
17. English Interest
18. Language Interest
19. Mathematics Interest
20. Social Studies Interest
21. Biology Interest
22. Physical Science Interest
23. Music Interest
24. Fine Arts Interest
25. Industrial Arts Interest
26. Business Interest
27. Home Economics Interest
28. Sports Interest
29. Manipulative Interest
30. Reading Interest
31. Surgency
32. Sociability
33. Self-sufficiency
34. Tolerance
35. Self-confidence
36. Persistence
37. Dominance
38. Autistic Tendency
39. Foresight
40. Gregariousness
41. Nervousness
42. Emotionality

The above tests and other measures were all obtained during three hours of testing time. Because of the shortness of the time limits some of the reliabilities run down into the fifties, perhaps lower for some of the personality variables. While this kind of reliability is not satisfactory for individual diagnosis, the battery was planned on the supposition that the reliability of the tests was satisfactory for group studies.

<u>General Nature of the Machine Program</u>

Since the purpose of the machine program was to explore all possibilities for moderator variables in data containing a large number of measurements, a very lengthy job was involved. It was anticipated that, out of many hundreds of possibilities, only a few valid moderator variables would be found. The most efficient program to fullfill such a purpose was considered to be one that would locate moderator variable situations rapidly among the great mass of possibilities without taking time to compute actual correlations. The expectation was that, once the relatively few moderator variables and the relationships affected by them were spotted, it would be a simple matter to compute the exact correlations.

In cognitive data, it was believed that the size of a validity coefficient might sometimes have a U-shaped or inverted-U-shaped relationship to a moderator variable. Therefore, the program was designed to compute an index of the validity of each predictor variable for subjects at each of three levels on each moderator: upper, middle, and lower thirds.

The program reported herein was designed for maximum speed with certain data having 42 predictors and one criterion. While a final, properly general program should eventually be set up for easy use with a flexible number of predictors, it was possible to speed up this trial program considerably by writing out instructions for exactly 42 predictors, so that it was not necessary for the machine to run repeatedly through loops designed to keep track of the number of predictors processed.

<u>The Indices that were Used in Place of Correlation Coefficients</u>

The first step was to trichotomize all variables into parts as equal as was permitted by tied values. Scores on all variables, including the criteria, then became 0, 1, or 2. The index used in place of a correlation coefficient was the sum over all people at one level on the moderator variable and over all predictors, X, of:

$1 - X$ when the criterion was 0

0 when the criterion was 1

$-1 + X$ when the criterion was 2 .

This resulted in the following values for the possible combinations of predictor and criterion:

Predictor \ Criterion	0	1	2
2	-1	0	1
1	0	0	0
0	1	0	-1

A positive correlation would yield high frequencies of cases in the lower-left to upper-right diagonal and would, therefore, be represented by a positive index. A negative correlation would yield high frequencies in the upper-left to lower-right diagonal, and so would be represented by a negative index.

If the distributions were symmetrical and neither the predictor nor the criterion was correlated with the moderator variable, the sum of predictor scores and the sum of criterion scores would both equal zero, and the index would be exactly proportional to the product-moment correlation. However, the moderator variable was sometimes rather unevenly trichotomized because of tied scores and was often considerably correlated with either or both of the two variables whose correlation was being considered. In such instances an uneven division of cases with respect to the high, middle, or low ranges of the moderator variable resulted in a skewing of the frequencies in the 3x3 chart above. In addition, some of the variables in the present study had distributions that were substantially non-normal. Depending in an unknown (or very complex) way on the frequencies of each variable in each cell, the index was subject to distortion from what it would have been had there been symmetric and even distributions. A few hypothetical frequencies such as might arise from correlated variables were tried in the 3x3 matrix, and the index was computed. Distortion did not seem great enough to interfere seriously with the usefulness of the method.

Steps in the Programmed Procedure

1. On the basis of known means and standard deviations of the variables, two cutting scores were selected that would divide the distribution into three maximally equal categories. For normally distributed variables, the cutting scores would be the mean plus or minus .43 times the standard deviation. The cutting scores for each variable were inserted into the program as constants.

2. The trichotomizing program substituted the values 0, 1, or 2 for each score and punched for each subject a card containing the trichotomized scores. This program operated at about 5 seconds per case for the 42-predictor problem.

3. If a count of these new scores made by a short supplementary program revealed gross imbalances, new cutting scores were used and the trichotomizing process was repeated.

4. The resulting cards were used in the main program, where they pass through the machine once for each moderator variable tried out. For the 42-predictor problem, the speed was about 75 cases per minute on each such pass. A loop at the end of each pass increased the variable-number designating the moderator variable by one.

