Cumhuriyet Sci. J., 45(1) (2024) 168-174
DOI: https://doi.org/10.17776/csj.1371399

Cumhuriyet Science Journal

| csj.cumhuriyet.edu.tr | Founded: 2002 ISSN: 2587-2680 e-ISSN: 2587-246X Publisher: Sivas Cumhuriyet University

Functional Analysis of Variance: An Application to Stock Exchange

Selin Öğütcü [1,a,*], **Nuri Çelik** [1,b]

[1] *Department of Mathematics, Faculty of Science, Gebze Technical University, Kocaeli, Türkiye.*
**Corresponding author*

Research Article

History
Received: 05/10/2023
Accepted: 05/02/2024

ABSTRACT

The concept of "functional data" allows for the representation of data collected repeatedly over a period of time as a continuous function within a specific range on the time axis, rather than as discrete measurement points. Traditional statistical analysis has been adapted to accommodate functional data. This paper discusses the adaptation of one-way analysis of variance for functional data, covering parameter estimations and obtaining test statistics. As a numerical example, stock exchange values from various countries across different continents are used. The aim is to discern potential differences among countries based on these stock exchange values during the Covid-19 pandemic, utilizing one-way analysis of variance for functional data.

Keywords: Functional data, One-way ANOVA, Covid.

[a] *selinogutcu8@gmail.com* *https://orcid.org/0000-0003-4568-9254*
[b] *nuricelik@gtu.edu.tr* *0000-0002-4234-2389*

Introduction

Functional data analysis (FDA) is a statistical technique employed to infer the features of an underlying function from data collected at multiple time points within the same observation. It facilitates statistical analysis and enables the comparison of entire functions with one another. Furthermore, FDA allows for parameter estimation during the analysis phase, noise reduction through curve smoothing, and the utilization of data collected at different times. The emergence of new technology over the years has underscored the necessity for FDA, which enables the modeling of data collected at various times, eliminating the need for evaluating observations simultaneously, as required in classical repeated measurement analyses.

When data is continuously gathered over a duration and represented as a continuous function on a specific interval on the time axis, rather than as discrete points, it is referred to as "Functional Data." The core concept of functional data analysis is to identify and assess statistical modeling or estimation techniques suitable for these functions. With the aid of advancing technology, computer programs have simplified the execution of this approach, leading to its increased popularity among researchers for analyzing repeated measurements. While several methods for longitudinal data analysis (LDA) exist in the literature for this type of data, functional data analysis holds an advantage over other methods due to its versatility, making it a preferred choice for analyzing repeated measurements.

Although Ramsay [1] and Ramsay & Dalzell [2] coined the term "functional data analysis," the roots of this discipline can be traced back to the works of Grenander [3] and Rao [4], which have a long history [5]. Going even further back, the evolution of this technique, with broad applications across various fields, can be dated to the 1800s. During this time, Gauss and the French mathematician Legendre endeavored to model and forecast the trajectory of a comet, which follows a curve [6].

Functional ANOVA (Analysis of Variance) is a statistical method employed to assess whether significant differences exist between the mean functions of two or more groups in a functional data set. It serves as an extension of the traditional ANOVA method, typically applied to non-functional data sets. In functional ANOVA, each observation in the data set is a function rather than a scalar value. The method decomposes the total variation in the data into various sources, including between-group and within-group variations. Additionally, the method can test the significance of interactions between different factors, such as time and group membership, on the variation in functional data. With applications in economics, biology, and engineering, Functional ANOVA proves to be a versatile tool in numerous fields.

Functional ANOVA is among the most preferred analysis methods, akin to traditional statistics. It is employed to scrutinize the variation in a set of functional data across various groups or conditions. This statistical approach proves particularly advantageous when data is collected over time and represented as a continuous function within a specific interval on the time axis. The primary objective of functional ANOVA is to identify and assess the sources of variability in the data, determining whether significant differences exist between groups or conditions.

This paper provides a detailed explanation of functional ANOVA, including the calculation of test statistics using the pointwise testing method. The choice of the pointwise testing method is based on its resemblance to traditional statistics and ease of calculation. In the application section, the objective is to investigate the impact on stock markets in different regions worldwide during the Covid-19 pandemic and to determine whether statistically significant differences exist among them. For detailed information, refer to [7].

