

基于模糊区间时间序列数据之相关性研究



杨志清*

东莞理工学院计算机科学与技术学院, 广东东莞 523109

摘要: 互联网的兴起, 不仅仅牵动着生活社会的便利改变, 同时也带动了数据科学的变革, 在数据型态日益多元与数据整合逐渐复杂等环境的冲击中, 相依性模糊区间数据俨然成为新时代的探讨方向之一, 同时由此延伸的方法探讨也逐渐成为探究的课题。本研究基于相依性模糊区间数据集的衍生, 创新提出相依性模糊区间相关系数的演算法, 藉以衡量相依性随机变量的相关性强弱, 并且经由蒙地卡罗的模拟分析与实际发生的数值分析等方式, 验证创新方法所获得的效果。在研究分析结果中, 模糊区间自相关系数演算法所得到的结果不仅可以体现变量之间的关系之外, 更能体现变量之间的相关性波动情况, 据以了解在变量相关性的条件下, 两者间数据的分散情况, 如此可以更精准的获得变量之间的相关信息, 提高数据隐藏信息的挖掘效能, 进而提供产业一个全新的相关性衡量方式, 据以作为决策者参考之用。

关键词: 相依数据; 模糊区间自相关系数; 模糊区间互相关系数; 时间序列

DOI: [10.57237/j.wjms.2026.02.001](https://doi.org/10.57237/j.wjms.2026.02.001)

Evaluation of the Fuzzy Correlation Coefficient for Interval Time Series Data

Chihching Yang*

School of Computer Science and Technology, Dongguan University of Technology, Dongguan 523109, China

Abstract: The development of the Internet has advanced data science while increasing the diversity and complexity of data types and integration processes. In this context, dependent fuzzy interval data has emerged as an important research topic. With the advent of big data, data generation has shifted from univariate observations to multi-valued representations. Accordingly, fuzzy data analysis methods based on multi-valued logic have been increasingly investigated for practical statistical applications. Since the advent of the big data era, the nature of data generation has gradually evolved from single-valued data to multi-valued data. To address the analytical needs of practical statistical applications, fuzzy data analysis methods based on multi-valued logic have been increasingly explored and developed. In this study, two algorithms are proposed for estimating correlation coefficients in interval time series data: the fuzzy interval autocorrelation coefficient and the fuzzy interval cross-correlation coefficient. These methods are developed to evaluate fuzzy correlations between random variables represented by count intervals, where count values are expressed as fuzzy interval data. Simulation experiments and numerical case analyses are conducted to assess the performance and accuracy of the proposed methods. The results indicate that the correlation coefficients obtained from fuzzy interval time series data are consistent with those

*通信作者: 杨志清, 3478978266@qq.com

derived from traditional univariate correlation methods. The proposed algorithms are also shown to effectively verify variations between two variables in fuzzy interval time series data.

Keywords: Dependent Data; Fuzzy Interval Autocorrelation Coefficient; Fuzzy Interval Cross-Correlation Coefficient; Time Series

1 前言

两随机变量的相关性旨在说明变量与变量之间存在关系，当关系系数愈大时，代表著两变量相互依存的现象愈重，反之则愈轻。基础的相关性研究中，在连续型随机变量的环境下，多数采用皮尔逊积差相关系数(Pearson product-moment correlation coefficient)演算法为主，探讨变量之间的相关性，并且以此为基础进行方法的延伸或应用，据以拓展相关系数的研究方向与领域，包括通信、销售、金融、科技、服务等相关产业所生成之数据，例如基于皮尔逊相关系数对于网络舆情评估指标与实时监测结果的验证 [1]，又或是利用皮尔逊相关系数分析高校科研和教学的相关性[2]，或是通过电商的交易数据分析不同时空下产品的销售相关性并对销售量进行预测，以及透过相关函数的建构，对于两种相关函数进行相关性的确立等[3, 4]；若在于间断型随机变量的环境下，一般会采用斯皮尔曼相关系数(Spearman's rank correlation coefficient)演算法，进行方法的延伸研究或应用，例如有学者利用斯皮尔曼相关系数衡量银行的货币供给量、利率、期限与收益率的关系[5]，又或者应用该相关系数算法分析脉冲噪声的稳健性与信号的关系[6]，或是透过个人的基本属性与位置偏好进行相关性分析等[7]；若在面对具有时间序列之随机变量的数据时，往往会采用时间序列方法所引用之自相关系数(Autocorrelation Coefficient)演算法，对其方法进行延伸或应用，例如有学者应用自相关系数分析台湾上市银行股票之报酬、市场指数、利率、汇率等因子之相关性[8]，又或者是藉由自相关系数分析地质坍塌跳动距离的影响[9]。由此可见，无论是相关性演算法理论的研究或随机变量之数据应用的实践均是学者所关注的课题研究方向之一。

随着物联网(Internet of Things; IoT)和互联网(Internet)的普及，逐渐冲击著国内外经济市场的演化，同时也逐渐改变交易市场的交易模式，在许许多多的瞬息万变的环境下，除了消费者的消费模式和消费习惯有所改变外，亦还有消费后所产生的庞大数据量与其生

成的速度，当数据在“质”方面有所改变之际，过去所引用的理论方法，可能无法满足当前数据的分析，甚至可能产生许多错误的信息解读，例如在百货公司的专柜营业中，其交易信息可以透过 POS (Point of Sales) 机进行纪录，其中信息纪录包括了交易编号、交易时间、金额、数量、商品名称、会员编号等数据集，若采用过去的数据分析方法可能会利用日平均值或周平均值进行探讨，而这样的方法无形之中，会失去许多的信息，好比日交易的波动情况，如那个时段的交易量或金额较高等。因此，在网络信息发达的环境中，完备了数据生成的完整性与多元性，故而导致数据采集型态俨然也从单元型态转换成多元型态，举例来说，零售店的交易金额会以日为单位来计算当日的交易额，如今在计算机系统的辅助，当日的交易细项可以更完整的被纪录，这样的现象足以说明实际数据型态也随着时代的演进悄然开始产生的变化。在相关性的研究领域，有许多学者致力于模糊数据的相关系数演算法研究[10]、混合型变量的相关系数演算法研究[11]、非参数量之相关性研究[12]等，有更多的学者则是从事自然、环境、经济、社会等领域之实际发生现况的相关性研究，例如在皮尔逊系数的基础上衡量多传感器之间的冲突程度，探讨在目标检测和故障诊断中的关系[13]，又或者如基于 Spearman 相关系数与有功分量相结合的故障选线方法，分析线路零序功率波形间的相关性来确定出具体发生故障的线路[14]，又如透过线性混合效应模型的相关系数研究，探讨葡萄糖数据的有效性[15]。

