Artículos de investigación

Kruskal-Wallis Test for Functional Data Based on Random Projections Generated from a Simulation of a Brownian Motion

Prueba de Kruskal-Wallis para datos funcionales basada en proyecciones aleatorias generadas a partir de una simulación de un movimiento browniano

Rafael Meléndez Surmay
Universidad de la Guajira, Colombia
rmelendez@uniguajira.edu.co
https://orcid.org/0000-0002-6449-0358
Ramón Giraldo Henao
Universidad Nacional de Colombia, Colombia
rgiraldoh@unal.edu.co
https://orcid.org/0000-0003-3010-169X
Francisco Rodríguez Cortes
Universidad Nacional de Colombia, Colombia
frrodriguezc@unal.edu.co
https://orcid.org/0000-0002-2152-8619

Recepción: 10 Enero 2024
Aprobación: 22 Abril 2024
Publicación: 29 Abril 2024

Acceso abierto diamante

Abstract

The *k*-sample problem for functional data has been widely studied from theoretical and applied perspectives. In literature, Gaussianity of the generating process is generally assumed, which may be impractical in some situations. This work proposes an extension of the Kruskal-Wallis test to the case of functional data as an alternative to the problem of non-Gaussianity. The methodology used consisted of transforming each group's functional data into scalars using random projections and subsequently performing classical Kruskal-Wallis tests. The main results were the extension of the Kruskal-Wallis test to the case of functional data and the verification of its unbiased and consistency properties. Reducing dimensionality from random projections allows us to extend the classical Kruskal-Wallis test to the functional context and solve problems of non-Gaussianity and atypical observations.
Keywords: Functional data, random projections, Kruskal-Wallis test, non-parametric statistics, brownian motion.

Resumen

El problema de *k* muestras de datos funcionales se ha estudiado ampliamente desde perspectivas teóricas y aplicadas. En la literatura se asume generalmente el supuesto de Gaussianidad del proceso generador, el cual puede ser impráctico en algunas situaciones particulares. Este trabajo tuvo como objetivo proponer una extensión de la prueba de Kruskal-Wallis al caso de datos funcionales, como alternativa al problema de no Gaussianidad. La metodología empleada consistió en transformar los datos funcionales de cada grupo en escalares empleando proyecciones aleatorias y en realizar posteriormente pruebas de Kruskal-Wallis clásicas. Los principales

Notas de autor

rmelendez@uniguajira.edu.co

resultados fueron la extensión de la prueba de Kruskal-Wallis al caso de datos funcionales y la comprobación de las propiedades de insesgadez y consistencia de esta misma. Se puede concluir que la reducción de la dimensionalidad a partir de las proyecciones aleatorias permite extender la prueba de Kruskal-Wallis clásica al contexto funcional y por ende solucionar problemas de no Gaussianidad y observaciones atípicas.

Palabras clave: Datos funcionales, proyecciones aleatorias, prueba de Kruskal-Wallis, estadística no paramétrica, movimiento Browniano.

Highllights inglés:

The Kruskal-Wallis test resolves the issue of non-normality in functional data.
The Kruskal-Wallis test addresses the treatment of outliers in functional data.
The use of random projections reduces the dimensionality of the problem of k samples in functional data.

Highlights español:

La prueba de Kruskal-Wallis resuelve el problema de no normalidad en datos funcionales.
La prueba de Kruskal-Wallis soluciona el tratamiento de datos atípicos en datos funcionales.
El uso de proyecciones aleatorias reduce la dimensionalidad del problema de k muestras en datos funcionales

1. INTRODUCTION

Advances in computational and analytical techniques allow for continuous monitoring of many processes. New statistical methods are needed to analyze large data sets arising from these processes. Functional data analysis (FDA) has emerged in recent decades as an alternative to statistical modeling of large data volumes. FDA is a framework for analyzing data consisting of random functions (usually curves) rather than observations of a few variables or random vectors [1]. New challenges have arisen in extracting meaningful information hidden in functional data [2]. As in classical statistics, in FDA data preprocessing, modeling, hypothesis testing, parameter estimation, and predictive analysis using parametric or nonparametric models are fields of interest. Many theoretical and applied contributions have been proposed in these areas [2], [3]. In the last decade, the FDA has already found applications in several areas of research, including ecology [4], epidemiology [5], remote sensing [4], outlier detection in environmental applications [6], and traffic volume forecasting [7].

To construct a functional observation X*ij*(.) from the discretely observed data one can employ a standard smoothing technique such as cubic B-splines [8]. The FDA package [9] implements the smoothing techniques in R [10].

