Studia Psychologiczne. t. 56 (2018), z. 1, s. 31–56
PL ISSN 0081-685X
DOI: 10.2478/V1067-010-0180-8

Tomasz Korulczyk

Institute of Psychology
The John Paul II Catholic University of Lublin

THE REVIEW OF DYADIC CONGRUENCY INDEXES

Dyadic research is becoming more common in social sciences, although there is a relatively small amount of publications in this field, written in Polish. In the article, I compared and contrasted the most common and useful indexes of congruence, which may be used to determine the degree of congruency of members of independent (uncorrelated) pairs. I also indicate usefulness, the conditions of use and the examples of research questions which may be answered using appropriate indexes. Mathematical assumptions, strengths, and weaknesses, a way of interpretation of each index have also been presented in the paper. The beneficial results of the review include making more conscious decisions due to getting familiar with using particular indexes in the correct context. SPSS syntax has been prepared with all indexes as a supplement to the article.

Keywords: dyadic research, similarity indexes, congruence indexes, SPSS syntax

In the course of both computerization of science, the development of computer technology, and statistical methods used to estimate dyadic congruence, significant progress has been made. Since the introduction of Pearson product moment correlation coefficient at the beginning of the 20th century, dozens of statistical methods have been described in the literature. In the further parts of this article, the most widely applied and useful coefficients in the dyadic research will be discussed. The characteristics of individual coefficients presented in the article may help to decide the researcher which one best suits the adopted research model. The subject of non-parametric measures of congruency (χ^2, φ^2, Kappa, biserial correlation, point-biserial, tetrachoric correlation) goes beyond the scope of this article and will not be taken into account. For more details, I refer to work of the researchers in this field (Baroni-Urbani, Buser, 1976; Cohen, 1988; Cureton, 1956; Divgi, 1979; Tate, 1954). There are many dyadic congruency estimation methods in the literature. Some researchers postulate to calculate differences in meta-traits, e.g. relating to personality as a whole, while the others use popular correlation coefficients or more complex ones specially developed for this particular purpose. Despite various approaches, some guidelines have been developed and will be presented later in the article.

INITIAL ANALYSIS AND CONGRUENCY LEVEL

Before starting to assess the compliance, the basic issue the researcher should start from is the analysis with using statistical tests and

* Correspondence concerning this article should be addressed to Tomasz Korulczyk: tomasz.korulczyk@gmail.com

visual methods as a scatterplot, giving the opportunity to “detect” characteristic tendencies, such as non-linearity of variables (Glicksohn, Golan, 2001). This assumption is crucial in a way that when assuming linearity, in the situation when it actually exhibits curvilinear features, its actual strength will be underestimated. This problem is observable in research relating to spousal fit and relationship satisfaction (see Figure 1).

The calculation of the compliance indexes presented in this article is considered to be correct if the members of the given dyad seem to be independent of each other in a statistical sense, in other words, they are not mutually correlated in the compared dimension. However, if the members of the pair share common features, it is more appropriate to carry out multilevel analyzes, considering them, on the one hand, as separate individuals (level 1), but also as constituting a single dyad (level 2). Failing to take this fact into account in the analysis may lead to unreliable results (Węziak, 2007).

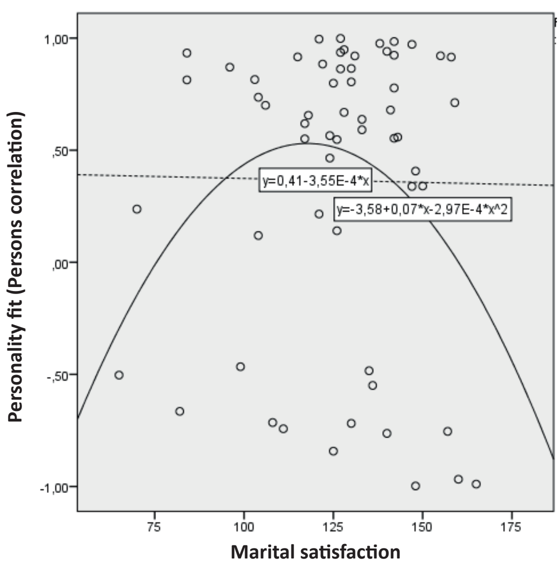

Figure 1. Linear regression line (dotted one) and quadratic line (a solid one) on the example of personality fit and marital satisfaction. Own elaboration.

Deliberating on the multilevel analyzes is beyond the purpose of this article thus for more details please review the following authors (De Leeuw, Meijer, 2008; Heck, Thomas, Tabata, 2013; Hox, 2010; Hox, Roberts, 2011).

After determining the independence of the components of the pair, the selected congruency index can be used. The difficulty may arise when choosing the right congruency index because each serves a different purpose, finally giving different results. It is worth paying attention to understanding the assumptions of each of them and examples of their application. Congruency indexes can be divided into: (1) differences (DI), (2) correlation (CI), (3) distances (DI), (4) heterogeneous (NHC) coefficients. A brief description of the group along with the characteristics of each of them are presented below. Formulas can be found in the supplemental material of this article but also in the form of ready-to-use SPSS syntax.

Compatibility can be estimated at various levels of detail: general (A), scale (B) and test items (C, Figure 2). In the literature, it may be calculated at the general level (Haselager, Hartup, van Lieshout, Riksen-Walraven, 1998), scale level (Blum, Mehrabian, 1999; Eysenck, Wakefield, 1981; Haselager et al., 1998; Kupersmidt, DeRosier, 1995) and test items (Gaunt, 2006; Luo, Klohnen, 2005). It is assumed that the researcher should strive to estimate dyadic congruency as much detail as possible in order not to reduce the variance of results (Klohnen, Mendelsohn, 1998; Luo, Klohnen, 2005).

In the studies on the compatibility of a person to the group before the calculation of the fit index, two additional steps are performed beforehand. The results of a group of people within a given dimension are correlated and if the correlation is high (there is group compliance), the results for all members of a given group are averaged, creating a variable for the whole group which is the base for the calculation of the congruence level (Sherony, Green,

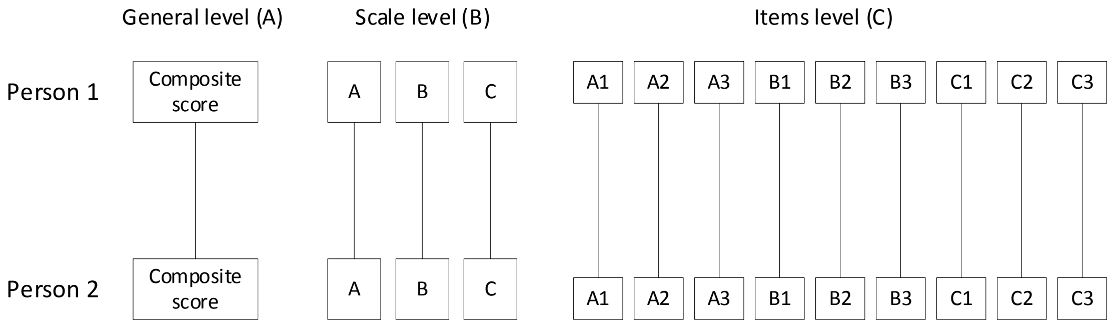

Figure 2. The level of detail of congruency measurement.

2002). The examples of such research are person-team fit or person-organization fit.

Compliance can be determined not only by congruency indexes. Another way presented in the literature is the comparison of means within individual variables with the *t*-test. In case of no differences, the similarity in a given area is assumed (Bleske-Rechek, Remiker, Baker, 2009). Another method is to create an interaction variable from the variables of both members of the pair, e.g. the openness of the husband and wife, and the inclusion of this variable in statistical analyzes (Bedyńska, Brzezicka, 2007). However, as research shows, this variable very often explains little, and most importantly, it is not necessarily the measure of similarity, which indicates that the self-description of one member of a couple is dependent of the self-description of the other one (Luo, Klohnen, 2005). In such situation, in order to interpret the meaning of interaction properly, it is necessary to use information from graphical charts of this interaction through its detailed analysis (Griffin, Murray, Gonzalez, 1999; Kenny, Cook, 1999)

DIFFERENCE INDEXES

DI are the most commonly used measures of congruence due to their simplicity. DI works well if the researcher needs simple information at the high level of generality. Therefore, a large amount of variance resulting from the diversity between particular test items is lost if it is calculated on scales or on a general level. There is also no need to verify complicated statistical assumptions. Due to its construction, DI does not investigate the relationship between the set variables. It can be a measure of similarity (if = 0) or dissimilarity, the size of which is difficult to interpret. It is almost impossible to properly interpret the correlations between the difference index and the outcome variable without analyzing the components of this index, and with unequal variances DI is distorted by the variance of these components (Luo, Klohnen, 2005).

