AJSIHD
GLOBAL RESEARCH NETWORK

AMERICAN Journal of Science on Integration and Human Development
Volume 3, Issue 11, 2025 ISSN (E): 2993-2750

Analysis of Correlations between Variables Using Control Charts Developed Based on Kendall's Criterion

Akhmedov S. A.
Ph.D., Andijan State University named after Z.M. Bobur

Dadamirzaeva O.
Ph.D student, Andijan State University named after Z.M. Bobur

Abstract: The article expands the potential applications of Kendall's criterion by using the control chart method to determine the correlation between two variables. This method of studying the correlation is particularly useful in longitudinal studies frequently encountered in pedagogical, psychological, medical, environmental research, statistical monitoring of natural phenomena, and forecasting.

The processing of experimental data and statistical conclusions in the control chart method are depicted in the diagram of this chart, which serves as a common language for various specialists. The hypothesis of independence or presence of correlation in diagrams is tested over time. The diagram not only assesses trends but also predicts changes in the process, which can be effective for data analysis. The diagram allows comparison of results from experiments on the same problem. Additionally, during the experiment, it is possible to adjust or refine the course of the experiment.

The control charts developed in this work take into account the sample size. For small samples, the binomial criterion is used, while for large samples, Kendall's criterion is applied. The characteristics of the new control chart are derived using Kendall's theorem. As an example, one problematic task from seismology was investigated, and the iterative process of drawing conclusions for this task is demonstrated using a control chart diagram.

Keywords: Rank correlation criteria, Kendall's criterion, binomial criterion, statistical hypothesis, control chart, confidence interval, longitudinal study.

Introduction

Rank-based correlation criteria are widely used in pedagogy, psychology, medicine, ecology, sociology, and other scientific disciplines where it is necessary to assess the relationship between variables [11,12,15,16,18,21,22,23,24]. In this work, we focus on problems addressed using Kendall's criterion [8,9,10,13] and these problems are solved by using a method based on control charts developed on the foundation of Kendall's criterion unlike the traditional approach. This method of studying correlations is especially beneficial in longitudinal (long-term) research such as quality of life, stress levels, and satisfaction with treatment. Therefore, these results may expand researchers' methodological toolkit and demonstrate the criterion's potential applications in this field.

We begin by presenting the essential background from the traditional method of determining rank correlation, as our results are based on these concepts and assertions. Assume that both X and Y variables are measured on one of three possible scales: ordinal, interval, or ratio. What is crucial in this context is that an order relationship is established between the paired values of X and Y.

Let n measurements be conducted to determine the correlation between variables X and Y. As a result, we obtain two independent samples: $(X_1, X_2, \dots, X_n)$ and $(Y_1, Y_2, \dots, Y_n)$. Next, we rank each sample. Denote R_i as the rank of X_i among the observations $(X_1, X_2, \dots, X_n)$, that is, the position occupied by X_i in the ordered sequence $X_{(1)}, X_{(2)}, \dots, X_{(n)}$. Similarly, let S_i be the rank of Y_i among the observations $(Y_1, Y_2, \dots, Y_n)$. Thus, the original paired data (X_i, Y_i), $i = 1,2, \dots, n$ generates a corresponding set of rank pairs (R_i, S_i). It is assumed that each possible permutation of ranks is equally likely and there are no multiple rank values.

Under the conditions described above, Kendall's criterion is used to test the following hypotheses at a given level of significance α. H_0: X and Y are independent; there is no association between them. Alternative hypothesis H_1: Depending on the context of the study, the alternative hypothesis can take one of the following forms:" X and Y are dependent, "There is a positive association between X and Y," or "There is a negative association between X and Y. "In this study, and throughout the remainder of the paper, we consider the Hypothesis: H_1: X and Y are dependent.

When $n > 10$, one of the following equivalent test statistics is used to test the hypothesis:

$\tau = \frac{P-Q}{N}, \tau = \frac{2P}{N} - 1$ or $\tau = 1 - \frac{2Q}{N}$

Where:

- ✓ P is the number of concordant pairs (R_i, S_i), for $i = 1,2, \dots, n$;
- ✓ Q is the number of discordant pairs (R_i, S_i), for $i = 1,2, \dots, n$;
- ✓ n is the sample size;
- ✓ $N = \frac{n(n-1)}{2}$ is the total number of pairwise comparisons.

