

# CORRELATION AND REGRESSION MODELS IN THE DYADIC DATA ANALYSIS

ŚLĄSKI  
PRZEGLĄD  
STATYSTYCZNY  
Nr 18(24)

Cyprian Kozyra

Wrocław University of Economics and Business, Poland

e-mail: cyprian.kozyra@ue.wroc.pl

ISSN 1644-6739  
e-ISSN 2449-9765

© 2020 Cyprian Kozyra

*This work is licensed under the Creative Commons Attribution-ShareAlike 4.0 International License. To view a copy of this license, visit <http://creativecommons.org/licenses/by-sa/4.0/>*

*Quote as:* Kozyra, C. (2020). Correlation and regression models in the Dyadic Data Analysis. *Śląski Przegląd Statystyczny*, 18(24).

DOI: 10.15611/sps.2020.18.17

JEL Classification: C20, C31

In this paper the Dyadic Data Analysis methods are presented with illustrations based on Polish official statistics data.

Dyadic Data Analysis (DDA) is dedicated only to data, which is organized as dyads, e.g. groups with exactly two members, where every dyad and each member of dyads is measured with the same variable.

**Table 1.** Presentation of row dyadic data

Observations	Variables	1. variable (X)	
		1. data ( $X_1$ )	2. data ( $X_2$ )
1. dyad		$x_{11}$	$x_{21}$
2. dyad		$x_{21}$	$x_{22}$
3. dyad		$x_{31}$	$x_{32}$

Source: (Dobos & Gelei, 2019).

DDA was introduced by the group of researchers in social sciences (R. Gonzales, D. Griffin, D.A. Kenny) in the 1990s, but it is still being developed and applied in different areas. There are two kinds of analysis, for distinguishable and exchangeable dyads, and a typical case of a distinguishable dyad is a married couple, i.e. husband and wife. Sometimes dyads in analysis are treated as indistinguishable and this is a reason for exchangeable dyadic data analysis. Datasets for dyadic data correlation and regression analysis are organized typically as double-entry tables.

**Table 2.** Dyadic data organized as double-entry table before analysis

Observations	Variables	
	$X$	$X'$
1. pair (initial order)	$x_{11}$	$x_{12}$
1. pair (changed order)	$x_{12}$	$x_{11}$
2. pair (initial order)	$x_{21}$	$x_{22}$
2. pair (changed order)	$x_{22}$	$x_{21}$
3. pair (initial order)	$x_{31}$	$x_{32}$
3. pair (changed order)	$x_{32}$	$x_{31}$

Source: (Dobos & Gelei, 2019).

Dependencies within dyads are investigated by means of the intraclass correlation coefficient (ICC)  $r(X, X')$ , which could be equivalently calculated from the one-way ANOVA for row dyadic data as  $(SSB - SSE)/(SSB + SSE)$ , where  $SSB$  is the sum of the squares between dyads and  $SSE$  is the sum of the squares within dyads. In the distinguishable case different than the classic Pearson's linear correlation, a coefficient equation is used where  $C$  is the code of dyad order – 1 initial or 2 changed:

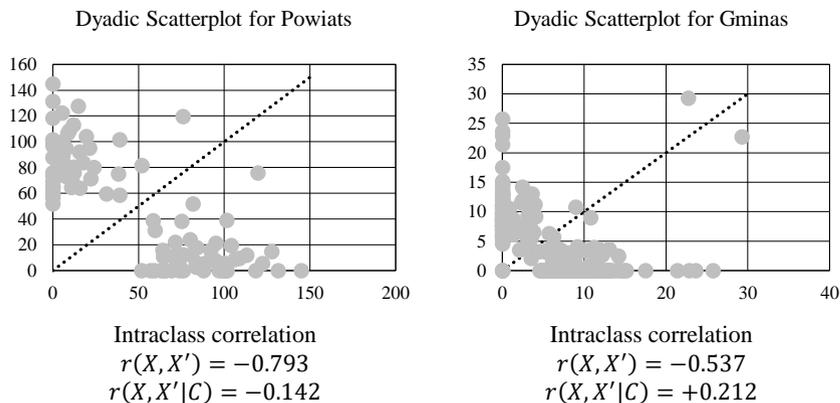
$$r_{XX':C} = \frac{r_{XX'} - r_{XC}r_{X'C}}{\sqrt{(1 - r_{XC}^2)(1 - r_{X'C}^2)}},$$

the equivalent to the two-way ANOVA for row dyadic data as  $(SSB - SSE)/(SSB + SSE)$ , where  $SSB$  is the sum of squares between dyads and  $SSE$  is the sum of squares within dyads.

DDA is illustrated by official statistics data concerning access to public services in 2017, for Polish local government units ('powiat' and 'gmina'), which are then divided into two separate units, urban and rural, with the seat in the same city or town. Such 46 powiat dyads (variable: number of beds in hospitals for 10 thousand inhabitants of 'powiat') and 143 'gmina' dyads (variable: number of classes in high schools for 1000 inhabitants of a 'gmina' aged 15-19) were analysed to illustrate DDA.