There are several calculation options in this group, although they are most often operationalized as the absolute difference (Weinberg, Scarr, Waldman, 1992). The first option concerns the estimation of the simple or absolute difference between variables at the scales level (Figure 2, B), however then we obtain a one-dimensional index, for each feature separately. For example, when combining the “personalities” of two people in the “Big Five” model, we will get five congruency indexes. Nevertheless, using it is not the best idea when one wants to combine several features within one construct, as it is necessary to determine which index is the most important. The application of DI may facilitate a correct answer to the research questions in the following research situations:

- if the researcher is interested in the compliance of relatively uncomplicated and non-complex phenomena (operationalized in a very general way, at scales or subscales)
- it works well in explaining specific variables (e.g. the absolute difference is characterized by the highest predictive validity of attitudes towards the brand, its preferences, and intentions of buying), yet it is applicable in other cases (e.g. satisfaction with marriage)
- if the researcher does not have the access to observable variables such as test items.

The examples of research questions:

- Is similarity in the dimension of extraversion in children associated with the increasing amount of relationships?
- To what extent does dissimilarity in the dimension of neuroticism of the patient and therapist affect the effectiveness of treatment?
- Does the similar level of emotional intelligence of the manager and the employee affect the burnout experienced by both of them?
- How is similarity of the level of laziness of the student and the tutor related to the student's learning performance?
- Does dissimilarity of the sensation-seeking of pilot 1 and pilot 2 affect flight safety?

The examples of research using DI can be found in various works (Blum, Mehrabian, 1999, Eysenck, Wakefield, 1981, Haselager et al., 1998, Kupersmidt, DeRosier, 1995, Robins, Caspi, Moffitt, 2000, Wells, 1991).

There are several calculation options in this group. The first one is to estimate the simple or absolute difference between variables at the scales level (Figure 2, B), but then we obtain a one-dimensional index, for each feature separately. For example, when combining the "personalities" of two people in the "Big Five" model, we will get five congruency indexes. Using it is not the best idea when we want to combine several features because it is then necessary to determine which index is more important.

Another computational variant that solves the problem of a separate coefficient for each variable is the creation of the so-called superordinate composite score, i.e. the component of several variables and the calculation of the simple difference between meta-traits for the members of the pair. In such situation, the researcher has an obligation to correctly compose such a meta-characteristic, which seems to be simple seemingly. As the research shows many mistakes can be made at this stage (Rożnowski, Korulczyk, 2018), so when deciding on the weights assigned to individual scales, it is worth using the well-established methodology as part of multi-criteria decision analysis (Multicriterion Decision Analysis, MCDA) which includes such popular methods as Analytic Hierarchy Process and Analytic Network Process. These methods are available as the part of the Super Decisions® and Expert Choice® computer programs. More on MCDA can be found in works of the following authors (Adamus, 2011, Rożnowski, Korulczyk, 2018, Saaty, Vargas, 2013).

The last calculation variant is the use of the differences models presented below which are considered to be the sum or mean of the differences between the features of two elements of a given dyad. One gets an index covering all the features. The obtained sum provides information about the size of the difference in relation to several dimensions, while the average informs about the average difference for all dimensions. It is the researcher who decides which information is more useful. The indexes from DI group (which can be expressed as a sum or average) will be presented as follows (Gorbaniuk, Stachoń-Wójcik, 2011):

1) A Simple Difference (SD, see Gaunt, 2006)
2) Absolute Difference (AD)

If the abovementioned ratios, expressed as the sum of both variables take values above or below 0, then SD and AD give different results. SD and AD as averages express the value weighted by the *k*-number of the dimensions being compared.

3) Divided simple difference, weighted through one of the variables (DSD)
4) Divided absolute difference, weighted through one of the variables (DAD)

If the above-mentioned ratios, expressed as the sum of both variables take values above or below 0, then DSD and DAD give different results. DSD and DAD as averages express the value weighted by the *k*-number of the compared dimensions, which causes their centralization around this number. Weighing by one of these variables, instead, centralizes them around the variable through which they were divided.

The above-mentioned indexes are applicable if the researcher is interested in the average for given dimensions, and the weighted measurements, and if the researcher wants to "clear" the difference from a given variable. However, the researcher must bear in mind that DSD and DAD are very difficult to interpret.

CORRELATION COEFFICIENTS

Correlation indexes (CI) have a long history in sciences and up to date are generally applied to determine the level of similarity between two features. They become widely used due to a relatively easy access, simple calculations, and interpretation. CI are perfect measures of similarity and co-occurrence of traits but also they provide detailed information on dissimilarity (if <0). They are also the only coefficients which give the information on the common explained variance of the variables. CI may also be calculated on each level of detail hence if the researcher is interested in congruence at the level of items, these measures enable it. More importantly, CI contrary to DI is able to depict the defined structure of data within compared variables. CI is a target measure of *monotonicity* that is - how given observations co-vary. It is worth mentioning that CI is not a measure of agreement of variables, although it may be the case that CI and scaled distance indexes (SDI) may take similar values. Typical measures of congruence are SDIs, for instance, Gower Agreement. To make sure that CI is an appropriate measure, it is necessary to visually analyze the scatterplot. The example presented below (Figure 3) is derived from the study of 51 participants, where raters assessed their agreement on the scale from 1 to 6. The data was presented on a line plot and scatterplot. As it can be seen the plot on the left-hand side, it presents some level of congruence between data, yet the scatterplot the regression line is almost parallel to X-axis and Pearson's coefficient equals –.04 which indicates dissimilarity. Therefore, CI should be applied carefully. Using CI requires meeting a few statistical conditions. In terms of parametric coefficients, such as *r*-Pearson, the conditions include (1) the correlated variables to be on an interval scale and (2) to be normally distributed, (3) the linear relationship between variables, (4) the lack of outliers, (5) the homoscedasticity of variables which means that the data is distributed evenly along the regression line.

When it comes to non-parametric coefficients such as Spearman's *rho* there are two conditions: (1) the variables must be at least on an ordinal scale and (2) the relationship between variables must be monotonic, that is linear or curvilinear but with the maximum of one curve point. Applying CI may enable providing correct answers to the research question which:

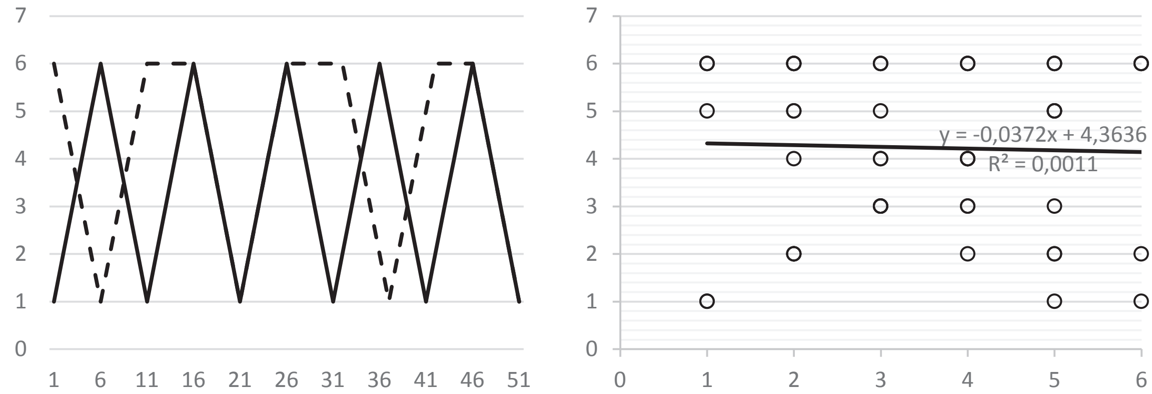

Figure 3. The problem of use of Pearson product-moment correlation as a measure of congruence/agreement.

• assumes a linear relationship between variables
• relating to both similarity and dissimilarity
• if the researcher is interested in the size of total variance explained
• if one is interested in congruence of complex phenomena (operationalized in a detailed way, on the level of observation)

The examples of research questions:
• What is the level of similarity of value systems between husband and wife?
• What level of similarity of temperament among pupils is related to displaying aggressive behaviors?
• What level of dissimilarity of goal attainment attitude between spouses is related to the risk of divorce?
• What level of dissimilarity within social intelligence between the client and the seller is related to the client's satisfaction?

The examples of using CI may be found in the following works: Pearson's product-moment correlation (Caspi, Herbener, 1990; Glicksohn, Golan, 2001; Haselager et al., 1998; Luo, Klohnen, 2005); Person's distance (Kiesler, Watkins, 1989; Robins et al., 2000); Spearman's rho (Hansen, Gold, 1977); Pearson's r for Qsort (Van Exel, De Graaf, 2005). The coefficients from this group include:

1) Pearson's product-moment correlation
The measure of linear relationship which is the quotient of covariance and the product of standard deviations. It is the measure commonly used in the majority of research investigating relationships between variables. More importantly, this coefficient is prone to extreme observations and also it decreases the size of relationship if it has a curvilinear shape. First and foremost, it is necessary to fulfill conditions before using it.