This work focuses mainly on proposing a methodology for comparing groups when the same functional variable has been observed in several individuals in each of these. Specifically, a traditional nonparametric tool to solve the .-sample problem for a functional response is adapted to the FDA scenario. Let $X_{i1}(t)$, $X_{i2}(t)$, $X_{in}(t)$ ... $i = 1,2, \ldots, k$ random set of functions defined over an interval $T = [a,b]$ which come from Gaussian processes GP $(\mu_k(t), \gamma_k(s,t))$ [8]. The hypothesis of interest is given in (1)

$$H_0: \mu_1(t) = \mu_2(t) = \cdots = \mu_k(t), \quad t \in T \tag{1}$$

Against the alternative that at least two functional means are different. The statistical literature has a widely considered hypothesis established in (1)The proposed approaches are proposed for point-wise t-tests, functional ANOVA, functional principal components analysis, and permutation tests.

Some authors have extensively studied the functional ANOVA problem. For example, [9] introduced an asymptotic version of the ANOVA F-test, and [2] considered asymptotic or bootstrapped versions of a L^2 norm based test, F-type statistic-based test, and globalizing pointwise F-test. Furthermore, [1] introduced a method based on a representation of a basis function, and [10] described a bootstrap procedure based on pointwise F-tests. However, Bayesian functional ANOVA has received less attention. But, [11] introduced a Gaussian process ANOVA modeling approach under a Bayesian framework.

Other approaches were considered by [12], [9], and [13]. Furthermore, [14] proposed a new method using a graphical interface based on the global rank test, and this procedure for functional ANOVA was applied using permutations. Other authors have proposed other approaches, such as that used the Westfall-Young randomization to correct for multiple tests. However, this method cannot obtain an overall p-value. Meanwhile, [15] divided the domain of interest into regions. However, a disadvantage is that the partition must be respected. Furthermore, [16] developed a multi-way functional ANOVA to determine rejection regions. Our interest is to provide an alternative to the case where the Gaussian assumption is unrealistic, and [17] presented a unified methodology for performing computation-free permutation tests for the testing of the *k* sample in commutative and noncommutative *Lq.* spaces, which includes multivariate and functional data.

This work is organized as follows. Sections 2.1 and 2.2 review the Kruskal-Wallis test and random projections. Section 3 presents an extension of the Kruskal-Wallis test for functional data and shows its respective pseudocode. In Section 4.1, we present the simulation study and in Section 4.2, we present the application with real data. Finally, we present the discussion and some conclusions.

2. BACKGROUND

2.1 Kruskal-Wallis test

This section briefly reviews the main statistical technique used in the analysis. Kruskal-Wallis [18] is a non-parametric statistical test that compares the median values of two or more independent samples. The null hypothesis for the Kruskal-Wallis test is that all the samples come from the same population, and the alternative hypothesis is that at least one group's sample comes from a population with a different median than the others. The test is based on the ranks of the observations within each group. It is an alternative to ANOVA when the normality assumption is unrealistic. The hypothesis of interest is shown in (2)

$$H_0: \tau_1 = \tau_2 = \cdots = \tau_k \quad vs \quad H_1: \tau_1, \cdots, \tau_k, \quad \text{not all equal} \tag{2}$$

Which establishes that there are no significant differences in the effects of the treatments. The null hypothesis states that the following distributions $F1 = F2 = \cdots = Fk$ are equal. To calculate the Kruskal-Wallis statistic, all . observations from the k-samples are combined and ordered from smallest to largest. Let r_{ij} be the rank of r_{ij} in this joint classification, and Rj defined as (3)

$$R_j = \sum_{i=1}^{n_j} r_{ij}, \quad R_{\cdot j} = \frac{R_j}{n_j}, \quad j = 1,2,\cdots,k \tag{3}$$

Thus, for example, R_1 is the sum of the ranks received by the observations of group 1 and R_1 is the average rank for these same observations. Kruskal-Wallis H statistics are given by [18] as shown in (4)

$$H = \frac{12}{N(N+1)} \sum_{j=1}^{k} n_j \left[R_{\cdot j} + \frac{N+1}{2} \right]^2 \tag{4}$$

At a significance level of α, H0 is rejected if H ≥ hα otherwise, do not reject. The values of hα are given in Table A.12 of [18]. When H0 is true, the statistic H has, as min(n., ⋯. nk) tends to infinity, an asymptotic chi-square X^2 distribution with *k* - 1 degrees of freedom. Under this assumption, the reject rule is.

Reject $H \geq \chi^2_{k-1,\alpha}$; otherwise, do not reject.

When the null hypothesis is rejected and it is concluded that at least one sample comes from a population with a different median, some post-hoc tests (e.g., Dunn's test) can be used to identify which samples differ significantly.