Functional Analysis of Variance

The problem of one-way ANOVA for functional data can be defined as follows:

Consider k independent samples denoted by

$$y_{i1}(t), \dots, y_{in_i}(t)\,, i = 1, \dots, k \tag{1}$$

Certain k samples meet the criteria of

$$y_{ij}(t) = \eta_i(t) + v_{ij}(t)\,, v_{ij}(t) \overset{i.i.d.}{\sim} SP(0,\gamma)\,, j = 1,2,\dots,n_i\,; i = 1,2,\dots,k\,, \tag{2}$$

where $\eta_1(t), \eta_2(t), \dots, \eta_k(t)$ represent the mean functions of the unidentified groups for the k samples, $v_{ij}(t)$, where $j = 1,2,\dots,n_i\,; i = 1,2,\dots,k$ denote the subject-effect functions, and $\gamma(s,t)$ denotes the shared covariance function for different times for s and t and SP can be described as a stochastic process. Our objective is to perform a one-way ANOVA test on the testing problem of

$$H_0 : \eta_1(t) \equiv \eta_2(t) \equiv \cdots \equiv \eta_k(t)\,, \qquad t \in \mathcal{T}, \tag{3}$$

where $\mathcal{T}$ is a time period of interest specified as $[a,b]$ with $-\infty < a < b < \infty$, as is often the case. The one-way ANOVA problem mentioned above is recognized as the k-sample problem for functional data, which expands on the two-sample problem for functional data presented in the preceding section.

Frequently, the objective of the one-way ANOVA testing problem (3) is to verify whether the effect of a treatment or factor is statistically significant. This treatment or factor is commonly utilized to divide the individual functions into various groups, categories, or samples. If the treatment or factor has a significant impact on the functional data, then the one-way ANOVA problem (3) will demonstrate statistical significance.

Define $\eta_i(t)$ as $\eta(t)$ added to $\alpha_i(t)$ ($\eta_i(t) = \eta(t) + \alpha_i(t)$) for all values of i ranging from 1 to k ($i = 1,2,\dots,k$), where $\eta(t)$ represents the mean function across all k samples, and $\alpha_i(t)$ denotes the main-effect function for each value of i. Subsequently, we can express the model (2) as a standard one-way ANOVA model for functional data by representing it as

$$y_{ij}(t) = \eta(t) + \alpha_i(t) + v_{ij}(t)\,, j = 1,2,\dots,n_i\,; i = 1,2,\dots,k. \tag{4}$$

By using this formulation, we can represent the null hypothesis (3) in an equivalent manner as a

$$\alpha_1(t) \equiv \alpha_2(t) \equiv \cdots \equiv \alpha_k(t) \equiv 0\,, t \in \mathcal{T} \tag{5}$$

In other words, the goal is to assess whether the main-effect functions are identical and have a value of zero. Using the k samples (1), we can obtain unbiased estimates for the group mean functions $\eta_i(t)$, where i ranges from 1 to k ($i = 1,2,\dots,k$), as well as the shared covariance function $\gamma(s,t)$, which can be represented as

$$\hat{\eta}_i(t) = \bar{y}_{i.}(t) = n_i^{-1} \sum_{j=1}^{n_i} y_{ij}(t)\,, i = 1,2,\dots,k\,, j = 1,2,\dots,n_i$$
$$\hat{\gamma}(s,t) = (n-k)^{-1} \sum_{i=1}^{k} \sum_{j=1}^{n_i} [y_{ij}(s) - \bar{y}_{i.}(s)][y_{ij}(t) - \bar{y}_{i.}(t)] \tag{6}$$

It is worth noting that, in this section, $n = \sum_{i=1}^{k} n_i$ refers to the total sample size. The approximated covariance function $\hat{\gamma}(s,t)$ is commonly known as the pooled sample covariance function. Keep in mind that $\hat{\eta}_i(t)$, where i takes values from 1 to k ($i = 1,2,\dots,k$), are independent, and

$$E\hat{\eta}_i(t) = \eta_i(t)\,, \qquad cov[\hat{\eta}_i(s), \hat{\eta}_i(t)] = \frac{\hat{\gamma}(s,t)}{n_i}, \quad i = 1,2,\dots,k. \tag{7}$$

Set $\hat{\eta}(t) = [\hat{\eta}_1(t), \hat{\eta}_2(t), \dots, \hat{\eta}_k(t)]^T$.