The results of the ICC analysis show an overall negative dependency in urban-rural dyads – the higher the level of access to public services in urban administrative units, the lower the level of access in rural administrative units, but it is true for the 'gminas' dataset only after removing the outlier dyad with high value for both local government units. The results also show that the negative value of Pearson correlation coefficient  $r(X, X')$  could be observed together with the positive value of ICC properly calculated for the distinguishable dyads. Dyad points with initial and changed order are located symmetrically to the diagonal dotted

line, and they are usually joined in smaller datasets with a line between two points on the scatterplot.



**Fig. 1.** Scatterplots for double-entry data concerning access to public services in urban and rural local government units with a common seat

Source: own elaboration.

DDA also introduced other types of the analysed correlation: cross-intraclass correlation, overall within-partner correlation, mean-level correlation, individual-level correlation, and dyad-level correlation.

Regression analysis in DDA can be simple, based on ICC, or extended to the actor-partner interdependence model (APIM) – see Table 4.

The basic coefficients and models of DDA could be estimated in every spreadsheet, e.g. MS Excel. Advanced analysis can be carried out in statistical software offering structural equations modelling (SEM) or hierarchical linear models (HLM). The code for many solutions is available in R language environment and its packages (also DyadR in github website and as a shiny application). Dyadic data analysis is a useful and interesting approach in the analysis of social problems when dyads are stable, with many potential areas of application.

**Table 3.** Equations of other correlation coefficients used in DDA

Type of correlation	Equation for $(X, X')$ , $(Y, Y')$ data (double entry)
1	2
Cross-intraclass correlation	$r(X, Y') = \frac{\text{cov}(X, Y')}{\sqrt{\text{var}(X)} \cdot \sqrt{\text{var}(Y)}}$
Overall within-partner correlation	$r(X, Y) = \frac{\text{cov}(X, Y)}{\sqrt{\text{var}(X)} \cdot \sqrt{\text{var}(Y)}}$

Tabela 3, cont.

Nr 18(24)

1	2
Mean-level correlation	$r_m(X, X', Y, Y') = \frac{r(X, Y) + r(X, Y')}{\sqrt{1 + r(X, X')} \cdot \sqrt{1 + r(Y, Y')}}}$
Individual-level correlation	$r_i(X, X', Y, Y') = \frac{r(X, Y) - r(X, Y')}{\sqrt{1 - r(X, X')} \cdot \sqrt{1 - r(Y, Y')}}}$
Dyad-level correlation	$r_d(X, X', Y, Y') = \frac{r(X, Y')}{\sqrt{r(X, X')} \cdot \sqrt{r(Y, Y')}}}$

Source: (Dobos &amp; Gelei, 2019).

Table 4. Equations of two main kinds of regression models in DDA

ICC (intraclass correlation coefficient)	APIM (actor-partner interdependence model)
$Y = \beta_0 + \beta_1 \cdot X + \beta_2 \cdot X' + \varepsilon$	$Y = \beta_0 + \beta_1 \cdot X + \beta_2 \cdot X' + \beta_3 \cdot (X \cdot X') + \varepsilon$

Source: (Dobos &amp; Gelei, 2019).

## References

- Dobos, I., & Gelei, A. (2019). Empirically Supported Methodological Critique of Double Entry in Dyadic Data Analysis. *Statistika*, 2, 198-2017.
- Griffin, D., & Gonzalez, R. (1995). Correlational analysis of dyad-level data in the exchangeable case. *Psychological Bulletin*, 118, 430-439.
- Gonzalez, R., & Griffin, D. (1999). The correlational analysis of dyad-level data in the distinguishable case. *Personal Relationships*, 6, 449-469.
- Kenny, D. A. (1995). The effect of nonindependence on significance testing in dyadic research. *Personal Relationships*, 2, 67-75.
- Kenny, D. A. (1996). Models of nonindependence in dyadic research. *Journal of Social and Personal Relationships*, 13, 279-294.
- Kenny, D. A., Kashy, D. A., & Cook, W. (2006). *Dyadic data analysis*. New York: Guilford.

## Source of data and other useful links

<https://bdl.stat.gov.pl/BDL/start> – Local Data Bank, Polish official statistics about local governments units: population and availability of public services – number of classes in high schools (grammar or secondary) and hospitals

<https://randilgarcia.github.io/week-dyad-workshop/>

<https://github.com/RandiLGarcia/dyadr/>

<http://davidakenny.net/DyadR/DyadRweb.htm>

<http://davidakenny.net/DyadR/DyadR.htm>

<https://bbs.utdallas.edu/pairlab/materials/>

<https://osf.io/prtyb/wiki/home/>

<https://onlinelibrary.wiley.com/doi/abs/10.1111/pere.12230>

<https://randilgarcia.github.io/week-dyad->

[workshop/APIM\\_SEM.html#apim\\_indistinguishable](https://randilgarcia.github.io/week-dyad-workshop/APIM_SEM.html#apim_indistinguishable)

[https://apimsem.ugent.be/shiny/apim\\_sem/](https://apimsem.ugent.be/shiny/apim_sem/)

**Keywords:** Dyadic Data Analysis, correlation and regression models, local government official statistics data.