2.2 Random Projections

The hypothesis of interest (see hypothesis in (1)) can be tested using the projections of the functions. These involve mapping high-dimensional data points into a lower-dimensional space using a randomly generated projection matrix [19]. The basic idea is to use a randomly generated projection matrix to map each high-dimensional data point onto a lower-dimensional space. By doing this, we can reduce the number of dimensions of the data while still retaining important information about the data structure.

Random projections are often used in situations where the dimensionality of the data makes it difficult to work with or analyze. In other words, random projections can be a handy tool for reducing the complexity of the data without losing important information. Given a set of data or a distribution in spaces of dimension greater than one, random projections consist of projecting the data or calculating the marginal of the distribution in a lower-dimensional subspace that has been chosen randomly [20]. Random projections preserve certain properties that are very important in the FDA. One of them is that it preserves distances with a high degree of probability if a projected subspace is the uniform distribution. This result is extended to the standard Gaussian distribution [10]. In this sense, [21] showed that if two distributions are defined in a separable Hilbert space and have finite moments of some order, then projecting the distributions onto a random one-dimensional subspace is sufficient to distinguish them with high probability, as long as the moments of one of the distributions match those of the random projection. In other words, if we have two distributions with similar moments up to some order, projecting them onto a random one-dimensional subspace will produce similar one-dimensional marginal distributions. However, if the moments of one of the distributions differ from those of the random projection, then the one-dimensional marginal distributions will be different, and the two distributions can be distinguished with high probability.

Once the functional data have been projected onto a lower-dimensional space, a hypothesis test can be performed to determine whether the functional means are equal. The choice of hypothesis test depends on the specific application, but a common approach is to use a t-test or an ANOVA test. One advantage of using random projections to test the equality of functional means is that it can be computationally efficient, mainly when dealing with high-dimensional functional data. It can also be robust to noise and outliers in the data, as random projections can help filter out some of the noise.

3. KRUSKAL-WALLIS TEST FOR FUNCTIONAL DATA

This research presents an extension of the Kruskal-Wallis test for functional data based on random projections.

We propose extending the Kruskal-Wallis test to the case of functional data (the observation for each individual in the sample corresponds to a functional datum). As in the univariate case, in the context of functional data analysis, statistical tests require the fulfillment of some assumptions. When the samples are small and the curves do not underlie a Gaussian stochastic process, the functional ANOVA could be inappropriate, and a non-parametric method may be used as a valid alternative. Specifically, a Kruskal-Wallis test for functional data based on random projections (KWFD) is proposed as an alternative methodology to the one-way functional ANOVA when the Gaussianity assumption is unrealistic. The KWFD is a non-parametric alternative for comparing the medians of functional data of three or more groups. We extended the KW test by randomly projecting the functional data onto a low-dimensional subspace.

Let $X_{ij}(t)$, $i = 1,2,\cdots, n_j$, j = 1, ⋯, k a functional random sample of curves, where . $t \in [a, b]$ is the domain (generally time), i correspond to an individual, and j the index for the level factor. The functional random variables are considered independent trajectories of the stochastic processes SP(μj (t),$\gamma(s,t)$),$j = 1,\cdots, k$ with a common covariance function $\gamma(s, t)$. Let $x_{ij}(t)$, $i = 1,2,\cdots, n$; $j = 1,\cdots, k$. be the recorded set of curves under the k. treatments. In the following, we describe the procedure for calculating the H statistic to test the null hypothesis in (1)

· Generate one Brownian motion $\upsilon(t)$ in the interval of interest $T \in \mathbb{R}$.

· Calculate the random projections $xij = \int xij(t)\upsilon(t)dt, ba$ $i = 1,\cdots, n; j = 1,2,\cdots, k$.

· Calculate the rank of each projected curve within its group.

· Using the random projections, proceed as in the usual way to calculate rij, Rij, and the statistic H in (3)

· Reject the null hypothesis in (2) at the level α if $Hc \geq \chi^2_{k-1;1-\alpha}$. An alternative is calculating the p-value using a permutation test.

The Kruskal-Wallis test for functional data based on random projections is calculated similarly to the univariate Kruskal-Wallis test. It is based on the sum of the ranks of the projected curves within each group. The test assumes no specific distribution for the functional data and can be robust to atypical curves.

4. RESULTS AND DISCUSSION

Section 4.1 presents a simulation study based on a single Brownian motion simulation. Section 4.2 shows the p-values obtained by generating 1000 random projections.