This estimator of $\eta(t)$ is impartial. As a result, we obtain $E\hat{\eta}(t) = \eta(t)$ and $Cov[\hat{\eta}(s), \hat{\eta}(t)] = \gamma(s,t)D$, where D is a diagonal matrix with diagonal entries $1/n_i$, $i = 1,2,\cdots,k$, and $D = diag(1/n_1, 1/n_2, \cdots, 1/n_k)$. This implies that the stochastic process $\hat{\eta}(t)$ follows the $SP_k(\eta, \gamma D)$ distribution, where $SP_k(\eta, \Gamma)$ represents a k-dimensional stochastic process with the vector of mean functions $\eta(t)$ and the matrix of covariance functions $\Gamma(s,t)$.

In order to examine techniques for carrying out main-effect, post hoc, it is necessary to explore the characteristics of $\hat{\eta}(t)$ and $\hat{\gamma}(s,t)$ in different scenarios. In pursuit of this objective, we outline the ensuing assumptions:

i. The k samples (1) are with $\eta_1(t), \eta_2(t), \dots, \eta_k(t) \in \mathcal{L}^2(\mathcal{T})$ and $tr(\gamma) < \infty$.
ii. The k samples (1) are normal distribution.
iii. As $n \to \infty$, the k sample sizes satisfy $n_i/n \to \tau_i$, $i = 1,2,\cdots,k$ such that $\tau_1, \tau_2, \dots, \tau_k \in (0,1)$.
iv. The subject-effect functions $v_{ij}(t) = y_{ij}(t) - \eta_i(t), j = 1,2,\dots,n_i\,; i = 1,2,\dots,k$ are identically and independently distributed.
v. The subject-effect function $v_{11}(t)$ satisfies $E \parallel v_{11} \parallel^4 < \infty$.
vi. The maximum variance $\rho = max_{t \in \mathcal{T}}\, \gamma(t,t) < \infty$.

vii. The expectation $E[v_{11}^2(s)v_{11}^2(t)]$ is uniformly bounded.

Theorem 2.1
Under above assumptions (i) and (ii), we have

$$D^{-\frac{1}{2}}[\hat{\eta}(t)-\eta(t)] \sim NP_k(0,\gamma I_k), \quad and \quad (n-k)\hat{\gamma}(s,t) \sim WP(n-k,\gamma). \tag{8}$$

The k-dimensional normality process (NP) of $\hat{\eta}(t)$ and the Wishart process (WP) of $(n-k)\hat{\gamma}(s,t)$ are both demonstrated by Theorem 2.1 assuming the normality assumption (ii). It is evident that Theorem 2.1 is fundamental in creating various tests for (3) if the k samples (1) are normal distribution. It's worth noting that even if the sample sizes $\eta_1,\eta_2,\dots,\eta_k$ are finite, Theorem 2.1 remains valid under the normality assumption (ii).

When conducting main-effect, post hoc, there's no requirement to identify the main-effect functions $\alpha_i(t)$, where $i = 1,2,\dots,k$, as stated in equation (4). In reality, these functions are not identifiable unless certain restrictions are enforced. Suppose we do aim to estimate these main-effect functions; in that case, the most widely employed constraint to ensure their identifiability is

$$\sum_{i=1}^{k} n_i\alpha_i(t) = 0\ , \tag{9}$$

which relates to the k sample sizes. Within this constraint, it's straightforward to demonstrate that

$$\hat{\alpha}_i(t) = \bar{y}_{i.}(t) - \bar{y}_{..}(t), i = 1,2,\dots,k, \tag{10}$$

provides unbiased estimators of the main-effect functions, where

$$\bar{y}_{..}(t) = n^{-1}\sum_{i=1}^{k}\sum_{j=1}^{n_i} y_{ij}(t) = n^{-1}\sum_{i=1}^{k} n_i\,\bar{y}_{i.}(t) \tag{11}$$

is the typical sample grand mean function. With constraint (12) enforced, the sample mean function $\bar{y}_{..}(t)$ becomes an unbiased estimator of the grand mean function $\eta(t)$ defined in equation (4).

