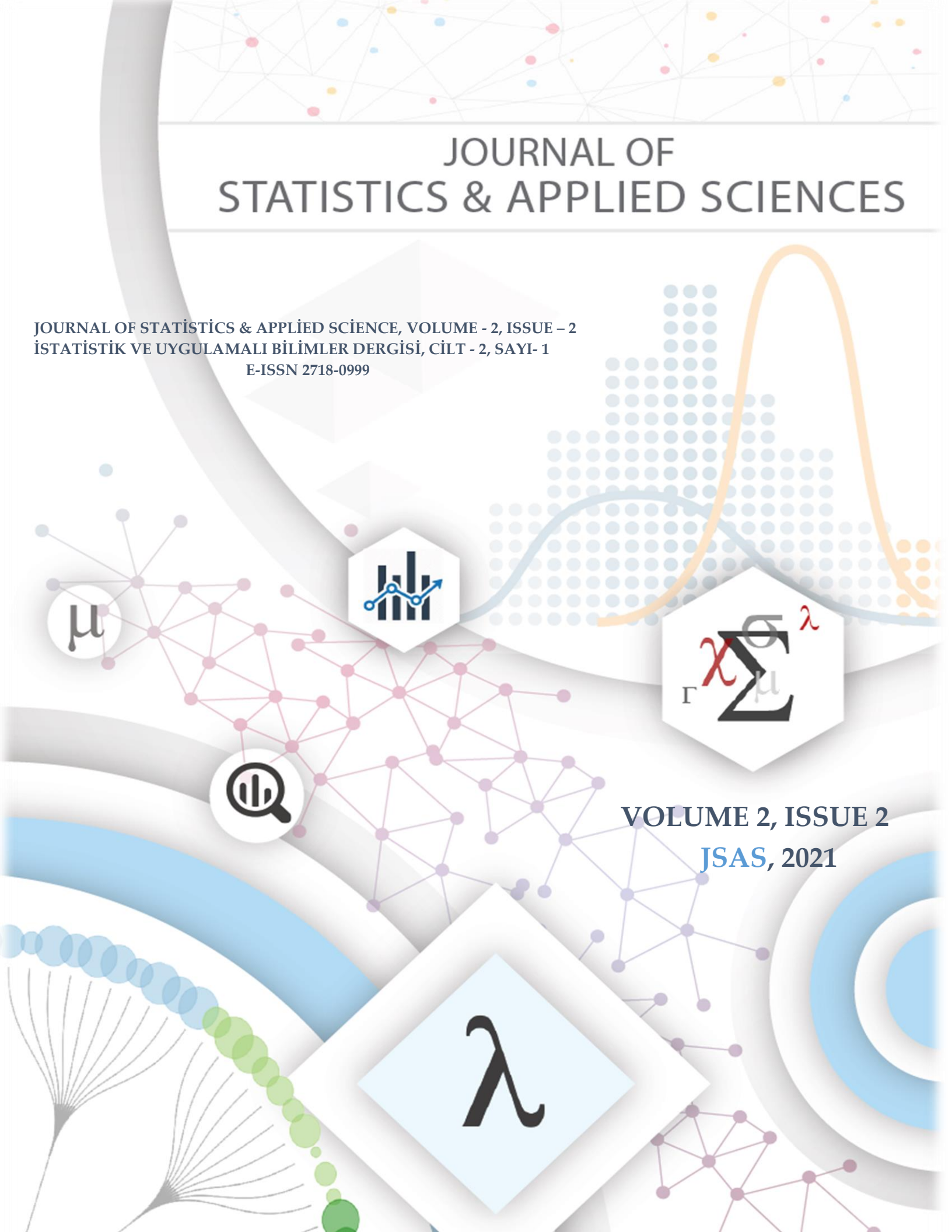
JOURNAL OF STATISTICS & APPLIED SCIENCES

JOURNAL OF STATİSTİCS & APPLİED SCİENCE, VOLUME - 2, ISSUE – 2
İSTATİSTİK VE UYGULAMALI BİLİMLER DERGİSİ, CİLT - 2, SAYI- 1
E-ISSN 2718-0999



VOLUME 2, ISSUE 2
JSAS, 2021

Journal of Statistics & Applied Science, Volume 2, Issue -2
ISSN 2718-0999

-JOURNAL OF STATİSTİCS & APPLİED SCİENCE-
-İSTATİSTİK VE UYGULAMALI BİLİMLER DERGİSİ-

Aralık 2021

Volume: 2 Issue: 2

Cilt: 2 Sayı: 2

Yayın Türü: 6 Aylık, Uluslararası, Hakemli

e-ISSN: 2718-0999

Dizinleme Bilgileri/ Abstracting and Indexing Services

İletişim / Contact

E-posta: jsas.journal@gmail.com

https://dergipark.org.tr/tr/pub/jsas

Journal of Statistics & Applied Science, Volume 2, Issue -2
ISSN 2718-0999

BİLİM VE DANIŞMA KURULU

EDİTÖR KURULU

İletişim E-posta: jsas.journal@gmail.com : https://dergipark.org.tr/tr/pub/jsas

Journal of Statistics & Applied Science, Volume -1 Issue -2
ISSN 2718-0999

İÇİNDEKİLER

Journal of Statistics & Applied Science, Volume -2 Issue -2
ISSN 2718-0999

Article

Received: date: 14.09.2021
Accepted: date: 24.10.2021
Published: date: 28.11.2021

Visualizations and Compositional Data Analysis for the European Time-Use

Cenk İÇÖZ[1]

[1]Eskişehir Teknik Üniversitesi, Fen Fakültesi, İstatistik Bölümü, Eskişehir Turkey, cicoz@eskisehir.edu.tr
Orcid: 0000-0002-0219-1088[1]
* Correspondence Author, e-mail: cicoz@eskisehir.edu.tr

Abstract: Time-use data is a specific data that is of interest to many researchers such as sociologists, psychologists, statisticians etc. It is multivariate by nature and can be seen as a subtype of multivariate data called compositional data. Thus, time-use data can also be analyzed and visualized by compositional data analysis techniques. This is rarely mentioned in literature; usually analyses and visualizations of time-use data are shown in multivariate real space. However, this may lead to false interpretation of the data and even some multivariate statistical analysis cannot be directly applied before a transformation is made. The main contribution of this study is to show how an analysis might be used for time-use data and what benefits might be gained from this new setting of the data set as compositional parts. Indeed, the results show that there are some differences between traditional techniques and compositional data analysis.

Keywords: Time-use data, Compositional data, Log-ratio analysis, Clustering Analysis, Ternary Diagrams

1. Introduction

Recent advances in computing and visualization software help data scientists working in several fields examine the data easily. Statisticians name the visual interpretation of the data as exploratory data analysis (EDA), which is pioneered by John W. Tukey [1] in his seminal book. EDA is the primary analysis for many statisticians to visually examine data first-hand before testing it. Using powerful statistical software, implementation of interactive visualizations, and statistical analyses are not bothersome anymore. Interactive visualizations are a handy tool and can give the researcher or data scientist especially a first impression. Interactive graphics on a computer screen allow a person to select individual observation values and the relationships between different variables. The adaptation of modern visualizations and analyses to time-use data may enlighten the patterns behind the time-use data and enable the scientist to understand what might be happening behind the curtain.

Time-use surveys have been conducted by the national statistics offices of many European countries. These surveys, which are designed to get information about several activities of people's daily life, are a great data source for investigating the differences of time-use according to gender and also the differences among countries. Time-use data that consist of several variables of activities of time-use for a sample of individuals chosen from different countries are naturally a form of multivariate data. Additionally, time-use data set are defined so that each row of the data adds up to 100% for the activities investigated. These type of data sets can be analyzed in a different space by using compositional data analysis techniques. In the literature, this property of time-use data sets is rarely mentioned. Compositional data has a correlation structure and other drawbacks for classic multivariate statistical analysis. Therefore, before applying multivariate analysis to compositional data some sort of transformations might be needed. Furthermore, visualizations like ternary diagrams, which are often used in the literature for other data types, may be used to subject time-use data to visual inference.

https://doi.org/10.52693/jsas.994860

The literature will be given below with the property of similar data use and the similar statistical analysis use with the aim of visualizing the data in lower dimension and clustering the objects (countries). Gálvez-Muñoz et. al. [2] examined Harmonised Time-Use Surveys (HETUS) from the perspective of gender inequality. They formed a difference data of gender by dividing women's time-use by men's time-use. The attractive property of this data is that it can be a subject for lower dimensionality visualization of gender inequality. By only looking at the ratio data one can make inferences about gender inequality. Higher ratios refer to higher differences for the concerned attribute. The target people of this study are in the age of 20-74 and under employment. Hierarchical clustering was performed using Euclidean distances as a proximity matrix and formed a grouping of countries by using this difference data. Furthermore, group mean differences were tested with the employment of analysis of variance. The study aimed to find out whether unpaid work is the main reason of gender inequality in all countries or not. Finally, they realized that Eastern European countries notably differ in clustering. Furthermore, in this study the differences among countries were not taken into account within each gender.

Moreno-Colom [3] examines gender segregation in domestic work. In this study, gender's role in contrasting the influence of welfare regimes and employment status on daily life organization is compared. First, they made country-wise comparisons of time-use patterns of European countries. Afterwards, they presented a special case about the effect of employment status on distribution of housework in Spain. One interesting result of the study is that daily task completion restricts equal share of housework between genders.

Coffe [4] examines the time-use of the Members of Parliament (MPs) of New Zealand. They claim the study of her is an innovative one, which is the first study ever to investigate the MPs' time-use. They asks to MPs both for completion of a time diary and a questionnaire. Using correspondence analysis as a main tool for cross tabular data, they find interesting results like differences according to gender and seniority between related activities of MP's and make inferences of MPs' time-use. For instance, Female MPs tend to spend more time on communication, meetings, travelling, and research and reading. On the other hand, male MPs decide to participate in social activities as well as attending House sessions.

Robinson and Gershuny [5] applied a multidimensional scaling (MDS) to diary collections of Oxford MTUS (Multinational Time-Use Study) data archive and compared these mappings with the 1965 Multinational Time-Budget Study. MDS is a technique that spatially maps the data using a proximity matrix.

Wight et al. [6] examines American teenagers' time-use by using ATUS data with a focus of variables that may affect children's well-being. They use OLS (Ordinary Least Squares) regression or logistic regression to find out the relationship between parental and household characteristics and teenagers' use of time.

None of the aforementioned studies [2-6] include any information about the space that the time-use data belongs to and all the analyses are performed without taking the compositional nature of the data into consideration. There are some recent studies that employs compositional data analysis for time-use data. Dumuid et al.[7] use compositional data analysis from Wave 6 of the Longitudinal Study of Australian Children diary data to explore the relationships between daily time-use. They also mentioned the drawbacks of using compositional data in time-use analysis. Gupta et. al. [8] compared time spent sedentary and in physical activity between age groups and sexes. They also showed and emphasized the difference of compositional data analysis results from the results that obtained through standard analytical procedures.

