

AD-A084 354

ARIZONA STATE UNIV TEMPE COLL OF EDUCATION F/G 12/1
AN ANOMALY IN THE PEARSON PRODUCT-MOMENT CORRELATION COEFFICIENT--ETC(U)
DEC 78 R C HAYGOOD, V S GERLACH AFOSR-76-2900

UNCLASSIFIED

TR-81229

AFOSR -TR-80-0333

NL

1 of 1
AD
A084354



END
DATE
FILMED
6-80
DTIC

AEOSR-TR- 80-0338

CENTER FOR INSTRUCTIONAL AND
TRAINING TECHNOLOGY

LEVEL II



An Anomaly in the Pearson Product-Moment
Correlation Coefficient

Robert C. Haygood

Vernon S. Gerlach

USAF Office of Scientific Research

Grant No. 76-2900



DTIC
ELECTE

MAY 16 1980

S D

ARIZONA STATE UNIVERSITY
EDUCATIONAL TECHNOLOGY

Technical Report #81229

Approved for public release;
distribution unlimited.

80 5 14 040

ADA084354

DC FILE COPY

1. REPORT DOCUMENTATION PAGE		READ INSTRUCTIONS BEFORE COMPLETING FORM
2. REPORT NUMBER 18 AFOSR TR-80-0333	2. GOVT ACCESSION NO. AD A084354	3. RECIPIENT'S CATALOG NUMBER
4. TITLE (and Subtitle) 6 An Anomaly in the Pearson Product-Moment Correlation Coefficient		5. TYPE OF REPORT & PERIOD COVERED 9 INTERIM rept.
7. AUTHOR(s) 10 Robert C. Haygood Vernon S. Gerlach		8. CONTRACT OR GRANT NUMBER(s) 15 AFOSR-76-2900
9. PERFORMING ORGANIZATION NAME AND ADDRESS Arizona State University Tempe, AZ 85281		10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS 61102F 2313/P2
11. CONTROLLING OFFICE NAME AND ADDRESS Office of Scientific Research United States Air Force Bolling AFB, DC		12. REPORT DATE 11 December 1978
14. MONITORING AGENCY NAME & ADDRESS (if different from Controlling Office) 12 10		13. NUMBER OF PAGES Nine
16. DISTRIBUTION STATEMENT (of this Report) Approved for public release; distribution unlimited.		15. SECURITY CLASS. (of this report) UNCLASSIFIED
17. DISTRIBUTION STATEMENT (of the abstract entered in Block 20, if different from Report) 16 2313 17 112		15a. DECLASSIFICATION/DOWNGRADING SCHEDULE
18. SUPPLEMENTARY NOTES 14 TR-81229		
19. KEY WORDS (Continue on reverse side if necessary and identify by block number) Statistics Correlation Measurement Pearson product-moment		
20. ABSTRACT (Continue on reverse side if necessary and identify by block number) An anomalous condition in the use of the Pearson product-moment correlation coefficient is reported in which decreasing variance of one or both variables is associated with increased absolute value of the coefficient, contrary to popular expectation. The reasons for this are discussed, along with a recommendation for a change in the method of presenting restriction-of-range phenomena in statistical instruction. 409989 <i>YU</i> <i>Unclassified</i>		

Rule Learning and Systematic Instruction In
Undergraduate Pilot Training

Vernon S. Gerlach, Principal Investigator

AN ANOMALY IN THE PEARSON PRODUCT-MOMENT
CORRELATION COEFFICIENT

Robert C. Haygood
Vernon S. Gerlach

Accession For	
NTIS GRA&I	☒
DOC TAB	☐
Unannounced	☐
Justification	
By _____	
Distribution/	
Availability Codes	
Dist.	Avail and/or special
A	

RESEARCH (AFSC)
and is
100-12 (7b).
A. D. BLOOM
Technical Information Officer

Technical Report #81229

Research sponsored by the Air Force Office of Scientific Research, Air Force Systems Command, USAF, under Grant No. AFOSR 75-2900. The United States Government is authorized to reproduce and distribute reprints for governmental purposes notwithstanding any copyright notation hereon.

College of Education
Arizona State University
Tempe, Arizona

December, 1978

DTIC
ELECTE
S MAY 16 1980 D
D

AN ANOMALY IN THE PEARSON PRODUCT-MOMENT
CORRELATION COEFFICIENT

An Anomaly in the Pearson Product-Moment Correlation Coefficient

When the range of one or both variables in a simple Pearson product-moment correlation is artificially restricted, the absolute value of the sample correlation coefficient is reduced below that of the true (population) coefficient. This phenomenon is responsible for the well known difficulty of finding useful correlations between grades and other indices of intellectual performance in such preselected populations as college students--especially graduate students (cf. Wallach, 1976).