然而，前述的课题研究方向多数是以单变量的统计量演算法为主，鲜少以多变量的统计量演算法为研究课题，更遑论模糊区间时间序列数据为基础的相关性演算法的研究。基此，本研究创新设计一个模糊区间时间序列相关系数演算法，藉以除了满足时间序列的相关性探讨外，亦能更进一步了解数据自身的变动现况，而此演算法可以适时解决许多产业在互联网科技的辅助下，所生成的庞大数据集，例如金融产业的股票

价格、汇率、报酬率等；又如医药生技产业的血压、血糖、体脂等；或如自然环境领域的空气品质监测、水质监测等。本文为能阐述演算法的实用性与有效性，分别透过模拟分析与数值分析的方式，加以说明演算法的衡量效果，进而可以有效掌握数据发生的实际概况，以达到数据分析辅以决策的结果。

2 相关函数演算法设计

时间序列的研究范畴中，自相关函数(Auto-Correlation Function; ACF)除了是决定时间序列模型的判定统计量之一，亦是一项决定时间序列变量前后期关系的重要指标，其中总体自相关函数表示为

$$\rho(h) = \frac{\text{cov}(Y_t, Y_{t-h})}{\sqrt{\text{Var}(Y_t)}\sqrt{\text{Var}(Y_{t-h})}}, \quad h = 1, 2, \dots, \quad (1)$$

在给定一组随机样本， $Y_t = [y_1, \dots, y_n]$ ，则样本自相关系数为

$$\hat{\rho}(h) = r(h) = \frac{\sum_{i=1}^n (y_i - \bar{y})(y_{i-h} - \bar{y})}{\sqrt{\sum_{i=1}^n (y_i - \bar{y})^2} \sqrt{\sum_{i=1}^n (y_{i-h} - \bar{y})^2}} \quad (2)$$

若存在两个时间序列变量 $X_t = [x_1, \dots, x_n]$ 和 $Y_t = [y_1, \dots, y_n]$ 时，互相关系数(Cross-Correlation Function; CCF)是一项决定两个时间序列变量相互滞后关系的重要指标之一，其函数表示为

$$r'(h) = \frac{\sum_{t=1}^n (x_t - \bar{x})(y_{t-h} - \bar{y})}{\sqrt{\sum_{t=1}^n (x_t - \bar{x})^2} \sqrt{\sum_{t=1}^n (y_{t-h} - \bar{y})^2}} \quad (3)$$

本研究改变过去传统样本自相关系数应用于单一观测值的计算方式，在模糊区间数据的基础条件下，创新设计模糊区间自相关系数与模糊区间互相关系数等算法，藉以探讨区间随机变量自我前后期的关系及两组区间随机变量之前后的相互影响关系，其演算方法的定义如下：

定义 1. 模糊区间自相关系数(Fuzzy Interval Autocorrelation Coefficient; FIAC)

假设 $X_t = [x_{1t}, x_{2t}]$ ， $t=1, 2, 3, \dots, n$ ，是来自于时间数列总体的模糊区间数据集，并结合时间数列之相依性的特性，导入自相关函数的公式，转换原有单变量的自相关系数为模糊区间自相关系数，改写为

$$r_{jk} = \frac{\sum_{t=1}^n (x_{jt} - \bar{x}_{jt})(x_{jt+k} - \bar{x}_{jt})}{\sqrt{\sum_{t=1}^n (x_{jt} - \bar{x}_{jt})^2} \sqrt{\sum_{t=1}^n (x_{jt+k} - \bar{x}_{jt})^2}}, \quad j = 1, 2 \quad (4)$$

$$\text{其中, } \bar{x}_{jt} = \frac{\sum_{t=1}^n x_{jt}}{n}, \quad j = 1, 2.$$

基于落后期数 k 的数值大小，影响自相关系数的大小，故定义二元模糊区间自相关系数(Binary Fuzzy Interval Autocorrelation Coefficient)为

$$r_B^{FIAC}(k) = [r_{low}, r_{up}], \quad (5)$$

其中， $r_{low} = \min(r_{1k}, r_{2k})$ 和 $r_{up} = \max(r_{1k}, r_{2k})$ 。

若来自于时间数列总体之模糊区间数据集呈现多元时， $X_t = [x_{1t}, \dots, x_{mt}]$ ， $t=1, 2, 3, \dots, n$ ，则多元模糊区间自相关系数可改写为

$$r_{jk} = \frac{\sum_{t=1}^n (x_{jt} - \bar{x}_{jt})(x_{jt+k} - \bar{x}_{jt})}{\sqrt{\sum_{t=1}^n (x_{jt} - \bar{x}_{jt})^2} \sqrt{\sum_{t=1}^n (x_{jt+k} - \bar{x}_{jt})^2}}, \quad j = 1, \dots, m, \quad (6)$$

$$\text{其中, } \bar{x}_{jt} = \frac{\sum_{t=1}^n x_{jt}}{n}, \quad j = 1, \dots, m.$$

基于落后期数 k 的数值大小，影响自相关系数的大小，故定义多元模糊区间自相关系数(Multiple Fuzzy Interval Autocorrelation Coefficient)为

$$r_M^{FIAC}(k) = [r_{low}, r_{up}], \quad (7)$$

其中， $r_{low} = \min(r_{1k}, \dots, r_{mk})$ 和 $r_{up} = \max(r_{1k}, \dots, r_{mk})$ 。

定义 2. 模糊区间互相关系数(Fuzzy Interval Cross-Correlation Coefficient; FICC)

假设 $X_t = [x_{1t}, x_{2t}]$ 和 $Y_t = [y_{1t}, y_{2t}]$ ， $t = 1, 2, 3, \dots, n$ ，同是来自于时间数列总体的模糊区间数据集，然为能深入了解两变量间的落后期数间的关系，进行应用互相关函数的公式，延伸创新设计模糊区间互相关系数为

$$r_{jk} = \frac{\sum_{t=1}^n (x_{jt} - \bar{x}_{jt})(y_{jt+k} - \bar{y}_{jt})}{\sqrt{\sum_{t=1}^n (x_{jt} - \bar{x}_{jt})^2} \sqrt{\sum_{t=1}^n (y_{jt+k} - \bar{y}_{jt})^2}}, \quad j = 1, 2 \quad (8)$$

$$\text{其中, } \bar{x}_{jt} = \frac{\sum_{t=1}^n x_{jt}}{n}, \quad \bar{y}_{jt} = \frac{\sum_{t=1}^n y_{jt}}{n}, \quad j = 1, 2.$$