$$SSH_n(t) = \sum_{i=1}^{k} n_i[\bar{y}_{i.}(t) - \bar{y}_{..}(t)]^2, \text{ and}$$
$$SSE_n(t) = \sum_{i=1}^{k}\sum_{j=1}^{n_i}[y_{ij}(t) - \bar{y}_{i.}(t)]^2, \tag{12}$$

refer to the pointwise between-subject and within-subject variations, respectively. $\bar{y}_{i.}(t)$, where $i = 1,2,\dots,k$ as defined in equation (6), and the sample grand mean function $\bar{y}_{..}(t)$ as defined in equation (11). When constraint (12) is in place, it's clear that

$$SSH_n(t) = \sum_{i=1}^{k} n_i\,\hat{\alpha}_i^{\,2}(t), \tag{13}$$

where the estimated main-effect functions $\hat{\alpha}_i(t), i = 1,2,\dots,k$ are given by equation (13), is easy to determine. We observe that when the null hypothesis (3) is true, $SSH_n(t)$ should be minimal, and when it is not valid, it should be substantial. It can be seen from (6) that

$$SSE_n(t) = (n-k)\hat{\gamma}(t,t). \tag{14}$$

The pointwise F-test, pointwise χ^2-test, and pointwise bootstrap test are all under consideration. Ramsay and Silverman (2005) introduced the pointwise F-test for (3) to extend the classical F-test into the domain of functional data analysis. At each $t \in \mathcal{T}$, the pointwise F-test is implemented for (3) using the pointwise F statistic:

$$F_n(t) = \frac{SSH_n(t)/(k-1)}{SSE_n(t)/(n-k)} \tag{15}$$

It is readily apparent from the classical linear model theory that, assuming the null hypothesis (3),

$$F_n(t) \sim F_{k-1,n-k}\ , t \in \mathcal{T} \tag{16}$$

is obtained when the k samples (1) are normal distribution.

Application to Stock Exchange

The news of the pandemic caused world stock markets to open with record declines, indicating that the impact of heightened volatility and negative perceptions towards the stock market may exceed expectations. Additionally, the economic repercussions of the coronavirus epidemic have also extended to the stock markets of affected nations. Presently, it is evident that the economic effects of the epidemic are multi-faceted [8]. The apprehension caused by the coronavirus outbreak rapidly spread to global financial markets. The Covid-19 aftermath, which was initially disregarded, has become a significant concern as it continues to spread swiftly beyond China [9].

COVID-19's appearance and subsequent proliferation to over 150 countries within two months have led to the cessation of commercial and economic operations, prompting concerns that it is not solely a health crisis but also harbors the potential for significant and far-reaching consequences for the global economy in the future [10]. The response of the stock markets has elicited significant apprehension, given that the world is presently grappling with the most severe economic downturn since the Great Depression [11]. Furthermore, Ashraf's study [12] discovered a correlation between the number of COVID-19 cases reported in a country and a corresponding decrease in stock market returns. Behavioral finance suggests that the rise in COVID-19 cases may have a substantial impact on global equity markets, making it challenging for individual investors to make informed investment decisions.

The main purpose of this application is to investigate how the stock market movements in different parts of the world move in the Covid-19 pandemic process, how they are affected and whether there is a statistically significant difference between them. For this reason, stock market opening data from various countries between 01/03/2020 – 01/03/2022 were taken from https://tr.investing.com/indices/world-indices.