Restriction of range is expressed relative to the population, with the most widely accepted index being the ratio of the sample standard deviation to the population standard deviation (Ghiselli, 1964; Guilford, 1954). Ghiselli has presented the mathematical proofs which establish the required correction for attenuation when the population variance is known or can be estimated accurately. In the limit, as the ratio of the sample standard deviation to population standard deviation approaches zero, the Pearson correlation coefficient approaches zero.

What is rarely made clear in classroom discussions is that correlation is not merely a function of the marginal distribution but of the conditional distributions as well. Perhaps for this reason, there appears to be a widespread misunderstanding of this phenomenon, leading to the belief (until recently shared by the authors) that any variable with a variance near zero would necessarily show a correlation coefficient near zero with any other variable, and that decreasing variance necessarily leads to decreasing absolute values of coefficients. For this reason, it seems useful to describe an anomaly recently encountered in our research.

The Anomaly

During the course of research on the development of flying training objectives for the U.S. Air Force (Barron, Gerlach & Haygood, 1976), we had occasion to collect ratings of the "observability" of training objectives using a standard 1 to 5 rating scale. In addition to ratings of complete statements of objectives, we also collected ratings of the three component parts of such statements--the condition, verb, and criterion (e.g., "given a pair of 9-digit numbers/add the two numbers/without error"). For a surprising number of statement components, there was near-universal agreement among the subjects participating, generating data such as those shown in Table 1. For example, all subjects but one rated "with 10% accuracy" as being highly observable--a 1 on the rating scale. On scanning the data, we were confident that the computed correlation coefficient for Table 1 would be near zero, because of the small variance; the discovery of a coefficient of +1.00 came as a distinct surprise.

In retrospect, it is clear why the coefficient must be 1.00: a straight line will fit the data perfectly, and increases or decreases in the variance resulting from the addition or elimination of pairs falling on either data point (1,1 or 2,3) will not change the data pattern. The correlation coefficients computed for varying numbers of 1,1 pairs and extra selected data points are shown in Table 2. Of special interest is the fact that addition of extra subjects at data point 1,1 decreases the variance but increases the correlation coefficient for the last three rows of Table 2. The actual coefficients in the last three rows are dependent on the arbitrary selection of the extra data points, but the pattern is the same regardless of which extra points are chosen.

Table 1
Pairs of scores found in Rating-Scale Research Data^a

<u>Subject No.</u>	<u>Rating on Scale A</u>	<u>Rating on Scale B</u>
1	1	1
2	1	1
3	1	1
4	1	1
5	1	1
6	1	1
7	1	1
8	1	1
9	1	1
10	1	1
11	1	1
12	1	1
13	1	1
14	2	3
15	1	1
16	1	1
17	1	1
18	1	1
19	1	1
20	1	1

^a
 $r_{AB} = +1.00$

Table 2

Pearson Correlation Coefficients for Selected Combinations of Scores

<u>Number of Extra Pairs</u>	<u>Actual Score Pairs</u>	<u>Number of 1,1 pairs</u>				
		<u>1_a</u>	<u>5</u>	<u>10</u>	<u>100</u>	<u>1000</u>
0	-	.00	.00	.00	.00	.00
1	2,3	1.00	1.00	1.00	1.00	1.00
2	2,3/2,2	.87	.93	.94	.95	.95
3	2,3/2,2/1,2	.71	.82	.84	.86	.87
4	2,3/2,2/1,2/2,1	.32	.57	.63	.70	.71

^aall non-zero coefficients are positive

Conclusion

When small numbers of subjects are used with discrete rating scales of limited range, it seems likely that data patterns of the type shown here will often be found. If all subjects use the same category, that rating necessarily correlates zero with any other scale. If one or two subjects choose divergent responses on each of two scales, the correlation may be quite high, even perfect, regardless of the variances obtained. Thus, the belief that decreased variance in one or both variables is always associated with reduced correlation is clearly incorrect. As an instructional note, we recommend that classroom presentations of the restriction of range problem be structured to avoid leaving such an impression with the student. While the data analyst himself is not likely to be misled, the risk is that those who do not have access to the raw data may take such correlations at face value and assume great predictive power where none exists.

References

Barron, M.A., Gerlach, V.S., Haygood, R.C. An empirical analysis of the component of behavioral objectives (Technical Report No. 60331). Tempe, Arizona: Arizona State University Instructional Resources Laboratory, 1976.

Ghiselli, E.E. Theory of Psychological Measurement. New York: McGraw-Hill, 1964.

Guilford, J.P. Psychometric Methods, second edition. N w York: McGraw-Hill, 1954.

Wallach, M.A. Tests tell us little about talent. American Scientist, 1975, 64, 57-63.