5. A branching on the score for the moderator variable determined, for each case, to which of three outputs the computations for the case would be applied.

6. One was subtracted from the criterion. If the criterion was then negative, each predictor score was subtracted from one and the result was added to the output for that predictor. If the criterion was then positive, each predictor score was added to -1 and the result was added to the output for that predictor.

7. The resulting totals when cumulated over all cases constitute the desired indices. These were punched at the end of the pass, so that each card, representing a single predictor, contained the index for the low, middle, and high ranges on the moderator variable.

Results

The program was run for three groups of subjects: 542 women at Smith College, 449 men at Cornell University, and 306 women at Cornell University. For each of these groups, two criteria were available: (1) average freshman grades, included as variable 43 and (2) grades in freshman English, included as variable 44. Variable 43 was not used in this study.

For each group of subjects the program output cards were listed, placing in each row three indices, one for each level on the moderator variable. There was a row devoted to each of the 42 predictors, and a page of 42 rows devoted to each of the 42 possible moderator variables. The 1764 rows for each of the three groups of subjects were inspected clerically for differences of 20 or more for any pair of indices in a row. Such differences occurred about 8 times on the average out of a possible 126 times on each page. For 34 of the 1764 rows there was confirmation of a difference of 20 or more in the same direction for either the two Cornell groups or the two women's groups or for all three groups. Of these 34 rows, Table 1 gives the indices for 13 rows selected as those of most psychological interest. Directly under each index appears the computed product-moment correlation. Table 2 gives the indices for the other 21 rows; product-moment correlations were not computed.

Inspection of Table 1 will show that, while the indices are related to the corresponding correlations, they are not related closely enough to be satisfactory for the present purpose. Evidently the method used for creating the index introduced too much distortion as a price for achieving what seemed to be the necessary speed. The IBM 650 is, in fact, not at its best in a program like this one where relatively little computing must be done on a relatively large mass of data. Satisfactory searching for moderator variables in this kind of data will require a larger machine or a more time-taking, more expensive computational procedure for the IBM 650.

The correlations of the moderator variables and skewness in the moderator and predictor variables were studied in order to locate the principal source of distortion, but no simple relationship to distortion was found. Unevenness in trichoto-

Table 1

Program-Computed Indices of Validity Compared with Product-Moment Correlations

Moderator	Predictor		Cornell Males N= 449			Cornell Females N= 306			Smith (Females) N= 542		
			L	M	H	L	M	H	L	M	H
Visual.	Verbal	Index	24	64	20	6	30	9	12	31	7
		r	.48	.44	.32	.21	.36	.27	.31	.32	.28
Visual.	Eng. Int.	Index	12	38	16	14	26	-1	12	0	-2
		r	.31	.23	.18	.39	.21	-.01	.27	.20	.02
Visual.	Lang. Int.	Index	21	33	12	14	26	0	5	14	-3
		r	.49	.19	.16	.34	.17	-.02	.21	.26	.08
Visual.	Read. Int.	Index	-1	30	5	13	28	5	4	19	1
		r	.16	.20	.13	.40	.21	.10	.09	.14	.17
Space	Eng. Int.	Index	24	35	7	7	26	6	6	8	-4
		r	.30	.27	.12	.18	.26	.10	.13	.29	.10
Mot. Sp.	Induc.	Index	24	17	58	10	6	26	-17	-5	-44
		r	.25	.31	.11	.05	.03	.20	-.13	.03	.08
Care.	Verbal	Index	15	57	36	2	27	16	13	22	15
		r	.24	.49	.43	.30	.31	.36	.32	.31	.32
Care.	Induc.	Index	12	16	50	3	19	23	-5	-15	-35
		r	.21	.18	.26	.03	.11	.13	-.01	-.04	.06
Verbal	Induc.	Index	25	7	29	25	-4	18	2	-7	-64
		r	.16	.16	.15	.25	-.14	.05	-.06	-.02	-.03
Induc.	Verbal	Index	30	56	22	26	13	6	26	21	3
		r	.41	.48	.30	.44	.22	.19	.32	.30	.39
Eng. Int.	Integ.	Index	20	30	9	16	24	1	11	-3	-2
		r	.28	.28	.12	.34	.32	.04	.20	.06	.27
Surgency	Induc.	Index	14	18	41	15	5	40	-22	0	-27
		r	.20	.17	.33	.13	.11	.09	-.10	.10	-.06
Surgency	Emot.	Index	19	8	-7	15	-8	5	15	-8	2
		r	.19	.10	.01	.26	-.05	.09	.12	.04	.05