根据两随机变量， X 和 Y ，的落后期数 k 的相关系数，从中取决其互相关系数的最大和最小系数，故而定义二元模糊区间互相关系数(Binary Fuzzy Interval Cross-Correlation Coefficient)为

$$r_B^{FICC}(k) = [r_{low}, r_{up}], \quad (9)$$

其中， $r_{low} = \min(r_{1k}, r_{2k})$ 和 $r_{up} = \max(r_{1k}, r_{2k})$ 。

若来自于时间数列总体之模糊区间数据集呈现多元时， $X_t = [x_{1t}, \dots, x_{mt}]$ 和 $Y_t = [y_{1t}, \dots, y_{mt}]$ ， $t=1, 2, 3, \dots, n$ ，则多元模糊区间互相关系数(Multiple Fuzzy Interval Cross-Correlation Coefficient)可改写为

$$r_{jk} = \frac{\sum_{t=1}^n (x_{jt} - \bar{x}_{jt})(y_{jt+k} - \bar{y}_{jt})}{\sqrt{(\sum_{t=1}^n (x_{jt} - \bar{x}_{jt})^2) \sum_{t=1}^n (y_{jt+k} - \bar{y}_{jt})^2}}, j = 1, \dots, m, \quad (10)$$

其中, $\bar{x}_{jt} = \frac{\sum_{t=1}^n x_{jt}}{n}$, $\bar{y}_{jt} = \frac{\sum_{t=1}^n y_{jt}}{n}$, $j = 1, \dots, m$ 。换言之, 当 $k=0$ 时, 式子(10)等价于皮尔逊(Pearson)积差相关系数的公式。根据随机变量 X 和落后 k 期之随机变量 Y 两者间的相关系数, 从中取决其互相关系数的最大和最小系数, 进而定义多元模糊区间互相关系数(Multiple Fuzzy Interval Cross-Correlation Coefficient)为

$$r_M^{FICC}(k) = [r_{low}, r_{up}], \quad (11)$$

其中, $r_{low} = \min(r_{1k}, \dots, r_{mk})$ 和 $r_{up} = \max(r_{1k}, \dots, r_{mk})$ 。

3 模拟分析

模拟分析主在透过多次随机样本的重复实验, 验证创新方法的准确性与可行性。本节应用蒙地卡罗(Monte Carlo)方法, 验证模糊区间自相关系数与模糊区间互相关系数等创新方法的效益, 其模拟过程与结果如下:

3.1 自相关系数模拟分析

考虑模糊区间相依数据的环境下, 透过相依性随机变量的样本观测值的生成, 验证二元和多元模糊区间自相关系数与单变量自相关系数间的差异, 其模拟过程如下:

步骤 1. 根据相依数据的特性, 生成一组四个元素的模糊区间之 MA(2) 时间序列随机变量的样本数据集, 即 $X_t = [x_{1t}, x_{2t}, x_{3t}, x_{4t}]$ 。

步骤 2. 以 $X_t = [x_{1t}, x_{2t}, x_{3t}, x_{4t}]$ 为基础, 分别四个元素的模糊区间数据转换为二元模糊区间数据集, 即 $X'_t = [x'_{1t}, x'_{2t}]$, 以及计算其平均值为 $\bar{X}_t$ 。

步骤 3. 利用生成的模糊区间随机变量之样本数据, 依定义 1 的模糊区间自相关系数的方法, 分别计算二元和多元模糊区间自相关系数, 并且以平均值来计算单变量自相关系数。

从步骤 1 至步骤 3 的程序, 在 MA(2) 相依数据的生成条件下, 分别计算二元和多元之模糊区间自相关系数及单变量自相关系数, 并且以 ACF 的函数值与延迟(Lag)期数为基准, 绘制如图 1 与图 2 所示。

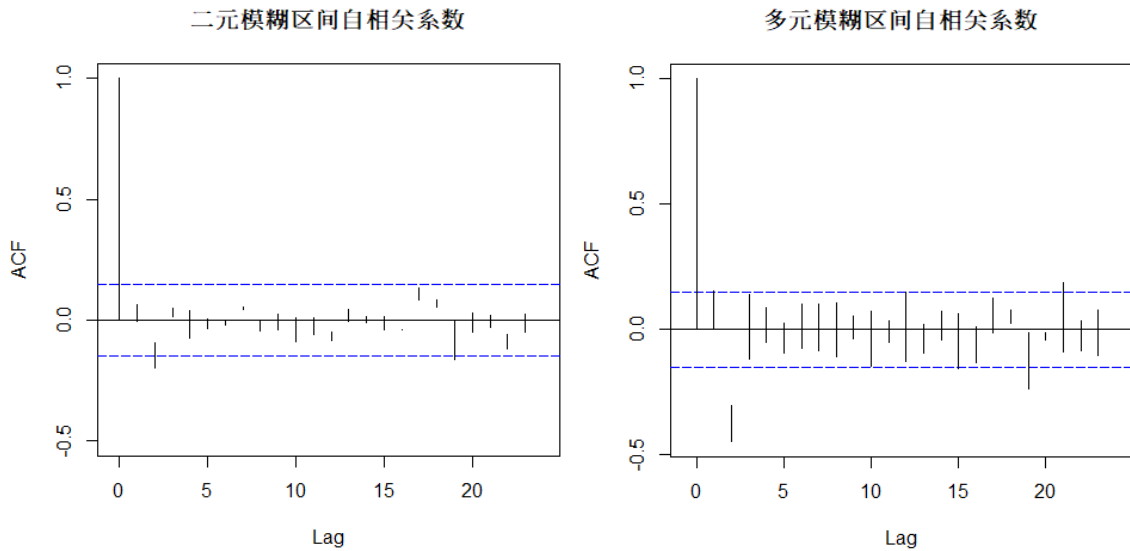


图 1 模糊区间自相关系数(ACF)

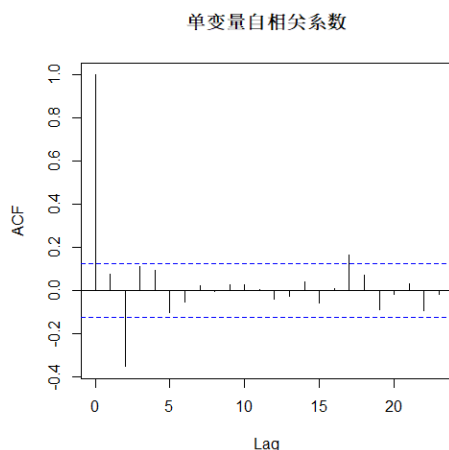


图2 单变量自相关系数(ACF)