2) Uncentered Pearson product-moment correlation
It is a measure of orientation, not a size so it may be useful if the researcher is interested in determining the direction of distance between variables. It is a variant of traditional Pearson coefficient with the assumption that the mean values for both variables equal 0. Similarly to *Cosine of the Angle,* this coefficient is equal to the cosine of the angle between two vectors in *n*-dimension space.

3) Pearson's Distance
It is computed by reversing traditional Pearson coefficient and it is used to obtain the coefficient which would not be on the scale below 0.

It would take values from 0 (reflecting +1) to 2 (reflecting –1).

4) Person's Absolute Distance
Using this coefficient may be useful if the research focuses on the strength of relation and its direction. It is also a simple transformation of Pearson by reversing its absolute value. It would then take values from 0 to 1. The higher value, then the coefficient shows the lack of correlation (0). The closer value is to 0 the stronger correlation becomes (+1).

5) Uncentered Pearson Distance
It is computed by reversing Uncentered Pearson. It is a measure of orientation and not the size that is why it may be applied when the researcher is interested in determining the direction of distance between variables on the scale which does not take value below 0 (precisely from 0 to 2).

6) Uncentered Pearson Absolute Distance
It is computed by reversing the absolute value of Uncentered Pearson. It may be applied if the researcher is interested in determining the size of distance direction between variables on the scale without values below 0. In this case, the higher the value of the coefficient is the lower size of direction (0). However, the closer value is to 0 the bigger it is (+1).

7) Spearman's *rho*
Spearman's *rho* is a rank-version of Pearson's product moment correlation coefficient, more resistant to outliers. It is also a better measure of curvilinear relationships, it deals better with outliers and it does not require a normal distribution of variables. It is usually applied when the researcher is interested in determining the similarity between variables yet he or she cannot use parametric measures.

8) Pearson's *r* for Q-sort
It is a simplified version of Pearson's product moment correlation which takes into account the fact that two profiles have the same distribution and there is the same amount of pool of items. It is used to assess congruence between two Q-sort profiles. For more information on the Q-sort technique, see Paczkiewicz (1978).

DISTANCE COEFFICIENTS

There are two kinds of DIs: scale ones (SDI) as a measure of congruence and non-scale ones (NDI) as a measure of dissimilarity. DI do not require verification of assumption except for data on an interval scale. They are usually computed on raw data (Cronbach, Gleser, 1953), which in turn has its advantages (resilience to new outliers) but also disadvantages (the difference in measurement units of variables has a strong impact on the values of distances). Thus, it is worth considering standardization of one scale in regard to the other one according to the formula no. 2 included in the supplemental material of this article.

SDI are the best-known measurements of congruence/agreement of compared variables. Therefore, applying SDI may enable providing correct answers to research questions which:

- Aim at determining the actual congruence within a given trait/construct
- Relate to complicated psychological phenomena such as e.g. personality fit
- Are operationalized as CI and contradict the latest reports from the literature or graphic analysis of the relationship (it may be the case when there is a lack of similarity but the agreement appears)
- Even after taking curve-linearity into account seem to be too much of simplification of a psychological phenomenon.

The examples of research questions:
- Is there a relationship between spouses' personality fit and their marital satisfaction?
- How is a congruence of goals of the subordinate and the superior related to intention to leave the company by the subordinate?
- How does congruence of value hierarchy of the patient and the therapist influence on treatment effectiveness?
- How does dissimilarity within narcissism trait of dating people transfer to their reciprocal first impression?

NDI possesses similar characteristics to DI as it does not investigate relations among variables and it is also difficult to interpret. Thus, applying NDI may enable providing the proper answer to the research question in a similar way to DI. The examples of research questions include:
- How does dissimilarity in an experienced level of organizational change between two team members transfer to their level of cooperation?
- How is similarity of perceiving candidate's suitability to work by the recruiter and the candidate related to the assessment of his or her competences?
- Is similarity of the level of impulsivity of two AA members addicted to alcohol-related to the risk of relapse?

SDI group involves:

1) Gower Agreement (GA)
GA is the most recommended measure of congruence (Barrett, 2010) and is applied in in different dyadic research context (Brudek, Korulczyk, Korulczyk, 2018; Robins et al., 2000). GA is a reversed version of the Gower's discrepancy coefficient (Gower, 1971), based on the maximal possible absolute discrepancy among all variables for both elements of the dyad. GA indicates the percentage of the average absolute level of similarity among all dyads of observations in the range between 0 and, indicating full congruence.

2) Double-Scaled Euclidean (DSE-s)
DSE-s is a modification of the traditional Euclidean Distance. Similarly to GA, it is applied in measuring agreement among variables. The measure is double scaled to enable comparing this coefficient with other agreement measures and to make it more resistant to measurement scales used in research (Barrett, 2010). Additionally, the linear distance between variables remains unchanged as scaling is also linear (in contrast to various non-linear types of scaling based on *z* normalization). DSE-s takes value from 0 - indicating maximal possible non-congruence, to +1- indicating complete congruence.

3) Kernel-Smooth Distance (KSD-s)
KSD-s is used in measuring non-linear congruence. Manipulating *s* parameter enables the researcher obtaining nearly perfect fit of the line to the data (Figure 4). The idea behind KSD-s is that the function of distance should be smoothed in a way that if a simple difference between the actual characteristics of a person and the estimated one is computed within a given range then the value of statistics should reflect a very small distance almost despite its actual size (Barrett, 2010). However, in the case when the distance increases, the estimated value of the index should increase in the size. The key to using these statistics correctly is to define accurately the fixed value of *s* smoother which causes an expected inertial effect (Barrett, 2010). Making the right choice depends on a very individual research case where the researcher should bear in mind the costs (difficulty in interpretation) and gains (more precise function) of more smoothed function. KSD-s oscillates between 0, indicating complete congruence, and +1indicating identicalness.

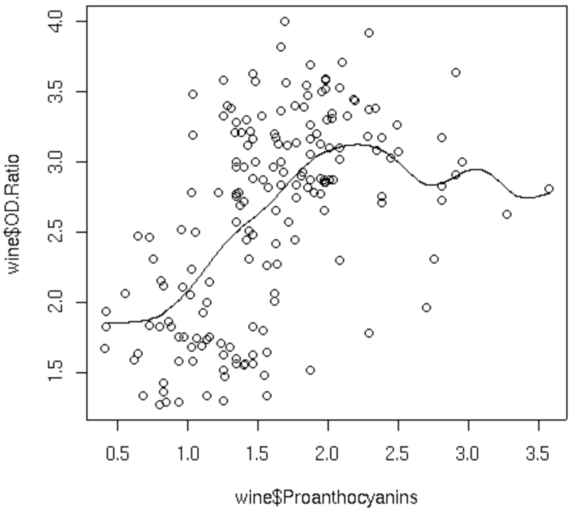

Figure 4. The nonlinear fit of KSD-s to the data. [*Source:* https://www.stat.berkeley.edu/~s133/wineno-rm.png]

NDI group involves:

1) Euclidian Distance (ED)
ED is a linear measure of dissimilarity. It is a straight geometrical distance in space and is commonly computed on raw data. Is it is applied in exploring dissimilarity among variables, like DI. It is resistant to outliers, but the difference of measurement units of variables has a strong impact on the value of distance.

2) The square of the Euclidean Distance (D^2, Cronbach, Gleser, 1953)
D^2 like ED is a measure of dissimilarity which takes into account the objects which are more distant than ED. It is applied in clinical and psychotherapy research (Kiesler, Watkins, 1989). D^2 is a variation of simple difference expressed as a squared sum. Using this measure is linked with some problems similar to those with ED.

3) Minkowski's Distance (MD)
MD is a generalized flexible measure of dissimilarity and is applied if the researcher aims to find the size of distance optimal for particular data. It is an absolute difference raised to the value of *p*, and then the extraction of *r* roots is performed. Manipulating the *p* parameter results in an increased weight of the difference between variables while changing the *r* parameter gives bigger or smaller meaning to more or less distant variables. MD may be equal other indexes: (1) ED, if *r* and *p*=1, (2) the City bock distance, if *r* and *p*=2.

4) City Block Distance
The City block distance (also called *Manhattan*, Taxicab distance) is a sum of absolute differences between variables. It perfectly describes distances in an urban space where getting from one point to another is possible only by moving towards north-south or east-west directions, but never diagonally. Like ED, it is a measure of dissimilarity, but it is more resistant to outliers as single differences are not squared. It is applied if the researcher is interested in the level of dissimilarity and the data is full of outliers.