For small samples $n \leq 10$, the test statistic is defined as: $\tau = P - Q$

In this case, the critical region is determined using the binomial distribution.

The hypotheses for the test can then be formally stated as:

H_0: $\tau = 0$; (no association between X and Y),

H_1: $\tau \neq 0$. (there is an association between X and Y).

In the traditional approach to hypothesis testing, the following statement from Kendall's theorem is used: As $n \to \infty$ (specifically, for $n > 10$) the test statistic τ approximately follows the standard normal distribution, i.e., $\tau \sim N(0; 1)$.

Accordingly, the critical region for the two-tailed test is given by [17]:

$|\tau| > \frac{2t_{\frac{\alpha}{2}}}{\sqrt{n}}$,

where $\Phi\left(-t_{\frac{\alpha}{2}}\right) = \frac{\alpha}{2}$ denotes the cumulative distribution function of the standard normal distribution.

In practice, after conducting the experiment, the observed value of τ is calculated based on the collected data. If this value falls within the critical region, the alternative hypothesis H_1 is

accepted; otherwise, the null hypothesis H_0 is retained. Typically, significance levels of 0.05 and 0.01 are used for decision-making. Hypothesis testing is carried out both at the beginning and at the conclusion of the experiment. However, traditional rank-based correlation methods do not provide insights into the temporal dynamics of the relationship between variables over time.

In the control chart method, the hypothesis of independence or the presence of correlation is tested over time. The control chart diagram not only allows for the evaluation of trends, but also enables the prediction of future changes in the process, making it a potentially powerful tool for data analysis and monitoring.

Methods

The control charts developed in this study are based on Kendall's rank correlation criterion and the binomial criterion, and are used to monitor the correlation between variables. The control limits of the chart determine whether the observed correlation value is statistically significant or not. These charts are applied to monitor the stability of processes over time, which is particularly useful for tracking changes in a controlled parameter G.

Two methods are primarily used in constructing control charts. In the Shewhart method, where knowledge of the distribution of the controlled parameter (G) is not required, the upper (UCL) and lower (LCL) limits of the control chart are determined using the six-sigma method [19]. In the confidence interval method, which requires knowledge of the exact (or limiting) distribution of G, the upper and lower limits of the control chart are taken as the boundaries of the confidence interval. In this case, we consider the sample size and significance level as constants and find the quantile of the G distribution from a table. In practice, both methods are used depending on the available data. In this paper, we will construct control charts using the confidence interval method.

The essence of the confidence interval method in constructing control charts is to determine the range of values within which the controlled parameter (the observed characteristic of the control chart, for example, the number of concordant or discordant pairs) of the process will fall with a given confidence probability. If the process is stable, then the G values will lie within the confidence interval with high probability.

If we know the distribution of the controlled parameter (exact or approximate), then we can determine its confidence interval - that is, the range to which its values will fall with a high probability $(0.99$ or $0.95)$ during a stable process. This interval then becomes the control limits of the chart. Using the statistical distribution, we find the quantiles and then determine the characteristics of the control chart: controlled parameter (G); upper (UCL) and lower (LCL) control boundaries. Next, a diagram is constructed, where the horizontal axis represents time or observation sequence numbers and the vertical axis represents the controlled parameter G. If the values of G are within the control limits, the correlation between the variables is insufficient (true $H_0{:}\tau = 0$), if the values exceed the limits, this is a signal about the sufficiency of the correlation (true $H_1{:}\tau = 0$

Control charts based on statistical criteria can be more sensitive than standard Shewhart charts, especially when analyzing complex processes. The choice of chart depends on the type of parameter being monitored, the nature of the data (normal or non-normal), and the need to track small changes or global shifts. These charts can be adapted for specific tasks, including pedagogical and psychological research, statistical control of natural phenomena, and forecasting [14,19,20].

Results.

The main results of the work are aimed at determining the characteristics of G control charts, including their Lower Control Limit (LCL) and Upper Control Limit (UCL). For this, we use the exact or limiting distribution of the controlled parameter, which was discussed in the introduction.