图1的模糊区间自相关系数主要是以二元和多元区间方式体现相依数据的滞后关系,图2的单变量自相关系数是以单一统计量为主讨论变量的滞后关系。依相依数据之随机变量的假设为MA(2),其中MA模型可以透过ACF予以判断滞后期数。换言之,本研究在假设MA(2)的条件时,同时也表示自相关系数的滞后期数应落在第2期,故以下分别叙述模拟结果:

(1)从滞后期数的角度来看,在单变量的ACF中, $r_2 = -0.3533$,显示滞后在第2期时的负自我相关影响程度较大。而二元模糊区间的ACF为 $r_B^{FIAC}(2) = [-0.1985, -0.0942]$,多元模糊区间的ACF为 $r_M^{FIAC}(2) = [-0.4495, -0.3030]$,其亦是同样滞后在第2期数,由此可知,在滞后期数的衡量上,模糊区间自相关系数的演算衡量效果与传统单变量自相关系数的衡量具有同样的效果。然而,在自相关系数的衡量上,FIAC的自相关系数算法可以呈现相关系数的变动概况,较单变量自相关系数显示的单一数值,更能具体掌握数据关系强弱的波动信息。

(2)从整体的角度来看,二元模糊区间自相关系数从滞后第2期之后,其相关系数则出现迅速下降的趋势,与单变量自相关所呈现的结果一致,而在多元模糊区间自相关系数虽然亦是自滞后第2期后出现下降趋势,但其自相关系数的区间变动是呈现偏大的情况,表示该相依数据的变动有明显偏大的现象。

总结上述自相关系数的模拟分析结果,FIAC算法可以确实衡量相依性之随机变量的自我相互影响关系,与单变量自相关系数具有相同的效果。FIAC算法在具

有单变量自相关系数之衡量效果外,更提供了相依性随机变量之自我相互影响的变动程度,也就是说,透过自相关系数的运算可以进一步了解相依性随机变量的前后期关系及其相关性的变动趋势。举例来说,若相依性随机变量 X 的自相关系数在滞后第2期区间系数为 $[-0.4495, -0.3030]$,而单变量自相关系数为 -0.3533 ,若从单变量自相关系数来看,代表 X 当期与滞后第2期具有低度的负相互影响关系,但仅局限于单一数值的表达方式,并未能进一步的体现 X 当期与第2期的进一步关系,而模糊区间自相关系数则能更加贴切的表达 X 当期与滞后第2期的关系俨然快达到负中度相关的水平,对此可以提供决策者更多信息,进而做出预防决策。

3.2 互相关系数模拟分析

互相关系数主要在于探讨两个相依性随机变量间的相互影响关系。在模拟分析中,是考虑经由两个相依性随机变量的样本观测值的生成,验证二元和多元模糊区间互相关系数的相互滞后关系,其模拟的过程如下:

步骤1. 采用ARMA的时间序列模型生成两组相依数据,而且为突显两随机变量的关系,其中一组以ARMA(0,0)模型生成四个元素的模糊区间随机变量的样本数据集,即 $X_t = [x_{1t}, x_{2t}, x_{3t}, x_{4t}]$, $t = 1, 2, \dots, n$,一组以ARMA(0,2)模型生成四个元素的模糊区间随机变量的样本数据集,即 $Y_t = [y_{1t}, y_{2t}, y_{3t}, y_{4t}]$, $t = 1, 2, \dots, n$ 。

步骤2. 分别以 $X_t = [x_{1t}, x_{2t}, x_{3t}, x_{4t}]$ 和 $Y_t = [y_{1t}, y_{2t}, y_{3t}, y_{4t}]$ 为基础,将四个元素的模糊区间数据转换为二元模糊区间数据集,即 $X'_t = [x'_{1t}, x'_{2t}]$ 和 $Y'_t = [y'_{1t}, y'_{2t}]$,以及分别计算其平均值为 $\bar{X}_t$ 和 $\bar{Y}_t$ 。

步骤3. 利用生成的模糊区间随机变量之样本数据,依定义2的模糊区间互相关系数的算法,分别计算二元和多元模糊区间互相关系数,并且以平均值来计算单变量互相关系数。

从步骤1至步骤3的程序,在ARMA(0,2)相依数据的生成条件下,所获得之二元和多元之模糊区间互相关系数及单变量互相关系数,并且以CCF的函数值与延迟(Lag)期数为基准,绘制如图3和图4所示。

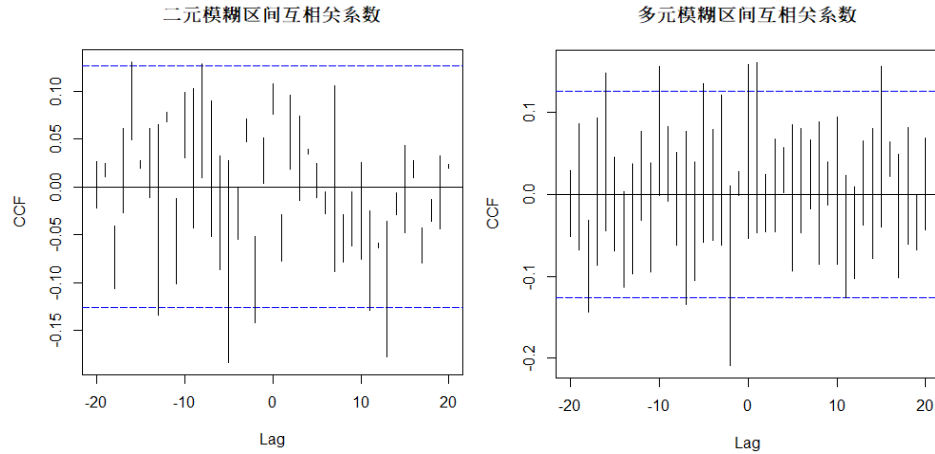


图3 模糊区间互相关系数(CCF)

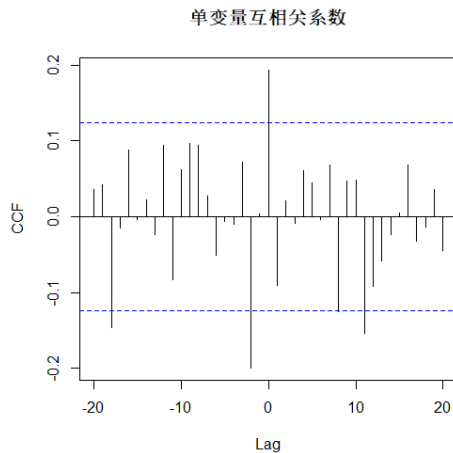


图4 单变量互相关系数(CCF)