Table 2

Program-Computed Indices of Validity Not Selected for Comparison with the Product-Moment Correlations

Moderator	Predictor	Cornell Males			Cornell Females			Smith (Females)		
		L	M	H	L	M	H	L	M	H
Math.	I.A. Int.	-26	-6	-1	-9	12	5	-18	-20	-7
Assoc. Mem.	Induc.	9	25	39	1	17	31	-8	-13	-38
Mech. Kn.	Induc.	13	19	51	10	4	39	-16	-5	-23
Number Sp.	Bes. Int.	26	-4	0	10	-12	-5	-6	-22	-5
Fl. of Exp.	Induc.	27	14	56	0	7	37	-10	-9	-38
Fl. of Exp.	Man. Int.	-16	9	3	-4	17	-1	-9	-11	-5
Sp. of Dis.	Induc.	22	22	53	0	16	34	-8	-9	-37
Math. Int.	Induc.	12	28	40	8	9	46	-7	-17	-24
H.E. Int.	Art Int.	3	28	2	3	19	-1	0	-4	0
Sports Int.	Induc.	6	31	36	8	9	41	-5	-17	-28
Read. Int.	Math. Int.	3	-9	8	3	-14	15	-9	-25	15
Read. Int.	P.S. Int.	1	-4	9	15	-2	18	-14	-12	21
Sociab.	Induc.	13	21	48	5	13	35	-8	-15	-20
Sociab.	Math. Int.	-18	8	12	11	-9	2	6	-14	-11
Sociab.	Self Sup.	22	20	-8	17	10	-10	6	12	-1
Self Sup.	Sociab.	4	14	-26	13	-5	-14	-3	-7	-10
Tolerance	Verbal	32	59	17	12	27	6	18	26	6
Self Con.	Induc.	12	21	48	5	17	31	-7	-9	-29
Foresight	Lang. Int.	26	35	5	14	26	0	7	6	3
Foresight	Sports Int.	6	-17	-10	9	-15	-3	2	-21	-2
Greg.	Induc.	23	15	42	13	7	35	-21	-6	-11

mizing the moderator variable because of tied scores was certainly affecting the index, but this in itself was insufficient to explain more than a minor part of the discrepancies that were observed.

A Moderator Variable

The significance or the importance of a moderator variable, once it is found, is, of course, not related to how good or poor was the method for spotting it. As clumsy as this method has proved to be in its present form, the tryout of it has managed to spot a rather convincing moderator variable.

By referring now only to the correlation coefficients in Table 1, a very consistent pattern can be found in the rows represented by the moderator-predictor pairs Visual.-Eng. Int., Visual.-Lang. Int., Visual.-Read. Int., and Space-Eng. Int., for each of which a high value on the moderator variable is associated with low validity for the predictor. The moderators, Visualization and Space, are related abilities. The predictors that are involved, interest in English, language, and reading, are alike in being interest scales and are all psychologically reasonable predictors of English grades.

It remains only to conjecture why high visual-spatial ability disrupts the correlation between English-related interests and grades in English. It was hoped that inspection of the scatter-plots would provide a clue. Figure 1 for males and Figure 2 for females show the scatter-plots for the two Cornell groups using English grades against interest in English with Visualization as the moderator. It can be seen in the plots (with some difficulty) that the minuses, representing cases with low Visualization, depict steeper regression than the pluses. However, there seem to be no striking characteristics in the groupings of points that can suggest a reason for the influence of the moderator. For the same two groups of Cornell students Table 3 shows some of the pertinent correlations. Low negative correlations are found between Visualization and English Interest, but that in itself cannot bring about a moderator effect. Visualization has a positive correlation with Mathematics Interest and Physical Science Interest. It is possible that students with high Visualization have interests that conflict with their interest in English, particularly in view of the crossover shown by the high correlation between Reading Interest and Physical Science Interest. For example, a student with high reading interest may express interest in both English and Science, but, if he is good at Visualization, his actual effort may be directed toward science and his English grades will not reflect his high interest in English.

