DOCUMENT RESUME

ED 231 885 TM 830 502

AUTHOR Broadbooks, Wendy J.; Elmore, Patricia B.
TITLE An Investigation of the Sampling Distribution of the Congruence Coefficient.
PUB DATE 15 Apr 83
NOTE 12p.; Paper presented at the Annual Meeting of the American Educational Research Association (67th, Montreal, Quebec, April 11-15, 1983).
PUB TYPE Speeches/Conference Papers (150) -- Reports - Research/Technical (143)

EDRS PRICE MF01/PC01 Plus Postage.
DESCRIPTORS *Factor Analysis; Goodness of Fit; Hypothesis Testing; *Research Methodology; *Sample Size; *Sampling; Simulation; Statistical Analysis; *Statistical Studies
IDENTIFIERS *Coefficients of Congruence; Magnitude Estimates; Population; *Replication; Statistical Simulation; Variables; Variance (Statistical)

ABSTRACT

This study developed and investigated an empirical sampling distribution of the congruence coefficient. The effects of sample size, number of variables, and population value of the congruence coefficient on the sampling distribution of the congruence coefficient were examined. Sample data were generated on the basis of the common factor model and principal axes factor analyses were performed. Sampling distributions were formed on the basis of 200 replications for each combination of psi, sample size, and number of variables. Characteristics of each distribution were described and goodness of fit to normality tests, t-tests, and chi square tests were performed to determine whether the sampling distributions behaved similarly to the Pearson product-moment correlation coefficient. The results indicated that when the population congruence coefficient is zero, the sampling distribution of the congruence coefficient is relatively stable and is similar to the sampling distribution of the correlation coefficient. The expected variance of each sampling distribution is influenced not only by sample size but also by the number of variables used in the factor analysis. Further research needs to be done to ascertain more specifically the effect of sample size and number of variables on the variance of the sampling distribution. (PN)

An Investigation of the Sampling Distribution of the Congruence Coefficient

Wendy J. Broadbooks and Patricia B. Elmore
Southern Illinois University at Carbondale

Paper presented at the annual convention of the American Educational Research Association, Montreal, Quebec, Canada, April 15, 1983.

An Investigation of the Sampling Distribution of the Congruence Coefficient

Wendy J. Broadbooks and Patricia B. Elmore
Southern Illinois University at Carbondale

The technique of factor analysis has attained considerable popularity as a method for analyzing the interrelationships within a set of data. Mulaik (1972) stated that "factor analysis as a scientific tool has as its goal the decomposition of variables into fundamental elements, the factors, which may then be used to explain the interrelationships among variables" (p. 337). Replications of studies utilizing factor analysis are just as necessary as replications of studies using any other statistical technique. In order to make comparisons of sets of factors and their interpretability across different studies, some objective basis is needed for the comparison in order to ascertain the degree of the relationship between corresponding factors in the studies. Several measures of factor similarity are available and commonly used. The present study focussed on one such similarity coefficient: the congruence coefficient (Burt, 1948; Tucker, 1951; Wrigley and Neuhaus, 1955). The congruence coefficient involves a comparison of two sets of factor loadings in terms of both the pattern and magnitude of the loadings. It has been used extensively as a descriptive statistic in research comparing factors across studies. In most of these studies the same set of variables but different subjects were used. The formula used to calculate the congruence coefficient is as follows:

$$\Psi_{pq} = \frac{\sum_{j=1}^{n} (a_{jp})(b_{jq})}{\sqrt{[\sum_{j=1}^{n} a_{jp}^2][\sum_{j=1}^{n} b_{jq}^2]}}$$

where n is the number of variables common to the two studies, p is the factor found in study one, q is the factor found in study two, and a_{jp} and b_{jq} are factor loadings. The value of the congruence coefficient can range from +1 (perfect agreement) to -1 (perfect inverse agreement). A zero value for the congruence coefficient reflects a lack of agreement between the two vectors of factor loadings. Some studies have examined the distribution of the congruence coefficients under various conditions (Cattell, 1978; Korth, 1973, 1978; Korth & Tucker, 1975; Nesselroade & Baltes, 1970; Nesselroade, Baltes, & Labouvie, 1971), however, none of these studies have examined the sampling distribution for varying values of the population congruence coefficient. For example, both Korth and Tucker (1975) and Korth (1978) reported results of Monte Carlo studies examining the distribution of chance congruence coefficients from simulated data. Korth (1978) constructed tables of the distribution of chance congruence coefficients and maintained that the results reported in the tables could be used to test the null hypothesis that two sets of random vectors have been matched. Korth stated that rejecting this null hypothesis still does not establish that two factors are identical; thus, the null hypothesis tested by these tables is still not strong (Korth, 1978, p. 428).