图3和图4在于呈现两组相依数据之间的互相关程度,前者以模糊区间相依数据为演算基础,后者是以单变量相依数据为计算基础。根据两组相依数据之随机变量,即 X_t 和 Y_t ,并且透过CCF予以判断落后期数之相关性的判断,兹将模拟分析的结果分别概述如下:

(1)从随机变量 Y_t 的落后期数来看,首先,当 $\text{Lag} = 0$ 时,单变量的CCF为0.1936,而二元和多元模糊区间的CCF分别为(0.0777, 0.1077)、(-0.0539, 0.1585),显示两变量在未考虑落后期数的条件下,其相关性达到弱低度相关的程度。其次,当 $\text{Lag} > 0$ 时,即探讨 Y_t 的落后期数对 X_t 的影响关系,在单变量CCF的部分,显示 Y_t 在落后第11期时对 X_t 出现负弱低度相关的影响,其系数 $r_{11} = -0.1535$ 。二元模糊区间CCF的部分,显示 Y_t 在落后第11、13期时对 X_t 出现负弱低度相关的影响,其系数分别为 $r_B^{FICC}(11) = [-0.1295, -0.0247]$ 、 $r_B^{FICC}(13) = [-0.17750, -0.0358]$,而多元模糊区间

CCF的部分,显示 Y_t 在落后第1、15期时对 X_t 出现正弱低度相关的影响,其系数分别为 $r_M^{FICC}(1) = [-0.0466, 0.1614]$ 、 $r_M^{FICC}(15) = [-0.0401, 0.1569]$ 。由此可知,在 Y_t 对 X_t 的互相关系数的衡量上,FICC的互相关系数算法可以呈现 Y_t 在落后期数对 X_t 产生相关影响时,利用模糊区间互相关系数的变动趋势,可以较单变量互相关系数值掌握更多相依数据的隐藏信息。

(2)从随机变量 X_t 的落后期数来看,当 $\text{Lag} < 0$ 时,即探讨 X_t 的落后期数对 Y_t 的影响关系,在单变量CCF的部分,显示 X_t 在落后第2、18期时对 Y_t 出现负弱低度相关的影响,其系数 $r_2 = -0.1995$ 、 $r_{18} = -0.1463$ 。二元模糊区间CCF的部分,显示 X_t 在落后第5和13期时对 Y_t 出现负弱低度相关的影响,而在落后第8和16期时则出现正弱低度相关的影响,其区间系数分别为 $r_B^{FICC}(5) = [-0.1836, 0.0269]$ 、 $r_B^{FICC}(8) = [0.0093, 0.1289]$ 、 $r_B^{FICC}(13) = [-0.1345, 0.0648]$ 、 $r_B^{FICC}(16) = [0.0493, 0.1306]$,而多元模糊区间CCF的部分,显示 X_t 在落后第2和18期时对 Y_t 出现负弱低度相关的影响,其区间系数分别为 $r_M^{FICC}(2) = [-0.2089, 0.1114]$ 和 $r_M^{FICC}(18) = [-0.1437, -0.0310]$, X_t 在落后第5、10、16期时对 Y_t 出现正弱低度相关的影响,其区间系数分别为 $r_M^{FICC}(5) = [-0.0586, 0.1358]$ 、 $r_M^{FICC}(10) = [-0.0014, 0.1564]$ 、 $r_M^{FICC}(16) = [-0.0448, 0.1479]$ 。由此可知,在 X_t 对 Y_t 的互相关系数的衡量上,FICC的互相关系数可以呈现更多元的相关信息,例如 $r_M^{FICC}(2) = [-0.2089, 0.1114]$ 的相关系数横跨正、负相关,意味着该时间序列变量的区间变动相当大,故而在获悉巨大变动的情况下,透过更多的隐藏信息的呈现提高决策的可行性。

总合上述互相关系数的模拟分析结果,FICC算法

可以衡量两组相依性随机变量之相互落后的影响关系与单变量互相关系数具有相同的效果, 此外, FICC 算法更能提供了两组相依性随机变量之相互影响的变动程度, 也就是说, 透过互相关系数的运算可以进一步了解两组相依性随机变量的前后期关系及其相互关性的变动趋势。

4 数值分析

为能验证模糊区间自相关系数与互相关系数的衡量可以落实于实际相依性数据之上, 本研究透过台湾 T(Transcend)公司和 U(United Microelectronics Corporation)公司等科技类股之股票交易价格为数值分析的实际相依性数据, 并且进一步利用本研究提出之创新模糊区间自相关系数与互相关系数等算法的结果, 以及导入单变量自相关系数与互相关系数的差异性。

4.1 股票日成交价之自相关系数

股票当日开始的交易价格为开盘价, 随着当日股票交易的现况, 其价格亦会随之变动且公布于交易中心的平台, 若从每日的股票交易的价格来看, 在数据库中的数据显示, 则会出现开盘价、最高价、最低价和收盘价等实际交易价格, 而这样的数据型态俨然形成一组区间型态数据集。此外, 股票交易价格从开盘价开始, 之后的交易价格会在此一基础上进行波动, 意味着交易价格的前后期具有相依性的关系, 故本研究搜集自 2022 年 1 月 1 日起至 2022 年 12 月 31 日止的股票日交易价格, 摒除非交易日之外, 整理 T 公司股票日交易数据共计 252 笔, 其中数据采取开盘价、最高价、最低价和收盘价等区间观测值方式进行整理, 对此区间数据分别进行自相关系数的探讨。