4.1 Simulation study indicators

We assess the power of the test to detect differences between medians of .-samples of functional data. To establish the performance, we show the results of a simulation study. We follow the procedure given in [15] to perform the analysis. For simplicity, just three groups of curves are considered.

$$\begin{aligned} X_{i1}(t) &= \mu(t) + \varepsilon_i(t) \\ X_{i2}(t) &= \mu(t) + \varepsilon_i(t) \\ X_{i3}(t) &= \mu(t) + \delta(t) + \varepsilon_i(t), \end{aligned} \qquad (5)$$

Where $\mu(t)=\sin(2\pi t), t \in (0,10)$, is the mean function and the errors $\varepsilon ij(t)=1,2,3$, follow a uniform distribution on $[-1,1]$. As an initial illustration, a graph of a Brownian motion and 120 simulated curves according to the equations given in (5) are shown in Figure 1.nbsp;The curves in red and green are very similar (these come from analogous models (rows 1 and 2 of the equations in 5, and the curves in blue involve an additional parameter $\delta(t)=\delta=1.2$ that makes these different from the previous ones. Notice in Figure 1 that the highest periodic peaks of the blue curves are close to 3, while in the other two cases (red and green curves), these are close to 2, i.e., the null hypothesis should be rejected. The errors are assumed to be uniform in the interval (1,1). Performing a hypothesis test on the means of functional data assuming that the processes are Gaussian with data such as those presented in Figure 1 would be inappropriate.

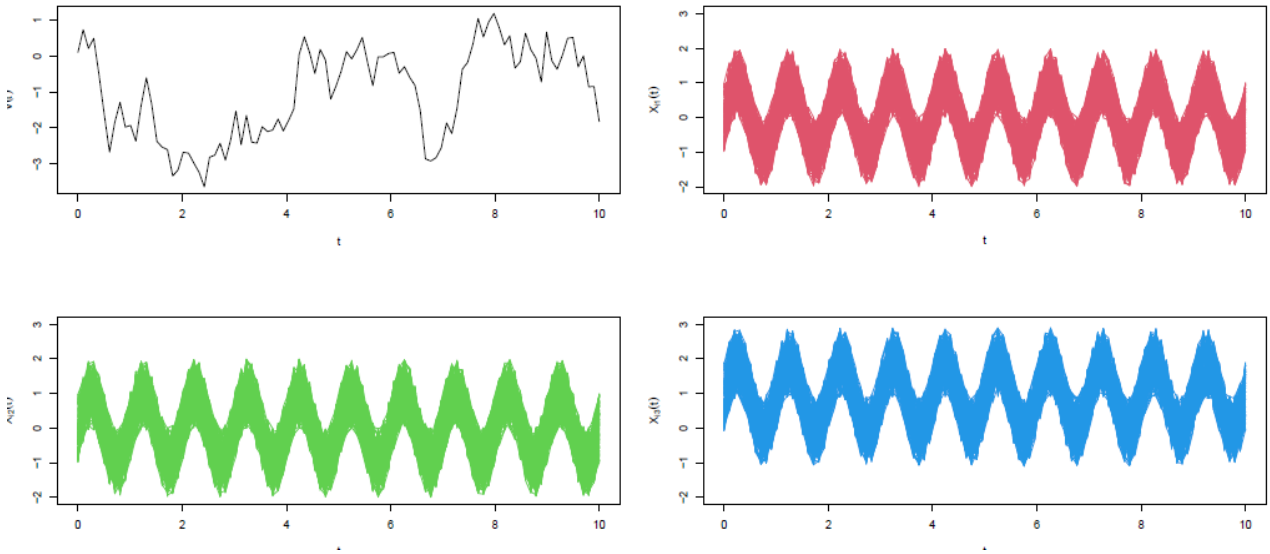

Figure 1.

Brownian motion v(t) = v(t - 1) + ϵ(t),ϵ(t) ~ Normal (0,0.5),t ∈ (0,10) (above left) and curves simulated under the models Xi1 (t) = μ(t) + εi (t) (above right), Xi2 (t) = μ(t) + εi (t) (below left), and Xi3 (t) = μ(t) + δ(t) + εi (t) (below right), with μ(t) = sin(2πt),δ(t) = 1.2 and ε(t) ~ uniform(-1,1).

Source: Created by the authors.

To evaluate the power of the test, we considered $\delta(t) = \delta$, for all $t \in [0,10]$, with $\delta = 0.0,\cdots,0.7$. Four sample size scenarios are considered (n = 10,30,80,120) for each sample group. In each case, 1000 realizations are generated. Based on each sample size, we performed a Kruskal-Wallis test as defined in Section 3. In each case, the power of the test is obtained as the percentage of $p-values$ less than 0.05. We used the libraries fda.usc and stats of R to perform the analysis [22]. Figure 2 shows the empirical power curves for each of the sampling sizes n and $\delta(t) = \delta$ values. Note that the power of the test increases when δ and n increase; that is, the simulation study provides evidence that the Kruskal-Wallis test for functional data is unbiased and consistent[1].

