Black Sea Journal of Statistics

Doi: 10.5281/zenodo.18234659

Open Access Journal
e-ISSN: 3108-4265

Research Article
Volume 1 - Issue 2: 27-30 / December 2025

ESTIMATING THE NONPARAMETRIC CONFIDENCE INTERVAL FOR CORRELATION COEFFICIENT ON ANIMAL DATA

Nursen KURDAL[1], Hasan ÖNDER[2]*

[1]*Ondokuz Mayıs University, Agricultural Faculty, Department of Animal Science, 55139, Samsun, Türkiye*

Abstract: Correlation coefficient is widely used in the all areas of science to establish the degree and direction of two variables. Confidence interval is a special form of estimating a certain parameter. With use of this method, a whole interval of acceptable values for the parameter is given instead of a single value, together with a likelihood that the real (unknown) value of the parameter will be in the interval. The confidence interval is based on the observations from a sample, and hence differs from sample to sample. In this study, nonparametric confidence interval estimation for Pearson correlation coefficient were shown using an animal data set.

Keywords: Correlation coefficient, Nonparametric confidence interval, Animal

***Corresponding author:** Ondokuz Mayıs University, Agricultural Faculty, Department of Animal Science, 55139, Samsun, Türkiye
E mail: hasanonder@gmail.com (H. ÖNDER)
Nursen KURDAL https://orcid.org/0000-0001-7198-2329
Hasan ÖNDER https://orcid.org/0000-0002-8404-8700

Received: June 26, 2025
Accepted: November 22, 2025
Published: December 15, 2025

Cite as: Kurdal, N., & Önder, H. (2025). Estimating the nonparametric confidence interval for correlation coefficient on animal data. *Black Sea Journal of Statistics, 1*(2), 27–30.

1. Introduction

Correlation is a statistical method that reveals the direction and degree of the relationship between variables. The measure of the relationship between the two variables is called the correlation coefficient (Arkin and Colton, 1939). The correlation coefficient is denoted by the small "*r*" and takes the value between -1 and +1 ($-1 \leq r \leq +1$). If *r* value takes values close to -1, it is determined that there is a negative relationship between variables, and if it takes values close to + 1, there is a positive relationship between variables. If the *r* value is close to 0, it means that there is no relationship between the two variables (Figure 1) (Kurdal and Önder, 2020).

In Figure 1; (a) the decrease in the other depending on the increase of one of the variables is a linear relationship, (b) there is no relationship between the two variables and (c) the increase in one of the variables due to the increase in the other is the linear relationship.

General comments regarding the strength of the correlation coefficient are given below (Köse, 2025):

- 0.00-0.25 Very poor relationship,
- 0.26-0.49 Weak relationship,
- 0.50-0.69 Moderate relationship,
- 0.70-0.89 High relationship,
- 0.90-1.00 Very high relationship.

The Pearson correlation coefficient can be calculated using the estimator (Equation 1):

$$r = \frac{n\sum xy - (\sum x)(\sum y)}{\sqrt{n(\sum x^2) - (\sum x)^2}\sqrt{n(\sum y^2) - (\sum y)^2}} \quad (1)$$

Correlation coefficients vary depending on the characteristics of the variables under investigation.

Correlation coefficients used to determine the relationship between classifiable qualitative variables:

- Phi coefficient,
- Cramer V coefficient,
- Ordinary coefficient,
- Lambda coefficient.

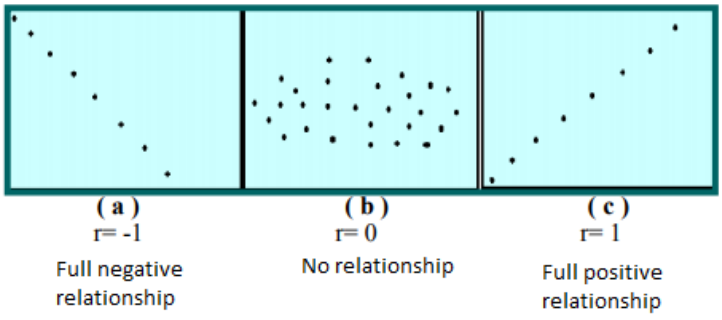

Figure 1. Scatter plots of extreme correlation coefficients.