The theorems presented below define the main characteristics of the new control charts. In experiments, after constructing the LCL and $UCL, G = \tau$ values are observed in time units, and during anomalies, intervention is made in the process to stabilize it.

Theorem 1. Let H_0 be true at the significance level α ($\alpha = 0.05$ or $\alpha = 0.01$). Then, for $n > 10$, the control chart characteristics have the following form:

$$G = P, LCL = n(n-1)\left(1 - \frac{2t_{\frac{\alpha}{2}}}{\sqrt{n}}\right);\ UCL = n(n-1)(1 + \frac{2t_{\frac{\alpha}{2}}}{\sqrt{n}}).$$

Proof. According to Kendall's theorem, the confidence interval for τ with a confidence probability of $1 - \alpha$ is determined from the following relationship:

$$P\left(-\frac{2t_{\frac{\alpha}{2}}}{\sqrt{n}} \le \tau \le \frac{2t_{\frac{\alpha}{2}}}{\sqrt{n}}\right) = 1 - \alpha, \text{ где } \Phi\left(-t_{\frac{\alpha}{2}}\right) = \frac{\alpha}{2}.$$

We take $\tau = \frac{2P}{N} - 1$, where $N = \frac{n(n-1)}{2}$, transforming the expression in parentheses identically with respect to P, we have:

$$P\left\{n(n-1)\left(1 - \frac{2t_{\frac{\alpha}{2}}}{\sqrt{n}}\right) \le P \le n(n-1)\left(1 + \frac{2t_{\frac{\alpha}{2}}}{\sqrt{n}}\right)\right\} = 1 - \alpha.$$

Then, for constant n and α, we have:

$$G = P, LCL = n(n-1)\left(1 - \frac{2t_{\frac{\alpha}{2}}}{\sqrt{n}}\right);\ UCL = n(n-1)(1 + \frac{2t_{\frac{\alpha}{2}}}{\sqrt{n}}).$$

The theorem is proved.

Conclusion:

For $\alpha = 0{,}05$, $LCL = n(n-1)\left(1 - \frac{3{,}29}{\sqrt{n}}\right);\ UCL = n(n-1)(1 + \frac{3{,}29}{\sqrt{n}})$

For $\alpha = 0{,}01$, $LCL = n(n-1)\left(1 - \frac{5{,}16}{\sqrt{n}}\right);\ UCL = n(n-1)(1 + \frac{5{,}16}{\sqrt{n}})$

In this case, $G = P$ and from the table of critical points for $N\ (0; 1)$: we find $t_{\frac{\alpha}{2}} = 1{,}645$ for $\alpha = 0{,}05$ and $t_{\frac{\alpha}{2}} = 2{,}58$ for $\alpha = 0{,}01$.

Theorem 2. Let H_0. be true at the significance level α ($\alpha = 0.05\ or\ \alpha = 0.01$). Then, at $n \le 10$, the characteristics of the control chart have the form:

$$G = \tau\ = P - Q, LCL = \tau_{\frac{\alpha}{2}};\ UCL = \tau_{1-\frac{\alpha}{2}},$$

where

$$\tau_{\frac{\alpha}{2}} = Bin^{-1}\left(N, \frac{\alpha}{2}, \frac{1}{2}\right), \tau_{1-\frac{\alpha}{2}} = Bin^{-1}(N, 1 - \frac{\alpha}{2}, \frac{1}{2}).$$

The proof of the theorem follows directly from the relation

$$P\left(\tau_{\frac{\alpha}{2}} \le \tau \le \tau_{1-\frac{\alpha}{2}}\right) = 1 - \alpha.$$

For example, we find the boundaries of the control chart from the table of critical points of the binomial distribution at $n = 5$ and $\alpha = 0.05$: $\tau_{\frac{\alpha}{2}} = 6$; $\tau_{1-\frac{\alpha}{2}} = 6$ and at n=5 and α=0.01: $\tau_{\frac{\alpha}{2}} = -8$; $\tau_{1-\frac{\alpha}{2}} = 8$.

To illustrate the results of the theorems, let's consider an example.

It is known that radon emissions from the Earth's crust can precede seismic activity. There is a hypothesis about a connection between radon emissions and animal behavior before earthquakes, but direct scientific evidence of radon's effects on animals under these conditions is still insufficient.