The aim of this study is to create interactive and classical visualizations of time-use data in a restricted space and search for patterns in data by using several statistical analyses techniques such as clustering analyses and log-ratio analysis (LRA). In this study, the compositional data analysis setting of the time-use data is considered. The statistical techniques used in this study get the benefits of using compositional data analysis approach as compared to the conventional data analysis. The compositional version of the related methods like LRA and distance measure is employed in clustering analysis. In doing so, this new setting might reveal completely different results and inferences compared to other studies using the traditional techniques for the same data. For clustering analysis, dissimilar formation of groups is obtained through two distance measures in different measurement spaces. LRA is an analysis based on log-ratios so it will give us a chance to make pairwise comparison of the activities of

countries' time-use neither principal component analysis (PCA) nor correspondence analysis (CA) will give. It is the only method among which has the sub-compositional coherence property. The article will be organized as follows. The theory of the statistical analyses is explained briefly in the next section of the study. Then, in the data section, categories of activities and the information about structure and gathering of the data are given. In the results section, the results of the analyses and interesting visualizations are shown, and some interpretations are given. Finally, a discussion is given about the results and findings, why to use compositional data analysis with time-use data, and other key points of the article.

2. Materials and Methods

A composition is often defined as a vector of D positive components $\mathbf{x} = [x_1, x_2, \dots, x_D]$whose total is equal to a constant K. Compositional data are strictly non-negative data that have a constant sum and its components consist of only relative information. Generally, the sum of the data is not of interest to researchers. Examples can be given as percentages of workers in different sectors, portions of the chemical elements in a mineral, concentration of different cell types in a patient's blood, portions of species in an ecosystem or in a trap, concentration of nutrients in a beverage, portions of working time spent on different tasks, portions of types of failures, percentages of votes for political parties, sources of pollution in air or in a water source, expenditures of households to different spending item categories, etc [9,10,11].

There are some disadvantages of using conventional multivariate analysis techniques for compositional data. Therefore, some transformations have to be applied before implementing multivariate analysis. These disadvantages can be found in van den Boogaart & Tolosana-Delgado [11]. For instance, variance-covariance matrices are singular so that some of the statistical analysis like Hotelling's T-square test and linear discriminant analysis cannot be applied directly; components do not fit into normal distribution which is a key assumption for many statistical analyses; and lastly, the data is correlated because of the restricted total sum. One can easily write a component in terms of other *(D-1)* components.

A subset of data that doesn't contain at least one component of the original compositional data is called a subcomposition. Closure operation is expressing each component of a compositional data in proportions, by just dividing them with the constant sum, or in percentages, by dividing them with the constant sum multiplied 100. In this way, the total is also transformed into 1 for proportions and 100 for percentages. A subcomposition has subcompositional coherence property after closure operation is performed if the calculated values (statistics or ratios) do not change when a subcomposition is used instead of full compositional data. Hence, the planned analysis results will not change for the subcomposition. Greenacre & Primicerio [12] proved numerically that log-ratios have the property of subcompositional coherence.

Clustering analysis is the grouping of similar objects according to proximity matrices into clusters; the objects in the same cluster are homogenous, and objects in different clusters are heterogeneous. There are several clustering methods. The most commonly used ones in statistical applications are k- means algorithm and hierarchical clustering.

In hierarchical clustering, as the name suggests, clusters are formed by a hierarchy at each level of the hierarchy by merging or dividing clusters or objects at the next lower level. Hierarchical clustering can be divided into two basic paradigms: agglomerative (bottom-up) and divisive (top-down). The agglomerative one begins at the bottom with single observation as a cluster and at each level recursively merges a selected pair of clusters or objects into a single cluster. The pair chosen for merging includes the two groups with the smallest intergroup dissimilarity. At the lowest level, each cluster contains a single observation. At the highest level, the final clusters are composed of all elements. Divisive strategy can be considered as the opposite of the agglomerative strategy [13].

The problem in the hierarchical clustering is the merging after the first iteration: merging a cluster and a single object. The distance matrix needs to be upgraded according to several rules. In the single linkage method, the minimum distance between a cluster and a single object is considered, while an average distance is taken into account in the average linkage method. Last of all, the maximum distance between a single object and a cluster is taken into account in the complete linkage method.

The average linkage method will be employed in this study for clustering analysis by using an adaptation of a Euclidean distance matrix for compositional data. Therefore, the results for the

clustering analysis will be completely different in real multivariate space. The clustering will be conducted in Aitchison's sample space called simplex [12]. Simplex is given with the constant sum, K, and the non-negativity constraint as in equation (1):

$$\mathbb{S}^D := \left\{ \boldsymbol{x} = [x_1, x_2, \dots, x_D] : \; x_i > 0 \; ; \sum_{i=1}^{D} x_i = K \right\} \quad [1]$$

To perform the specific operations in the simplex, the data must be converted via a transformation called centered log-ratio (clr) transformation. This transformation can be thought of as reweighting the data according to its geometric mean *g(x)*, which is considered a general tendency measure used for ratio-scale data. The clr transformation is defined as in equation (2):

$$clr(\boldsymbol{x}) = \; log\left(\frac{x_i}{g(\boldsymbol{x})}\right)_{i=1,2,\dots,D} \quad [2]$$

The distance matrix will be organized between composition "ith" and composition "jth" through Aitchison distance given in equation (3). Furthermore, it can be easily interpreted as Euclidean distances of clr transformed data.

$$d(\boldsymbol{x}_i, \boldsymbol{x}_j) = \left[\sum_{k=1}^{D} \left(log\left(\frac{x_{ik}}{g(\boldsymbol{x}_i)}\right) - log\left(\frac{x_{jk}}{g(\boldsymbol{x}_j)}\right) \right)^2 \right]^{\frac{1}{2}} \quad [3]$$

$g(\mathbf{x}_i)$ in equation (3) is the geometric mean of the component ith. Martin-Fernandez et. al [15]compared the measures of difference for compositional data in use for hierarchical clustering and mentioned that the unit constraint has to be taken into account while performing hierarchical clustering on compositional data. The Aitchison distance has also got the subcompositional coherence property and other properties of compositional data according to the study.
Martin-Fernandez et al. [15] gave an important example of the difference between Aitchison distance and the Euclidean distance. The paper shows that a three-part (D=3), compositional data consist of only 4 compositions. These are given in equation (4):

$$\boldsymbol{x_1} = [0.1, 0.2, 0.7] \, , \boldsymbol{x_2} = [0.2, 0.1, 0.7], \boldsymbol{x_3} = [0.3, 0.4, 0.3], \qquad \boldsymbol{x_4} = [0.4, 0.3, 0.3] \quad [4]$$

Euclidean distances between the first two and last two compositions are equal according to the compositional data set given. On the other hand, their compositional distances are not equal, although they have the same amount of differences, ±0.1, in the first two components. This is the result of the compositional nature of the data that one part is dependent on the other parts. In the calculation of the distance between composition 1 and composition 2, the residual part is 0.7 while it is 0.3 between compositions 3 and 4. It can be interpreted that the first distance is produced over a residual of 0.7 that could produce a greater distance, while the second distance is produced over 0.3 residual. Therefore, the distance between the first two compositions is greater than the last two. Thus, the two dendrograms that are obtained through Euclidean and Aitchison distance will be presented to examine the differences between them. Different clustering occurs for the same data because of the reasons mentioned above.
Ternary diagrams are the fundamental graphs to visualize a 3-part compositional data. They are the analogy of the scatterplots for the three-dimensional display. All points will be located in a planar triangle in a ternary diagram with the edges at (1,0,0), (0,1,0), and (0,0,1). The interpretation of ternary diagrams is not complicated, and ternary diagrams can be used as a tool of visual analysis for compositional data. Through an edge of the triangle, the value in that edge can increase up to 1 for the associated component, whereas the other two components' values are approximately zero (because of the constant sum property). Furthermore, through the middle of the triangle, each components' value approximates to each other like (0.33, 0.33, 0.33) for a composition.

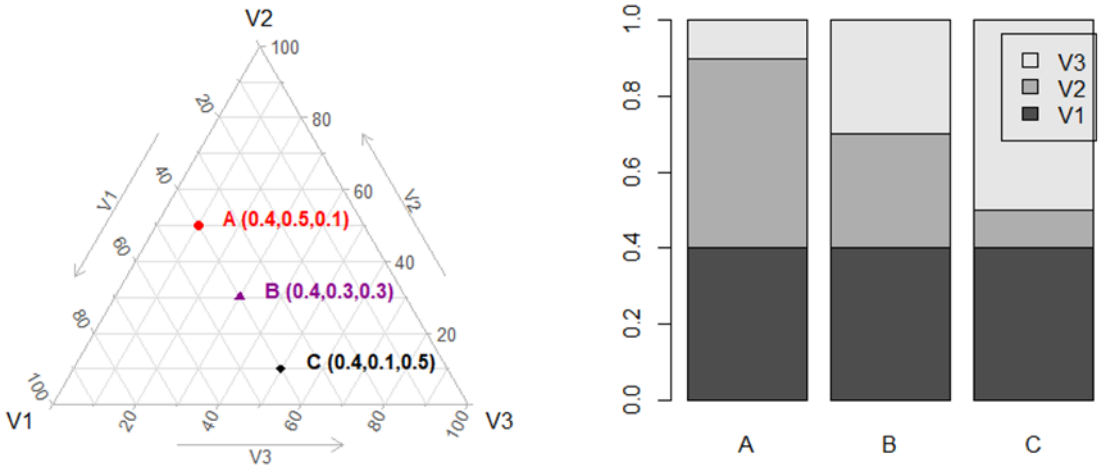


Figure 1. Ternary diagram example and bar graph

In Figure 1, some compositions are given for the interpretation of a ternary diagram in the left and it's counterpart as a stacked barplot in the right. A component's value remains the same along a line parallel to the side of a triangle, which is also located opposite to an edge. To illustrate this property, three compositions (A, B, and C) are given in Figure 1. All of them have equal component values in V1 component so they are in the same line that is parallel to the side of triangle located opposite of V1 edge. The arrows outside the triangle show the direction of the increase through the edge of the related component's value. If the composition gets closer to an edge, it gets a higher value on that component up to 1. Let's say that V1 is gainful work activity and V3 is domestic work activity. Compositions A, B, and C are in a line parallel to V2-V3 side and opposite to the V1 edge. Therefore, the values on gainful work component for these three composition stay the same with an amount of 0.4, and only the other two components' values change. Furthermore, when going down from composition A through composition C along the same line, the location of a component is getting closer to the V3 edge. Therefore, from composition A to composition C, domestic work's value is increasing from 0.1 to 0.5. Figure 1 is plotted by using the "ggtern" package in R.

In the bar graph, it may be inferred that the exact values from the scale of the axis on the other hand for a three part composition it could be known for sure in a ternary diagram with given scales. As the size of the variable number increase for a compositional data, the 3-part combinations also will increase and it will hard to interpret the data with ternary diagrams either.

In Figure 2, the stacked barplots according to genders with values given as labels in each individual bar representing each variable. By this way, it is easy to see the whole variable values together in a visualization without inferring the variable values from the scales of the graph.

