DERLEME REVIEW

Usage of the Intraclass Correlation Coefficient As a Measure of Dependence in Dyadic Data: Review

Diyadik Verilerde Bağımlılığın Bir Ölçüsü Olarak Sınıf İçi Korelasyon Katsayısının Kullanımı

İsmet DOĞAN,[a]
Nurhan DOĞAN[a]

[a]Department of Biostatistics and Medical Informatics,
Afyon Kocatepe University
Faculty of Medicine,
Afyonkarahisar

Geliş Tarihi/*Received:* 27.05.2015
Kabul Tarihi/*Accepted:* 02.07.2015

Yazışma Adresi/*Correspondence:*
Nurhan DOĞAN
Afyon Kocatepe University
Faculty of Medicine,
Department of Biostatistics and
Medical Informatics, Afyonkarahisar,
TÜRKİYE/TURKEY
nurhandogan@hotmail.com

ABSTRACT Dependence is an important issue because traditional analysis approaches, such as analysis of variance and ordinary least square regression, assume independence of individual scores and they can produce biased parameter estimates and standard errors if applied incorrectly to dyadic data. A number of analytical techniques that estimate the degree of dependence in dyadic data are available. Perhaps the most widely used approach is the intraclass correlation which quantifies the proportion of response variable variability caused from mean differences across dyads. The primary purpose of this article is to provide a review for alternatives to intraclass correlation coefficient (ICC) approach for analyzing dyadic data. This procedure is limited to ICCs developed data with just one systematic variance (one way model), because it is only ICCs for dyadic relationship that has been derived. This article discusses the strength and weakness of each ICC methods in the analysis of dyadic data. The importance of appropriate ICC in the analysis of dyadic relationship in dyadic data is also highlighted in this article. When investigators concerned with the dyadic relationship among multiple observations made on randomly selected objects of measurement and the error variance for measurement is uniform across the condition measurement, ICC provide appropriate measure.

Key Words: Dyadic relationship; dependence; intraclass correlation coefficient; analysis of variance

ÖZET Varyans analizi ve regresyon analizi gibi klasik analiz yöntemlerindeki gözlem değerlerinin bağımsızlığı varsayımından dolayı bağımlılık önemli hale gelmekte ve diyadik verilerin değerlendirilmesinde klasik yöntemlerin yanlışlıkla uygulanması yanlı parametre tahminlerinin ortaya çıkmasına neden olmaktadır. Diyadik verilerde bağımlılığın derecesini tahmin eden bir dizi analitik yöntem bulunmaktadır. Yanıt değişkenindeki değişimi oran olarak ifade eden sınıf içi korelasyon katsayısı diyadik verilerdeki bu farklılıkları anlamak için en yaygın kullanılan yaklaşımdır. Bu çalışmanın temel amacı diyadik verilerin analizi için bir yaklaşım olan sınıf içi korelasyon katsayısı yaklaşımını gözden geçirmektir. Diyadik ilişkilerin türetildiği sınıf içi korelasyon katsayısı ile ilgili prosedür, sistematik varyans (tek yönlü model) kaynağının bir tane olduğu veriler ile ilgili yöntemle sınırlandırılmıştır. Bu makalede diyadik veri analizinde kullanılan sınıf içi korelasyon katsayısı yöntemlerinin her birine ilişkin güçlü ve zayıf yönler tartışılmıştır. Diyadik verilerde diyadik ilişkinin analizinde uygun sınıf içi korelasyon katsayısının kullanılmasının önemi de bu makalede vurgulanmıştır. Sınıf içi korelasyon katsayısı, ölçüm değerlerine ait hata varyansının tekbiçimli olduğu ölçüm koşullarının söz konusu durumlarda rasgele seçilmiş nesnelerin ölçümünün yapıldığı çoklu gözlemler arasındaki diyadik ilişkiler ile ilgilendiklerinde araştırmacılara uygun ölçüm sağlamaktadır.