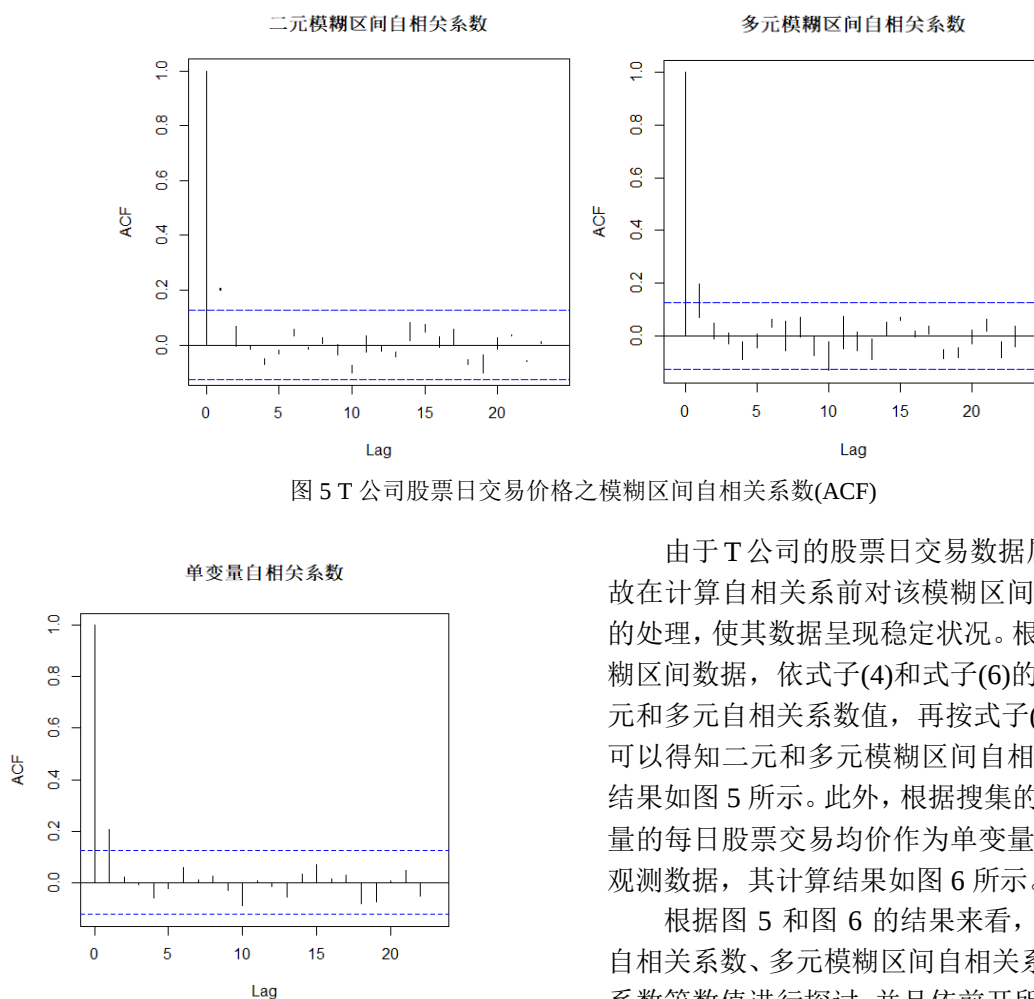


图 5 T 公司股票日交易价格之模糊区间自相关系数(ACF)

由于 T 公司的股票日交易数据属于非稳定的趋势, 故在计算自相关系数前对该模糊区间数据进行一阶差分处理, 使其数据呈现稳定状况。根据稳定状况下的模糊区间数据, 依式子(4)和式子(6)的算法, 分别计算二元和多元自相关系数值, 再按式子(5)和式子(7)的公式可以得知二元和多元模糊区间自相关系数区间, 计算结果如图 5 所示。此外, 根据搜集的区间数据计算单变量的每日股票交易均价作为单变量自相关系数的演算观测数据, 其计算结果如图 6 所示。

根据图 5 和图 6 的结果来看, 分按二元模糊区间自相关系数、多元模糊区间自相关系数、单变量自相关系数等数值进行探讨, 并且依前开所提之方法衡量 T 公司的股票交易价格与落后期数的交易价格之间所存在

图 6 T 公司股票日交易价格之单变量自相关系数(ACF)

着的关系，并将分析结果叙述如下：

(1)从自相关系数值的角度分析：T 公司的股票交易价格中，以日均价为基础所计算之单变量自相关系数，在落后 1 期时的系数值为 0.2067，显示该公司当期的股票交易价格与落后 1 期具有低度正相关的影响关系。若以模糊区间为基础所得到之二元模糊区间自相关系数在落后 1 期时呈现显著的影响关系，其系数值为(0.201, 0.2057)，显示其自相关程度是介于低度正相关的关系，而多元模糊区间自相关亦是在落后 1 期时出现显著关系，其系数值为(0.0705, 0.1964)，由此显示其自相关程度介在无相关至低度正相关之间，整体模糊区间自相关的影响关系与单变量所获得的结果一致。

(2)从自相关系数趋势的角度分析：有鉴于 T 公司股票的日常交易价之单变量自相关系数显示与落后 1 期之自相关系数值约为 0.2，表示该股票日常交易价有约 20%的可能性与落后 1 期有影响关系，相对有 80%的可能性与其他因素有影响关系。而多元模糊区间自相关系数则是显示其与落后 1 期的相关性介于 0.07 与 0.2 之间，意味着当期与落后 1 期的关系强度并未太高，这也表示了从自相关系数区间可以获得更多信息且进行探讨。

(3)从验证的角度分析：本研究搜集 T 公司于 2023 年 1 月 3 日至 1 月 10 日的股票日交易数据(如表 1 所示)进行说明，依据 2022 年 12 月 27 日至 30 日的区间数据来看，从开盘至收盘之间，这几日的跌幅情况比涨幅多，比如 27 日当日从开盘则是呈现跌幅至收盘，30 日当日虽然收盘比开盘高了 0.2，但其间整体最高价格也仅比收盘多出 0.2。若按与落后 1 期的自相关影响的程度来看，由于 2022 年 12 月 30 日的整体模糊区间呈现小幅上涨趋势，故而可以判断落后 1 期应该有两成机会出现上涨的可能性，而实际上 2023 年 1 月 3 日则与落后 1 期的收盘价呈现持平现象，而这样的结果与自相关系数衡量相似。又如 2023 年 1 月 6 日的区间呈现下跌的走势，故而判断 1 月 9 日的结果可能会出现下跌，然与落后 1 期的收盘相较却是上涨，而 1 月 9 日当日的区间却是持续呈现下跌走势，故而再次判断 1 月 10 日的结果可能会出现下跌，然而 1 月 10 日当日收盘价与落后 1 期的收盘相较却出现下跌的现象，因此可以了解当模糊区间内的数据波动，辅以自相关系数的关系，对于时间序列相依模糊区间数据的相关性会更具有其分析价值。

表 1 T 公司股票日交易信息

日期	模糊区间日交易价	与前 1 期涨跌	均价
2022/12/27	(66.8, 66.8, 66.4, 66.8)	—	66.7
2022/12/28	(66.3, 66.8, 66.1, 66.3)	↓	66.4
2022/12/29	(66.3, 66.3, 65.7, 65.7)	↓	66.0
2022/12/30	(65.8, 66.2, 65.8, 66.0)	↑	66.0
2023/1/3	(66.0, 66.0, 65.4, 66.0)	—	65.9
2023/1/4	(66.0, 66.4, 65.9, 66.2)	↑	66.1
2023/1/5	(66.4, 66.8, 66.3, 66.7)	↑	66.6
2023/1/6	(66.7, 66.7, 66.3, 66.7)	—	66.6
2023/1/9	(67.3, 67.5, 67.0, 67.1)	↑	67.2
2023/1/10	(67.3, 67.3, 66.9, 67.0)	↓	67.1