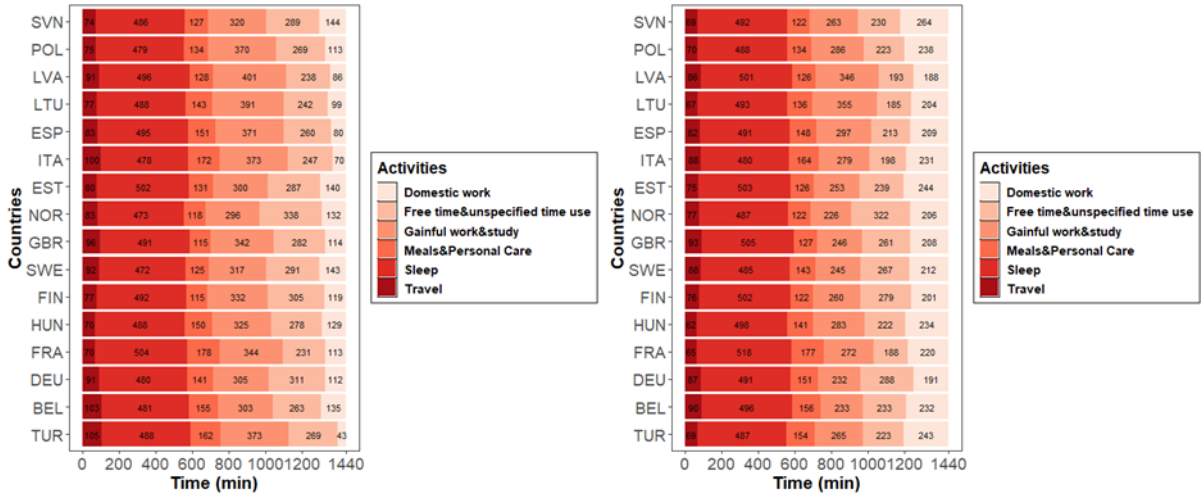


Figure 2. Stacked barplots according to genders with labels

The same kind of relationship between scatterplots and scatterplot matrices is also valid for ternary diagrams and ternary diagram matrices. For a compositional data that is composed of more than three components, ternary diagram matrices are employed in literature. The "ggtern" package provides a margin for the third component on the top edge, which can be both a fixed component itself or the geometric mean of the rest of components. Likewise scatterplot matrix ,the part in the right of the diagonal are the same except for the replacement of the two components located in the bottom side of the triangle.

Log-ratio analysis (LRA) is a method which is used to visualize compositional data in lower dimensions. When all the variables of the data are in interval scales, PCA is the preferred analysis. However, for the ratio scale data, multiplicative differences can be the subject of the analysis. Log-ratio transformation can be used to convert multiplicative differences to the additive differences and enable to application of PCA to ratio-scale data.

LRA is similar to PCA and CA that is used to visualize pairwise log-ratios in a biplot along with the sample points. Usually, standardization should be applied before PCA for data which is measured in different units, and the variables need to be reweighted through their standard deviations. When all variables are measured in the same scale as in our case, log-ratios are the perfect standardization. They have a special feature of comparability between both within variables and both within objects like in our case. Thus, no further standardization is necessary when log-ratio transformation is conducted once. The main difference between LRA and CA/PCA is that it has subcompositional coherence, because log-ratios remain the same for a subcomposition. Instead of biplot axes in PCA and CA link vectors are used for interpretation of biplot in LRA [10]. These link vectors demonstrate the outperforming objects in each side. When a composition is located close to a component, it means that it has got a higher ratio when compared to other objects in that link. For instance, if a composition has a higher ratio of component A over B, it is located closer to the component A side of the link vector as compared to other compositions, whereas it is far away from side B of the link vector and vice versa.

The algorithm of the weighted version of LRA is given in Greenacre [14] below:

1) Calculate the row and column margins, **r** and **c**, respectively, when $n = \sum_i \sum_j n_{ij}$ is the grand total of the compositional data matrix **N**: $\mathbf{r} = (1/n)\mathbf{N1}$, $\mathbf{c} = (1/n)\mathbf{N}^{\mathrm{T}}\mathbf{1}$
2) Logarithmic transformation of the elements of the matrix **N**: $\mathbf{L} = log(\mathbf{N})$
3) Weighted double-centering of L: $\mathbf{Y} = (\mathbf{I} - \mathbf{1r}^{\mathrm{T}})\mathbf{L}(\mathbf{I} - \mathbf{1c}^{\mathrm{T}})^{\mathrm{T}}$
4) Weighted singular value decomposition (SVD) of **Y**: $\mathbf{S} = \mathbf{D}_{\mathrm{r}}^{1/2}\mathbf{Y}\mathbf{D}_{\mathrm{c}}^{1/2} = \mathbf{U}\mathbf{D}_{\varphi}\mathbf{V}^{\mathrm{T}}$
5) Calculation of the coordinates

The weighted version of LRA has several advantages like giving a different weight for each row and column (proportional to row and column margins). The positions of the objects in the biplot are determined by the log-ratio distances given in equation (3).

2.1. Data

The sources of the data are given in Aliaga [17]. Data for Turkey is taken from the Turkish Statistical Institute (Turkish Statistical Institute 2006). The data which consist of 16 countries and 6 variables of aggregated activity categories from time-use diaries. Components represent average time spent on each activity for the countries given. Time-use data is a multivariate data, which includes 6 different activities concerning the time-use as variables (components in compositional data case). Each variable has strictly positive values, and each of them carry relative information. The total time is 1440 minutes in a day, which is the constant sum of the each composition. Therefore, the data fits into the classic definition of compositional data.

Time Use Surveys provide statistics about differences between women and men in gainful and domestic work, their participation in educational and cultural activities, and other aspects of life. A representative sample of individuals completes a diary during one weekday, and one weekend day is distributed over the whole year [17]. Daily time is divided into 6 categories, which includes same kind of tasks for ease of interpretation. These are gainful work & study, domestic work, travel, sleep, meals & personal care, and free time and unspecified time-use. Explanations of these categories are given in[15], as written below:

- Gainful work and study includes time spent on primary and secondary jobs and related activities, breaks, and travel during working hours, and during job seeking. The time spent on study at school and during free time is combined with gainful work.
- Domestic work includes housework, child and adult care, gardening and pet care, construction and repairs, shopping and services, and household management.
- Travel includes commuting and trips connected with all kinds of activities, except travel during working hours.
- Sleep includes sleep during night or daytime, waiting for sleep, naps, as well as passive lying in bed because of sickness.
- Meals, personal care includes meals, snacks and drinks, dressing, personal hygiene, making up, shaving, sexual activities, and personal healthcare.
- Free time and unspecified time use includes all other kinds of activities, e.g. volunteer work and meetings, helping other households, socializing and entertainment, sports and outdoor activities, hobbies and games, reading, watching TV, resting, or doing nothing.

There are some methodological notes concerning how the surveys were conducted. The age range of the respondents differs between countries. Long time periods spent on resting is counted as sleep in France, whereas in other countries, rest is included as free time. Also, the national data is rounded, which may result in some discrepancies (see methodological notes [17] for more information).

Although the data is outdated, the results and interpretations regarding to it may differ slightly when compositional analysis methods are employed. Furthermore, the collection of this sort of data is not an easy task as completion of a diary is needed for citizens chosen as a sample for each country. Indeed, great collaboration is necessary between national statistics offices of countries to fulfill the instructions which are constituted by EUROSTAT. There are still nuances in the application of the instructions which may affect the studies conducted with the data and must be taken into consideration.

3. Results

Descriptive graphics like boxplots, bar plots, and thematic maps can be examined by using the link [18]. All the interactive plots were drawn using Tableuau software. The dashboard in the link below can be examined online, or the tableau workbook can be downloaded to examine each sheet separately. To draw a different thematic map according to a time-use activity, a user must select “variable section” to desired time-use activity. One can freely download Tableuau software’s free trial version or student version by proving a valid student identity.

Compositional calculations and ternary diagram matrixes are made by using the package “compositions” in R. Ternary diagram matrix is given in Figures 3. In the first row and first column of the ternary diagram matrix, free time, domestic work, and gainful work are the selected activities. The genders are shown with different colors to be separated from each other in each ternary diagram, respectively blue and red for women and men. Also, It is seen that there are obvious differences between

men and women in spent time on domestic work and gainful work. In some of the ternary diagrams, the points in the diagram for the men and women coincides which means that there no obvious difference in this 3-part variable selection. However, for other ternary diagrams including 3 variable parts like domestic work, gainful work, free time and meals&personalcare, domestic work and free time the separation is more noticeable between genders because women spent more time on meals&personalcare and domestic work when compare to men had more working hours when compare to women. A clustering dendrogram shows how the units are separated from each other according to a similarity measure. When the height of the dendrogram is high, the dissimilarity of units increases in the dendrogram. In Figures 4 and 5, cluster dendrograms of the hierarchical clustering analysis, which are obtained by using two different distance measures, are given for both genders' time-use. Turkey is seen as a single cluster according to dendrogram of Aitchison distance and is very far away from other countries in Figure 4(a). On the other hand, it is located with Italy and Spain in the same branch at a lower height of the dendrogram in Figure 4(b).

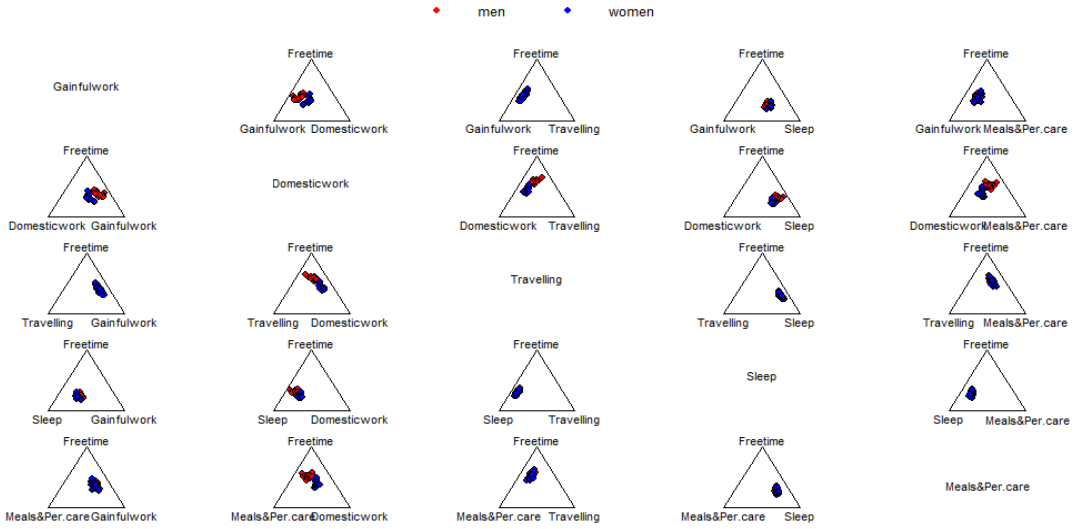


Figure 3: Ternary diagram Matrix of men and women time-use

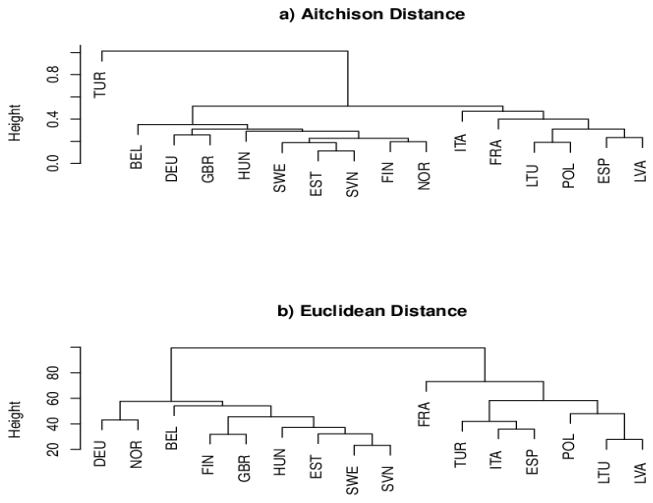


Figure 4. Cluster dendrogram of men's data according to two distance measures.