Correlation coefficients used to determine the relationship between sortable qualitative variables:

- Spearman correlation coefficient,
- Gamma coefficient,
- Kendall's tau-b coefficient,
- Kendal's tau-c coefficient,
- Somer's d coefficient.

Correlation coefficients used in determining the relationship between discrete / continuous qualitative variables:

- Pearson correlation coefficient (if both variables show normal distribution),
- Spearman correlation coefficient (if at least one of the variables is not normally distributed).

Correlation coefficients used to determine the relationship between a classifiable qualitative variable and a discrete / continuous quantitative variable:

- Double series correlation coefficients,
- Point double series correlation coefficients.

Correlation coefficient used to determine the relationship between a sortable qualitative variable and a discrete / continuous quantitative variable:

- Multiple series correlation coefficient (Köse, 2025).

For the animal science as in all the other branches of the science the correlation coefficient (CC) is a basic relation statistics. In many studies confidence interval of the correlation coefficient is ignored. But interval estimation is very essential for statistical prediction.

Recent studies have shown that imprecise estimates of correlation coefficients can result in a series of problems, including an increase in multicollinearity in multivariable regression analyses, as well as overestimating direct effects and increasing noise in path analysis. In this sense, it is essential for experimental planning to ensure sufficient n to estimate correlation coefficients with an acceptable level of precision (Olivoto et al., 2018).

In this study, nonparametric and parametric 95% confidence interval (CI) for correlation coefficient were compared with using goat kid growth data.

2. Materials and Methods

This study was carried out at the private dairy goat farm in Bafra province of Samsun, Turkey (40°31'N, 36°53'E and 650 m above the sea level). Data was collected from 82 Saanen kids from birth (W0) to six month of age (W6).

The well-known parametric confidence interval of Pearson correlation can be calculated as given in Equation 2 (Mudelsee, 2003):

$$CI = t_{n-2,\,\alpha/2}.S_r$$

$$S_r = \sqrt{\frac{1-r^2}{n-2}} \qquad (2)$$

The nonparametric confidence interval of Pearson correlation can be calculated as given in Equation 3 (Olivoto et al., 2018):

$$CI = 0.45304^{r} * 2.25152 * n^{-0.50089} \qquad (3)$$

3. Results and Discussion

From birth (W0) to six month of age (W6) of Saanen kids monthly live weight, the minimum correlation coefficient was obtained as 0.7239 between W0 and W3, and the maximum was obtained as 0.9902 between W4 and W5 (Table 1).

The correlation coefficients showed high relations as expected for live weights. To demonstrate the confidence intervals correlation coefficients was sorted ascending. Correlation coefficients and its nonparametric 95%confidence intervals were given in Figure 2. Correlation coefficients and its parametric 95% confidence intervals were given in Figure 3.

The range of nonparametric confidence intervals was 0.026542 when the range of parametric confidence intervals was 0.102553. It means that nonparametric confidence intervals was so close to each other when parametric confidence intervals was not. With the increasing correlation coefficient parametric confidence intervals decreased but nonparametric confidence intervals. Its underlying reason is the estimator of parametric CI when the CC increases the CI decreases because of the standard error of correlation coefficient as seen in Figure 3.