5) The Canberra Distance
The Canberra distance is a weighted version of the City block distance (Lance, Williams, 1967) but it is more resistant to variables with high values (Krebs, 1989). The difference between these two is that the absolute difference between variables is divided by the absolute sum. If the index takes values > 0 then it shows the lack of congruence between the variables and 0 shows the perfect congruence. This measure is more responsive to the type of used scale (Lance, Williams, 1967) and as a more sensitive measure, it is applied in research calculating the precise deviations of similarity (Emran, Ye, 2001).

6) Chebyshev Distance
It is also called chessboard distance as it is a distance between two spaces on a chess board expressed in moves of a king. In statistics, it is

a distance between two variables expressed as a maximal difference of values between the variables. It is applied in a research case when dissimilarity is assumed in advance.

7) Power Distance
Alike MD, it is a generalized flexible measure of dissimilarity through adjusting weights of observations and variables by manipulating the *r* and *p* parameters. It is applied in the same research cases as MD.

8) Cosine similarity
It is also referred to as *the Cosine of the angle* which evaluates the cosine of the angle between two variables/vectors (Tan, Steinbach, Kumar, 2005). It is a measure of an orientation and not a size, therefore, it is applied when the researcher is interested in determining the direction of distance between variables. It may be interpreted through CI categories where -1 indicates an opposite direction, 0 is the 90 ° angle and 1 indicates the same direction.

Non-homogeneous coefficients

The group of non-homogeneous coefficients (NHC) consists of the indexes which could not be categorized in any other group due to their unique features: intraclass correlation coefficients, McCrae's - the coefficient of profile agreement, sociometric measures, integration quotient, and C-index.

Each of these coefficients is designed to measure congruence in particular conditions and is interpreted in an individual way. Coefficients from this ground are hardly available in most statistical packages; however, they are included in the SPSS syntax included in the supplemental material of this article.

Some researchers suggest that if there is a strong or at least moderate correlation between variables then it is worth to perform scale-to-scale conversion to get the most reliable results of ICC coefficients (Barrett, correspondence with the author, May 18, 2017). It may be performed according to the formula no. 2 included in the supplemental material. Using NHC is possible after fulfilling following conditions: variables (1) must be on an interval scale, and in the case of ICC, variables (2) must be normally distributed. This group comprises:

1) McCrae's coefficient (McCrae, 1993)
It is designed to assess profiles to increase the impact of congruence at the level of extreme values. Thus, by employing weight it underestimates the values located in the middle of the distribution. It is widely used in clinical research. The coefficient is interpreted as CI (−1 indicates complete mismatch and +1 indicates a complete match).

2) Sociogram
In sociometric research, congruence is measured in a three-fold way. The most common one includes asking all group members to indicate three persons who they like the most (LM) and three who they like the least (LL). Then, the nominations (indications) are summed up for each person, separately for LM and LL dimensions and then the results are standardized on *z* scale (getting zLM and zLL). Following that, by subtracting zLL from zLM the social preference indicator (SP) is obtained, and by adding zLL to zLM the social impact indicator (SI) is obtained. The *z* standardization is repeated on both coefficients (to obtain zSP and zSI, Coie, Dodge, Coppotelli, 1982). On the basis of the above-mentioned indicators, it may be concluded whether a group member is popular (zSI > 1; zLM > 0; zLL < 0), rejected (zSP < −1; zLM < 0; zLL > 0), sociographically neglected (zSI < −1; LM = 0), or controversial (zLM > 0; zLL > 0). In order to ease the interpretation, the above-mentioned variables may be converted into the *r*-Person scale (from −1 to 1) by using SPPS syntax or the formula no. 30 from the supplemental material of this article.

3) Integrativeness quotient (Tripp, Sondak, 1992)
It is the best-known measure of negotiation effectiveness and it is used primarily in such research context. It is based on Pareto distribution and it comprises the ratio of favorable to unfavorable agreements. The measure was initially introduced by Tripp, Sondak (1992), however, it was modified by arcsine transformation to get the effect of reducing negative skewness (Weingart, Hyder, Prietula, 1996).

4) Congruence index (Brown, Gore, 1994)
It is the best measure of the level of fit between Holland's occupational types. The fit is represented by describing a distance in Holland's hexagonal model of career fields between first, second or third letters of the highest score for a person and occupational environmental. A distance may take value from 3 to 0 in the following way: 3 if the person and organization have the same letters; 2 if the person and organization have adjoined letters; 1 if the person and organization are two letters away; 0 if the person and organization have opposite letters. The distance is computed for the first three letters with the highest score. The index may take values from 0 to 18 reflecting the level of fit. For instance, the person with ACS profile who works (or intends to work) in an organization with ERS profile may fit moderately to such setting (Figure 5: C = 3(0) + 2(3) + 2 = 8)

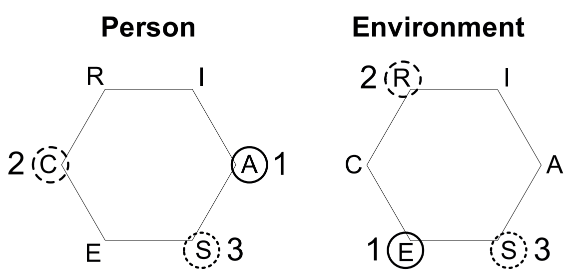

Figure 5. The example of two profiles of occupational preferences of person and environment in of Holland's theory.

5) ICC-1
ICC-1 also referred to as ICCde computed by double entry method, alike CI is the measure of monotonicity that is co-variance (Barrett, 2010). When it comes to ICC-1, variables are treated as random samples from hypothetic populations and their attributes are also independently selected. The measure assumes that each case and the position has separate and unique rater assigned who performs a single rating and is randomly selected from a wider population (Shrout, Fleiss, 1979). If we have 20 observations and 5 variables, we need 100 raters. ICC-1 fluctuates from 0, which means non-congruence, to 1 which means complete congruence. ICC-1 works best with simple research questions when the researcher is interested in congruence understood as independence of each rating e.g. when we assess the pair where each member provided several independent responses, e.g. measurement over time (Lakey et al., 2002). The examples of research questions include:

- Is agreement between the therapist and the patient regarding the daily rating of psychiatric treatment over last week related to intensification of suicidal thoughts?

6) ICC-2
ICC-2 is a mixed measure of agreement and monotonicity of variables. It assumes that each rater rates each case and each variable, and that the rater is randomly selected from a wider population of raters (Shrout, Fleiss, 1979). If we have 20 observations and 5 variables, we need at least 2 raters who would make 100 ratings. ICC-2 ranges between 0 which indicates non-congruence and 1 which indicates complete congruence. ICC-2 works best with the majority of research questions if the researcher is interested in congruence understood in the way that any two subjects could be replaced by other subjects (the possibility of generalization onto the population, Sherony, Green,

2002). The examples of research questions are as follows:

- Is congruence between assessors in the method of assessment center related to latter work performance of the new employee?
- Does non-congruence between sales team members in the realm of common targets transfer into their intention of leaving the company?

7) ICC-3

ICC-3 just like ICC-1 is a measure of monotonicity of variables. It presumes that each rater rates each case and each variable, but contrary to ICC-s it is assumed that raters comprise the population of all raters (Shrout, Fleiss, 1979)**.** If we have 20 cases and 5 variables then we need at least 2 raters that would perform 100 ratings. ICC-3 ranges from 0 indicating non-congruence to 1 indicating complete congruence. ICC-3 works best with the majority of research questions which investigate congruence between measurements of one person, that is with qualitative and idiographic research. The examples of research questions include:

- Is congruence within the perceived sense of living before and after experienced trauma related to participant's perceived stress?
- Are neuropsychological examination results congruent before and after past injury which was subject to judgment in a compensation case?

INTERPRETATION OF INDEXES

A correct interpretation of index is as crucial as its estimation and although it is difficult to find clear guidelines for particular groups in the literature, they may be interpreted by analogy.

In the case of DI and NDI, in order to assess the size of difference for a particular scale (B level), before computing, regardless of an adopted variant, one should standardize variables onto a chosen scale, e.g. z or sten scale. If the researcher is interested in the size of difference on an overall level of the set and not a single scale, then one should compute Cohen's d statistic by using attached SPSS syntax or the formula no. 1 from the supplemental material of this article. Dissimilarity is interpreted in the following way (Cohen, 1988): [0 - 0,2] – insignificant; [0,2 - 0,5] – small; [0,5 - 0,8] – moderate; [0,5 - 0,8] – large; [0,8 - 1,3] – very large. If DI has been computed only in one dimension and two compared variables were standardized into z scale beforehand, then this value may be interpreted as Cohen's d.

In the case of SDI, CI and NHC, one can use guidelines included in the literature, mentioned by Cohen (1988), Evans (1996) and Guilford (1965), which are presented below (Table 1). Rescaled indexes of distance may sometimes take values over the maximum range for a given scale, e.g. +1 and it is a special situation. It results from the fact that the index easily acquires extreme values (> 3 in z scale). Similarly, ICC coefficients may very rarely, in some situations take negative values and it is connected with calculations performed on random data (Barrett, 2010).