注：1.与前 1 期涨跌是以当日收盘价与前 1 期收盘价为衡量基准。

4.2 股票日成交价之互相关系数

为能体现互相关系数的演算效果，本研究除引用前述的 T 公司股票的日常交易价格外，另搜集 U 公司的自 2022 年 1 月 1 日起至 2022 年 12 月 31 日止的股票日常交易价格，同时亦摒除非交易日之外，整理 U 公司股票日常交易数据共计 252 笔，其中亦以采取开盘价、最高价、最低价和收盘价等区间观测值方式进行整理。然而，由于 U 公司的股票日常交易数据属于非稳定的趋势，故对其模糊区间数据进行一阶差分的处理，使其数据

呈现稳定状况，并且结合差分后之 T 公司的股票日常交易数据，进行互相关系数的演算。依式子(8)和式子(10)的模糊区间互相关系数算法，并且结合 T 公司股票交易价格之模糊区间数据，分别计算二元和多元之互相关系数值，再按式子(9)和式子(11)的公式可以得知两公司之间的二元和多元模糊区间互相关系数区间，计算结果如图 7 所示。此外，根据搜集的 U 公司之区间数据，计算 U 公司每日股票交易的均价且结合 T 公司之每日股票交易的均价，作为两公司之间的单变量互相关系数的演算观测数据，其计算结果如图 8 所示。

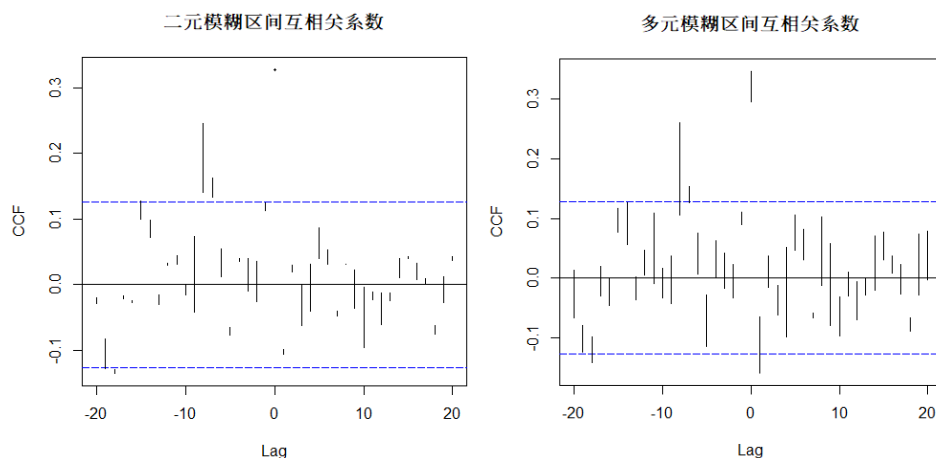


图7 T公司与U公司股票日交易价格之模糊区间互相关系数(CCF)

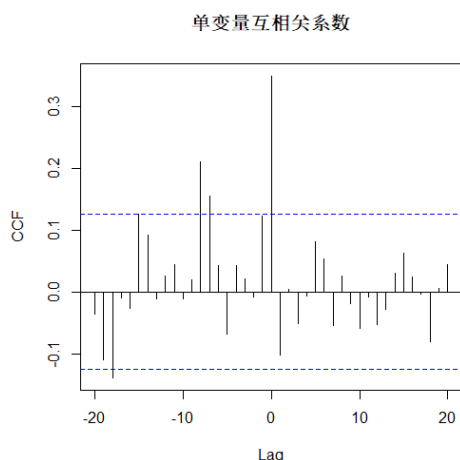


图8 T公司与U公司股票日交易价格之单变量互相关系数(CCF)

从图7和图8的结果可以得知在T公司股票日交易价格落后期数与U公司股票日交易价格之间的关系,以及U公司股票日交易价格落后期数与T公司股票日交易价格之间的关系,并两者之分析结果分别叙述如下:

(1)从T公司股票日交易价格落后期数的角度分析:首先,在 $Lag = 0$ 时,显示T公司与U公司的股票日交易价格具有低度的相互影响关系,其中单变量的互相关系数为0.3488,二元模糊区间互相关系数为(0.3276, 0.3278),多元模糊区间互相关系数为(0.2952, 0.3471)。其次,在 $Lag > 0$ 时,二元模糊区间与单变量所计算之互相关系数与U公司股票日交易价格无显著的相互影响关系,而多元模糊区间互相关系数则是呈现在落后1期(即 $Lag = 1$)时,其系数值为(-0.1589, -0.0647),表示落后1期的T公司股票之日交易价对U公司股票日交易价呈现弱低度至无相关的关系,意味

着T公司股票日交易对U公司股票价格影响有限。

(2)从U公司股票日交易价格落后期数的角度分析:在 $Lag > 0$ 时,单变量互相关系数计算得知在落后7、8期时,呈现与T公司股票日交易价格具有显著的相互影响关系,其系数值为0.1547、0.2106。二元模糊区间互相关系数亦是出现U公司股票日交易价格在落后7、8期时与T公司股票日交易价有显著相互影响关系,其区间系数值分别为(0.1333, 0.1628)、(0.1403, 0.2465),而多元模糊区间互相关系数则是在U公司股票日交易价格落后7、8期时与T公司股票日交易价呈现显著相互影响关系,其区间系数值分别为(0.1261, 0.1546)、(0.1057, 0.2601)。由此可见,虽然U公司股票之日交易价格在落后7、8期时与T公司股票日交易价呈现弱低度相关,这也意味着在考虑T公司股票时,可以参酌U公司股票的日交易价格。

5 结论

随着互联网的兴起,网络系统所覆盖的范围已经从经济、环境逐渐扩充至社会、生活、交通等层面,然而在覆盖层面愈广的情况下,同时也反应各个层面的网络传输、浏览、交易等行为都可以迅速汇聚成数据且数量可能极其庞大,而数据型态更有可能一改过去单一数值的呈现形式,演变成多元的区间数值形式呈现,甚至在这些生成的数据中,不遑有许多属于相依性的数据型态,例如空气质量监测数据、股票交易价格、家户水电用度等。基于数据的相依性与多元性,如果采用过去单变量的方法对区间数据进行分析,可能会遗漏数据所隐藏的信息,例如区间数据间的波动、极大与极小的差距等,而这些信息可能直接或间接影响后续的决策结果。除此