Anahtar Kelimeler: Diyadik ilişki; bağımlılık; sınıf içi korelasyon katsayısı; varyans analizi

Turkiye Klinikleri J Biostat 2015;7(2):119-25

doi: 10.5336/biostatic.2015-46361

The dyad, which consists of two members, is arguably the fundemental unit of interpersonal interaction and relations. People involved in dyadic relationships often influence each other's behaviours, cognitions and emotions.[1] The study of behavior and emotion naturally benefits from gathering information about the most influential factors. For many individuals, characteristics of their close relationships may extensively influence their behavior and affect. For this reason, it is desirable to study couples or dyads together. Methodologically, these data are more complete in terms of potentially influential factors, because interactions between partners (such as disagreements or sexual intercourse), couple level characteristics (such as relationship duration) and individual-level characteristics (such as age) may all play a role in determining behavior and mood. Dyadic data, therefore, can answer complex questions about behavior and mood. Analytically, dyadic data presents challenges when compared with data from independent individuals.[2]

The intrinsically dyadic nature of many of the measurements in social and behavioral science research means that they are often linked to other measurements in the study, and the strength of these links may be one of the most important research questions to be examined. Consider the following examples:

a) Both members in a romantic relationship evaluate whether they are satisfied with the relationship,

b) The amount of self-disclosure made by two people interacting is measured to ascertain whether there is reciprocity,

c) Two persons are asked to describe a common target person to determine whether there is agreement in person perception,

d) Members of a family describe their attachment relationships with one another.

In each of these cases, the issues of stability, consistency, and correlation between related measurements are interesting phenomena worth studying in their own right. However, none of them can be addressed easily by standard methods developed for the study of individuals.[3]

The study of relationships often involves collecting data from more than one partner or group member (e.g., siblings, parents and children, etc.). Such data violate the assumption of statistical independence inherent in traditional analytic methods, and so more complex analytic methods are required. The relationship research often requires analysis of "mixed independent variables," or variables that vary both within and between dyads.[4]

Dependence is an important issue because traditional analysis approaches that assume independence of individual scores, such as analysis of variance and ordinary least square regression, can produce biased parameter estimates and standard errors if applied incorrectly to dyadic data. Dyadic dependence refers to the fact that the variable scores collected from individuals interacting within dyads are not independent, but are likely to be more correlated than scores from individuals in different dyads. A number of analytic techniques (ICC, hierarchical linear modeling, actor-partner interdependence model, structural equation modeling, within and between analysis, random group resampling etc.) are available that estimate the degree of dependence in dyadic data. Perhaps the most widely used approach is the intraclass correlation which quantifies the proportion of response variable variability that is due to mean differences across dyads.[5]

Agreement between fixed observers or methods that produce readings on a continuous scale is usually evaluated via one of several intraclass correlation coefficients.[6] When one is interested in the relationship among variables of a common class, which means variables that share both their metric and variance, ICCs are alternative statistics for measuring homogeneity, not only for pairs of measurements but for larger sets of measurements as well. The most fundamental interpretation of an ICC is that it is a measure of the proportion of a variance (variously defined) that is attributable to objects of measurement, often called targets. The objects might be gymnastics contestants, litters, twin pairs, or students, and the corresponding measurements might be Judges' ratings, IQs of the twins, weights of the littermates, or test scores of the students.[7]

The ICC is one of the oldest, as well as one of the most versatile, statistics. The original computational method for the ICC was proposed by Karl Pearson, in 1901.[8] Although also called a correlation, the intraclass correlation is really not a correlation at all. The label is an unfortunate misnomer that disguises the fact that the intraclass correlation is a univariate statistic (unlike the Pearson correlation, which is a bivariate statistic). It measures association within groups (in this case, dyads) but when only a single variate (e.g., turn length) is involved. It was originally developed to analyze twins and subsequently extended to other cases of matched members of a given sampling unit, in this case, members of a dyad. It is most appropriate when there is no basis for distinguishing Person A from Person B so that there are no separate X and Y variates, only a Y variate obtained from two or more group members. The intraclass can be conceptualized as a repeated measures analysis of variance with dyad or group member being the repeated factor. It can also be viewed as a reliability problem, with the focus being on consistency between dyad members. ICC considers both similarities in means and the shapes of distributions, but it ignores individual differences between partners.[9]