The women's dendrogram obtained through Aitchison distance in Figure 5(a) is more homogenous compared to mens' when the dendrogram is divided from lower height. When the dendrograms of each gender in Figures 4 and 5 are cut through at a lower height with greater similarity, different clusters show up according to the Aitchison and Euclidean distances.

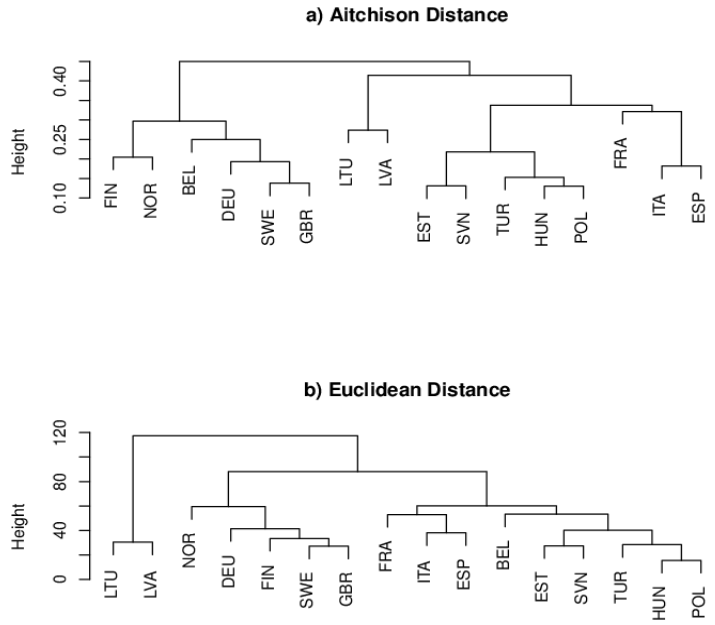


Figure 5. Cluster dendrogram of women's data according to two distance measures.

In Tables 1 and 2, clusters which are the results of the hierarchical cluster analysis using Aitchison distance are given for men and women, respectively. Countries are divided into three clusters by hierarchical clustering analysis according to their time-use for men while for women they are divided into four clusters. Eastern European countries seem to fall into same clusters for both genders. Furthermore, Scandinavian countries are in the same clusters so one can interpret that being spatially close can be a factor influencing the cluster formation. Therefore, a spatial relation can also be considered in clustering time-use. Turkey is listed as a single cluster in men's time-use because its domestic time use is so low as compared to others, and it might be an outlier compare to an average composition.

Dendrograms have different structures according to both distances with major and minor key points. For instance, the time-use of Turkish men completely differs according to two distance measures that would most probably result in a different cluster after the analysis. Therefore, the clusters formation by using different distance measures also differs from one another.

Table 1. Clusters of men according to time-use

Clusters	Countries
1	Turkey
2	Belgium, Germany, Hungary, Finland, Sweden, Great Britain, Norway, Estonia, Slovenia
3	France, Italy, Spain, Lithuania, Latvia, Poland

Table 2. Clusters of women according to time-use

Clusters	Countries
1	Turkey, Hungary, Estonia, Poland, Slovenia
2	Belgium, Germany, Finland, Sweden, Great Britain, Norway
3	France, Italy, Spain,
4	Lithuania, Latvia

In Figures 6 and 7, boxplots of clusters for genders are given. The difference in domestic work and gainful work can also be recognized after the cluster analysis. They can be the underlying cause for the clustering with free time activities. In both genders, the 2nd clusters contain the countries which have got more free time than others. The leading countries in these clusters are the Scandinavian countries of Finland, Norway, and Sweden. The most homogenous clusters are seen in terms of sleep activity,

according to the boxplots. In Norway, diary construction is also considered as socializing, a subcategory of free time, and may be the reason of its higher value in that category.

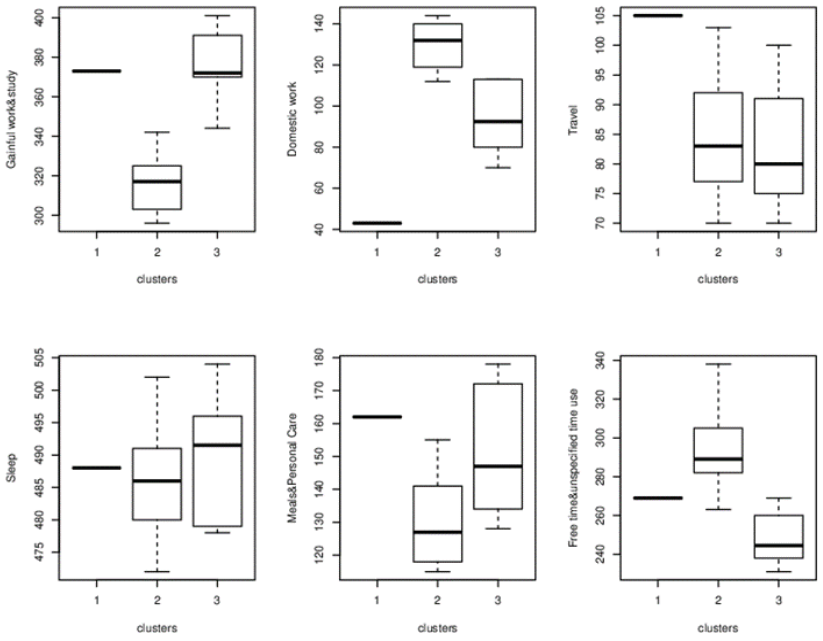


Figure 6. Boxplot of men's time-use data according to clusters

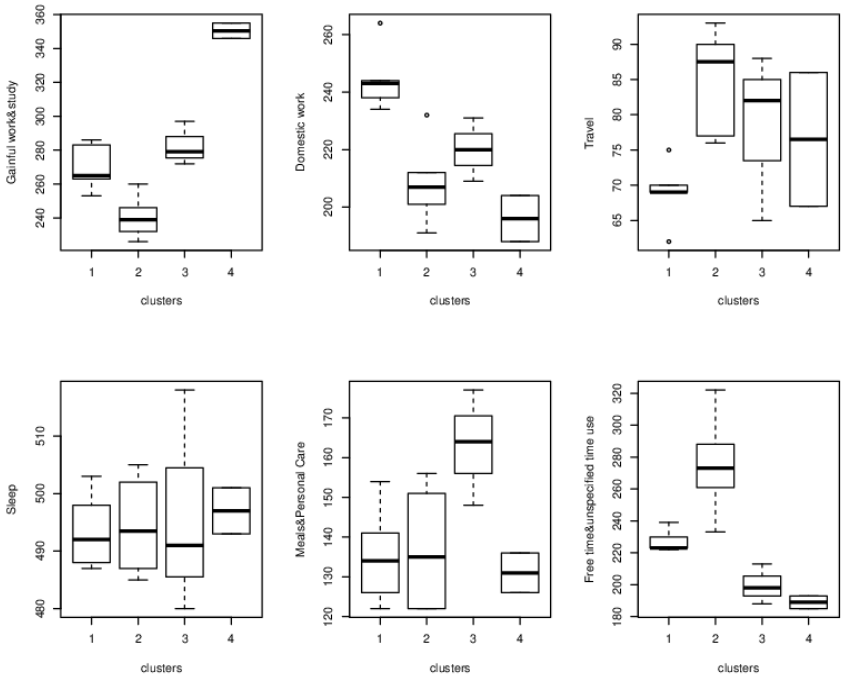


Figure 7. Boxplot of women's time-use data according to clusters

To interpret the log-ratio biplots, the link vectors that connect all the combinations of pairwise components must firstly be drawn. In our example, only three of the pairwise components that are thought of as the most important are taken in consideration. These pairwise components are also selected for an easier interpretation when there are 15 pairwise components in total. A link vector takes the place of the biplot axes in CA but with a slight difference in meaning. A link vector is a two-sided arrow whose component value dominates the other or vice versa through the concerned variable. Therefore, the ratios are reversed according to direction of the arrow.

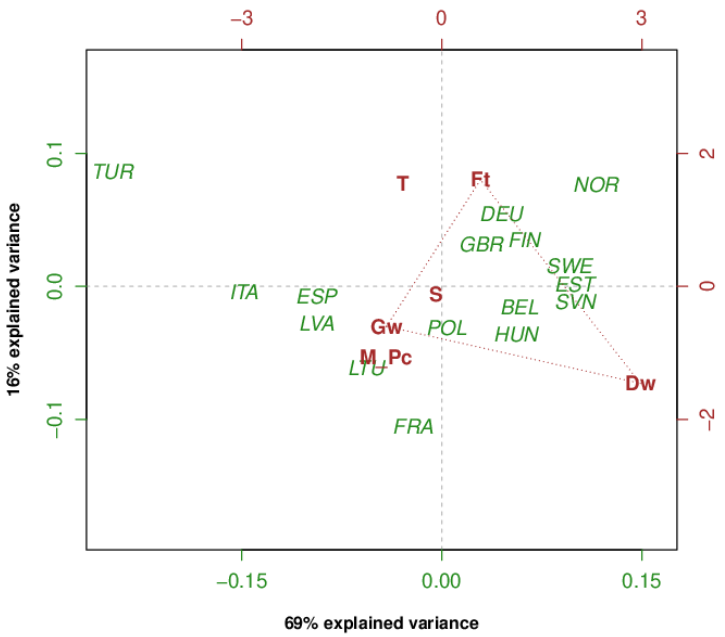


Figure 8. Log-ratio biplot of men's time-use data

In Figure 8, the scattering of the units in log-ratio biplot is obtained. Turkey is located in the upper left portion of the graph, which is obtained as a single cluster in Table 1. Thanks to its lower domestic work value and high gainful work value, the log-ratio of gainful work to domestic work is higher than any other country. It can be inferred that when moving from domestic work to gainful work component, different clusters are revealed. One cluster is located close to free time because of their high values in this component. Examples are Norway, Germany and Finland.

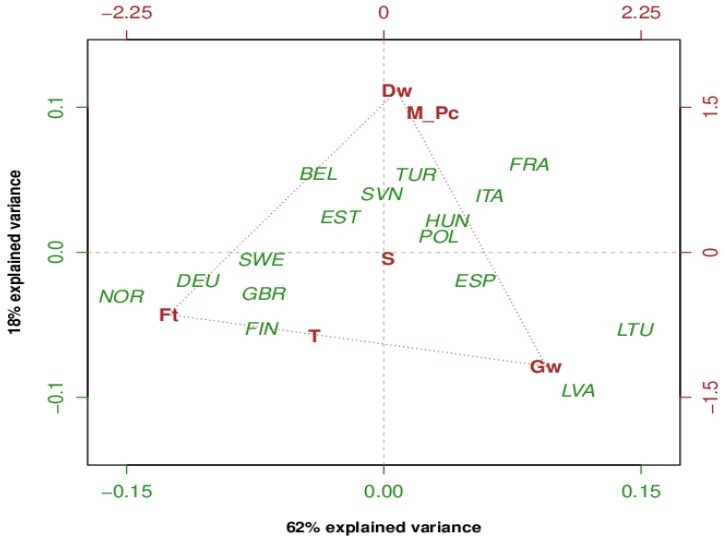


Figure 9. Log-ratio biplot of women's time-use data

In Figure 9, gainful work to domestic work link and gainful work to free time link are the link vectors that affects the cluster formation for women's time use. There are not so diverging countries from components of time as compared to those of men. There are many countries located near the domestic work component. Important difference can be noticed that Latvian and Lithuanian women has got a higher working hours than any other country because they are located in the edge of gainful work variable in the gainful work- domestic work link vector. Furthermore, Scandinavian countries and Germany has got much higher free time when compared to other countries as they are located closer to the edge free time variable on the link vector freetime and domestic work. Turkey, Belgium Slovenia

and other have have high domestic value to freetime and they are located on the other edge of the same link vector.