SUMMARY

Although dyad research becomes increasingly popular it is still difficult to find polish publications including guidelines of how to correctly compute congruence in a given paradigm. Moreover, there might be some difficulties with defining a term and choosing a relevant index. The following article reviewed indexes from different groups and discussed its usefulness for psychological research. Both advantages and disadvantages have been explored along with conditions of employing these measures. Finally, some recommendations as for the ways of interpreting them have been presented.

Table 1 The interpretation of congruence coefficients.

<table>
<tr><th rowspan="2">Range
0 to 1</th><th rowspan="2">Range
+1 to –1</th><th rowspan="2">Range
0 to 2</th><th rowspan="2">Pearson absolute distance</th><th colspan="3">Interpretation of the strength</th></tr>
<tr><th>Cohen (1988)</th><th>Evans (1996)</th><th>Guilford (1965)</th></tr>
<tr><td>—</td><td>–1</td><td>2</td><td>0</td><td rowspan="5">strong</td><td rowspan="3">very strong</td><td>perfect</td></tr>
<tr><td>—</td><td>–0,9</td><td>1,9</td><td>0,1</td><td rowspan="2">very high</td></tr>
<tr><td>—</td><td>–0,8</td><td>1,8</td><td>0,2</td></tr>
<tr><td>—</td><td>–0,7</td><td>1,7</td><td>0,3</td><td rowspan="2">strong</td><td rowspan="2">high</td></tr>
<tr><td>—</td><td>–0,6</td><td>1,6</td><td>0,4</td></tr>
<tr><td>—</td><td>–0,5</td><td>1,5</td><td>0,5</td><td rowspan="2">medium</td><td rowspan="2">moderate</td><td rowspan="2">moderate</td></tr>
<tr><td>—</td><td>–0,4</td><td>1,4</td><td>0,6</td></tr>
<tr><td>—</td><td>–0,3</td><td>1,3</td><td>0,7</td><td rowspan="3">small</td><td rowspan="2">weak</td><td rowspan="2">low</td></tr>
<tr><td>—</td><td>–0,2</td><td>1,2</td><td>0,8</td></tr>
<tr><td>—</td><td>–0,1</td><td>1,1</td><td>0,9</td><td>very weak</td><td>very low</td></tr>
<tr><td>0</td><td>0</td><td>1</td><td>1</td><td>lack</td><td>lack</td><td>no relationship</td></tr>
<tr><td>0,1</td><td>0,1</td><td>0,9</td><td>0,9</td><td rowspan="3">small</td><td>very weak</td><td>very low</td></tr>
<tr><td>0,2</td><td>0,2</td><td>0,8</td><td>0,8</td><td rowspan="2">weak</td><td rowspan="2">low</td></tr>
<tr><td>0,3</td><td>0,3</td><td>0,7</td><td>0,7</td></tr>
<tr><td>0,4</td><td>0,4</td><td>0,6</td><td>0,6</td><td rowspan="2">medium</td><td rowspan="2">moderate</td><td rowspan="2">moderate</td></tr>
<tr><td>0,5</td><td>0,5</td><td>0,5</td><td>0,5</td></tr>
<tr><td>0,6</td><td>0,6</td><td>0,4</td><td>0,4</td><td rowspan="5">large</td><td rowspan="2">strong</td><td rowspan="2">high</td></tr>
<tr><td>0,7</td><td>0,7</td><td>0,3</td><td>0,3</td></tr>
<tr><td>0,8</td><td>0,8</td><td>0,2</td><td>0,2</td><td rowspan="3">very strong</td><td rowspan="2">very high</td></tr>
<tr><td>0,9</td><td>0,9</td><td>0,1</td><td>0,1</td></tr>
<tr><td>1</td><td>1</td><td>0</td><td>0</td><td>perfect</td></tr>
</table>

Note. **Coefficients with a range from 0 to 1** = Gower Agreement, Double Scaled Euclidian Distance, KSD-s, ICC-1, ICC-2, ICC-3. The following are recommended to be rescaled: Euclidian Distance, Squared Euclidian Distance, Minkowski Distance, City Distance, Canberra Distance, Chebyshev Distance, Power Distance. **Coefficients with a range from +1 to –1** = Pearson product-moment, Pearson Uncentred, Spearman's *rho*, The Cosine of Angle; **Coefficients with a range from 0 to 2** = Pearson's Distance, Pearson's Distance Uncentred.

REFERENCES

Adamus, W. (2011). Metody ilościowe w organizacji, zarządzaniu i psychologii. In B. Rożnowski, M. Łaguna (Eds.), *Człowiek w pracy i w organizacji* (pp. 145-168). Lublin: Wydawnictwo KUL.

Baroni-Urbani, C., Buser, M. W. (1976). Similarity of binary data. *Systematic Zoology*, *25*(3), 251-259. DOI: 10.2307/2412493

Barrett, P. (2010). Test reliability and validity: The inappropriate use of the Pearson and other variance ratio coefficients for indexing reliability and validity. Retrieved from: *http://www.pbarrett.net/techpapers/correlations_reliability_validity_Rev_1_July_2010.pdf*.

Bedyńska, S., Brzezicka, A. (2007). *Statystyczny drogowskaz: praktyczny poradnik analizy danych w naukach społecznych na przykładach z psychologii.*

Warszawa: Wydawnictwo SWPS Academica. DOI: 10.12775/rp.2005.012

Bleske-Rechek, A., Remiker, M. W., Baker, J. P. (2009). Similar from the start: assortment in young adult dating couples and its link to relationship stability over time. *Individual Differences Research*, *7*(3), 142-158.

Blum, J. S., Mehrabian, A. (1999). Personality and temperament correlates of marital satisfaction. *Journal of Personality*, *67*(1), 93–125. DOI: 10.1111/1467-6494.00049

Brown, S. D., Gore, P. A. (1994). An evaluation of interest congruence indices: Distribution characteristics and measurement properties. *Journal of Vocational Behavior*, *45*(3), 310-327. DOI: 10.1006/jvbe.1994.1038

Brudek, P., Korulczyk, T., Korulczyk, N. (2018). Dopasowanie w parze małżeńskiej a satysfakcja ze związku u osób starszych. *Roczniki Psychologiczne/ Annals of Psychology*, 21 (0), 00-00.

Caspi, A., Herbener, E. S. (1990). Continuity and change: assortative marriage and the consistency of personality in adulthood. *Journal of Personality and Social Psychology*, *58*(2), 250-258.

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences (2*nd *ed.).* Hillsdale, NJ: Lawrence Erlbaum Associates. DOI: 10.4324/9780203771587

Coie, J. D., Dodge, K. A., Coppotelli, H. (1982). Dimensions and types of social status: A cross-age perspective. *Developmental Psychology*, *18*(4), 557–570. DOI: 10.1037/0012-1649.18.4.557

Cronbach, L. J., Gleser, G. C. (1953). Assessing similarity between profiles. *Psychological Bulletin*, *50*(6), 456-473. DOI: 10.1037/h0057173

Cureton, E. E. (1956). Rank-biserial correlation. *Psychometrika*, *21*(3), 287–290. DOI: 10.1007/ bf02289138

De Leeuw, J., Meijer, E. (2008). *Handbook of multilevel analysis.* New York: Springer.

Divgi, D. R. (1979). Calculation of the tetrachoric correlation coefficient. *Psychometrika*, *44*(2), 169-172. DOI: 10.1007/bf02293968

Emran, S. M., Ye, N. (2001). *Robustness of canberra metric in computer intrusion detection.* Presented at the 2001 IEEE Workshop on Information Assurance and Security United States Military Academy, West Point, NY. DOI: 10.1002/qre.441

Evans, J. D. (1996). *Straightforward statistics for the behavioral sciences.* Pacific Grove, CA: Brooks/Cole Pub Co. DOI: 10.2307/1165199

Eysenck, H. J., Wakefield, J. A. (1981). Psychological factors as predictors of marital satisfaction. *Advances in Behaviour Research and Therapy*, *3*(4), 151-192. DOI: 10.1016/0146-6402(81)90002-3

Gaunt, R. (2006). Couple similarity and marital satisfaction: are similar spouses happier? *Journal of Personality*, *74*(5), 1401-1420. DOI: 10.1111/j.1467-6494.2006.00414.x

Glicksohn, J., Golan, H. (2001). Personality, cognitive style and assortative mating. *Personality and Individual Differences*, *30*(7), 1199-1209. DOI: 10.1016/S0191-8869(00)00103-3

Gorbaniuk, O., Stachoń-Wójcik, M. (2011). Zgodność obrazu siebie i wizerunku konsumenta-wartość predykcyjna różnych metod pomiaru kongruencji. *Marketing i Rynek*, *6*, 17-26.