Nesselroade and Baltes (1970) and Nesselroade, Baltes, and Labouvie (1971) also reported results of simulation studies of the congruence coefficient based on the generation of random data. The research performed in the present study is an extension of what has previously been done in that a systematic examination of the distribution of the congruence coefficient was done for varying values of the population congruence coefficient; that is, not only for the situation where one can assume that there is no relationship between two factors (which would theoretically result from a population congruence coefficient of zero), but also for situations where one can assume that there is some relationship between two factors (which would result from a population congruence coefficient different from zero).

The purpose of the present research was to develop and investigate an empirical sampling distribution of the congruence coefficient. The effects of sample size, the number of variables, and the magnitude of the population congruence coefficient were considered in the analyses.

Method

For the purposes of the present study, simulated data were generated. Population congruence coefficients were specified by determining pairs of population factor loading vectors that resulted in the following values for the congruence coefficient: .00, .10, .20, .30, .40, .50, .60, .70, .80, .90, and .99. Sample data were generated from these population factor loading vectors on the basis of the common factor model. Sample sizes examined were: 50, 100, and 200. Number of variables used were: 10, 30, and 50. The data were generated using Fortran computer programs.

Principal axes factor analyses were performed using squared multiple correlations as the communality estimates and one factor was extracted in each analysis. A sample congruence coefficient and a sample Fisher z-transformed congruence coefficient was computed for each pair of sample factor loadings. Sampling distributions of the congruence coefficients and of Fisher z-transformed congruence coefficients were formed on the basis of 200 replications for each combination of psi (population congruence coefficient), sample size, and number of variables.

Characteristics of the empirical sampling distribution were examined (e.g., mean, variance, selected percentiles) along with the results of goodness of fit to normality tests. T-tests and chi square tests were performed in order to determine whether the sampling distribution of the congruence coefficient and of the Fisher z-transformed congruence coefficient behaved similarly to the sampling distribution of the correlation coefficient and of the Fisher z-transformed correlation coefficient, respectively. T-tests were done to test for equal means and chi square tests were done to test for equal variances of the desired distributions. All statistical tests performed used a .01 level of significance.

Results and Discussion

The following results were found concerning the characteristics of the sampling distributions of the congruence coefficient and the Fisher z-transformed congruence coefficient: (1) the congruence coefficient is positively biased for values of psi from .10 to .40; (2) the congruence coefficient is negatively biased for values of psi from .50 to .60 and from .90 to .99; (3) the skewness and kurtosis values of most of the sampling distributions seemed to be near those expected for a normal distribution; (4) almost all of the sampling distributions examined did approximate a normal distribution (of the 88 goodness of fit to normality tests only six tests led to a rejection of the null hypothesis); (5) regardless of the value of psi, as sample size became larger the variance of each distribution either stayed the same or decreased; (6) regardless of the value of psi, as the number of variables increased the variance of each distribution either stayed the same or decreased; (7) the sampling distributions of the congruence coefficient behave similarly to the correlation coefficient only when psi is equal to zero (on the basis of the results of the t-tests, goodness of fit to normality tests, and chi square tests performed in the present study).

Table 1 presents a summary of the three statistical tests that were performed on each of the sampling distributions. If the congruence coefficient behaved similarly to the correlation coefficient then there would be no "X's" on the table and all the statistical tests would have led to a retention of the null hypothesis.

When psi is equal to zero, the sampling distribution of the congruence coefficient behaves similarly to the sampling distribution of the correlation coefficient. When 10 variables were used the obtained variances were equal to the expected variances. However, when the number of variables used was larger than 10, the obtained

TABLE 1

Summary of Statistical Tests Performed on Each Distribution[a]