The ICC which is a reproducibility criterion giving the proportion of variance attributable to differences between methods. The ICC was developed to deal with several measurement methods and has emerged as a universal and widely accepted reproducibility index.[10] The earliest ICCs were modifications of the Pearson Correlation Coefficient. There is no ordering of the repeated measures in the ICC, and it can be applied to more than two repeated measurements. However, the modern version of ICC is now calculated using variance estimates, obtained from the analysis of variance, through partitioning of the total variance between and within subject variance.[11]

ICCs are used when one is interested in the relationship among variables of a common class, which means variables that share both their metric and variance. ICCs are based on variance partitioning and therefore are subject to essentially the same assumptions as analysis of variance. These include homogeneity of variance (the variances within the units are statistically the same), normality (the population scores are normally distributed), statistical independence (the observations are independent), and measures that are of equal psychological intervals.[12] If data were collected from dyadic relationships, three additional considerations need to be addressed before the correct statistical analysis technique can be identified and applied: dependence, distinguishability, and the type of dyadic variables to be analyzed.[5]

There are different types of ICC. Researcher are sometimes confused and unsure which type of ICC to use. Muller and Buttner (1994), demonstrated that the different types of ICC may result in quite different values for the same data set, under the same sampling theory. So it is important to determine which type of ICC is suitable, depending on the purpose of the analysis. Therefore it is important to report the type of ICC that has been used in a reliability study. There has been considerable debate regarding the most appropriate type of ICC to be used in measuring reliability. Several versions of the ICC were derived depending on the study scheme.[13] Potential alternatives were presented in Bartko (1966), Shrout and Fleiss (1979), McGraw and Wong (1996).[7,14,15] It is shown that care must be taken in choosing a suitable ICC with respect to the underlying sampling theory. For this purpose a decision table was given in Cook (2000). It may be used to choose a coefficient which is appropriate for a specific study setting (Table 1).[16]

The intraclass correlation, provides a unique estimate of the relationship between scores from indistinguishable dyad members. Intraclass correlations for dyads are interpreted in the same fashion as Pearson correlations. Thus, if a dyad member has a high score on a measure and the intraclass correlation is positive, then the other dyad member also has a relatively high score; if the intraclass correlation is negative, then the other dyad member has a relatively low score. A common alternative interpretation of a positive intraclass correlation is the proportion of variation in the outcome measure that is accounted for by dyad. That is, if the intraclass correlation equals 0.40, then 40% of the

TABLE 1: Intraclass correlation coefficients and conditions for their use.[16]

ICC	Analysis of Variance	Model	Unit of Analysis	Formula
ICC(1,1)	One-way	Random targets Random judges	Individual	$\frac{BMS-WMS}{BMS+(k-1)WMS}$
ICC(2,1)	Two-way	Random targets Random judges	Individual	$\frac{BMS-EMS}{BMS+(k-1)EMS+k(JMS-EMS)/n}$
ICC(3,1)	Two-way	Random targets Fixed judges	Individual	$\frac{BMS-EMS}{BMS+(k-1)EMS}$
ICC(1, k)	One-way	Random targets Random judges	Mean	$\frac{BMS-WMS}{BMS}$
ICC(2, k)	Two-way	Random targets Random judges	Mean	$\frac{BMS-EMS}{BMS+(JMS-EMS)/n}$
ICC(3, k)	Two-way	Random targets Fixed judges	Mean	$\frac{BMS-EMS}{BMS}$

n: number of targets (subjects); k: number of judges (trials); BMS: between targets mean square; WMS: within target mean square; JMS: between judges mean square; EMS: error mean square

variation in the scores is accounted for by the particular dyad to which individuals belong. The common-variance explanation of the nonindependence becomes problematic when the correlation is negative.[3]