4. Conclusions

In this study, a bunch of visualizations and related statistical analyses are given to examine the time-use data of European countries. Data includes average time-use of the residents of EU countries and Turkey, which is obtained through surveys (time-use diaries). Even though the surveys are conducted according to some rules, there are still methodologic differences between countries because of the segmentation of time-use Also, the sample size and sample diversity (spatial sampling of the respondents) for a country are the key issues for national representability of the data. Another issue is the homogeneity of the data. The age gap of residents who participate in the survey is so large that it may play a role in a completely different time-use data set.

A pioneering part of this study is that it may be a study in which interactive visualizations is mentioned and the importance of the need to treat the time-use data as a compositional data. It should treat the data according to its own space rather than treating every multivariate data in the multivariate real space.

Furthermore, visualizations resulting from compositional thinking called "ternary diagrams" and log-ratio biplots for time-use compositional data are given. Ternary diagrams can be easily interpreted for three-part, compositional data. However, when the component number increases, the interpretation would be difficult for the researcher. Moreover, LRA interpretation is different from and more complicated that of CA and PCA, which results in a difficult understanding of the findings; however, it is the true visualization tool for the compositional data. Also, it reflects the distance measure given in equation 3, in the determination of the country locations in the biplots. Parallel results are found with a study in literature, mostly giving importance to the gender inequality of time-use.

The differences of using two distinct distance measures are demonstrated with the dendrograms derived from the cluster analysis, and the major differences of the clusters are shown. The proof of getting different results from implementing the statistical analysis in different measurement space is also given in our study. It is important to use an adaptation of conventional multivariate statistical analysis of compositional data, because it fits with the compositional setting. In this way, one may treat a country like a composition as a whole, rather than an object that has independent attributes. Also, a lower dimensionality visualization method which is suitable for compositional data is given for the visual inference of the data.

Clusters obtained in our study have some similarities with those in the study by Gálvez-Muñoz et. al. [2]. However, the cluster analysis is applied to each gender in this study, whereas Gálvez-Muñoz et al. [2] applied cluster analysis on the difference data to investigate gender equality. Furthermore, Gálvez-Muñoz et al. relates countries in the same cluster have the same level of GDP. This inference can also be valid for our study. In this study, I tried to give all the visualizations and analyses separately for each gender so the reader can also see the differences of countries within gender.

Usage of compositional data analysis reveals that Turkey's time-use among men slightly differs with the conventional multivariate counterparts. It can be classified as an outlier after compositional data analysis, but it appears as an ordinary observation in clustering in multivariate real space.

The quantitative analysis part of the study is far more developed than the institutional setting builds upon only gender equality in the given literature. When the social researchers and data scientists get together as a team to investigate time-use, much more trustworthy and consistent results can be obtained. It is hoped that all the visualizations and analyses can be a reference guides to researchers working with time-use data.

Author Contributions:

Funding: No financial resources were provided for this study.

Conflicts of Interest: The authors declare no conflict of interest.

References

[1] J.W. Tukey,. Exploratory data analysis. Reading, Mass, Addison-Wesley Pub. Co, 1977

[2] L. Gálvez-Muñoz, P. Rodríguez-Modroño, and M. Domínguez-Serrano, "Work and Time Use By Gender: A New Clustering of European Welfare Systems". Feminist Economics, 17(4), pp.125–157,2011

[3] S. Moreno-Colom, "The gendered division of housework time: Analysis of time use by type and daily frequency of household tasks". Time & Society, pp.1–25,2015, Available at: http://tas.sagepub.com/cgi/doi/10.1177/0961463X15577269.

[4] H. Coffe, "Time use among New Zealand Members of Parliament." Time & Society, 2015 Available at: http://tas.sagepub.com/cgi/doi/10.1177/0961463X15579578

[5] J.P. Robinson and J. Gershuny, "Visualizing multinational daily life via multidimensional scaling (MDS)", Electronic International Journal of Time Use Research, 2013, 10, issue 1, p. 76-90, Available at : http://eijtur.org/pdf/volumes/eIJTUR-10-1-5_Robinson_Gershuny.pdf#page=76

[6] V. R. Wight, J. Price, S. M. Bianchi, & B. R. Hunt "The time use of teenagers". Social Science Research, 2009. 38(4), pp.792-809.

[7] D. Dumuid, Ž. Pedišić, J. Palarea-Albaladejo, J.A. Martín-Fernández, Hron, K. and T. Olds, "Compositional data analysis in time-use epidemiology: what, why, how". International journal of environmental research and public health, 2020. 17(7), p.2220.

[8] N. Gupta, S.E. Mathiassen, G. Mateu-Figueras, M. Heiden,, D.M. Hallman, M.B. Jørgensen, and A. Holtermann, "A comparison of standard and compositional data analysis in studies addressing group differences in sedentary behavior and physical activity." International Journal of Behavioral Nutrition and Physical Activity, 2018. 15(1), pp.1-12.

[9] J. Aitchison,, A Concise Guide to Compositional Data Analysis, 2005. Available at :http://ima.udg.edu/activitats/codawork05/A_concise_guide_to_compositional_data_analysis.pdf

[10] Bacon-Shone J., A short history of compositional data analysis: In: Pawlowsky-Glahn V and Buccianti A (eds) *Compositional Data Analysis Theory and applications.* 2011, New Delhi: John Wiley & Sons, Ltd

[11] K.G. van den Boogaart, R. Tolosana and M. Bren, Compositions: Compositional Data Analysis. R package version 1.40-1. 2014, Available at: https://CRAN.R-project.org/package=compositions

[12] M. Greenacre, and R., Primicerio, Multivariate Analysis of Ecological Data. 2013. Bilbao: Fundación BBVA

[13] T. Hastie, R. Tibshirani, and J. Friedman, The Elements of Statistical Learning 2nd ed. ,2008 ,Springer New York Inc. New York NY USA

[14] J. Aitchison, "The Statistical Analysis of Compositional Data". Journal of the Royal Statistical Society, 1982, 44(2), pp.139–177.

[15] J.A. Martın-Fernandez, C. Barcelo-Vidal, and V. Pawlowsky-Glahn,Measures of difference for compositional data and hierarchical clustering methods In: Buccianti A, Nardi G and Potenza R (eds.) Proceedings of IAMG'98, The Fourth Annual Conference of the International Association for Mathematical Geology: De Frede, 1998 Naples, p. 526–531

[16] M Greenacre, *Biplots in Practice,* 2010, Bilbao: Fundación BBVA

[17] C.Aliaga, 2006, How is the time of women and men distributed in Europe?, EUROSTAT

[18] https://public.tableau.com/app/profile/cenk.i.z/viz/EuropeanTimeUseData/Story1 (accessed Oct. 19, 2021)

Journal of Statistics & Applied Science, Volume -2 Issue -2
ISSN 2718-0999

Article

Received: date: 06.03.2021
Accepted: date: 06.10.2021
Published: date: 28.11.2021

Doğrusal Fiziki Programlama Yaklaşımı ile Kapalı Döngü Tedarik Zinciri Optimizasyonu

Gamze ÖZTÜRK[1,*], Hasan Kıvanç AKSOY[2]

[1]Eskişehir Osmangazi Üniversitesi, Fen Edebiyat Fakültesi, İstatistik Bölümü, gamzeozturk787@hotmail.com
[2]Eskişehir Osmangazi Üniversitesi, Fen Edebiyat Fakültesi, İstatistik Bölümü, hkaksoy@ogu.edu.tr
Orcid: 0000-0002-6368-2753[1] Orcid: 0000-0001-7954-9327[2]
* Correspondence Author, e-mail: gamzeozturk787@hotmail.com

Öz: Kapalı Döngü Tedarik Zinciri (KDTZ) sistemi, ileri ve tersine tedarik akışlarını bütün bir sistem üzerinde görmeyi sağlayarak üretici-müşteri ve müşteri-üretici ilişkilerini bir araya getirmektedir. Kapalı Döngü Tedarik Zinciri'nde ağ yapısı karmaşık olup burada iade edilen ürünlerin geri kazanım miktarı, maliyeti, süresi ve geri dönüşüm oranı gibi belirsizlikler yer almaktadır. Kapalı Döngü Tedarik Zinciri'nde bu belirsizliklerden yola çıkarak toplam kâr ve geri kazanım oranının maksimizasyonu, toplam geri kazanım süresinin minimizasyonunu hedefleyen çok amaçlı optimizasyon problemi ele alınmıştır. Çok amaçlı optimizasyon modelinin çözümü için karşılaşılan zorluklardan biri de amaç fonksiyonundaki hedeflerin ağırlıklarının belirlenememesi ve buna bağlı olarak karar vericinin tutarsız sonuçlarla karşılaşabilmesidir. Bu noktada Doğrusal Fiziki Programlama yöntemi karar vericilerin tercih aralıklarını belirleyerek hedefler için kriter ağırlıklarını tutarlı bir şekilde hesaplanmasına olanak sağlamaktadır. Bu çalışmada Kapalı Döngü Tedarik Zinciri problemi için yeniden üretilebilir ve ortak parçalara sahip 3 farklı ürün ele alınmıştır. Uygulaması yapılan çok amaçlı optimizasyon modelinde optimal ürün seçeneği Doğrusal Fiziki Programlama yaklaşımı ile belirlenmiş ve elde edilen sonuçlar iade edilen ürün sayısı, satış geliri, geri kazanım süresi ve geri kazanım oranındaki değişiklikler üzerinden üç farklı senaryo kullanılarak analiz edilmiştir.
Anahtar Kelimeler: Kapalı döngü tedarik zinciri, Doğrusal fiziki programlama, Çok amaçlı optimizasyon

Close Loop Supply Chain Optimization with Linear Physical Programming Approach

Abstract: The Closed-Loop Supply Chain (CLSC) system brings together producer-customer and customer-producer relationships, by enabling the forward and reverse supply flows to be seen on a whole system. In the CLSC, network structure is complex, there are uncertainties such as the recovery amount, cost, time and recycling rate of the returned products. Based on these uncertainties in the CLSC, multi-objective optimization problem aimed at maximizing the total profit, recovery rate, minimizing the total recovery time was discussed. One of the difficulties encountered in the solution of the multi-objective programming model is that the weights of the goals in the objective function which aren't known and consequently the decision maker may encounter inconsistent results. At this point, Linear Physical Programming method allow decision makers to calculate the weights of the goals in consistent manner, taking into account their preference ranges. In this study, for the CLSC problem, 3 different remanufacturable products with common parts are considered. In the multi-objective optimization model applied, the optimal product option was determined with the LPP approach, the results were analyzed changes in the number of returned products, sales revenue, recovery time and recovery rate using three different scenarios.
Keywords: Closed loop supply chain, Linear physical programming, Multi-objective optimization

1. Giriş

Günümüzdeki işletmelerin çalışma süreçlerine bakıldığında tedarik zinciri kavramının işletmeler için çok önemli bir hal aldığı görülmektedir. Hızla ilerleyen teknoloji, değişen çevre koşulları ve artan rekabet durumu karşısında zorlu koşullarda ayakta kalabilen işletmeler ise Kapalı Döngü Tedarik Zinciri konusunda farklı yaklaşımlara yönelmişlerdir. Kimi işletmeler uygun yerlere geri kazanım tesisi açmaya yönelirken kimi işletmeler tedarik zincirindeki taşıma maliyeti ve taşıma süresini optimum seviyeye çekmeye çalışmıştır. Fakat ulaşılması istenen ortak nokta müşteri taleplerinin en doğru şekilde karşılanabilmesidir.