[1] The R code used is available at https://github.com/frajaroco/KWfdRP/blob/main/KWtest.R

[2] See Canada's Climate Regions at the link https://sites.google.com/a/ocsb.ca/cgc-1d/a-unit-4-climate/1-canadas-climate-regions).

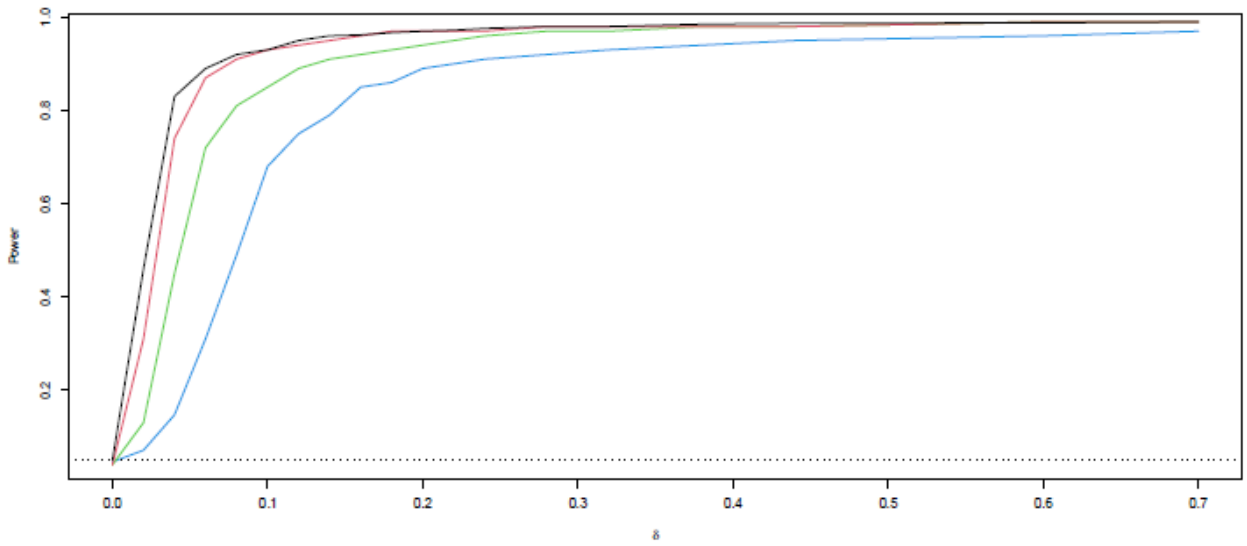

Figure 2.

Empirical power curves of the Kruskal-Wallis test according to the variation function $\delta(t) = \delta$ and the sample size n. n = 10 (blue line), n = 30 (green line) n = 80 (red line), and n = 100 (black line) for each sample group. The bottom dashed line corresponds to the significance level $\alpha = 5\,\%$. Created by the authors.

Source: Created by the authors.

4.2 Real data analysis: Temperature curves in Canada

We apply the Kruskal-Wallis test for functional data from Section 3 to a widely used meteorological data set in the context of the FDA [23]. This corresponds to the average daily (30-year) temperature (in degrees Celsius) at each of the 35 weather stations located in four climatic zones of Canada (in brackets the number of stations in each zone): Arctic (4) Pacific (7), Continental (9), and Atlantic (15) (see Figure 3). The Pacific zone is located on the west coast of Canada, including British Columbia and parts of Yukon and the Northwest territories. This area is defined by mild, rainy winters and cool, dry summers. The continental region covers the central parts of Canada, including Manitoba, Saskatchewan, and parts of Alberta and Ontario. Its climate is marked by cold winters and short and hot summers. The Atlantic zone covers the eastern parts of Canada, including Nova Scotia, New Brunswick, and Prince Edward Island. It has mild, wet winters and cool, moist summers. The Arctic region covers the northernmost parts of Canada, including Nunavut, the northwest territories, and parts of Yukon, Quebec, and Labrador. This zone has long, harsh winters and short, cool summers (see Canada's Climate Regions at the link https://sites.google.com/a/ocsb.ca/cgc-1d/a-unit-4-climate/1-canadas-climate-regions). The daily temperature data for the four climatic zones were smoothed using a Fourier basis function. The curves obtained after smoothing are shown in Figure 3.The interest is to determine whether there are significant differences between the mean (median) curves of these areas. For this purpose, we apply the Kruskal-Wallis test presented in Section 3. We generate random projections using (6) with i the index corresponding to the weather station in each one of the four climatic zones (j=1 (Arctic), 2 (Pacific), 3 (Continental), 4 (Atlantic)) and $\nu(t)$ a Brownian motion. The number of stations in each zone is 4 (Arctic), 7 (Pacific), 9 (Continental), and 15 (Atlantic).