Table1. Pearson correlation coefficients for monthly live weights of Saanen kids

	W1	W2	W3	W4	W5	W6
W0	0.8527	0.7908	0.7239	0.7558	0.7641	0.7588
W1		0.9705	0.9402	0.9112	0.8867	0.8600
W2			0.9827	0.9599	0.9317	0.8972
W3				0.9671	0.9356	0.8948
W4					0.9902	0.9686
W5						0.9842

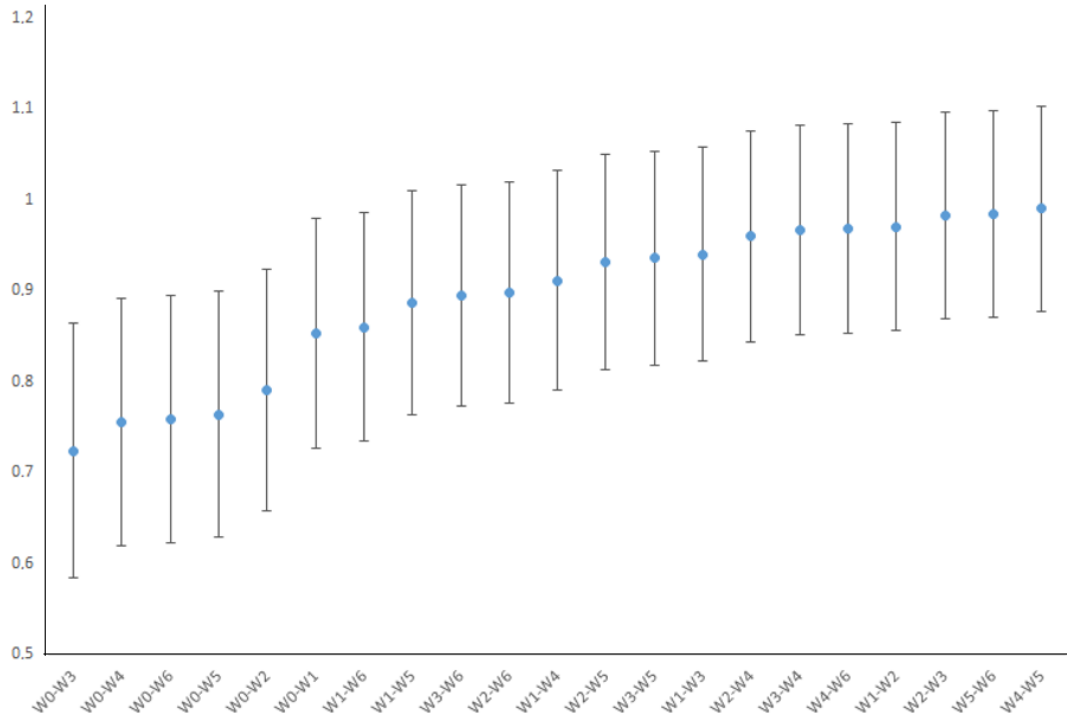

Figure 2. Correlation coefficients and its nonparametric 95%confidence intervals.