Figure I (males)

Scatter Plot of English Interest and English Grades Moderated on Three Levels by Visualization

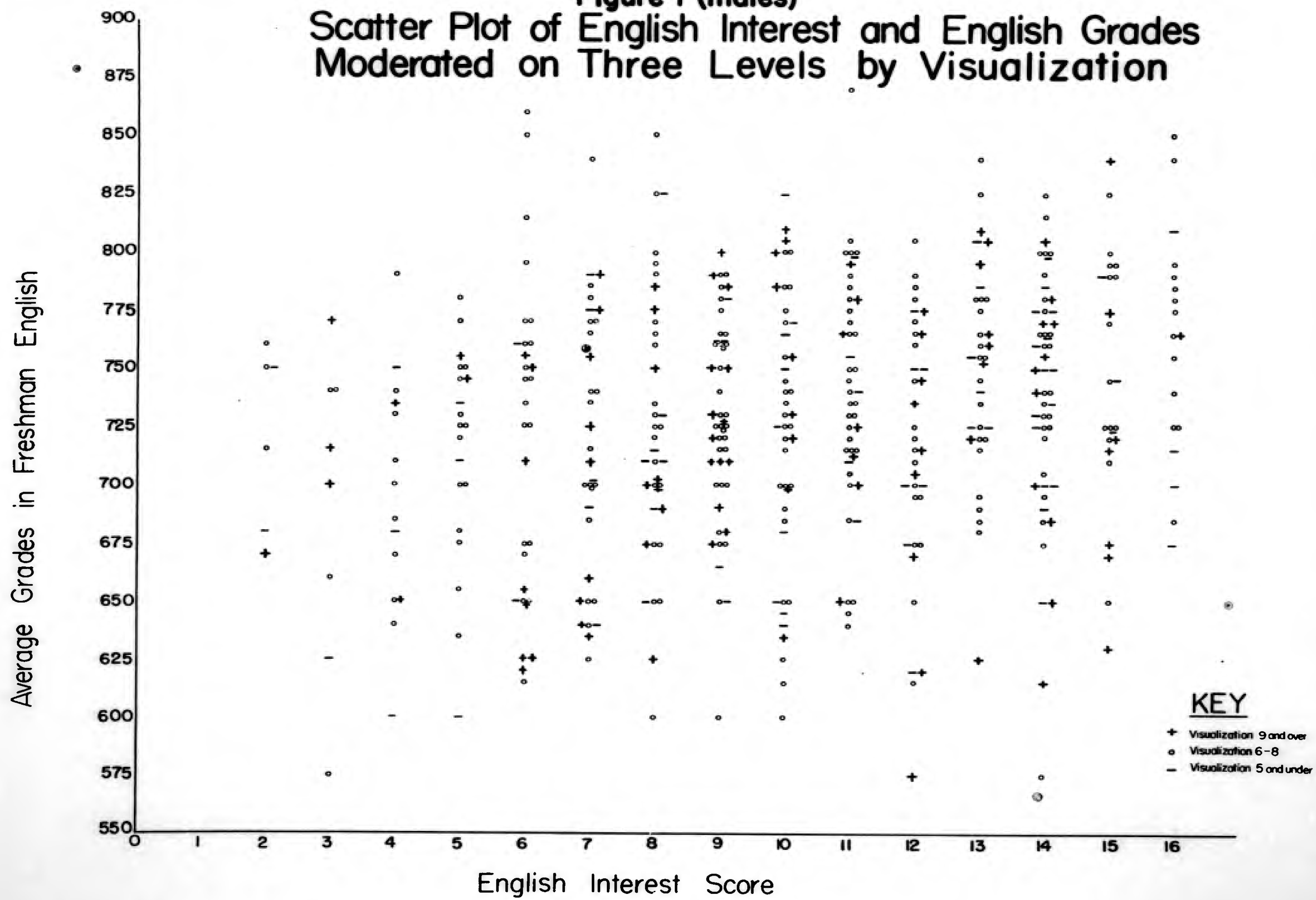

Figure 2 (Females)

Scatter Plot of English Interest and English Grades Moderated on Three Levels by Visualization