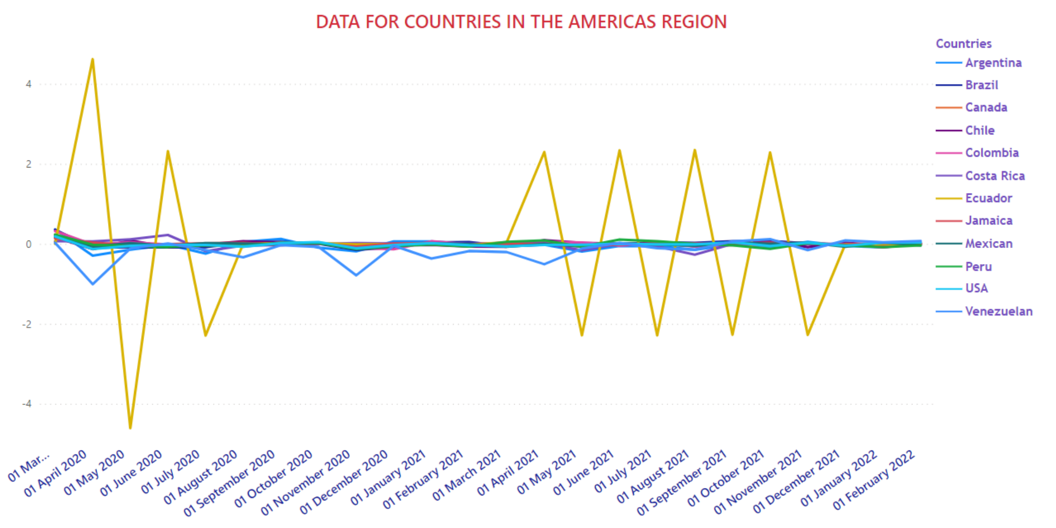

Figure 1. Data for countries in the Americas region.

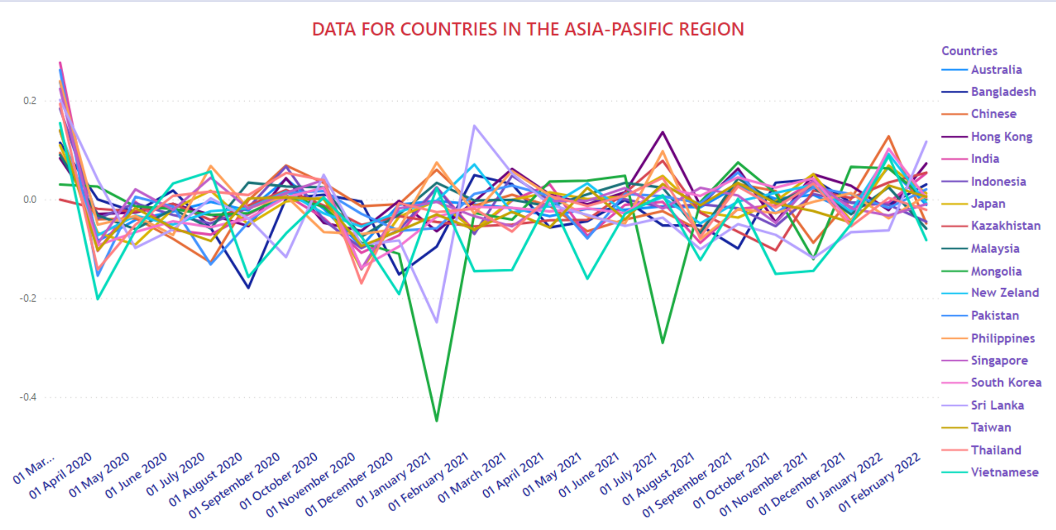

Figure 2. Data for countries in the Asia-Pacific region.

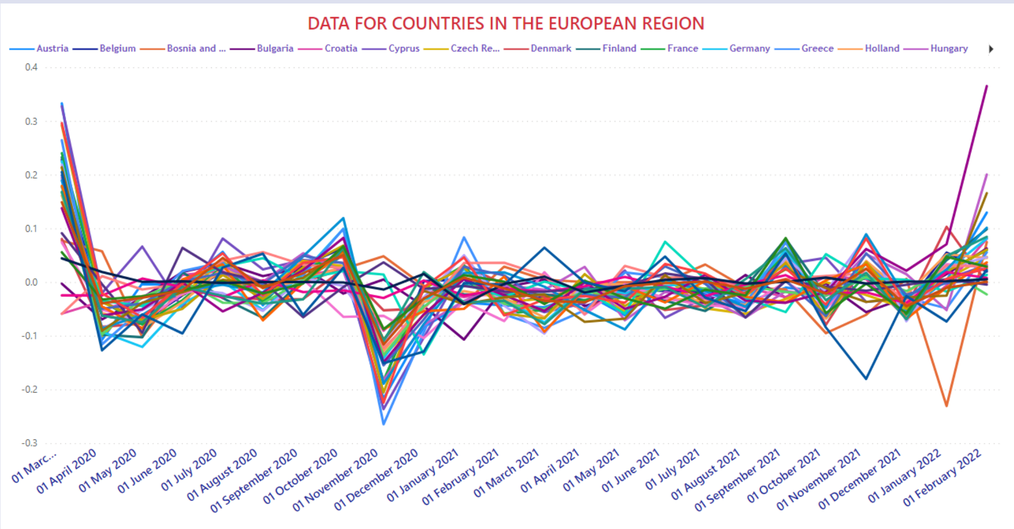

Figure 3. Data for countries in the European region.