The mathematical model for solving this problem can be constructed using Theorem 1. Since the task relates to longitudinal (long-term) research, we will conduct an experiment in a specific area to determine, for example, the correlation between dog barking and radon concentration in an enclosed area. In this case, we will define a unit time of measurement by studying the dogs' behavior in relation to radon concentration in the air, ensuring that the experimental results reflect real data.

If we mark the beginning of the foreshock with a bold point on the horizontal time axis, then before a strong earthquake, the diagram of the expected control chart may have the following appearance:

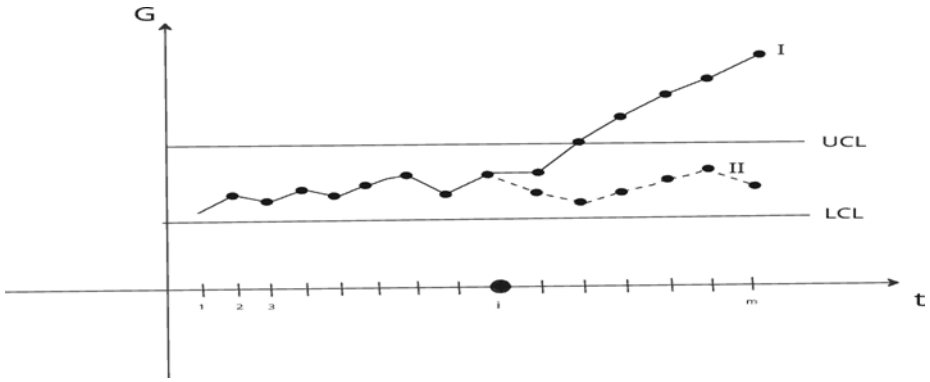

Рис. 1. Диаграмма

$$\text{КК} - \tau: I)\ P > UCL, \tau > 0{,}70;\ II)\ LCL < P < UCL, \tau < 0{,}49.$$

Discussion

The main objective of this article was to study correlations using the control chart method in longitudinal studies. This is due to the fact that research in this area, at the initial stage, relies mainly on iterative conclusions. The control chart method is an iterative mathematical model of such phenomena. On the other hand, we note the known advantages of rank criteria: Independence from data distribution; Resistance to outliers; Ability to work with ordinal data; Simplicity of calculations. Kendall's correlation coefficient and its advantages: Accuracy of assessment in small samples; More stringent consideration of the order of values; Interpretation in terms of probability. Kendall's criterion is applied in various fields: In pedagogy and psychology: when analyzing the impact of teaching methods on academic performance (for example, studying the relationship between active teaching methods and the level of material comprehension); when researching psychological factors (for example, evaluating the relationship between motivation level and students' academic performance). In medicine: connection between symptoms and diagnoses; evaluation of treatment effectiveness. In ecology: research on pollution factors; analysis of climate change. In seismology: the relationship between animal behavior and earthquake precursors; the relationship between earthquake magnitude and precursor parameters. [1-7]. The above-mentioned tasks and studies are longitudinal (long-term) in nature, and studying correlation using control charts allows for the identification of deviations and assessment of process stability more effectively than the traditional method. Additionally, this method enables comparison of results from identical experiments conducted in different locations.

In tasks where the Kendall criterion does not provide a complete answer, other ranking criteria can be used to develop additional control charts. As a result, various control charts can be

compiled for different purposes. For non-specialists in statistics, software tools can be prepared to obtain statistical conclusions in such situations.