		Ψ/Z_Ψ = .00/.00			.10/.100			.20/.203			.30/.310			.40/.424			.50/.549			.60/.693			.70/.867			.80/1.099			.90/1.472			.99/2.647		
NV	NS	D	t	χ^2_v	D	t	χ^2_v	D	t	χ^2_v	D	t	χ^2_v	D	t	χ^2_v	D	t	χ^2_v	D	t	χ^2_v	D	t	χ^2_v	D	t	χ^2_v	D	t	χ^2_v	D	t	χ^2_v
10	50					X			X			X			X	X		X			X			X				X	X	X	X	X	X	X
10	100					X			X			X			X						X				X		X	X	X	X	X		X	X
10	200					X			X			X			X			X			X			X			X	X		X	X	X	X	X
30	50			X		X	X		X	X		X	X		X	X		X	X		X	X		X	X		X	X		X	X		X	X
30	100			X		X	X		X	X		X	X		X	X		X	X		X	X			X		X	X		X		X	X	X
30	200			X		X	X		X	X		X	X		X	X		X	X		X	X		X	X		X			X			X	X
50	100			X		X	X		X	X		X	X		X	X		X	X		X	X		X	X		X	X		X	X		X	X
50	200			X		X	X		X	X		X	X		X	X		X	X		X	X		X	X		X	X		X	X	X	X	X

NOTE: For each distribution 200 replications were performed.

[a]Where NV=number of variables, NS=number of subjects, D=goodness of fit to normality test, t=test for equal means, and χ^2_v=chi square test of equal variances. An "X" in this table means that the null hypothesis for the test was rejected, p<.01.

variances were less than the expected variances. Therefore, a researcher could use the hypothesis-testing procedure available to test the null hypothesis that rho is equal to zero to test whether psi is equal to zero. Since the obtained variance for the sampling distributions (when psi is equal to zero) were equal to or less than the expected variances, this statistical test would be a conservative test of the null hypothesis. If the null hypothesis that psi is equal to zero was rejected, a researcher could conclude that there is some relationship between the two factors on which the congruence coefficient is calculated. The magnitude of the sample congruence coefficient is an indication of the strength of the relationship between the two factors.

The findings of the present study have implications for researchers who wish to use the congruence coefficient in order to determine whether two factors have been replicated. Tables (which are available from the senior author) were constructed that present the characteristics of the distributions of the congruence coefficient and of the distributions of the Fisher z-transformed congruence coefficient for the selected values of the population coefficients and across all combinations of sample size and number of variables used in this study. Further investigation of the sampling distribution under more combinations of sample size and number of variables would provide information as to whether the results reported here are representative and generalizable. The results found in this study do indicate that the stability of the congruence coefficient increases as number of variables and subjects increase.

Further research might be done to (1) examine the effect on the sampling distribution of the congruence coefficient of varying the distributions of the population factor loadings; (2) examine the distributions that result from using negative values for the population congruence coefficient; (3) examine the generalizability of the results of the present study to studies in which more than one factor has been extracted from the data and several factor comparisons are to be made; (4) examine more levels of sample size and number of variables; (5) examine other possible transformations of sample congruence coefficients in an attempt to discover some transformation that would result in a sampling distribution with known properties.

REFERENCES

Burt, C. Factor analysis and canonical correlations. British Journal of Psychology, Statistical Section, 1948, 1, 95-106.

Cattell, R. B. Matched determiner's vs. factor invariance: A reply to Korth. Multivariate Behavioral Research, 1978, 13, 431-448.

Korth, B. A. Analytic and experimental analysis of factor matching methods. Unpublished doctoral dissertation, University of Illinois at Urbana-Champaign, Psychology Department, 1973.

Korth, B. A. A significance test for congruence coefficients for Cattell's factors matched by scanning. Multivariate Behavioral Research, 1978, 13, 419-430.

Korth, B. A., and Tucker, L. R. The distribution of chance congruence coefficients from simulated data. Psychometrika, 1975, 40, 361-372 (erratum in Psychometrika, 1979, 44, 365).

Mulaik, S. A. The foundations of factor analysis. New York: McGraw Hill, 1972.

Nesselroade, J. R., and Baltes, P. B. On a dilemma of comparative factor analysis: A study of factor matching based on random data. Educational and Psychological Measurement, 1970, 30, 935-948.

Nesselroade, J. R., Baltes, P. B., and Labouvie, E. W. Evaluating factor invariance in oblique space: Baseline data generated from random numbers. Multivariate Behavioral Research, 1971, 6, 233-241.

Tucker, L. R. A method of synthesis of factor analysis studies. Personnel Research Section Report, No. 984. Washington, D. C.: Department of the Army, 1951.

Wrigley, C., and Neuhaus, J. O. The matching of two sets of factors. American Psychologist, 1955, 10, 418.