By dividing the world geography into three as America, Asia-Pacific and Europe, 12 countries such as Argentina, Brazil, Ecuador, Colombia, Peru, Chile, Venezuela, USA, Canada, Mexico, Jamaica, Costa Rica were taken from the Americas region. In the Asia -Pacific region, 19 countries were included in Australia, Bangladesh, China, Indonesia, Philippines, South Korea, India, Hong Kong, Japan, Kazakhstan, Malaysia, Mongolia, Pakistan, Singapore, Sri Lanka, Thailand, Taiwan, Vietnam, New Zealand. Similarly, a total of 33 countries including Germany, Austria, Belgium, the United Kingdom, Bosnia and Herzegovina, Bulgaria, the Czech Republic, Denmark, Finland, France, Croatia, the Netherlands, Ireland, Spain, Sweden, Switzerland, Italy, Iceland, Montenegro, Cyprus, Hungary, Malta, Norway, Poland, Portugal, Romania, Russia, Serbia, Slovakia, Slovenia, Turkey, Ukraine, and Greece are included from the European region. The African region has not been taken into account due to its already underdeveloped status and its economic situation not matching that of the other three regions.

Between 01/03/2020 – 01/03/2022, the data were prepared by applying the $log(P_t/P_{t-1})$ transformation in order to smooth the stock market opening data in different regions and different countries. Figure 1, Figure 2 And Figure 3 show graphs of the transformed data in all three regions.

Before moving on to functional ANOVA, data were tested for normality. The results according to the Shapiro-Wilk test of normality, test statistic is 0.944 and p-value is 0.197 and it has been determined that each data set is normally distributed. $(p - value > 0.05)$

For the functional analysis of variance, firstly the average functions of all three regions were calculated. Figure 4 shows the mean functions.

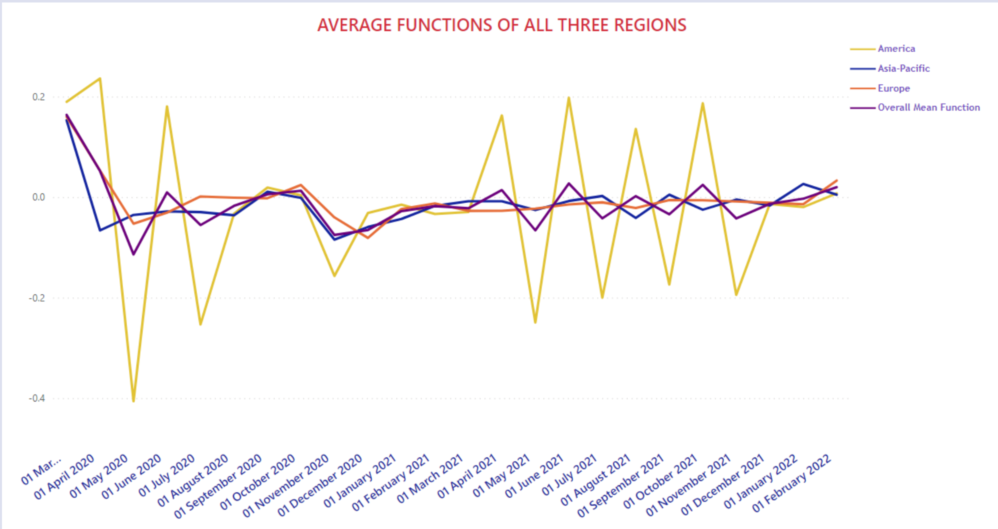

Figure 4. Average functions of all three regions.

