Multivariate Behavioral Research

Routledge
Taylor & Francis Group

ISSN: 0027-3171 (Print) 1532-7906 (Online) Journal homepage: https://www.tandfonline.com/loi/hmbr20

A COMPARISON OF THE CLASSIFICATION ACCURACY OF PROFILE SIMILARITY MEASURES

ROBERT M. CARROLL & JOHN FIELD

To cite this article: ROBERT M. CARROLL & JOHN FIELD (1974) A COMPARISON OF THE CLASSIFICATION ACCURACY OF PROFILE SIMILARITY MEASURES, Multivariate Behavioral Research, 9:3, 373-380, DOI: 10.1207/s15327906mbr0903_12

To link to this article: https://doi.org/10.1207/s15327906mbr0903_12

Published online: 10 Jun 2010.

A COMPARISON OF THE CLASSIFICATION ACCURACY OF PROFILE SIMILARITY MEASURES

ROBERT M. CARROLL

and

JOHN FIELD
University of Maryland

ABSTRACT

Thirteen profile similarity measures were compared, using generated data. Profiles were generated from sets of three standards by adding random and normally distributed error components to the profile points of the standards. The three standards within each set were varied systematically, altering the elevation, scatter, and shape similarities between the standards. A correct classification occurred if the generated profile was most similar to the standard from which it was generated. Significant differences were found between the proportions of correct classifications for the 13 profile similarity measures under all conditions. Osgood and Suci's D and Cattell's r_p were superior to or equal to all other measures under all conditions.

INTRODUCTION

The number of theoretical papers discussing profile similarity measures far outweighs the number of papers presenting empirical evaluations. Some of the earliest attempts at comparing profile similarity measures empirically were made by Mosel and Roberts (1954), Muldoon and Ray (1958), and Helmstadter (1957).

Mosel and Roberts used D (Osgood & Suci, 1952), Dumas' r_{ps} (Dumas, 1946, 1947, 1949, 1950, 1953), Meehl's *DI* (Mosel & Roberts, 1954), Cattell's r_p (Cattell, 1949), and Pearson's product moment correlation (Hays, 1963) in their comparisons. They compared the similarity of 30 profiles to four standards selected arbitrarily from the original 30. The measures differed widely in rank-ordering the 30 profiles with regard to similarity to the standard. The rank-order intercorrelations among the five measures which were computed for each of the four standards, were found to vary depending upon which standard was used. Mosel and Roberts concluded that the measures could not be compared in absolute terms. They also noted that Cattell's r_p correlated most highly with the criterion of clinical judgment.

Muldoon and Ray (1958) compared the similarity of 19 profiles to an arbitrarily chosen standard using the mean, Spearman's *rho* (Hays, 1963), Cattell's r_p, Osgood and Suci's D, Meehl's *DI*, and

Dumas' r_{ps}. In contrast to Mosel and Roberts, Muldoon and Ray found that Dumas' r_{ps} was most highly correlated with the criterion of clinical judgment.

Helmstadter compared the effectiveness of 10 profile similarity measures, using generated data. Eight measurements on arbitrarily chosen dimensions were obtained from each of three standards consisting of a sphere, a right circular cylinder, and a regular tetrahedron. A normally distributed error component was added to these measurements in order to generate 90 profiles from each standard. Each of these 270 generated profiles was then compared to each of the standards using the 10 profile similarity measures. The profile was classified into the group represented by the standard most similar to the profile. The proportion of correct classifications was then computed for each similarity measure. Helmstadter found that all of the measures were able correctly to classify profiles significantly better than chance and with a high degree of accuracy. However, he also found little difference between the success rates for the measures. The high success rates and lack of differentiation among the success rates for the different measures might be attributable to the large interstandard differences.

The present study is an extension of Helmstadter's method but attacks the problem in a somewhat more systematic fashion by introducing several sets of three standards whose differences with respect to elevation, shape, and scatter vary in different combinations. In addition, the variance of the normally distributed error component which was added to the profile points of the standards was varied. The purpose of the present research is to compare the success rates of 13 profile similarity measures and to determine which of the measures have the highest probability of correctly classifying a profile into its appropriate group under the various conditions of inter-standard similarity.

METHOD

The similarity relationships between the standards were systematically varied by altering the elevation, shape, and scatter of the standards. Seven sets of three standards with four profile points per standard were created by successively varying one, then two, then all three of the characteristics among the standards. Table 1 shows the standards used for each of the seven cases.

In the cases where elevation was varied (Cases 1, 4, 5, 7), the three standards had mean values of 30, 50, and 70. In the cases where scatter was varied (Cases 3, 5, 6, 7), the standard deviations of the three standards were 10, 15, and 20. When shape was varied (Cases 2, 4, 6, 7) the intercorrelations between the standards were equal to zero except in Cases 2 and 7 where $r_{12} = -1.0$. When elevation was constant, the mean of each standard was 50. When scatter was constant, the standard deviation of each standard was 10 and when shape was constant, the intercorrelation between standards was 1.0.