$$x_{ij} = \int x_{ij}(t)v(t)\,dt, \quad j = 1,\cdots,4, \quad and \ t \in (0,365) \tag{6}$$

After obtaining the random projections, we conduct a classical Kruskal-Wallis test with these values. For this case, a p - value = 0.00361 was obtained, and consequently, in concordance with Canada's Climatic description above, the null hypothesis is rejected. Note that there are some atypical curves in each panel of Figure 3.Using a classical ANOVA test based on random projections can be limited in this case. A robust methodology, as proposed here, could be more appropriate. Wilcoxon's post-hoc tests [24] (Table 1) at a 10 % significance level of 10 % show that the medians of the Atlantic and Pacific zones are significantly different from the median of the Arctic region. At the same level, there are differences between the medians of the Atlantic and Continental regions. A graphical comparison (Figure 3) indicates marked differences between the curves of these regions.

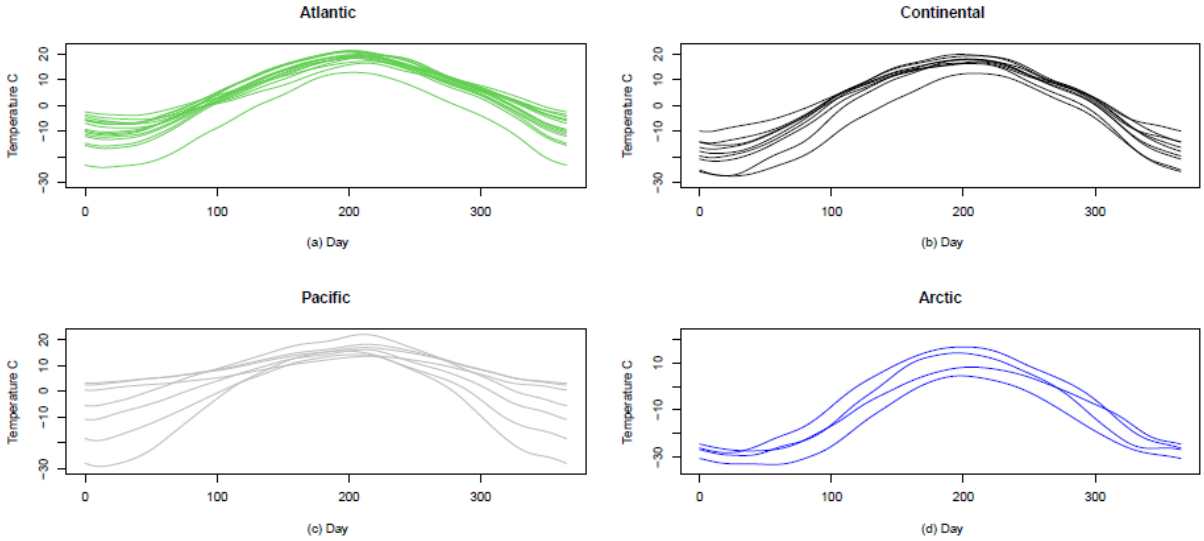

Figure 3.

Temperature curves ($x_{ij}(t)$) for the Atlantic, Continental, Pacific, and Arctic climate zones obtained after daily data (averages of 30 years) are smoothed using Fourier basis functions. Created by the authors.

Source: Created by the authors.

Table 1

Wilcoxon posthoc tests.

	Atlantic	Continental	Pacific
Continental	0.09	--	--
Pacific	0.95	0.25	--
Artic	0.01	0.20	0.07

Source: Created by the authors.

The results described above are based on random projections from a particular BM. The attached R code[3], shows the values found with 1000 Brownian motions, and the general conclusion is the same.

[3] https://github.com/frajaroco/KWfdRP/blob/main/KWCanadianWeather.R

4.3 Discussion

ANOVA for functional data has been widely discussed, and several approaches have been considered [1], [2]. Many of these are based on the Gaussianity assumption [8, 10]. Here, we adapt a classical non-parametric test to this scenario. The strength of the Kruskal-Wallis test for functional data proposed here lies in its

versatility. It does not depend on the assumption of Gaussianity, thus extending its applicability to various real-world scenarios where data may deviate from a Gaussian distribution. This test is flexible and can be used with various types of functional data, including curves and time series. It does not impose strict assumptions on the data distribution, making it suitable for analyzing diverse datasets. This approach is particularly advantageous when dealing with data that may not conform to normality or have unknown distributions. Like other statistical tests, the Kruskal-Wallis test assumes the independence of observations within and between groups. Violations of this assumption could potentially affect the accuracy of the test results. If the Kruskal-Wallis test indicates significant differences between groups, post-hoc tests can be conducted to identify differences between groups. Many other non-parametric methods are available for post-hoc testing, each with strengths and limitations.