Tedarik zincirinde ileri ve tersine olmak üzere iki tip akış vardır. İleri tedarik zincirinde akış yönü tedarikçi-müşteri iken tersine tedarik zincirinde bu yön müşteri-üretici şeklindedir. Kapalı Döngü Tedarik Zinciri, ileri ve tersine tedarik zinciri akışları arasındaki ilişkiyi ortaya çıkararak geleneksel tedarik zinciri kavramını genişletir [3]. Tedarik zincirinde ağ yapısı üreticiden tüketiciye doğrusal bir yol izlerken Kapalı Döngü Tedarik Zinciri'nde bu yol tüketiciden tekrar üreticiye ulaşıp döngü halini almaktadır.

Kapalı Döngü Tedarik Zinciri ağ yapısı bu sistemi oluşturan en önemli unsurdur. Çünkü optimum maliyet ve zaman kullanarak işletmeye fayda sağlayacak olan döngü bu ağ yapısında şekillenecektir. Diğer bir yandan Kapalı Döngü Tedarik Zinciri'nde ağ yapısı karmaşık olup burada müşteriler tarafından iade edilen ürünlerin geri kazandırılma miktarı, geri kazandırılma maliyeti, geri kazandırılma süresi ve geri dönüşüm oranı gibi belirsizlikler yer almaktadır. Kapalı Döngü Tedarik Zinciri'nde bu belirsizliklerden yola çıkarak toplam geri kazanım kâr miktarının maksimizasyonu ve toplam geri kazanım süresinin minimizasyonunu hedefleyen bir optimizasyon problemi ortaya çıkmaktadır. Bu problem çok amaçlı doğrusal programlama modeli olarak da düşünülebilir. Çok amaçlı doğrusal programlama modelinin çözümü için karşılaşılan zorluklardan biri de amaç fonksiyonundaki hedeflerin ağırlıklarının bilinmemesi ve buna bağlı olarak karar vericinin tutarsız sonuçlarla karşılaşabilmesidir. Bu noktada Doğrusal Fiziki Programlama (LPP) yaklaşımı bu probleme alternatif bir çözüm önerisi sunmaktadır. LPP yönteminde ulaşılması istenen hedefler 4 farklı sınıf fonksiyonunda kategorize edilerek her bir hedefin tercih aralıkları tanımlanmaktadır. Sonuçta Doğrusal Fiziki Programlama yaklaşımı, karar vericilerin tercih aralıklarını kullanarak hedefler için kriter ağırlıklarını tutarlı bir şekilde hesaplanmasına olanak sağlamaktadır.

Bu çalışmada yeniden üretilebilir ve ortak parçalara sahip 3 ürününün Kapalı Döngü Tedarik Zinciri'nde geri kazanım süresinin minimizasyonu, toplam kârın ve geri kazanım oranının maksimizasyonunun sağlanmasıyla birlikte optimal ürün seçeneğinin belirlenmesi amaçlanmaktadır. Bu çalışmanın geri kalan bölümleri aşağıdaki şekilde düzenlenmiştir:

İkinci bölümde Kapalı Döngü Tedarik Zinciri ile ilgili literatür araştırmasına değinilecektir. Üçüncü bölümde Kapalı Döngü Tedarik Zinciri ağ akışlarının belirlenmesiyle birlikte uygulanması öngörülen LPP yaklaşımı detaylı olarak açıklanacaktır. Dördüncü bölümde problem tanımına ve müşteriden iade edilen ürünlerin geri kazanım süresi, maliyet ve geri dönüşüm oranı belirsizliklerin bulunduğu çok amaçlı doğrusal programlama modeline yer verilecektir. Modelle ilgili zaman, maliyet kısıtları ve kalite düzeyleri sayısal bir örnek üzerinde analiz edilecektir. Beşinci bölümde ise uygulama sonuçları ve gelecekte yapılacak çalışmalar ile ilgili öneriler sunulacaktır.

2. Literatür Taraması

Avrupa' da 18. yy' da gelişen Endüstri Devrimi birçok teknik buluşun etkisini arttırarak sanayileşmeye önem veren ülkelere ekonomik güç kazandırmıştır. Sermaye gücüne sahip olanlar ise hammadde ve pazar arayışına yönelmişlerdir. Üretim alanında yaşanan teknolojik gelişmelerle ürün çeşitliliği artmış ve seri üretim yapısı ortaya çıkmıştır. Seri üretim ile hayatımızda tedarik zinciri, lojistik, pazarlama gibi kavramlar yer almıştır. Neticede müşteri ihtiyaçlarının karşılanmasında en önemli unsurlardan biri iyi bir tedarik zinciri yönetiminin sağlanmasıdır.

İşletmelerin tüm çalışma süreçlerini yakından ilgilendiren tedarik zinciri; tüm ürün ve hizmetlerin tedarikçiden başlayıp en son aşamada müşteriye gidene kadarki süreci, bu yolda yer alan tüm faaliyetleri, insan kaynağını, teknolojiyi, firma yapılarını ve kaynakları kapsayan kavramın adı olarak açıklanabilir [2]. Tersine tedarik zincirinde hammaddenin kaynağı son kullanıcı yani müşteridir ve

müşteriden üreticiye doğru bir malzeme/enerji akışı vardır. Kapalı Döngü Tedarik Zinciri ise üretici-müşteri ve müşteri-üretici akışlarını bir bütün haline getirmektedir. Son yıllarda çevre koruma bilincinin daha fazla hissedilmesi, kaynakların daha verimli kullanılması, israfın önüne geçilerek maliyeti azaltma düşüncesi ve sıfır atık kampanyaları geri kazanım tesislerinin her alanda yaygınlaşmasını sağlamıştır. Bütün bu sebeplerle beraber müşteri ihtiyaçlarını daha fazla karşılayabilmek, KDTZ ağ yapısındaki belirsizlikleri giderebilmek ve yeniden üretimde ortaya çıkan sorunları çözümlemek için Kapalı Döngü Tedarik Zinciri ile ilgili birçok çalışma mevcuttur. Mevcut çalışmalarda yer alan çözüm önerileri şu şekilde özetlenebilir:

Marin ve Pelegrin [10], geri dönüş tesisi için yerleşim problemini incelemiş, lagrange yöntemine dayalı sezgisel çözüm yöntemine yer vermiştir. Bunun yanı sıra Jayaraman vd. [8], dağıtım merkezlerinin konumu, geri kazandırılmış ürünlerin optimal miktarlarının aktarılması, üretimi ve stoklanması için aynı anda çözülen 0-1 karma tamsayı problemi sunmuştur. Kapalı çevrim tedarik zincirlerinin verimli bir şekilde uygulanması ve geri kazanılan ürünlerin ortaya çıkan akışları için uygun lojistik yapıların kurulmasını gerektiğini belirten Fleischmann vd. [5], bu konuyu dikkate alarak ürünün geri kazanımı birçok durumda mevcut lojistik yapılarına etkin bir şekilde entegre edilerek lojistik ağını bütüncül bir şekilde yeniden tasarlamak için daha kapsamlı bir yaklaşıma yer vermiştir.

İyi yönetilen tedarik zinciri yönetimini; atık, envanter ve lojistikte önemli tasarruf olarak tanımlayan Kannan vd. [9], sipariş üzerinden entegre edilen çok kademeli ve çok ürünlü kapalı döngü ağ modeli geliştirmiştir. Hindistan'ın güney kısmında bulunan plastik eşya üreten firmasında iki örnek olay için uygulaması yapılmıştır. Karma tamsayılı doğrusal programlama modeli genetik algoritma ile çözülmüştür. Pishvaee ve Torabi [11], belirsizlik ve risk altında Kapalı Döngü Tedarik Zinciri ağı tasarımı için olasılıklı bir programlama yaklaşımı geliştirmişlerdir. Yapılan çalışmada çok amaçlı olasılıklı karma tamsayılı programlama modeli önermişlerdir. Önerilen modelde ileri ve tersine tedarik zinciri stratejik ağ tasarımına entegre edilmiştir ve taktik malzeme akışına da karar verilmiştir. Karma tamsayılı programlama modelini çözmek için etkileşimli bulanık çözüm yaklaşımı geliştirilmiştir. Çözüm yaklaşımının uygulanabilirliğini göstermek için sayısal verilerle deneyler yapılmıştır.

Son dönemde firmaların klasik üretim planlama tekniğinden farklı olarak orta vadeli planlama faaliyetlerine yöneldiğini söyleyen Subulan vd. [12], yeniden üretime dikkat çekmişlerdir. Bu kapsamda Kapalı Döngü Tedarik Zinciri'nde orta vadeli programlama ile ilgili bulanık karma tamsayı programlama modeli önermiştir. Önerilen model için iki üretim seçeneği; yeni üretilen ürünleri direkt olarak üretim tesislerinde üretmek veya müşteriden toplanan ürünlerin yeniden üretim tesislerinde iyileştirilerek geri kazanılması değerlendirilmiştir. Ayrıca geliştirilen modelde depoların kapasite durumları, toptancı ve perakendeci talepleri, iade ve kabul oranları, haftalık direkt üretim süreleri/ geri kazanımla üretim süreleri ve amaç fonksiyonundaki parametreler bulanık olarak hesaplanmıştır. Modelin çözümü ise ILOG OPL Studio 6.3 programı ile yapılmıştır. Zhang vd. [13], ürün iadelerini ve yeniden üretimi dikkate aldığında Kapalı Döngü Tedarik Zinciri'nde sınırlı kapasitede parti büyüklüğü problemini incelemiştir. Karma tamsayılı doğrusal programlama modelinde, çözüm önerisi olarak lagrange gevşetmeye dayalı bir yaklaşım sunmuştur. Kaliteli çözüm olduğu düşünülen bu yöntem amaç fonksiyonundaki değerin alt sınırının daha düşük değerde olmasını sağlayabilmiştir.