Table 1
The Seven Cases of Interstandard Similarity

Case 1:	Elevation varies, shape constant, scatter constant			
	$S1$ = [20,40,20,40]	E = 30,	S = 10	r_{12} = 1.0
	$S2$ = [40,60,40,60]	E = 50,	S = 10	r_{13} = 1.0
	$S3$ = [60,80,60,80]	E = 70,	S = 10	r_{23} = 1.0
Case 2:	Elevation constant, shape varies, scatter constant			
	$S1$ = [60,40,40,60]	E = 50,	S = 10	r_{12} = -1.0
	$S2$ = [40,60,60,40]	E = 50,	S = 10	r_{13} = 0.0
	$S3$ = [60,60,40,40]	E = 50,	S = 10	r_{23} = 0.0
Case 3:	Elevation constant, shape constant, scatter varies			
	$S1$ = [40,60,40,60]	E = 50,	S = 10	r_{12} = 1.0
	$S2$ = [35,65,35,65]	E = 50,	S = 15	r_{13} = 1.0
	$S3$ = [30,70,30,70]	E = 50,	S = 20	r_{23} = 1.0
Case 4:	Elevation varies, shape varies, scatter constant			
	$S1$ = [80,80,60,60]	E = 70,	S = 10	r_{12} = 0.0
	$S2$ = [40,60,60,40]	E = 50,	S = 10	r_{13} = 0.0
	$S3$ = [20,40,20,40]	E = 30,	S = 10	r_{23} = 0.0
Case 5:	Elevation varies, shape constant, scatter varies			
	$S1$ = [60,80,60,80]	E = 70,	S = 10	r_{12} = 1.0
	$S2$ = [35,65,35,65]	E = 50,	S = 15	r_{13} = 1.0
	$S3$ = [10,50,10,50]	E = 30,	S = 20	r_{23} = 1.0
Case 6:	Elevation constant, shape varies, scatter varies			
	$S1$ = [40,60,60,40]	E = 50,	S = 10	r_{12} = 0.0
	$S2$ = [65,65,35,35]	E = 50,	S = 15	r_{13} = 0.0
	$S3$ = [30,70,30,70]	E = 50,	S = 20	r_{23} = 0.0
Case 7:	Elevation varies, shape varies, scatter varies			
	$S1$ = [60,80,60,80]	E = 70,	S = 10	r_{12} = -1.0
	$S2$ = [65,65,35,35]	E = 50,	S = 15	r_{13} = 0.0
	$S3$ = [10,10,50,50]	E = 30,	S = 20	r_{23} = 0.0

E = Elevation or mean
S = Scatter or standard deviation
r = Shape or correlation

One thousand profiles were generated from each standard by adding to each profile point of the standard a computer-generated, normally-distributed random error component. The relative sizes of the error components were varied as a function of the pooled within-profile variance of the three standards. The standard deviations of the normally distributed error component were set at 1/2 σ_p, 1 σ_p, and 2σ_p where σ_p is the pooled within-profile standard deviation. Thus there were a total of 21 conditions, as each of the seven cases of interstandard similarity had associated with it three standard deviations of the normally-distributed error component.

After the 3,000 profiles were generated, 1,000 from each standard, each was compared to the three standards and categorized into the group of the standard to which the profile was most similar as indicated by the profile similarity measure. A correct classification occurred if the similarity measure indicated the profile to be most similar to the standard from which it was generated. Thirteen different profile similarity measures were compared in this study. All have been recommended at one time or another for use in assessing the similarity of test profiles. The indices used were: Osgood and Suci's D, D^{I} and D^{II} (Cronbach & Gleser, 1953), Dumas' r_{ps}, DuToit's index (1954), Cattell's r_p, Pearson's r, Spearman's *rho*, intraclass r (Webster, 1952), Cohen's r_c (1969), Dumas' X^2 (1947), *AD* (Ellson, 1947; Kogan, 1949), and Meehl's *DI*.

If more than one standard had equally high similarity to the profile, the decision as to which group to classify the profile in was made on a chance random process. For example, if a similarity measure classified a generated profile equally into two groups, a decision was made on the basis of a random process with the probability of classification into either group set at 1/2. Helmstadter (1957) justified this procedure on the assumption that if any indeterminancy occurred in any practical situation, the final classification would be made on chance factors.