Gower, J. C. (1971). A general coefficient of similarity and some of its properties. *Biometrics*, *27*(4), 857-871. DOI: 10.2307/2528823

Griffin, D., Murray, S., Gonzalez, R. (1999). Difference score correlations in relationship research: A conceptual primer. *Personal Relationships*, *6*(4), 505-518. DOI: 10.1111/ j.1475-6811.1999.tb00206.x

Guilford, J. P. (1965). *Fundamental Statistics in psychology and education (4*th *ed.)*. New York: McGraw-Hill.

Hansen, R. M., Gold, I. K. (1977). Blacktail prairie dogs, desert cottontails and cattle trophic relations on shortgrass range. *Journal of Range Management*, *30*(3), 210-214. DOI: 10.2307/3897472

Haselager, G. J. T., Hartup, W. W., van Lieshout, C. F. M., Riksen-Walraven, J. M. A. (1998). Similarities between friends and nonfriends in middle childhood. *Child Development*, *69*(4), 1198-1208. DOI: 10.2307/1132369

Heck, R. H., Thomas, S. L., Tabata, L. N. (2013). *Multilevel and Longitudinal Modeling with IBM SPSS*. New York: Routledge. DOI: 10.4324/9780203701249

Hox, J. J. (2010). *Multilevel Analysis (2*nd *ed.).* New York: Routledge.

Hox, J. J., Roberts, J. K. (2011). *Handbook of Advanced Multilevel Analysis.* New York: Routledge. DOI: 10.4324/9780203848852

Kenny, D. A., Cook, W. (1999). Partner effects in relationship research: Conceptual issues, analytic difficulties, and illustrations. *Personal Relationships*, *6*(4), 433-448. DOI: 10.1111/j.1475-6811.1999.tb00202.x

Kiesler, D. J., Watkins, L. M. (1989). Interpersonal complementarity and the therapeutic alliance: A study of relationship in psychotherapy. *Psychotherapy: Theory, Research, Practice, Training*, *26*(2), 183-194. DOI: 10.1037/h0085418

Klohnen, E. C., Mendelsohn, G. A. (1998). Partner selection for personality characteristics: A couple-centered approach. *Personality and Social Psychology Bulletin*, *24*(3), 268-278. DOI: 10.1177/0146167298243004

Krebs, C. J. (1989). *Ecological Methodology.* New York: Harper Collins Publishers.

Kupersmidt, J. B., DeRosier, M. E. (1995). Similarity as the basis for children's friendships: The roles of sociometric status, aggressive and withdrawn behavior, academic achievement and demographic characteristics. *Journal of Social and Personal Relationships*, *12*(3), 439-452. DOI: 10.1177/0265407595123007

Lakey, B., Adams, K., Neely, L., Rhodes, G., Lutz, C. J., Sielky, K. (2002). Perceived support and low emotional distress: The role of enacted support, dyad similarity, and provider personality. *Personality and Social Psychology Bulletin*, *28*(11), 1546-1555. DOI: 10.1177/014616702237582

Lance, G. N., Williams, W. T. (1967). Mixed-data classificatory programs I - agglomerative systems. *Australian Computer Journal*, *1*(1), 15–20.

Luo, S., Klohnen, E. C. (2005). Assortative mating and marital quality in newlyweds: a couple-centered approach. *Journal of Personality and Social Psychology*, *88*(2), 304-326. DOI: 10.1037/0022-3514.88.2.304

McCrae, R. R. (1993). Agreement of personality profiles across observers. *Multivariate Behavioral Research*, *28*(1), 25-40. DOI: 10.1207/s15327906mbr2801_2

Paszkiewicz, E. (1978). *Technika Q-sort*. Warszawa: PWN.

Robins, R. W., Caspi, A., Moffitt, T. E. (2000). Two personalities, one relationship: both partners' personality traits shape the quality of their relationship. *Journal of Personality and Social Psychology*, *79*(2), 251-259. DOI: 10.1037/0022-3514.79.2.251

Rożnowski, B., Korulczyk, T. (2018). Synthetic-singular index of Multidimensional Job Analysis. In A. Biela (Ed.), *Theoretical and methodological base for job analysis* (pp. 286-300). Frankfurt: Peter Lang Verlag.

Saaty, T. L., Vargas, L. G. (2013). *Decision making with the analytic network process. Economic, political, social and technological applications with benefits, opportunities, costs and risks (2 ed.).* New York: Springer.

Sherony, K. M., Green, S. G. (2002). Coworker exchange: relationships between coworkers, Leader–Member Exchange, and work attitudes. *Journal of Applied Psychology*, *87*(3), 542–548. DOI: 10.1037//0021-9010.87.3.542

Shrout, P. E., Fleiss, J. L. (1979). Intraclass correlations: uses in assessing rater reliability. *Psychological Bulletin*, *86*(2), 420-428. DOI: 10.1037/0033-2909.86.2.420

Tan, P.-N., Steinbach, M., Kumar, V. (2005). *Introduction to Data Mining.* New York: Pearson Education. DOI: 10.1002/0471687545.ch1

Tate, R. F. (1954). Correlation between a discrete and a continuous variable. Point-biserial correlation. *The Annals of Mathematical Statistics*, *25*(3), 603-607. DOI: 10.1214/aoms/ 1177728730

Tripp, T. M., Sondak, H. (1992). An evaluation of dependent variables in experimental negotiation studies: Impasse rates and Pareto efficiency. *Organizational Behavior and Human Decision Processes*, *51*(2), 273-295. DOI: 10.1016/0749-5978(92)90014-X

Van Exel, J., De Graaf, G. (2005). Q methodology: A sneak preview. *PsycEXTRA Dataset*. DOI: 10.1037/e518022006-003

Weingart, L. R., Hyder, E. B., Prietula, M. J. (1996). Knowledge matters: The effect of tactical descriptions on negotiation behavior and outcome. *Journal of Personality and Social Psychology*, *70*(6), 1205-1217. DOI: 10.1037//0022-3514.70.6.1205

Wells, P. A. (1991). Personality similarity and quality of marriage. *Personality and Individual Differences*, *12*(5), 407–412. DOI: 10.1016/0191-8869(91)90057-i

Węziak, D. (2007). Wielopoziomowe modelowanie regresyjne w analizie danych. *Wiadomości Statystyczne*, *9*, 1-12. DOI: 10.12775/pbe.2013.003

Tomasz Korulczyk

Instytut Psychologii
Katolicki Uniwersytet Lubelski Jana Pawła II

PRZEGLĄD WSPÓŁCZYNNIKÓW ZGODNOŚCI PAR

STRESZCZENIE

Choć w literaturze można znaleźć znaczącą ilość sposobów liczenia zgodności, niewiele jest prac jasno określających, w których kontekstach badawczych zastosowanie wskaźników z poszczególnych grup jest niewłaściwe. Ponadto, poza opisem samego wskaźnika trudno jest znaleźć informacje na temat jego optymalnej użyteczności. Celem tego artykułu jest zestawienie i uporządkowanie najbardziej znanych wskaźników wykorzystywanych do określania zgodności w badaniach par. Opisana została ich użyteczność oraz na jakie pytania badawcze pomagają odpowiedzieć poszczególne grupy współczynników. Podane zostały ich założenia matematyczne, słabe i mocne strony oraz warunki i przykłady zastosowania w psychologii. Na koniec zaprezentowane zostały wskazówki do interpretacji poszczególnych współczynników w oparciu o doniesienia z literatury. Warunkiem koniecznym zastosowania zaprezentowanych współczynników jest statystyczna niezależność (wzajemne nieskorelowanie) elementów w parze. Do artykułu przygotowany został także syntaks programu SPSS pozwalający obliczyć opisane wskaźniki. Zapoznanie się z artykułem może pomóc czytelnikowi podjąć bardziej świadome decyzje badawcze.

Słowa kluczowe: współczynniki zgodności, współczynniki podobieństwa, zgodność par, syntaks SPSS

SUPPLEMENTARY MATERIALS OF THE ARTICLE THE REVIEW OF DYADIC CONGRUENCY INDEXES

Table 1 Summary of indexes from various groups, conditions of their application, their advantages and disadvantages.

Order number	No. of formula	Name	Conditions of use	Advantages and disadvantages
Difference Indexes				
1	3	Simple Difference	• Variables measured on an interval scale	+ Ease of calculation - Interpretation only in terms of the size of the difference
2	4	Absolute Difference		
3	5, 6	Divided Simple Difference		
4	7, 8	Divided Absolute Difference		
Correlation Coefficients				
1	9	Pearson's product-moment correlation	• Variables measured on an interval scale • Variables should have a normal distribution • The relationship between variables should be linear • No outliers • Homoscedasticity of variables	+ High availability + Computation of congruence for a whole set of variables + Easy interpretation – Correlation coefficients are designed to the monotonicity of variables, not their agreement
1	10	Uncentered Pearson product-moment correlation		
2	11	Pearson's Distance		
3	12	Person's Absolute Distance		
4	13	Uncentered Pearson Distance		
5	14	Uncentered Pearson Absolute Distance		
7	16	Pearson product-moment correlation for Q-sort		
6	15	Spearman's rho	• Variables measured on an ordinal scale • The relationship between variables should be monotonic	

Table 1 Cont.