5. CONCLUSIONS

We propose a non-parametric method for the *k*-functional problem, which is useful when the sample size is small, the assumption of normality is not reasonable, or when there are atypical curves. We propose the use of one-dimensional random projections to solve the problem. After obtaining scalars from functions using random projections, a classical Kruskal-Wallis test can be used to test the hypothesis. The results obtained from the simulated and real data show a good performance of the methodology. The results (Figure 2) illustrate that the Kruskal Wallis test extension performs well under the null hypothesis. Power increases for larger sample sizes and distance parameter. This plot allows us to validate that the proposed test is unbiased and consistent. Some authors consider using points-wise test statistics for functional data problems with two samples and similarly for the .-sample problem, although they are not global tests. Our approach is a helpful alternative when the sample is small, and the Gaussian assumption is inappropriate.

REFERENCES

[1] T. Górecki and Ł. Smaga, "A comparison of tests for the one-way ANOVA problem for functional data," *Comput. Stat.*, vol. 30, no. 4, pp. 987–1010, Dec. 2015. https://doi.org/10.1007/s00180-015-0555-0

[2] J. T. Zhang, *Analysis of variance for functional data*, 1st ed. New York, NY, USA: Chapman and Hall/CRC, 2013. https://doi.org/10.1201/b15005

[3] F. Ferraty, P. Vieu, and S. Viguier-Pla, "Factor-based comparison of groups of curves," *Comput. Stat. Data Anal.*, vol. 51, no. 10, pp. 4903–4910, Jun. 2007. https://doi.org/https://doi.org/10.1016/j.csda.2006.10.001

[4] M. L. Bourbonnais et al., "Characterizing spatial-temporal patterns of landscape disturbance and recovery in western Alberta, Canada using a functional data analysis approach and remotely sensed data," *Ecol. Inform.*, vol. 39, pp. 140–150, May. 2017. https://doi.org/https://doi.org/10.1016/j.ecoinf.2017.04.010

[5] A. Roy, T. Nelson, and P. Turaga, "Functional data analysis approach for mapping change in time series: A case study using bicycle ridership patterns," *Transp. Res. Interdiscip. Perspect.*, vol. 17, p. 100752, Jan. 2023. https://doi.org/https://doi.org/10.1016/j.trip.2022.100752

[6] J. M. Torres, P. J. G. Nieto, L. Alejano, and A. N. Reyes, "Detection of outliers in gas emissions from urban areas using functional data analysis," *J. Hazard. Mater.*, vol. 186, no. 1, pp. 144–149, Feb. 2011. https://doi.org/https://doi.org/10.1016/j.jhazmat.2010.10.091

[7] M. Tang, Z. Li, and G. Tian, "A Data-Driven-Based Wavelet Support Vector Approach for Passenger Flow Forecasting of the Metropolitan Hub," *IEEE Access*, vol. 7, pp. 7176-7183, Jan. 2019. https://ieeexplore.ieee.org/abstract/document/8600312

[8] Z. Jin-Ting, and X. Liang, "One-way ANOVA for functional data via globalizing the pointwise F-test," *Scand. Stat. Theory Appl.,* vol. 41, no. 1, pp. 51–71, Mar. 2014. https://doi.org/10.1111/sjos.12025

[9] A. Cuevas, M. Febrero, and R. Fraiman, "An anova test for functional data," *Comput. Stat. Data Anal.*, vol. 47, no. 1, pp. 111–122, Aug. 2004. https://doi.org/https://doi.org/10.1016/j.csda.2003.10.021

[10] J. O. Ramsay, and B. W. Silverman, *Functional Data Analysis*, 2nd ed. New York, NY, USA: Springer-Verlag New York, 2005. https://doi.org/10.1007/b98888

[11] C. G. Kaufman, and S. R. Sain, "Bayesian Functional ANOVA Modeling Using Gaussian Process Prior Distributions," *Bayesian Anal.*, vol. 5 no. 1, pp. 123–149, Mar. 2010. https://doi.org/10.1214/10-BA505

[12] Q. Shen, and J. J. Faraway, "An F test for linear models with functional responses," *Statistica Sinica,* vol. 14, pp. 1239–1257, 2004. https://api.semanticscholar.org/CorpusID:55106079

[13] P. Delicado, "Functional k-sample problem when data are density functions," *Comput. Stat.*, vol. 22, no. 3, pp. 391–410, Sep. 2007. https://doi.org/10.1007/s00180-007-0047-y

[14] M. Myllymäki, T. Mrkvička, P. Grabarnik, H. Seijo, and U. Hahn, "Global envelope tests for spatial processes," *J. R. Stat. Soc. Series B Stat. Methodol.,* vol. 79, no. 2, pp. 381–404, Mar. 2017. https://doi.org/10.1111/rssb.12172