References

1. Chen, Z. S., Kulkarni, P., Galatzer-Levy, I. R., Bigio, B., Nasca, C., & Zhang, Y. (2022). Modern Views of Machine Learning for Precision Psychiatry. archival preprint archival:2204.01607. https://arxiv.org/abs/2204.01607
2. Zhang, T., Yang, K., Ji, S., & Ananiadou, S. (2023). Emotion Fusion for Mental Illness Detection from Social Media: A Survey. arXiv preprint arXiv:2304.09493. https://arxiv.org/abs/2304.09493.
3. Parkes, L., Satterthwaite, T. D., & Bassett, D. S. (2020). Towards Precise Rest-State fMRI Biomarkers in Psychiatry: Synthesizing Developments in Transdiagnostic Research, Dimensional Models of Psychopathology, and Normative Neurodevelopment. arXiv preprint arXiv:2006.04728. https://arxiv.org/abs/2006.04728
4. Correia, R. B., Wood, I. B., Bollen, J., & Rocha, L. M. (2020). Mining Social Media Data for Biomedical Signals and Health-Related Behavior. arXiv preprint arXiv:2001.10285. https://arxiv.org/abs/2001.10285.
5. Turayev, J., & Ganihanov, A. (2022). Analysis of the prevalence of mental disorders in dynamics in 2018-2021 in Uzbekistan. Modern medicine through the eyes of young scientists, (1). https://inlibrary.uz/index.php/medicine-eyes-young-scientists/article/view/22562
6. Lunev, V. E. (2021). Psychological medicine: a summary review of contemporary research and perspectives. Hippocrates. https://www.hip-med.com/psihologicheskaya-mediczina-referativnyj-obzor-sovremennyh-issledovanij-i-perspektivy/
7. Katz, M. J., Derby, C. A., Wang, C., Zimmerman, M. E., & Lipton, R. B. (2021). Longitudinal assessment of perceived stress and memory complaints in the Einstein aging study. Journal of Alzheimer's Disease, 79 (4), 1571-1580. https://doi.org/10.3233/JAD-200989
8. METHODS OF ESTIMATING CORRELATION COEFFICIENTS IN THE PRESENCE OF INFLUENTIAL OUTLINE (S) Etaga Harrison O., Okoro Ifeanyichukwu, Aforka Kenechukwu F. and Ngonadi Lilian O. African Journal of Mathematics and Statistics Studies ISSN: 2689-5323 Volume 4, Issue 3, 2021 (pp. 157-185) DOI: 10.52589/AJMSSLLNZXUOZ.
9. A Comparison of the Pearson, Spearman Rank and Kendall Tau Correlation Coefficients Using Quantitative Variables Asian Journal of Probability and Statistics 20 (3): 36-48, 2022; Article No.AJPAS.92398 ISSN: 2582-0230 DOI: 10.9734/AJPAS/2022/v20i3425
10. Katerina Zlatkova "Anxiety and Aggressiveness in Children with Emotional Behavioral Disorders - Presented in an Individual Case" Research insights Research penetration https://doi.org/ 10.53656/ped2021-1.06
11. Shishlyannikova L. M. Application of Correlation Analysis in Psychology © Moscow City Psychological and Pedagogical University © PsyJournals.ru, 2008
12. Gibbons, J. d. (1997), Nonparametric Methods for Quantitative Analysis, 3rd edn., American Sciences Press, Syracuse, NY.
13. Kendall M., Stuart A. Statistical Conclusions and Connections. "Science" M.1973.
14. Solonin S. I. The Method of Control Cards. Yekaterinburg. UrSU. 2014.
15. A.N.Krichevets, E.V. Shikin, A.G. Dyachkov Mathematics for Psychologists Moscow Psychological and Social Institute Moscow 2003

16. Shelexova. Mathematical methods in pedagogy and psychology. Maykop - 2010.
17. G.I., Ivchenko, Yu.I. Medvedev. Introduction to Mathematical Statistics Textbook. M.: Iz-vo LKI, 2010-358p.
18. V.K. Romanko. Statistical analysis of data in psychology. Moscow 2015 - 169 p.
19. H.J. Mittag, H. Rinne. Statistical methods of quality assurance. M.: Mashinostroyeniye. 1995.
20. D. Wheeler, D. Chambers. Statistical Process Management. M.: Alpina Business Books, 2009.
21. Glass J. Statistical Methods in Pedagogy and Psychology / J. Gass, J. Stanley. Translated from English by L.I.Khairusova. - M.: Progress, 1976. - 494 p.
22. Yermolayev, O.Yu. Mathematical Statistics for Psychologists / O.Yu.Yermolayev. - M.: MPSI, 2006. - 336 p.
23. Heritage. A.D. Mathematical methods of psychological research. Data Analysis and Interpretation: Textbook. - SPb: Speech, 2006.- 392 p.
24. Sidorenko. E.V. Methods of mathematical processing in psychology / E.V.Sidorenko. - SPb: Speech, 2007. - 349 p.