Order number	No. of formula	Name	Conditions of use	Advantages and disadvantages
Distance indexes				
1	17	Gower Agreement	• Variables measured on an interval scale	+ Computation of congruence for a whole set of variables + Can be used to measure non-compliance – Limited availability – Need to rescale some coefficients – Interpretation is relatively difficult without rescaling
2	18	Double-Scaled Euclidean		
3	19	Kernel-Smooth Distance		
1	20	Euclidian Distance		
2	21	Squared Euclidean Distance		
3	22	Minkowski's Distance		
4	23, 24	City Block Distance		
5	25	Canberra Distance		
6	26	Chebyshev Distance		
7	27	Power Distance		
8	28	Cosine Similarity		
Non-homogeneous coefficients				
1	29	McCrae's Rpa coefficient	• Variables measured on an interval scale	+ Suitable for profiles with extreme results – Limited availability
2	30	Sociogram	• Variables measured on an interval scale • Variables should have a normal distribution	+ Designed for measurement of agreement + Easy interpretation - Limited availability – It is difficult to choose the right one in advance
3	31	Integrativeness quotient		
4	32	C-index		
5	33	ICC-1	• Variables measured on an interval scale • Variables should have a normal distribution	+ Adapted to different research situations
6	34	ICC-2		
7	35	ICC-3		

INITIAL TRANSFORMATIONS

Calculation of Cohen's d value

First, we subtract the mean for a group of traits in one person from the mean for a group of traits in another person in a dyad and divide it by a pooled standard deviation. The formula for Cohen's d is as follows:

$$\boldsymbol{d} = \frac{\boldsymbol{M_1} - \boldsymbol{M_2}}{\boldsymbol{SD}} \tag{1}$$

A scaling variable x to a second variable y

Variables can be rescaled using the following formula (Luchonacho, 2017):

$$\mathrm{Y}_{\mathrm{Xscaled}} = \frac{\mathrm{Xmax} - \mathrm{Xmin}}{\mathrm{Ymax} - \mathrm{Ymin}} * \left(\mathrm{Y} - \mathrm{Ymax}\right) + \mathrm{Xmax} \tag{2}$$

where
X, Y = entered variables
Xmax = the maximum value of the scaled first variable
Ymax = the maximum value of the scaled second variable
Xmin = the minimum value of the scaled first variablej
Ymin = the minimum value of the scaled second variable

DIFFERENCE INDEXES

Indexes from this group will be presented below (Gorbaniuk, Stachoń-Wójcik, 2011):

1) A simple difference model, which can be expressed as a sum or mean (see Gaunt, 2006):

$$\mathrm{D} = \sum_{i=1}^{n} \left(A_{1i} - B_{2i}\right) \quad \text{or} \quad \mathrm{D} = \sum_{i=1}^{n} \left(A_{1i} - B_{2i}\right) / n \tag{3}$$

2) Model of absolute difference, which can be expressed as a sum (and is then identical to the City Distance) or as mean:

$$\mathrm{D} = \sum_{i=1}^{n} \left|A_{1i} - B_{2i}\right| \quad \text{or} \quad \mathrm{D} = \sum_{i=1}^{n} \left|A_{1i} - B_{2i}\right| / n \tag{4}$$

3) A model of a divided simple difference, which can be expressed as a sum or mean, where the simple difference of variables is weighted by one of these variables:

$$\mathrm{D} = \sum_{i=1}^{n} \left(A_{1i} - B_{2i}\right) / A_{1i} \quad \text{or} \quad \mathrm{D} = \left[\sum_{i=1}^{n} \left(A_{1i} - B_{2i}\right) / A_{1i}\right] / n \tag{5}$$

$$\mathrm{D} = \sum_{i=1}^{n} \left(A_{1i} - B_{2i}\right) / B_{2i} \quad \text{or} \quad \mathrm{D} = \left[\sum_{i=1}^{n} \left(A_{1i} - B_{2i}\right) / B_{2i}\right] / n \tag{6}$$

4) Model of the divided absolute difference, which can be expressed as a sum or mean, where the absolute difference of variables is weighted by one of these variables:

$$\mathrm{D} = \sum_{i=1}^{n} \left| A_{1i} - B_{2i} \right| / A_{1i} \quad \text{or} \quad \mathrm{D} = \left[\sum_{i=1}^{n} \left| A_{1i} - B_{2i} \right| / A_{1i} \right] / n \tag{7}$$

$$\mathrm{D} = \sum_{i=1}^{n} \left| A_{1i} - B_{2i} \right| / B_{2i} \quad \text{or} \quad \mathrm{D} = \left[\sum_{i=1}^{n} \left| A_{1i} - B_{2i} \right| / B_{2i} \right] / n \tag{8}$$

where:
A_i, B_i = first and second variable for i observation
n = number of observations

CORRELATION INDEXES

The following indicators will be presented below:

1) Pearson's product-moment correlation

The formula for the coefficient is as follows (Rodgers, Nicewander, 1988):

$$r = cov_{AB} / \left(s_A * s_B \right) \tag{9}$$

where:

$$cov_{AB} = \frac{\sum_{i=1}^{n} A_i B_i - \left[\left(\sum_{i=1}^{n} A_i * \sum_{i=1}^{n} B_i \right) / n \right]}{n-1}$$

$$s_A = \left(\sum_{i=1}^{n} A^2 - n * \bar{A}^2 \right) / (n-1)$$

i s_B– analogously to s_A

A_i, B_i = first and second variable for i observation

2) Uncentered Pearson product-moment correlation

The formula for the coefficient in this form is (Bandyopadhyay, Saha, 2012):

$$r_U = cov_{AB} / \left(s_A^{(0)} * s_B^{(0)} \right) \tag{10}$$

where:

$$cov_{AB} = \frac{\sum_{i=1}^{n} A_i B_i}{n-1}$$

$$s_A^{(0)} = \frac{\sum_{i=1}^{n} A^2}{n-1}$$ i $s_B^{(0)}$ – analogously to s_A

A_i, B_i = first and second variable for i observation

3) Pearson's Distance

The formula for conversion is as follows:

$$d = 1 - r \tag{11}$$

4) Person's Absolute Distance

The formula for conversion is as follows:

$$d = 1 - |r| \tag{12}$$

5) Uncentered Pearson Distance

The formula for conversion is as follows:

$$d = 1 - r_U \tag{13}$$

6) Uncentered Pearson Absolute Distance

The formula for conversion is as follows:

$$d = 1 - |r_U| \tag{14}$$

7) Spearman's *rho*

Wzór na współczynnik jest następujący (Gibbons, Chakraborti, 2003):

$$r = 1 - \left[6 \sum_{i=1}^{n} \left(A_i - B_i \right)^2 \right] / \left[n\left(n^2 - 1\right) \right] \tag{15}$$

where:

A_i, B_i = rank for individual variables

8) Pearson product-moment correlation for Q-sort

The formula is as follows (Cohen, 1957):

$$\mathrm{r} = 1 - \frac{\sum D^2}{K}, \quad \text{where} \quad \mathrm{K} = 2\mathrm{N}\sigma 2 \tag{16}$$

DISTANCE INDEXES

The following distance indexes will be listed below:

1) Gower Agreement (GA)

It is expressed in the form of formula (Gower, 1971):

$$ZG = 1 - \left[\left(\sum_{i=1}^{n} \frac{|A_i - B_i|}{Rs} \right) / n \right] \tag{17}$$

where:

n = number of observations

A_i, B_i = first and second variable for *i* observation

Rs = range, the maximum possible difference (max - min) taken from the variable A_i lub B_i

2) Double-Scaled Euclidean (DSE-s)
It is expressed in the form of formula (Barrett, 2010):

$$\text{DSEs} = 1 - \sqrt{\left(\sum_{i=1}^{n} \frac{\left(A_i - B_i\right)^2}{RS_i^2}\right) / n} \tag{18}$$

where:
n = number of observations
A_i, B_i = first and second variable for i observation
Rs = range, the maximum possible difference (max - min) taken from the variable A_i lub B_i

3) Kernel-Smooth Distance (KSD-s)
The formula of the statistic is as follows (Barrett, 2010):

$$KSD = \frac{1}{n}\left(\sum_{i=1}^{n}\left(\frac{1}{s\sqrt{2\pi}} * \exp\left(-\frac{\left(A_i - B_i\right)^2}{2s^2}\right)\right) * \left(100 * (s\sqrt{2\pi}\right)\right) \tag{19}$$

where:

$$s = \frac{RS_i}{SP}$$

SP = constant, *smoothing parameter*
n = number of observations
A_i, B_i = first and second variable for i observation
Rs = range, the maximum possible difference (max - min) taken from the variable A_i or B_i