Results

Table 2 presents the success rates of each similarity measure according to case and error condition. A chi-square goodness-of-fit test for uniformity with 12 degrees of freedom was performed on each of the seven cases and for each error condition. All of the

Table 2
Proportion of Correct Classifications

Similarity Measure	Error Condition	Case 1	2	3	4	5	6	7
	$\frac{1}{2}\ \sigma_p$	.992*	.996*	.988*	.992*	.988*	.998*	.998*
D	$1\ \sigma_p$	.838*	.907*	.863*	.875*	.862*	.911*	.906*
	$2\ \sigma_p$	.634*	.694*	.628*	.645*	.635*	.598*	.683*
	$\frac{1}{2}\ \sigma_p$	.992*	.996*	.988*	.992*	.988*	.998*	.998*
Cattell's r_p	$1\ \sigma_p$	.838*	.907*	.863*	.875*	.862*	.911*	.906*
	$2\ \sigma_p$	.634*	.694*	.628*	.645*	.635*	.598*	.683*
	$\frac{1}{2}\ \sigma_p$	.991*	.996*	.985*	.990*	.985*	.998*	.998*
AD	$1\ \sigma_p$	.805	.891	.823	.862	.829	.896	.896*
	$2\ \sigma_p$	.602	.668	.596	.617	.597	.567	.662
	$\frac{1}{2}\ \sigma_p$	.838*	.907*	.836*	.849*	.986*	.994*	.991*
Intra-class r	$1\ \sigma_p$	.992*	.996*	.986	.989	.846	.879	.873
	$2\ \sigma_p$	.634*	.694*	.628	.622	.635	.598*	.655
	$\frac{1}{2}\ \sigma_p$	.983*	.996*	.985*	.979	.959	.998*	.972*
X^2 Dumas	$1\ \sigma_p$	.802	.906*	.849	.828	.771	.913*	.794
	$2\ \sigma_p$	.554	.691*	.608	.547	.511	.590*	.531
	$\frac{1}{2}\ \sigma_p$	.434	.996*	.988*	.878	.636	.998*	.970
D^{I}	$1\ \sigma_p$	.382	.907*	.863*	.664	.485	.911*	.770
	$2\ \sigma_p$	.361	.694*	.628*	.494	.421	.598*	.573
	$\frac{1}{2}\ \sigma_p$	.336	.996*	.987*	.871	.633	.998*	.968
D^{II}	$1\ \sigma_p$	.334	.907*	.863*	.648	.482	.912*	.766
	$2\ \sigma_p$	.335	.694*	.629*	.468	.389	.599*	.563
	$\frac{1}{2}\ \sigma_p$	.927	.996*	.281	.956	.958	.991*	.985
Cohen's r_c	$1\ \sigma_p$	.774	.907*	.302	.837	.802	.886	.885
	$2\ \sigma_p$	.606	.694*	.354	.638	.622	.592*	.672
	$\frac{1}{2}\ \sigma_p$	.326	.996*	.974	.878	.587	.994*	.945
DI	$1\ \sigma_p$	.340	.907*	.787	.664	.587	.883	.752
	$2\ \sigma_p$	.352	.694*	.594	.494	.406	.589*	.560
	$\frac{1}{2}\ \sigma_p$	.362	.996	.339	.878	.334	.991*	.941
RHO	$1\ \sigma_p$	.336	.902*	.330	.664	.335	.886	.761
	$2\ \sigma_p$	.319	.680	.333	.494	.340	.592*	.565
	$\frac{1}{2}\ \sigma_p$	.226	.996*	.246	.878	.311	.991*	.947
Pearson's r	$1\ \sigma_p$	.200	.907*	.255	.664	.310	.886	.775
	$2\ \sigma_p$	.331	.694*	.364	.494	.336	.592*	.567
	$\frac{1}{2}\ \sigma_p$	.332	.691	.340	.613	.322	.668	.860
Dumas' r_{ps}	$1\ \sigma_p$	.336	.648	.330	.500	.327	.627	.701
	$2\ \sigma_p$	.333	.522	.320	.425	.335	.486	.531
	$\frac{1}{2}\ \sigma_p$	.328	.914	.346	.782	.343	.869	.870
DuToit's Index	$1\ \sigma_p$	.335	.830	.345	.612	.332	.796	.707
	$2\ \sigma_p$	.334	.638	.327	.472	.319	.553	.534

*Correctly classified profiles as well as any other measure to within one percentage point for the given case and error condition.

obtained chi-squares were significant at the .005 level, indicating the 13 similarity measures did differ in their ability to classify the profiles correctly. Chi-square goodness-of-fit tests were also performed using only the similarity measures which were also used by Helmstadter (1957), namely, intra-class *r*, *rho*, Dumas' X^2, Pearson *r*, Meehl's *DI*, D^I, and *AD*. These tests were significant in all cases except Cases 2 and 6 of $1/2\ \sigma_p$ and Case 2 of $1\ \sigma_p$. This finding contrasts with Helmstadter's result that the methods had similar hit rates with respect to classifying profiles successfully.