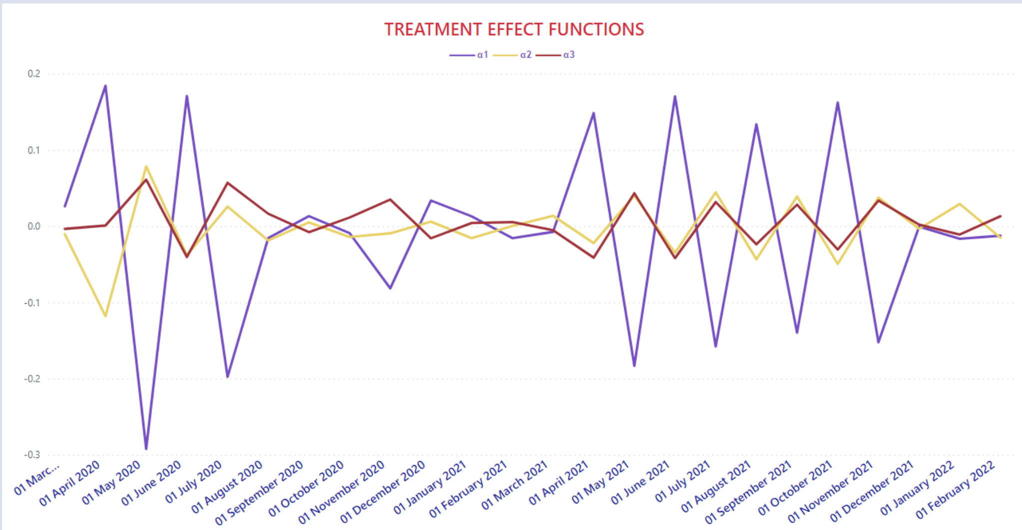

Figure 5. Treatment effects.

Three treatment effect functions were obtained by using these mean functions. Figure 5 shows the treatment effects mentioned.

Finally, using the $SSH_n(t)$ and $SSE_n(t)$ values, the test statistics ($F_n(t)$ values) were calculated as in Figure 6.

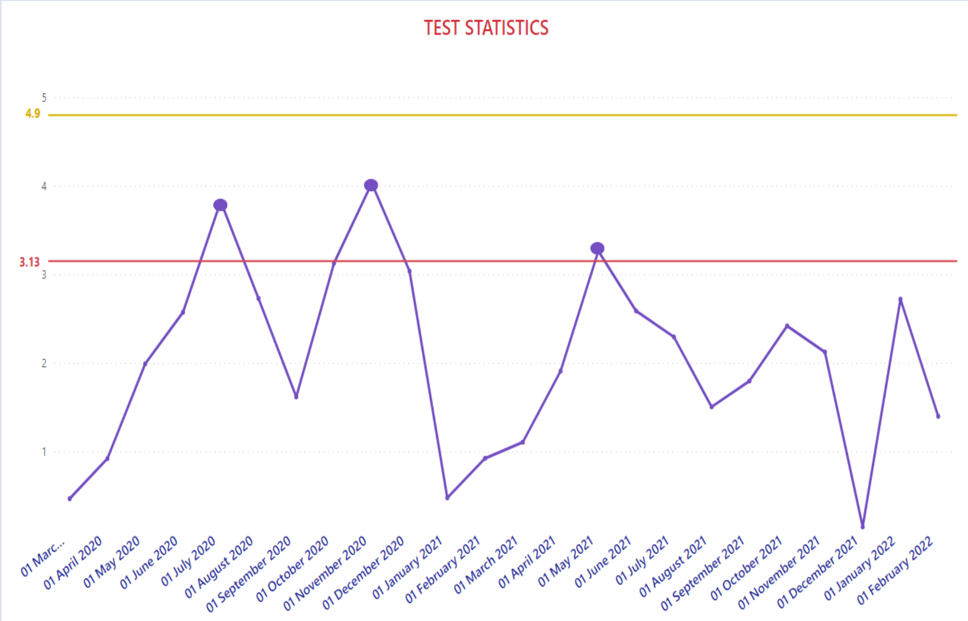

Figure 6. Test statistics ($F_n(t)$ values).

After the critical table value was determined as $F_{2,61} = 3.13$ for $\alpha = 0.05$ and $F_{2,61} = 4.9$ for $\alpha = 0.01$, the analysis result was started.

For $\alpha = 0.01$, stock market data during the Covid-19 process do not differ between continents.

For $\alpha = 0.05$, although we observe that it exceeds $F_{2,61} = 3.13$ on 3 separate dates, 01/07/2020 , 01/11/2020 and 01/05/2021, in the Covid-19 process, stock market data does not generally differ between continents.