Unscaled Distance Indexes:

1) Euclidian Distance
The coefficient is expressed as a formula (Cronbach, Gleser, 1953):

$$D = \sqrt{\sum_{i=1}^{n} \left(A_{1i} - B_{2i}\right)^2} \tag{20}$$

2) Squared Euclidean Distance
The coefficient is expressed as a formula (Cronbach, Gleser, 1953):

$$D^2 = \sum_{i=1}^{n} \left(A_{1i} - B_{2i}\right)^2 \tag{21}$$

3) Minkowski's Distance
The coefficient is expressed as a formula (Bandyopadhyay, Saha, 2012):

$$\mathrm{D}_{\mathrm{mink}} = \sqrt[r]{\sum_{i=1}^{n} \left|A_{1i} - B_{2i}\right|^p} \tag{22}$$

4) City Block Distance
The coefficient is expressed as a formula (Bandyopadhyay, Saha, 2012):

$$\mathrm{D}_{\mathrm{man}} = \sum_{i=1}^{n} \left| A_i - B_i \right| \tag{23}$$

After rescaling:

$$\mathrm{Man} = 1 - \left(\sum_{i=1}^{n} \frac{\left| A_i - B_i \right|}{RS_i^2} \right) / \, n \tag{24}$$

where:
n = number of observations
RS = range (max-min) for observation i
A_i, B_i = first and second variable for i observation

5) Canberra Distance
The coefficient is expressed as a formula (Lance, Williams, 1967):

$$\mathrm{D}_{\mathrm{can}} = \sum_{i=1}^{n} \frac{A_i - B_i}{\left| A_i \right| + \left| B_i \right|} \tag{25}$$

6) Chebyshev Distance
The coefficient is expressed as a formula (Bandyopadhyay, Saha, 2012):

$$\mathrm{D}_{\mathrm{czeb}} = Max\left(\left| A_i - B_i \right| \right) \tag{26}$$

7) Power Distance
The coefficient is expressed as a formula (Bandyopadhyay, Saha, 2012):

$$\mathrm{D}_{\mathrm{po}} = \sum_{i=1}^{n} \left(\left| A_{1i} - B_{2i} \right|^{p} \right)^{\frac{1}{r}} \tag{27}$$

8) Cosine similarity
The coefficient is expressed as a formula (Bandyopadhyay, Saha, 2012):

$$\cos(\theta) = \left(\sum_{i=1}^{n} A_{1i} * B_{2i} \right) / \sqrt{\sum_{i=1}^{n} A_i^2 * B_i^2} \tag{28}$$

NON-HOMOGENEOUS COEFFICIENTS

The following are non-homogeneous coefficients (NHC):

1) McCrae's r_{pa} coefficient
The coefficient is expressed as a formula (McCrae, 1993):

$$r_{pa} = I_{pa} / \sqrt{(n-2) + I_{pa}^2} \tag{29}$$

where:

$$I_{pa} = \left(n + 2 * \sum_{i=1}^{n} \left(\frac{(A_i + B_i)}{2} \right)^2 - \sum_{i=1}^{n} \left| A_i - B_i \right|^2 \right) / \sqrt{10 * n}$$

2) Sociogram

For easier interpretation, the above variables can also be rescaled on the Person's product-moment correlation score (from –1 to 1) based on the following formula (Luchonacho, 2017):

$$r_x = \frac{2}{\mathrm{Xmax} - \mathrm{Xmin}} * (\mathrm{X} - \mathrm{Xmax}) + 1 \qquad (30)$$

where:
X = scaled variable
Xmax = the maximum value of the scaled variable
Xmin = the minimum value of the scaled variable

3) Integrativeness quotient

The coefficient is expressed as a formula (Schweitzer, Gomberg, 2001):

$$\mathrm{PE}_{\mathrm{score}} = \arcsin\left(1 - \frac{PS(x_i, y_i)}{PS(x_i, y_i) + PI(x_i, y_i)}\right) \qquad (31)$$

where:
$PS(x_i, y_i)$ = number of favourable Pareto agreements (settlements)
$PI(x_i, y_i)$ = number of unfavorable Pareto agreements (settlements)

4) C-index

The coefficient is expressed as a formula (Brown, Gore, 1994):

$$\mathrm{C}_{\mathrm{index}} = 3(\mathrm{X}) + 2(\mathrm{Y}) + (\mathrm{Z}) \qquad (32)$$

where:
X = distance (in the hexagonal model) between the first letters with the highest result
Y = distance (in a hexagonal model) between the second letters with the highest result
Z = distance (in a hexagonal model) between the third letters with the highest result

5) ICC-1

The coefficient is expressed as a formula (Barrett, 2010):

$$ICC1 = (MS_p - WMS) / \left[MS_p + (k-1) * WMS \right] \qquad (33)$$

where:
MS_p = The average square of deviations between observations
MS_r = The average square of deviations between variables
MS_{res} = The average square of residual deviations
k = number of variables
n = number of observations

WMS = The average square of deviations inside the observation:

$$WMS = \frac{\left[\left(MS_r * (k-1)\right) + \left(MS_{res} * (n-1) * (k-1)\right)\right]}{n * (k-1)}$$

6) ICC-2

The coefficient is expressed as a formula (Barrett, 2010):

$$r_{icc2} = \frac{MS_p - MS_{res}}{MS_p + (k-1) * MS_{res} + \left(\left(k * \left(MS_r - MS_{res}\right)\right)/n\right)} \tag{34}$$

where:
MS_p = The average square of deviations between observations
k = number of variables
n = number of observations
MS_{res} = The average square of residual deviations (interactions)
MS_r = The average square of deviations between variables

7) ICC-3

The coefficient is expressed as a formula (Barrett, 2010):

$$r_{icc2} = \left(MS_p - MS_{res}\right) / \left[MS_p + (k-1) * MS_{res}\right] \tag{35}$$

where:
MS_p = The average square of deviations between observations
MS_{res} = Average square of residual deviations (interactions)
k = number of variables

REFERENCES

Bandyopadhyay, S., Saha, S. (2012). *Unsupervised classification: similarity measures, classical and metaheuristic approaches, and applications.* Berlin: Springer-Verlag.

Barrett, P. (2010). *Test reliability and validity: The inappropriate use of the Pearson and other variance ratio coefficients for indexing reliability and validity.* Pozyskano z: *http:// www.pbarrett.net/techpapers/correlations_reliability_validity_Rev_1_July_2010.pdf.*

Brown, S. D., Gore, P. A. (1994). An evaluation of interest congruence indices: Distribution characteristics and measurement properties. *Journal of Vocational Behavior*, *45*(3), 310–327. DOI:10.1006/jvbe.1994.1038

Cohen, J. (1957). An aid in the computation of correlations based on Q sorts. *Psychological Bulletin*, *54*(2), 138-139. DOI:10.1037/h0046220

Cronbach, L. J., Gleser, G. C. (1953). Assessing similarity between profiles. *Psychological Bulletin*, *50*(6), 456-473. DOI:10.1037/h0057173

Gaunt, R. (2006). Couple similarity and marital satisfaction: are similar spouses happier? *Journal of Personality*, *74*(5), 1401-1420. DOI:10.1111/j.1467-6494.2006.00414.x

Gibbons, J. D., Chakraborti, S. (2003). *Nonparametric Statistical Inference (4 wyd.).* New York: Marcel Dekker, Inc. DOI:10.1007/978-3-642-04898-2_420

Gorbaniuk, O., Stachoń-Wójcik, M. (2011). Zgodność obrazu siebie i wizerunku konsumenta-wartość predykcyjna różnych metod pomiaru kongruencji. *Marketing i Rynek*, *6*, 17-26.

Gower, J. C. (1971). A general coefficient of similarity and some of its properties. *Biometrics, 27*(4), 857–871. DOI:10.2307/2528823

Lance, G. N., Williams, W. T. (1967). Mixed-Data Classificatory Programs I - Agglomerative Systems. *Australian Computer Journal*, *1*(1), 15-20.

Luchonacho. (2017). *Stackexchange*. Pozyskano z: *https://stats.stackexchange.com/questions /25894/changing-the-scale-of-a-variable-to-0-100*

McCrae, R. R. (1993). Agreement of personality profiles across observers. *Multivariate Behavioral Research*, *28*(1), 25-40. DOI:10.1207/s15327906mbr2801_2

Rodgers, J. L., Nicewander, W. A. (1988). Thirteen ways to look at the correlation coefficient. *The American Statistician*, *42*(1), 59-66. DOI:10.1080/00031305.1988.10475524

Schweitzer, M. E., Gomberg, L. E. (2001). The Impact of Alcohol on Negotiator Behavior: Experimental Evidence. *Journal of Applied Social Psychology*, *31*(10), 2095-2126. DOI:10.1111/j.1559-1816.2001.tb00165.x