The basic idea for dyadic data analysis is that when members of the same dyad are analyzed together, their data are dependent on each other in much the same way that two observations in a two-time-point repeated measures design are related to each other.[17] A convenient arrangement for the measurements is reproduced in Table 2, where each row represents a dyad (which vary in number from 1 to n) and each column represents a variable measuring some characteristic for each member of the dyads (which vary in number from 1 to 2) (Table 2).

The intraclass correlation can be computed via several approaches, the first of which is based on analysis of variance (ANOVA) techniques. The calculation of dyadic relation starts with the performance of a repeated-measures ANOVA. The ICC is calculated using variance estimates, obtained from the repeated-measures ANOVA, through partitioning of the total variance between and within dyad variance. The computational formulas are summarized in Table 3.

The intraclass correlation for dyads is then defined as:[3]

$$ICC = \frac{BMS - WMS}{BMS + WMS}$$

Consider as an example the data in Table 4 from a fictitious study of liking between same gen-

TABLE 2: Data Matrix for dyadic data used in calculating ICC.

Dyads	Members 1	2	Total
1	x_{11}	x_{12}	P_1
2	x_{21}	x_{22}	P_2
3	x_{31}	x_{32}	P_3
.	.	.	.
.	.	.	.
i	x_{i1}	x_{i2}	P_i
.	.	.	.
.	.	.	.
n	x_{n1}	x_{n2}	P_n
Total	T_1	T_2	G

der roommates. In this dyad structured data set there are 10 pairs of roommates and individuals are asked to rate how much they like their roommates on a 9 point scale. The following demonstration illustrates why a Pearson correlation would be the wrong approach for estimating the correspondence between liking scores when dyad members are indistinguishable (Table 4).

The analysis results are given in Table 5.

In the case of dyadic and group designs, the ICC has a special meaning because it assesses the degree of agreement within group members. For example, if we assess how often two strangers speak, the ICC provides a measure of agreement within dyads, and so it provides a natural measure of interdependence. If each individual vocalizes at a rate that is equal to his dyadic partner's, but different dyads have different mean levels of vocalization, then the ICC will be a perfect 1 because pairs are maximally similar (i.e. all the variance is between couples). There are two measures used in determining ICC, namely inter-dyad variation (variation between dyads) and intra-dyad variation (variation within dyads). When inter-dyad and intra-dyad variations are equal, ICC becomes 0, since there is no evidence of similarity or distinction across coupled individuals. If intra-dyad variation is bigger, ICC will be negative, indicating that individuals within groups are behaving more distinctly than expected by chance. Hence, this suggests that these individuals are acting in a complementary fashion.[8] However, the ICC values for the reproducibility suggested by Rosner (2010) can be used for dyadic relation.[18] According to these considerations, it suggests that ICC < 0.4 indicate poor dyadic relationship, 0.4 ≤ ICC < 0.75 as fair to good dyadic relationship and ICC ≥ 0.75 as

TABLE 3: ANOVA table for dyadic data.

Source of Variation	Degrees of Freedom	Sum of Square	Mean Square
Between dyads	$n-1$	$(4)-(1)$	BMS
Within dyads	n	$(2)-(4)$	WMS
Between members	1	$(3)-(1)$	MMS
Residual	$n-1$	$(2)-(3)-(4)+(1)$	RMS
$(1)=G^2/2n$	$(2)=\sum\sum X_{ij}^2$	$(3)=\left(\sum T_j^2\right)/n$	$(4)=(\sum P_i^2)/2$

TABLE 4: ICC example using data from fictitious study of roommates.

	Scores	
Dyads	1	2
1	8	6
2	5	3
3	7	2
4	8	5
5	8	7
6	5	6
7	3	4
8	8	9
9	6	7
10	2	3

TABLE 5: ANOVA table for fictitious study of roommates data.