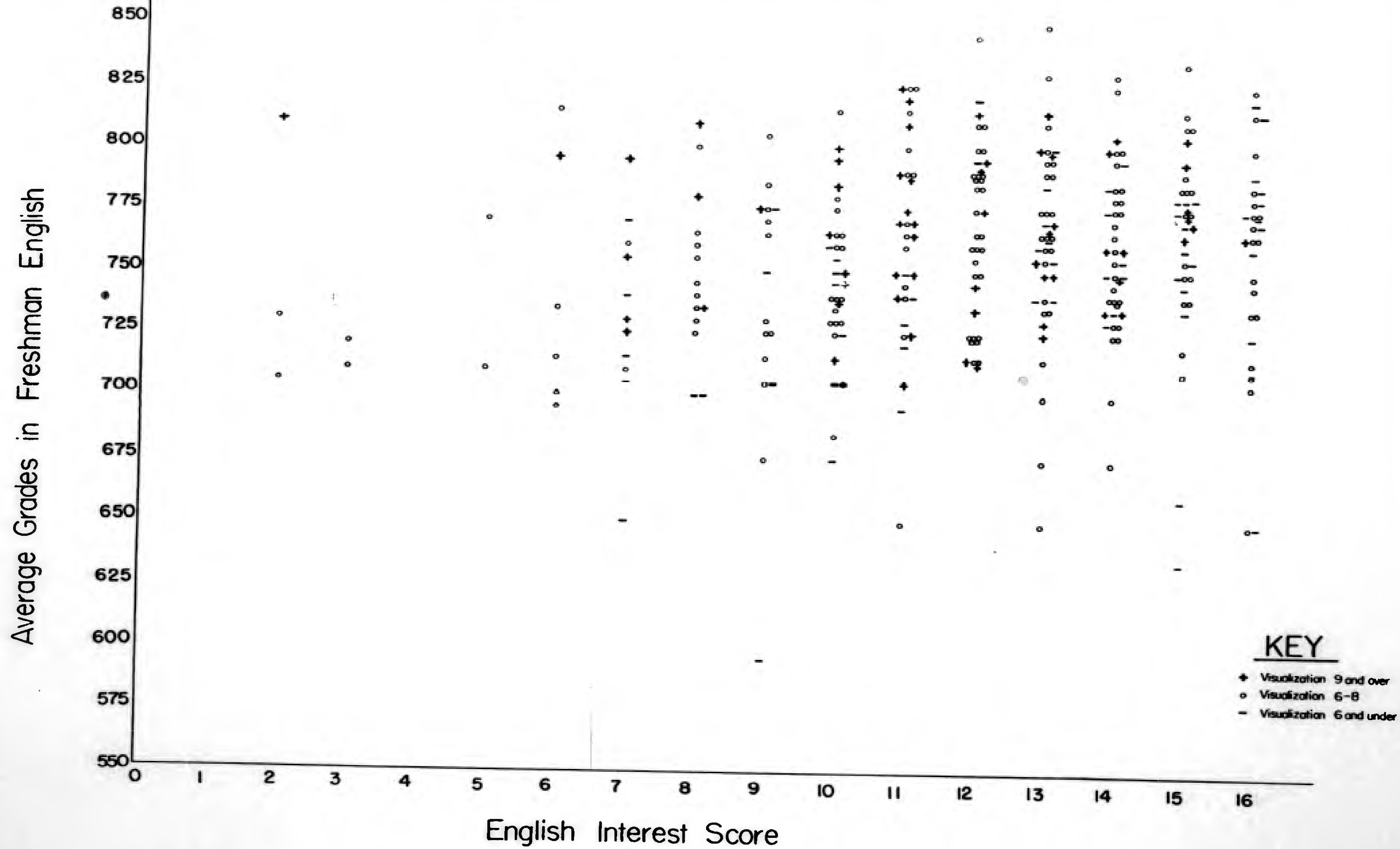

Table 3

Intercorrelations of English Grades and
Some of the Predictors

Males

	Vis.	Eng.	Math.	P.S.	Read.	Grades
Visualization		-06	27	15	04	03
English Interest	-06		-03	09	69	23
Math. Interest	27	-03		36	27	-06
Phys. Sci. Interest	15	09	36		43	00
Reading Interest	04	69	27	43		17
English Grades	03	23	-06	00	17	

Females

	Vis.	Eng.	Math.	P.S.	Read.	Grades
Visualization		-15	23	13	-03	16
English Interest	-15		-20	03	50	19
Math. Interest	23	-20		35	18	-01
Phys. Sci. Interest	13	03	35		48	13
Reading Interest	-03	50	18	48		22
English Grades	16	19	-01	13	22	

References

Frederiksen, N. and Melville, S. D. Differential predictability in the use of test scores. Educ. Psychol. Meas., 1954, 14, 647-656.

Frederiksen, N. and Gilbert, A. C. F. Replication of a study of differential predictability. Educ. Psychol. Meas., 1960, 20, 759-767.

French, J. W. Comparative prediction of success and satisfaction in college major fields, Part I. Research Bulletin 59-10, Princeton, N.J.: Educational Testing Service, 1959.

Ghiselli, E. E. The prediction of predictability. Educ. and Psychol. Meas., 1960, 20, 3-8.

Hoyt, D. P. and Norman, W. T. Adjustment and academic predictability. J. couns. Psychol., 1954, 1, 96-99.

Saunders, D. R. Moderator variables in prediction. Educ. Psychol. Meas., 1956, 16, 209-222.

UNCLASSIFIED