之外,在生成数据的同时,对于数据的产生有时会出现不确定性的误差现象,致使数据实际发生值出现偏差,例如股票交易过程中,可能在挂牌购买的价格与实际购买的价格会有所误差,故而产生了期望价格与交易价格的差异化,若依单变量的统计分析方法可能直接引用交易价格,如此将会遗失期望价格的信息。

在监督式或非监督式模型的建构过程中,变量与变量之间的相关性往往是模型构建的第一步骤,也是重要的步骤之一,而变量之间的关系会影响到模型的预测效能,更遑论是相依性数据的相关系数,毕竟相依性数据本身就带着自我前后期的相关性,故而时间序列相依数据的相关系数的测量算法探讨一直是值得深入研究的课题之一。一般而言,单变量的相关系数算法会随着随机变量的型态不同采用不同的算法,例如连续型随机变量会采用皮尔逊积差相关系数算法、间断型随机变量会采用斯皮尔曼等级相关系数算法,而相依性随机变量会采用自我相关系数算法等。然而,当所观察随机变量存在于模糊的区间性且据有前后期之相依性关系时,此时对该变量使用单变量的相关系数算法,势必会导致数据分析的误差现象发生,故此,本研究提出一个维持模糊区间数据的条件下,对相依区间数据的相关系数进行演算的方法,进而获得区间相关系数的结果,而且从研究的模拟分析与数值分析中,可以得到本文提出的FIAC和FICC等算法不仅可以达到与单变量相关系数的效果外,更能提供区间相关系数的变动趋势,足以变量自我的短、中、长期的变动关系与其趋势,据此可以提升对相依变量的区间数据的剖析功能,降低单变量算法所可能遗失的效果,得以透过区间数据估计的方式予以且从中获得更多的隐藏讯息。

总而言之,在时间与科技的推演时代中,相依模糊区间数据俨然逐渐成为经济、社会、环境等各大领域所面临的问题,虽然单变量的数据分析方式亦能将区间数据简化,但终究可能会遗漏部分数据分析的结果及扩大误差的现象,本研究考虑在相依区间数据生成的环境下,将自相关系数与互相关系数等算法的概念导入于研究课题中,据以因应相依性与模糊区间性的数据型态,并且藉由二元和多元的区间数据型态之模糊区间自相关系数算法的思维,延伸创新估计二元和多元的模糊区间互相关系数算法,试图解决多元且真实之相依模糊区间数据的问题,藉此从中挖掘出更多的相依区间数据的隐藏讯息,期望可以被应用于经济、社会、企业经营等领域之政策拟订参考。

参考文献

- [1] 覃玉冰, 邓春林, 杨柳., 基于皮尔逊相关系数的网络舆情评估指标体系构建研究 [J]. 情报探索, 2018(10): 5.
<https://link.cnki.net/doi/10.13603/j.cnki.51-1621/z.2019.10.009>
- [2] 程娟娟. 高校科研与教学关系实证研究——基于皮尔逊相关系数的分析 [J]. 中国高校科技, 2022(10): 7.
<https://link.cnki.net/doi/10.16209/j.cnki.cust.2022.10.016>
- [3] 畚梓航, 徐嘉桦, 姚志玉, 梁伟典. 基于皮尔逊相关系数的网购大数据分析—以天猫佰润居旗舰店交易记录为例 [J]. 韩山师范学院学报, 2020, 41(3): 7.
<https://link.cnki.net/doi/10.16775/j.cnki.44-1302/G4.2021.03.012>
- [4] Ildar Z. Batyrshin, 2019, Constructing Correlation Coefficients from Similarity and Dissimilarity Functions [J], Acta Polytechnica Hungarica, 16(10), p191-204.
<https://doi.org/10.12700/aph.16.10.2019.10>
- [5] 邓蓉婕, 方兆本. 基于斯皮尔曼相关分析的理财产品收益率分析 [J]. 统计与决策, 2019(16): 164-167.
<https://doi.org/10.13546/j.cnki.tjyjc.2019.16.036>
- [6] 张伟锋, 徐维超. 斯皮尔曼简捷相关系数与基尼伽玛相关系数对脉冲噪声的稳健性分析 [J]. 电子世界, 2020(10): 81-82.
<https://link.cnki.net/doi/10.19353/j.cnki.dzsj.2020.10.042>
- [7] Song, Ha Yoon; Park, Seongjin, An Analysis of Correlation between Personality and Visiting Place using Spearman's Rank Correlation Coefficient [J], KSII Transactions on Internet and Information Systems, 2018, 14(5): 1951-1966.
<https://doi.org/10.3837/tiis.2018.12.027>
- [8] Jost P J, Wagenhofer A. Special Section: Economic Analyses in Business Administration [J]. Schmalenbach Business Review, 2018, 70(1): 1-2.
<https://doi.org/10.1007/s41464-018-0047-9>
- [9] Guotao Ma. and Mohammad Rezaia, Mohaddeseh Mousavi Nezhad, 2022, Effects of spatial autocorrelation structure for friction angle on the runout distance in heterogeneous sand collapse [J], Transportation Geotechnics, 33: 100705.
<https://doi.org/10.1016/j.trgeo.2022.100705>
- [10] Yang, C. C., 2016, Correlation Coefficient Evaluation for the Fuzzy Interval Data [J], Journal of Business Research, 69(6), 2138~2144.
<https://doi.org/10.1016/j.jbusres.2015.12.021>
- [11] 杨志清. 计数区间之模糊相关系数研究 [J], 统计与决策, 2020, 36(20): 28-32.
<https://link.cnki.net/doi/10.13546/j.cnki.tjyjc.2020.20.006>

- [12] 王鑫蕊, 窦鑫莹, 杨志清, 施建华. 模糊区间点二列相关系数研究 [J]. 高师理科学刊, 2022, 42(11): 8-15.
<https://link.cnki.net/doi/10.19986/j.cnki.23-1418/n.2022.11.002>
- [13] 陶洋, 祝小钧, 杨柳. 基于皮尔逊相关系数和信息熵的多传感器数据融合 [J]. 小型微型计算机系统, 2023, 44(5): 1075-1080.
<https://link.cnki.net/doi/10.19979/j.cnki.issn.1000-7180.2023.05.015>
- [14] 季鹏, 陈芳芳, 徐天奇, 等. 基于 Spearman 相关系数法与有功分量法的高阻接地故障选线方法研究 [J]. 山东电力技术, 2022, 49(12): 8-13.
<https://link.cnki.net/doi/10.19457/j.cnki.sddlj.2023.10.008>
- [15] 孙慧慧. 基于 M 估计的具有一致相关线性混合效应模型中相关系数的齐性检验 [J]. 应用概率统计, 2013, 29(5): 469-479.
<http://www.chinadoi.cn/10.3969/j.issn.1001-4268.2013.05.003>