Source of Variation	Degrees of Freedom	Sum of Square	Mean Square
Between dyads	9	66.800	7.422
Within dyads	10	24.000	2.400
Between members	1	3.200	3.200
Residual	9	20.800	2.311

ICC = 0.511
Pearson Correlation = 0.525

excellent dyadic relationship. ICC is a ratio of variances derived from analysis of variance, so it is unitless.

The ICC is maximally positive when within-dyad scores are identical and scores differ between dyads. It is maximally negative when dissimilarity within dyads is high (i.e., there is high within-dyad variance) and differences across dyads are small (i.e., there is little variance). Hence, a large, positive ICC may be interpreted as evidence of reciprocity and a large, negative ICC as compensation. However, the latter interpretation is clouded by the fact that ICC is highly sensitive to sources of heterogeneity between dyad members (e.g., sex, age, status, role, or communication competence) that might influence the behavior under examination. Consequently, a negative correlation may be due either to actual adaptation between interactants or to heterogeneity on other variables. It is therefore best reserved for use when dyad members are known to behave similarly prior to the interaction (i.e., their baseline behaviors are nearly the same) or when "investigators assume that reciprocal and compensatory interaction patterns are contingent upon similarities and differences among behavioral levels" and investigators wish to document that the between-subjects factors account for the patterns.[9]

CONCLUSION

Many processes under study in the health sciences, such as treatment delivery, child care, and disease transmission, involve interpersonal relationships and mutual influence involving two persons (e.g., physician–patient, parent–child, wife–husband). Conventional methods for inferential data analyses, including analysis of variance and general linear regression, assume that observations obtained from each individual are independent. When such analyses are applied to data obtained from interacting dyads, the assumption of independent observations may be violated, leading to underestimation of standard errors and invalid inferences (i.e., increased Type I error). To overcome the problem of nonindependence in the case of distinguishable dyad members, such as female–male couples, researchers often conduct separate analysis for each member class.[19]

Several methods have been proposed for analyzing dyadic data, among them are methods based on mixed effects or multi-level models and structural equation models. Many of these methods, however, require each observation to have a measure from one member of the dyad paired with a measure from the other member of the dyad so that the data consists of multiple paired observations. This is not always the case depending on how data is collected.[2]

The ICC takes into account the level and variability of each member's responses and emphasizes differences between individuals. The ICC is the measure of the relative homogeneity of the scores within the classes in relation to the total variation among all the scores. Thus, as homogeneity increases, ICC values increase; as homogeneity decreases, ICC values decrease.[20]

The ICC is influenced greatly by between-subjects variability. If the ICC is applied to data from a group of individuals with a wide range of the measured characteristics, the value of the ICC will indicate higher reproducibility, compared to the same analysis when applied to a group of data with a narrow range of the same characteristic. However, this is an unfair criticism, because the ICC is not meant to provide an index of absolute measurement error. In general, the ICC is a ratio and does not quantify precision.[21]

Pearson product-moment coefficient of correlation ignores both level and variability of the couple's responses. A Pearson correlation coefficient essentially represents the degree to which two respondents are a predictable distance from each other, i.e,, the similarity of response pattern. Thus, if two respondents consistently respond to items with scores that differ by the same amount, the Pearson correlation would be unity, even though there is no real agreement. Although the correlation equals 1, true agreement is not represented; instead, the correlation is a measure of predictability. The ICC takes level and variability of scores into account. There is perfect correlation (predictability) and perfect agreement when the ICC equals 1. Further, the ICC assesses the degree of response agreement for multiple observers without limiting who the observers might be (e.g., self, child, parent, friend, trained coder, etc.).[20]

There is no single perfect method to assess dyadic relationship; however researchers should be aware of the inappropriate ICC methods that they should avoid when analyzing dyadic data. Inappropriate ICC will lead to invalid conclusions and thus validated instrument might not be accurate or reliable.