[15] O. A. Vsevolozhskaya, M. C. Greenwood, and D. B. Holodov, "Pairwise comparison of treatment levels in functional analysis of variance with application to erythrocyte hemolysis," *Ann. Appl. Stat.*, vol. 8, pp. 905–925, Jun. 2014. https://api.semanticscholar.org/CorpusID:38476665

[16] A. Pini, S. Vantini, B. M. Colosimo, and M. Grasso, "Domain-selective functional analysis of variance for supervised statistical profile monitoring of signal data," *J. R. Stat. Soc. Ser. C Appl. Stat.*, vol. 67, no. 1, pp. 55–81, Jan. 2018. https://doi.org/10.1111/rssc.12218

[17] A. B. Kashlak, S. Myroshnychenko, and S. Spektor, "Analytic Permutation Testing for Functional Data ANOVA," *J. Comput. Graph. Stat.*, vol. 32, no. 1, pp. 294–303, May. 2023. https://doi.org/10.1080/10618600.2022.2069780

[18] M. Hollander, D. A. Wolfe, and E. Chicken, "The onw-Way Layout Introduction," in *Nonparametric Statistical Methods*, D. J. Balding *et al.,* Eds., Hoboken, New Jersey: John Wiley & Sons, 2013.

[19] D. Achlioptas, "Database-friendly random projections," in *Proceedings of the twentieth ACM SIGMOD-SIGACT-SIGART symposium on Principles of database systems,* New York, NY, USA, 2001. https://api.semanticscholar.org/CorpusID:2640788

[20] A. Nieto-Reyes, "Random Projections: Applications to Statistical Data Depth and Goodness of Fit Test," *BEIO Rev. Of. la Soc. Estadística e Investig. Oper.*, vol. 35, no. 1, pp. 7–22, Mar. 2019. https://www.seio.es/beio/BEIOVol35Num1.pdf#page=13

[21] J. A. Cuesta-Albertos, R. Fraiman, and T. Ransford, "Random projections and goodness-of-fit tests in infinite-dimensional spaces," *Bull. Brazilian Math. Soc.*, vol. 37, no. 4, pp. 477–501, Dec. 2006. https://doi.org/10.1007/s00574-006-0023-0

[22] R. Ihaka, R. Gentleman. *The R Project for Statistical Computing.* (V R.4.2.1 2022). Accessed: Apr.. 16, 2023. [Online]. Available: https://cran.r-project.org/bin/windows/base/old/4.2.1/

[23] J. Ramsay, G. Hooker, and S. Graves, *Functional Data Analysis with R and MATLAB.* New York, NY, USA: Springer New York, 2009. https://doi.org/10.1007/978-0-387-98185-7

[24] T. Pohlert, *The Pairwise Multiple Comparison of Mean Ranks Package (PMCMR) v4.4.* 2016. Accessed: Apr.16, 2023. [Online]. Available: http://cran.r-project.org/package=PMCMR

Notes

ACKNOWLEDGEMENT AND FUNDING :
The authors thank the Editor and reviewers for their constructive comments, which improved the article's presentation. Francisco J. Rodríguez-Cortés and Ramón Giraldo has been partially supported by Universidad Nacional de Colombia, HERMES projects, Grant/Award Number: 612113.

CONFLICTS OF INTEREST :
The authors declare no conflict of interest.

AUTHORS CONTRIBUTIONS :
Rafael Meléndez Surmay, **Ramón Giraldo Henao**, and **Francisco Rodríguez Cortes**, performed data processing, formal analysis, investigation, methodology, and original draft writing.

Información adicional

How to cite / Cómo citar: R. Meléndez Surmay, R. Giraldo Henao, and F. Rodríguez Cortes, "Kruskal-Wallis Test for Functional Data Based on Random Projections Generated from a Simulation of a Brownian Motion," *TecnoLógicas*, vol. 27, no. 59, e2986, Apr. 2024. https://doi.org/10.22430/22565337.2986

Enlace alternativo

https://revistas.itm.edu.co/index.php/tecnologicas/issue/view/135 (html)

Rafael Meléndez Surmay, Ramón Giraldo Henao,
Francisco Rodríguez Cortes

Kruskal-Wallis Test for Functional Data Based on Random Projections Generated from a Simulation of a Brownian Motion

Prueba de Kruskal-Wallis para datos funcionales basada en proyecciones aleatorias generadas a partir de una simulación de un movimiento browniano

TecnoLógicas
vol. 27, núm. 59, p. 1 - 11, 2024
Instituto Tecnológico Metropolitano, Colombia
tecnologicas@itm.edu.co

ISSN: 0123-7799 **/ ISSN-E:** 2256-5337

DOI: https://doi.org/10.22430/22565337.2986