Amin ve Zhang [1], birden fazla ürün içeren Kapalı Döngü Tedarik Zinciri ağı tasarımını yapılandırıp belirsizlik altındaki durumları değerlendiren üç aşamalı model geliştirmişlerdir. Yeniden üretim ve yenileme işlemlerinde QFD modelini örnek alarak müşteri, parça ve süreç arasındaki ilişkileri analiz edip çok amaçlı karma tamsayılı doğrusal olmayan programlama modelini bulanık kümeler teorisi yaklaşımıyla ele almışlardır. Hajipour vd. [6], stokastik çok ürünlü Kapalı Döngü Tedarik Zinciri'nde karma tamsayılı doğrusal olmayan programlama modeli önermiştir. Kapalı döngü tedarik ağında Radyo Frekansı Tanımlama (RFID) sistemi, satış sonrası elde edilen kârı arttırarak ürün kayıplarını ve genel teslim süresini azaltmıştır. Önerilen modeli çözmek için iki meta-sezgisel algoritma geliştirmiştir. GRASP (Açgöz Rasgele Adaptif Arama Prosedürü) ve PSO (Parçaçık Sürü Optimizasyonu) olan bu iki algoritma kullanılarak farklı boyutlarda hesaplamalar yapılmıştır. Sonuçta, GRASP' ın hem kâr hem zaman açısından PSO' dan daha iyi bir performans gösterdiği belirlenmiştir.

Tüm bu incelemeler kapsamında yapılan araştırmalarla, Kapalı Döngü Tedarik Zinciri konusunun işletmeler için önemli hale geldiğinin ispatını mevcut çalışmalar ifade etmektedir. Mevcut çalışmalara

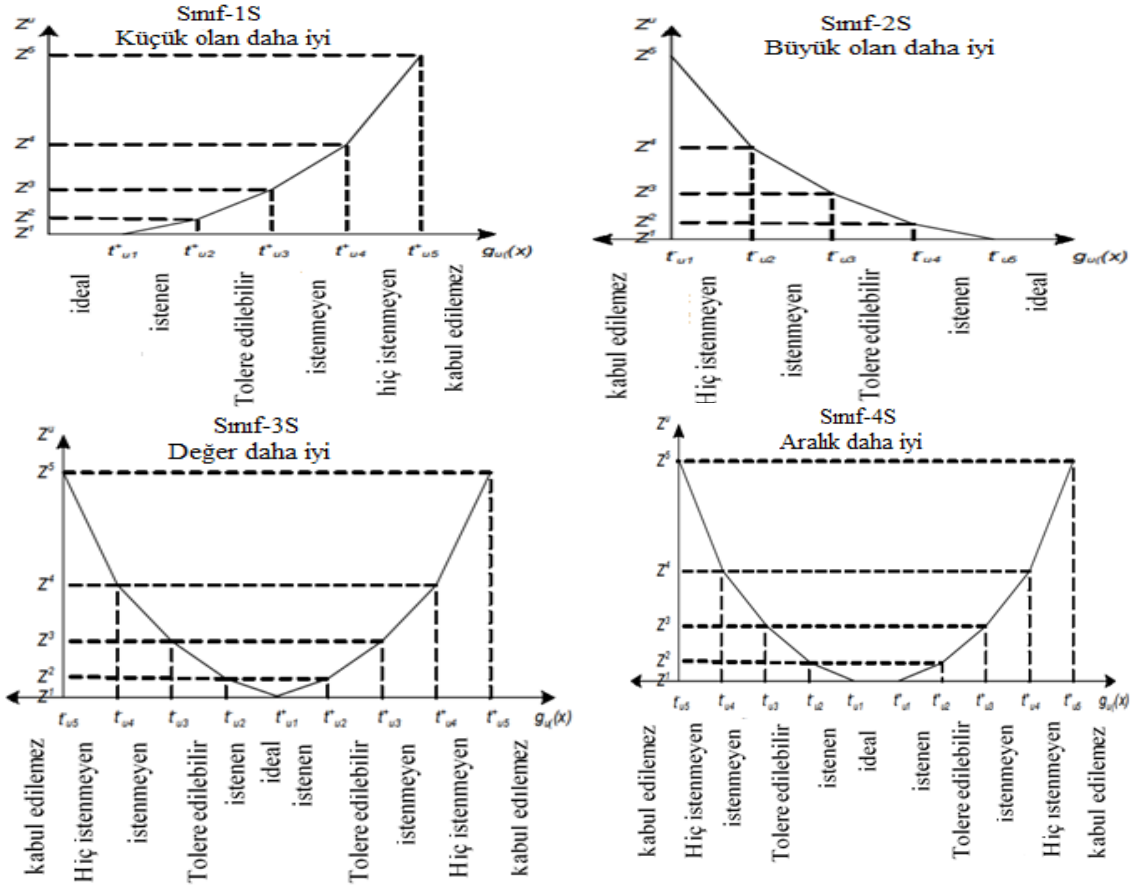
farklı bir bakış açısı getirerek 4. Bölümde sunulan KDTZ modelinin analizinde kullanılan Doğrusal Fiziki Programlama yaklaşımı sonraki bölümde açıklanmıştır.

3. Doğrusal Fiziki Programlama (LPP) Yaklaşımı

Çok amaçlı optimizasyon problemlerinin çözümünde birçok alternatif yaklaşım geliştirilmiştir. Bunlar arasında; VEGA (Vektör Değerlendirmeli Genetik Algoritma), SMEA (Kendi Kendini Düzenleyen Çok Amaçlı Evrimsel Algoritma), MOEA/D (Ayrıştırmaya Dayalı Çok Amaçlı Evrimsel Algoritma), PSA (Pareto Benzetimli Tavlama), MOSA (Çok Amaçlı Benzetimli Tavlama), Hedef Programlama, Dinamik Programlama ve LPP (Doğrusal Fiziki Programlama) yaklaşımları mevcuttur.
LPP yaklaşımı çok amaçlı optimizasyon probleminde karar vericiye ulaşılması istenen hedefler için kriter ağırlıklarını tutarlı bir şekilde hesaplamaktadır. LPP yaklaşımında her kriter için Soft durumda 1S, 2S, 3S ve 4S ve Hard durumda 1H, 2H 3H ve 4H olmak üzere 4 farklı sınıf fonksiyonu bulunmaktadır. Soft ve Hard durumları için sınıf fonksiyonları şu şekilde ifade edilir[7]:

Soft:	Hard:
Sınıf 1S: Küçük olan daha iyi (min)	Sınıf 1H: Daha küçük olmalı
Sınıf 2S: Büyük olan daha iyi (max)	Sınıf 2H: Daha büyük olmalı
Sınıf 3S: Değer daha iyi	Sınıf 3H: Eşit olmalı
Sınıf 4S: Aralık daha iyi	Sınıf 4H: Aralık içinde olmalı

Soft durumdaki sınıf fonksiyonları için Gupta ve Imtanavanich [4]' in yaptığı bir çalışmada yer alan grafikler Şekil 1'de gösterilmiştir. Bu grafiklerde yatay eksendeki $g_u(x)$ kriterler için tercih aralıklarını, dikey eksende yer alan $z_u(x)$ ise minimize edilmesi istenen sınıf fonksiyonunu ifade etmektedir.



Şekil 1. LPP yaklaşımı için Soft sınıf fonksiyonları

Şekil 1'de görüldüğü gibi her sınıf fonksiyonunun ideal değeri 0'dır yani sınıf fonksiyonunun en küçük değerler alması beklenir. Tercih aralıkları u. kritere ilişkin değerleri kategorize etmektedir. Bu tercih aralıkları ideal, istenen, tolere edilebilir, istenmeyen, hiç istenmeyen ve kabul edilemez olarak

tanımlanıp 1S, 2S, 3S, 4S sınıf fonksiyonları için aşağıda verilmiştir:

1S sınıf fonksiyonu için
$g_u \geq t_{u5}^+$ (kabul edilemez aralık)
$t_{u4}^+ \leq g_u \leq t_{u5}^+$ (hiç istenmeyen aralık)
$t_{u3}^+ \leq g_u \leq t_{u4}^+$ (istenmeyen aralık)
$t_{u2}^+ \leq g_u \leq t_{u3}^+$ (tolere edilebilir aralık)
$t_{u1}^+ \leq g_u \leq t_{u2}^+$ (istenen aralık)
$g_u \leq t_{u1}^+$ (ideal aralık)

2S sınıf fonksiyonu için
$g_u \leq t_{u5}^-$ (kabul edilemez aralık)
$t_{u5}^- \leq g_u \leq t_{u4}^-$ (hiç istenmeyen aralık)
$t_{u4}^- \leq g_u \leq t_{u3}^-$ (istenmeyen aralık)
$t_{u3}^- \leq g_u \leq t_{u2}^-$ (tolere edilebilir aralık)
$t_{u2}^- \leq g_u \leq t_{u1}^-$ (istenen aralık)
$g_u \geq t_{u1}^-$ (ideal aralık)

3S sınıf fonksiyonu için
$g_u \leq t_{u5}^-$ & $g_u \geq t_{u5}^+$ (kabul edilemez aralık)
$t_{u5}^- \leq g_u \leq t_{u4}^-$ & $t_{u4}^+ \leq g_u \leq t_{u5}^+$ (hiç istenmeyen aralık)
$t_{u4}^- \leq g_u \leq t_{u3}^-$ & $t_{u3}^+ \leq g_u \leq t_{u4}^+$ (istenmeyen aralık)
$t_{u3}^- \leq g_u \leq t_{u2}^-$ & $t_{u2}^+ \leq g_u \leq t_{u3}^+$ (tolere edilebilir aralık)
$t_{u2}^- \leq g_u \leq t_{u1}^-$ & $t_{u1}^+ \leq g_u \leq t_{u2}^+$ (istenen aralık)
$g_u = t_{u1}$ (ideal değer)

4S sınıf fonksiyonu için
$g_u \leq t_{u5}^-$ & $g_u \geq t_{u5}^+$ (kabul edilemez aralık)
$t_{u5}^- \leq g_u \leq t_{u4}^-$ & $t_{u4}^+ \leq g_u \leq t_{u5}^+$ (hiç istenmeyen aralık)
$t_{u4}^- \leq g_u \leq t_{u3}^-$ & $t_{u3}^+ \leq g_u \leq t_{u4}^+$ (istenmeyen aralık)
$t_{u3}^- \leq g_u \leq t_{u2}^-$ & $t_{u2}^+ \leq g_u \leq t_{u3}^+$ (tolere edilebilir aralık)
$t_{u2}^- \leq g_u \leq t_{u1}^-$ & $t_{u1}^+ \leq g_u \leq t_{u2}^+$ (istenen aralık)
$t_{u1}^- \leq g_u \leq t_{u1}^+$ (ideal aralık)

Karar verici u. kriter için t_{u5}^- 'den t_{u5}^+ 'e kadar bütün tercih değerlerini belirlemektedir. Örneğin 1S sınıfı için değerlendirilen kriter "maliyet" ve t_{u1}^+ 'den t_{u5}^+ 'e kadar olan tercih değerleri de sırasıyla 5000$, 10000$, 15000$, 20000$ ve 25000$ olarak tanımlanırsa maliyet tutarı 3965$ olan alternatif seçenek ise ideal aralıkta, maliyet tutarı 7000$ olan alternatif seçenek istenen aralıkta, maliyet tutarı 13986$ olan seçenek tolere edilebilir aralıkta, maliyet tutarı 18750$ olan alternatif seçenek istenmeyen aralıkta, maliyet tutarı 22500$ olan alternatif seçenek hiç istenmeyen aralıkta ve maliyet tutarı 30000$ olan alternatif seçenek kabul edilemez aralıktadır.