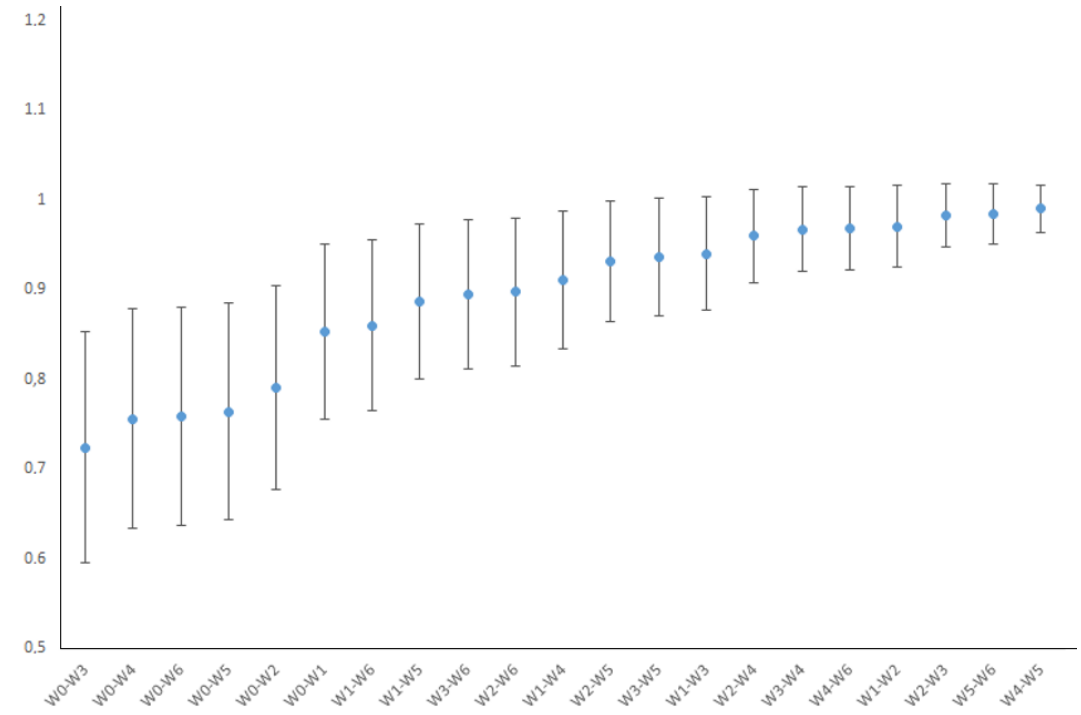

Figure 3. Correlation coefficients and its parametric 95%confidence intervals.

4. Conclusion

The nonparametric confidence interval for Pearson correlation coefficient is more reliable than parametric confidence interval. Use of nonparametric confidence interval can be preferred instead of parametric confidence interval for animal studies.

Author Contributions

The percentages of all authors' contributions are presented below. All authors reviewed and approved the final version of the manuscript.

	N.K.	H.Ö.
C	70	30
D	70	30
S	70	30
DCP	70	30
DAI	70	30
L	70	30
W	70	30
CR	70	30
SR	70	30
PM	70	30
FA	70	30

C=Concept, D= design, S= supervision, DCP= data collection and/or processing, DAI= data analysis and/or interpretation, L= literature search, W= writing, CR= critical review, SR= submission and revision, PM= project management, FA= funding acquisition.

Conflict of Interest

The authors declare that there is no conflict of interest.

Ethical Consideration

Ethics committee approval was not required for this study because of there was no study on animals or humans.

Acknowledgements

This study has been published in the proceeding book of the IV. International Congress on Domestic Animal Breeding, Genetics and Husbandry-2020 (ICABGEH-20), Online, August 12-14, 2020.

References

Arkin, H., & Colton, R. R. (1939). *Statistical methods*. Barnes & Noble Inc.

Köse, K. (2025). *Correlation and regression analysis* (Sunum slaytları). Toraks Derneği. Erişim adresi: https://file.toraks.org.tr/TORAKSFD23NJKL4NJ4H3BG3JH/mse-ppt-pdf/Kenan_KOSE3.pdf (Alıntı tarihi: 5 Nisan 2020)

Kurdal, N., & Önder, H. (2020). Estimating the nonparametric confidence interval for correlation coefficient on animal data. In *International Congress on Animal Breeding, Genetics and and Husbandry (ICABGEH-20)* (pp. 89–92).

Mudelsee, M. (2003). Estimating Pearson's correlation coefficient with bootstrap confidence interval from serially dependent time series. *Mathematical Geology, 35*(6), 651–665.

Olivoto, T., Lúcio, A. D. C., Souza, V. Q., Nardino, M., Diel, M. I., Sari, B. G., Krysczun, D. K., Meira, D., & Meier, C. (2018). Confidence interval width for Pearson's correlation coefficient: a Gaussian-independent estimator based on sample size and strength of association. *Agronomy Journal, 110*(1), 1–8. https://doi.org/10.2134/agronj2017.09.0566