REFERENCES

1. Cappelleri JC, Stecher VJ. Analysis of dyadic data: a guideline applied to erectile dysfunction. J Eval Clin Pract 2010;16(6):1340-4.
2. Blood EA, Kalish LA, Shrier LA. Estimating heterogeneous intra-class correlation coefficients in dyadic ecological momentary assessment. J Mod App Stat Meth 2013;12(1):207-19.
3. Kenny DA, Kashy DA, Cook WL. Basic definitions and overview. Dyadic Data Analysis. 1st ed. New York: The Guilford Press; 2006. p.1-24.
4. Isaacs C. He says, she says: a dyadic approach to understanding safer sex behaviour in intimate heterosexual relationships, PhD. Thesis, Department of Psychology, The University of Western Ontario, London, Ontario, Canada; 2013. p.17-8.
5. Peugh JL, DiLillo D, Panuzio J. Analyzing mixed-dyadic data using structural equation models. Structural Equation Modeling 2013;20(2):314-37.
6. Haber M, Barnhart HX. Coefficients of agreement for fixed observers. Statistical Methods in Medical Research 2006;15(3):255-71.
7. McGraw KO, Wong SP. Forming inferences about some intraclass correlation coefficients. Psychological Methods 1996;l(1):30-46.
8. Griffin D, Gonzalez R. Models of dyadic social interaction. Philosophical Transactions of the Royal Society B: Biological Sciences 2003;358(1431):573-81.
9. Burgoon JK, Stern LA, Dillman L. Issues in studying interaction adaptation, analyzing adaptation patterns, Intraclass correlation. Interpersonal Adaptation: Dyadic Interaction Patterns. Cambridge, Angleterre: Cambridge University Press; 1995. p. 156-8.
10. Vanbelle S. Agreement between raters and groups of raters, PhD. Thesis, Department de Mathematique, Universite De Liege, Liege, Wallonia, Belgium; 2009. p.6.
11. Rafdzah Z, Awang B, Noorhaire N, Noor AI. A systematic review of statistical methods used to test for reliability of medical instruments measuring continuous variables. Iran J Basic Med Sci 2013;16(6):803-7.
12. Castro SL. Data analytic methods for the analysis of multilevel questions: A comparison of intraclass correlation coefficients, rwg(j), hierarchical linear modeling, within-and between-analysis, and random group resampling. The Leadership Quarterly 2002;13(1):69-93.
13. Müller R, Büttner P. A critical discussion of intraclass correlation coefficients. Stat Med 1994;13(23-24):2465-76.
14. Bartko JJ. The intraclass correlation coefficient as a measure of reliability. Psychol Rep 1966;19(1):3-11.
15. Shrout PE, Fleiss JL. Intraclass correlations: uses in assessing rater reliability. Psychol Bull 1979;86(2):420-8.
16. Cook C. A review of intraclass correlation. Technical Report. No: TM030336. Texas: Texas A&M University; 2000; p.29.
17. Gonzales R, Griffin D. Dyadic data analysis. In: Cooper HM, ed. APA Handbook of Research Methods in Psychology: Data Analysis and Research Publication. Washington, DC: American Psychological Association; 2012. p. 439-50.
18. Rosner B. (Multisample Inference) Fundementals of Biostatistics. 7th ed. Boston: Brooks/Cole, Cengage Learning Inc; 2010. p.569.
19. McMahon JM, Pougeta ER, Tortu S. A guide for multilevel modeling of dyadic data with binary outcomes using SAS PROC NLMIXED. Comput Stat Data Anal 2006;50(12):3663-80.
20. Porter LC. Generating a couple score: Insider and insider-outsider reports of a marriage, PhD. Thesis, Texas Tech University, Lubbock, Texas; 1996; p.25.
21. Weir JP. Quantifying test-retest reliability using the intraclass correlation coefficient and the SEM. J Strength Cond Res 2005;19(1):231-40.