According to the Pointwise test, for alpha = 0.05, Table 1 was prepared in order to determine from which regions the difference originated on the said dates, and it was determined that the difference originated from the Americas region. $(p - value < 0.05)$

Table 1. The p-values of the t-tests.

	July/2020	November/2020	May/2021
		America	
Asia-Pacific	0.009	0.014	0.011
Europe	0.002	0.008	0.012

Conclusion

Functional analysis of variance (ANOVA) is a statistical technique used in functional data analysis to investigate the variability in a set of functional data across different groups or conditions. This approach is particularly useful when the data is collected over time and expressed as a continuous function along a specific interval on the time axis. The goal of functional ANOVA is to identify and evaluate the sources of variability in the data and to determine whether there are significant differences between groups or conditions. This method involves decomposing the functional data into a sum of orthogonal basis functions and then testing the differences between groups or conditions based on the coefficients of these basis functions. Functional ANOVA is a powerful tool for analyzing functional data and can provide valuable insights into the underlying patterns and trends in the data.

In the application part of the paper, it was investigated how the stock market movements in different parts of the world moved during the Covid-19 pandemic process, how they were affected and whether there was a statistically significant difference between them. Adhering to this purpose, stock market opening data recorded by https://tr.investing.com/indices/world-indices from various countries between 01/03/2020 and 01/03/2022 were used. By dividing the world geography into three as America, Asia-Pacific and Europe, 12 countries were taken from the Americas region and from the Asia -Pacific region 19 countries and a total of 33 countries from the European region. In addition, the reason for excluding the African region is that it is already considered underdeveloped and its economic situation does not align with the other three regions.

After the necessary statistical analyzes and examinations were completed, the analysis was concluded. For α = 0.05, it was observed that there was a significant difference on 3 different dates, 01/07/2020, 01/11/2020 and 01/05/2021, and according to the Pointwise test, it was determined that the difference originated from the Americas region. However, since 3 months is very short for 2 years of data, we conclude that it was concluded that the stock market data in the Covid-19 process did not differ between continents.

Conflicts of interest

There are no conflicts of interest in this work.

References

[1] Ramsay J., When the data are functions*, Psychometrika*, 47 (4) (1982) 379-396.

[2] Ramsay J., Dalzell C., Some tools for functional data analysis, *Journal of the Royal Statistical Society, Series B (Statistical Methodology)*, (1991) 539-572.

[3] Grenander U., Stochastic processes and statistical inference, *Ark. Mat.*, 1 (3) (1950) 195-277.

[4] Rao C.R., Some statistical methods for comparison of growth curves, Biometrics, 14 (1) (1958) 1-17.

[5] Wang S., Sobel A., Fridlind A., Feng Z., Comstock J.M., Minnis P., and Nordeen M., Simulations of cloud-radiation interaction using large-scale forcing derived from the CINDY/DYNAMO northern sounding array, *J. Adv. Model. Earth Syst.*, 7 (3) (2015) 1472-1498.

[6] Yaraee K., Functional Data Analysis with Application to MS and Cervical Vertebrae Data, Master thesis, University of Alberta, 2011.

[7] Ogutcu S., Functional Data Analysis and An Application to Analysis of Variance, Mater thesis, Gebze Technical University, 2023.

[8] Demirdöğen O., Yorulmaz R., Kovid-19 Salgınının Dünya Ekonomilerine Etkileri, Ortadoğu Araştırmaları Merkezi, Ankara, (2020).

[9] Albulescu C., Coronavirus and Financial Volatility: 40 Days of Fasting and Fear, *Quantitative Finance*, (2020).

[10] Khan S., Siddique R., Li H., Ali A., Shereen M.A., Bashir N., Xue M., Impact of coronavirus outbreak on psychological health, *J Glob Health*, 10 (2020) 1-16.

[11] IMF Blog, The Great Lockdown: Worst Economic Downturn Since the Great Depression, NOS. Available at: https://www.imf.org/en/Blogs/Articles/2020/04/14/blog-weo-the-great-lockdown-worst-economic-downturn-since-the-great-depression. Retrieved March 14, 2023.

[12] Ashraf B.N., Stock Markets Reaction To Covid-19: Cases Or Fatalities?, *Research in International Business and Finance*, (2020).