The success rate for all similarity measures decreased as the size of the error component increased. Spearman's *rho*, Pearson's *r*, DuToit's index, and Dumas' r_{ps} all showed a consistent pattern of performing best in the cases where shape was varied (Cases 2, 4, 6, 7) and no better than chance in the cases where shape was held constant. Cohen's r_c did no better than chance in Case 3 for all error conditions, indicating it failed to pick up scatter differences among the standards. D^I and D^{II} disregard elevation differences and did least well under Case 1 where the standards had only elevation varied. One result contrary to what was expected was the finding that D^{II} and *DI* did well under Case 3, where only scatter varied among the standards, as these measures were developed primarily to detect shape differences. Osgood and Suci's *d* and Cattell's r_p were superior or equal to all other measures in all cases and under all error conditions.

Discussion

If a psychologist or counselor is confronted with a classification problem based upon a profile of test scores, the data of Table 2 might be of some assistance. The measures which have superior successful classification proportions for each of the 21 conditions have asterisks (*) following their proportions in Table 2. Many of the measures proved to be superior in several cases. The obvious practical implication of this study is that the measures with asterisks in Table 2 represent the recommended methods to use in any applied situation which approximates any of the 21 conditions studied here. This would be particularly relevant where there were standard profiles which were thought to be representative of the classification types and where individuals within a classification type could be thought of as having normally-distributed scores about each profile point of the standard.

If one assumes that each profile score is a projection upon orthogonal axes, then the distance between any two profiles is given by D, using a Euclidean metric, and by AD, using a city-block metric. It is interesting to note by inspecting Table 2 that using the city-block metric correctly classified profiles as well as any measure for all seven cases but only for the 1/2 σ_p error condition. Using the Euclidean metric resulted in classifications as accurate as any other measure for all seven cases and all error conditions.

Most of the measures were not designed to maximize classification accuracy under each of the seven population conditions studied here. Therefore the failure of a measure to classify profiles accurately under all conditions does not imply it is unacceptable for use under circumstances for which it might be specifically derived. To cite one instance, Cohen's r_c was derived for instances where elevation, scatter, and direction of measurement of profile elements are to be ignored.

The generality of these findings is of course limited by certain parametric restraints which were imposed upon the conditions, such as the four-point profiles, homogeneity of variance of the error component per profile point, normally distributed error component, and the specific manner in which elevation, scatter, and shape of the standards were varied. Further research could be directed at other conditions.

References

Cattell, R. B. r_p and other coefficients of pattern similarity. *Psychometrika*, 1949, *14*, 279-298.

Cohen, J. r_c: A profile similarity coefficient invariant over variable reflection. *Psychological Bulletin*, 1969, *71*, 281-284.

Cronbach, L. J., & Gleser, G. C. Assessing similarities between profiles. *Psychological Bulletin*, 1953, *50*, 456-473.

Dumas, F. M. A quick method of analyzing the similarity of profiles. *Journal of Clinical Psychology*, 1946, *2*, 80-83.

Dumas, F. M. On the interpretation of personality profiles. *Journal of Clinical Psychology*, 1947, *3*, 57-65.

Dumas, F. M. The coefficient of profile similarity. *Journal of Clinical Psychology*, 1949, *5*, 123-131.

Dumas, F. M. A note on the coefficient of profile similarity. *Journal of Clinical Psychology*, 1950, *6*, 300-301.

Dumas, F. M. Quick methods for the analysis of the shape, elevation, and scatter of profiles. *Journal of Clinical Psychology*, 1953, *9*, 345-348.

DuToit, J. M. A modification of the index of profile similarity. *Journal of Clinical Psychology*, 1954, *10*, 384-386.

Ellson, D. G. A method for identifying pattern clusters of test score profiles. *American Psychologist*, 1947, *2*, 425.

Hays, W. L. *Statistics for psychologists*. New York: Holt, Rinehart, & Winston, 1963.

Helmstadter, G. C. An empirical comparison of methods for estimating profile similarity. *Educational and Psychological Measurement*, 1957, *17*, 71-82.

Kogan, W. S. *An investigation into the relationship between psychometric patterns and psychometric diagnosis*. Unpublished Ph.D. dissertation, University of Pittsburgh, 1949.

Mosel, J. N., & Roberts, J. B. The comparability of measures of profile similarity: An empirical study. *Journal of Consulting Psychology*, 1954, *18*, 61-66.

Muldoon, J. F., & Ray, O. S. A comparison of pattern similarity as measured by six statistical techniques and eleven clinicians. *Educational and Psychological Measurement*, 1958, *18*, 775-781.

Osgood, C. E., & Suci, G. J. A measure of relation determined by both mean difference and profile information. *Psychological Bulletin*, 1952, *49*, 251-262.

Webster, H. A note on profile similarity. *Psychological Bulletin*, 1952, *49*, 538-539.