3.1. LPP Adımları

LPP (Doğrusal Fiziki Programlama) yöntemi uygulanırken 4 adım izlenir[7]:

1. Her bir kriterin Soft ya da Hard durumu için 4 sınıf fonksiyonundan biri belirlenir.
2. Her bir kriter için farklı derecelerdeki aralıkların hedef değerleri (yatay eksendeki $\mathbf{g_u(x)}$ değerleri) tanımlanır.
   I. Ağırlıkları oluşturmak için LPPW (Doğrusal Fiziki Programlama Algoritması) kullanılır. Başlama: $\boldsymbol{\beta = 1.1}$, $\mathbf{w_{u1}^+} = 0$, $\mathbf{w_{u1}^-} = 0$, $\tilde{\boldsymbol{z}}^{\boldsymbol{2}}$= küçük pozitif sayı (örneğin 0.1)
   u=0; s=1, $\boldsymbol{n_{sc}}$= kriter sayısı (soft)
   II. u = u+1
   III. s = s+1
   sırasıyla $\tilde{\boldsymbol{z}}^{\boldsymbol{s}}$, $\tilde{\boldsymbol{t}}_{\boldsymbol{us}}^{\boldsymbol{+}}$, $\tilde{\boldsymbol{t}}_{\boldsymbol{us}}^{\boldsymbol{-}}$, $\mathbf{w_{us}^+}$, $\mathbf{w_{us}^-}$, $\tilde{\boldsymbol{w}}_{\boldsymbol{us}}^{\boldsymbol{+}}$, $\tilde{\boldsymbol{w}}_{\boldsymbol{us}}^{\boldsymbol{-}}$, $\tilde{\boldsymbol{w}}_{\boldsymbol{min}}$ parametrelerini değerlendir.
   Eğer $\tilde{\boldsymbol{w}}_{\boldsymbol{min}}$, küçük pozitif sayıdan daha küçük değerde ise $\boldsymbol{\beta}'$ yı arttır ve II. Adıma git.
   IV. Eğer s ≠ 5 ise III. Adıma git.
   V. Eğer u ≠ $\boldsymbol{n_{sc}}$ ise II. Adıma git.

Algoritmada u soft kriter indeksini, $\boldsymbol{\beta}$ dışbükeylik parametresi, s aralık indeksini, $\mathbf{w_{us}^+}$ u. kriterin s. aralıktaki pozitif ağırlığını, $\mathbf{w_{us}^-}$ u. kriterin s. aralıktaki negatif ağırlığını, $\tilde{\boldsymbol{z}}^{\boldsymbol{s}}$ s. aralıktaki meydana gelen değişimi, $\tilde{\boldsymbol{t}}_{\boldsymbol{us}}^{\boldsymbol{+}}$ u. kriter için pozitif kenarındaki s. aralık uzunluğunu, $\tilde{\boldsymbol{t}}_{\boldsymbol{us}}^{\boldsymbol{-}}$ u. kriter için negatif kenarındaki s. aralık uzunluğunu, $\tilde{\boldsymbol{w}}_{\boldsymbol{us}}^{\boldsymbol{+}}$ pozitif normalleştirilmiş ağırlığı, $\tilde{\boldsymbol{w}}_{\boldsymbol{us}}^{\boldsymbol{-}}$ negatif normalleştirilmiş ağırlığı, $\tilde{\boldsymbol{w}}_{\boldsymbol{min}}$, $\tilde{\boldsymbol{w}}_{\boldsymbol{us}}^{\boldsymbol{+}}$ ve $\tilde{\boldsymbol{w}}_{\boldsymbol{us}}^{\boldsymbol{-}}$ değerlerinin minumum değerini göstermektedir. $\mathbf{w_{us}^+}$, $\mathbf{w_{us}^-}$, $\tilde{\boldsymbol{z}}^{\boldsymbol{s}}$, $\tilde{\boldsymbol{t}}_{\boldsymbol{us}}^{\boldsymbol{+}}$, $\tilde{\boldsymbol{t}}_{\boldsymbol{us}}^{\boldsymbol{-}}$, $\tilde{\boldsymbol{w}}_{\boldsymbol{us}}^{\boldsymbol{+}}$, $\tilde{\boldsymbol{w}}_{\boldsymbol{us}}^{\boldsymbol{-}}$ değerlerinin hesaplanması aşağıda verilmiştir:

$$\mathbf{w_{us}^-} = \frac{\tilde{z}^s}{\tilde{t}_{us}^-} \; ; \boldsymbol{s} = [\mathbf{2}, \mathbf{5}] \quad (1)$$

$$\mathbf{w_{us}^+} = \frac{\tilde{z}^s}{\tilde{t}_{us}^+} \; ; \boldsymbol{s} = [\mathbf{2}, \mathbf{5}] \quad (2)$$

$$\tilde{\boldsymbol{z}}^{\boldsymbol{s}} = \boldsymbol{\beta}\,(\boldsymbol{n_{sc}} - \mathbf{1})\,\tilde{\boldsymbol{z}}^{\boldsymbol{s-1}} \; ; \; \boldsymbol{s} = [\mathbf{3}, \mathbf{5}]; \; (\boldsymbol{n_{sc}} \geq \mathbf{1}); \; \boldsymbol{\beta} > \mathbf{1} \quad (3)$$

$$\tilde{\boldsymbol{t}}_{\boldsymbol{us}}^{+} = \boldsymbol{t}_{\boldsymbol{us}}^{+} - \boldsymbol{t}_{\boldsymbol{u(s-1)}}^{+} \; ; \boldsymbol{s} = [\mathbf{2}, \mathbf{5}] \quad (4)$$

$$\tilde{\boldsymbol{t}}_{\boldsymbol{us}}^{-} = \boldsymbol{t}_{\boldsymbol{us}}^{-} - \boldsymbol{t}_{\boldsymbol{u(s-1)}}^{-} \; ; \boldsymbol{s} = [\mathbf{2}, \mathbf{5}] \quad (5)$$

$$\tilde{\boldsymbol{w}}_{\boldsymbol{us}}^{+} = \frac{w_{us}^{+}}{\sum_{s=2}^{5} w_{us}^{+}} ; \boldsymbol{s} = [\mathbf{2}, \mathbf{5}] \quad (6)$$

$$\tilde{\boldsymbol{w}}_{\boldsymbol{us}}^{-} = \frac{w_{us}^{-}}{\sum_{s=2}^{5} w_{us}^{-}} ; \boldsymbol{s} = [\mathbf{2}, \mathbf{5}] \quad (7)$$

3.Her bir alternatif için tüm kriterler ve aralıklarına göre toplam sapma değeri hesaplanır.

$$\min \boldsymbol{J} = \sum_{u=1}^{n_{sc}} \sum_{s=2}^{5} (\tilde{\boldsymbol{w}}_{\boldsymbol{us}}^{-} . \boldsymbol{d}_{\boldsymbol{us}}^{-} + \tilde{\boldsymbol{w}}_{\boldsymbol{us}}^{+} . \boldsymbol{d}_{\boldsymbol{us}}^{+}) \quad (8)$$

$$\boldsymbol{g_u} - \boldsymbol{d}_{\boldsymbol{us}}^{+} \leq \boldsymbol{t}_{\boldsymbol{u(s-1)}}^{+}; \boldsymbol{d}_{\boldsymbol{us}}^{+} \geq \mathbf{0} \quad (9)$$

$$\boldsymbol{g_u} \leq \boldsymbol{t}_{\boldsymbol{u5}}^{+} \text{ ; 1S, 3S ve 4S sınıflarındaki her } \boldsymbol{u} \text{ için } \boldsymbol{u} = \mathbf{1}, \mathbf{2}, \ldots ; \boldsymbol{s} = [\mathbf{2}, \mathbf{5}] \quad (10)$$

$$\boldsymbol{g_u} + \boldsymbol{d}_{\boldsymbol{us}}^{-} \leq \boldsymbol{t}_{\boldsymbol{u(s-1)}}^{-} \; ; \boldsymbol{d}_{\boldsymbol{us}}^{-} \geq \mathbf{0} \quad (11)$$

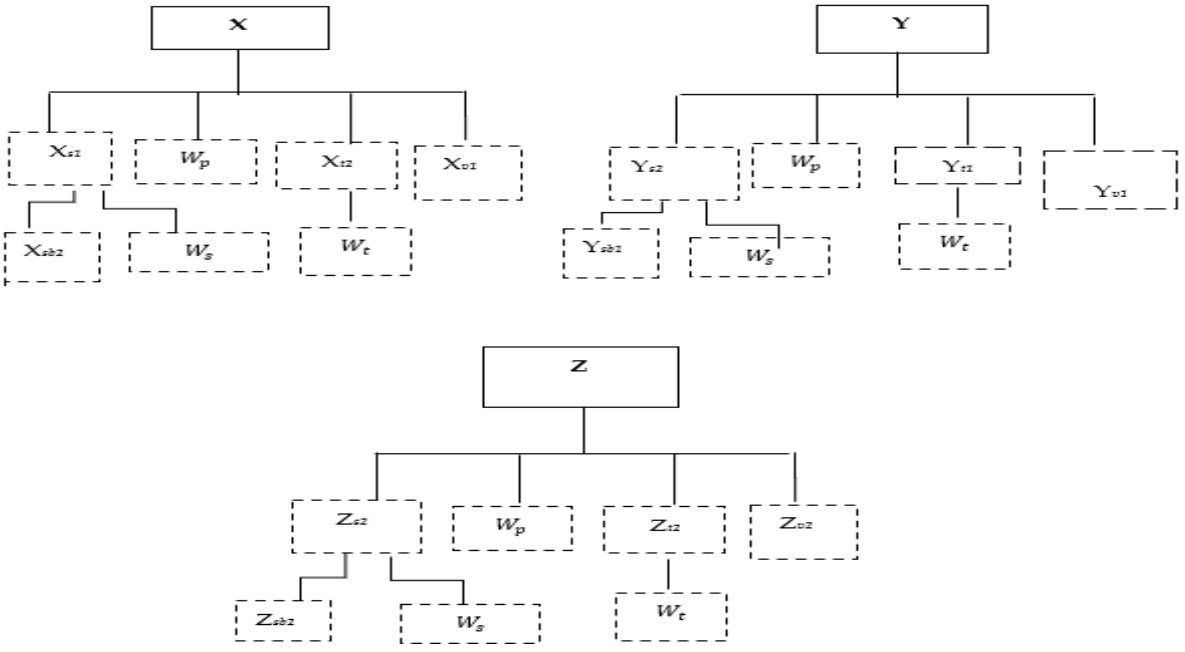
$$\boldsymbol{g_u} \leq \boldsymbol{t}_{\boldsymbol{u5}}^{-}; \text{ 2S, 3S ve 4S sınıflarındaki her } \boldsymbol{u} \text{ için } \boldsymbol{u} = \mathbf{1}, \mathbf{2}, \ldots ; \boldsymbol{s} = [\mathbf{2}, \mathbf{5}] \quad (12)$$

Alternatif seçeneğinin u. kriter için tercih aralığındaki kriter değerlerinin negatif ve pozitif yönündeki sapmaları sırasıyla d_{us}^{-} ve d_{us}^{+} ifade etmektedir. Mutlak değeri en küçük sapma değerine eşit olan alternatif en iyi seçenek olarak belirlenir.

4. Problem Tanımı

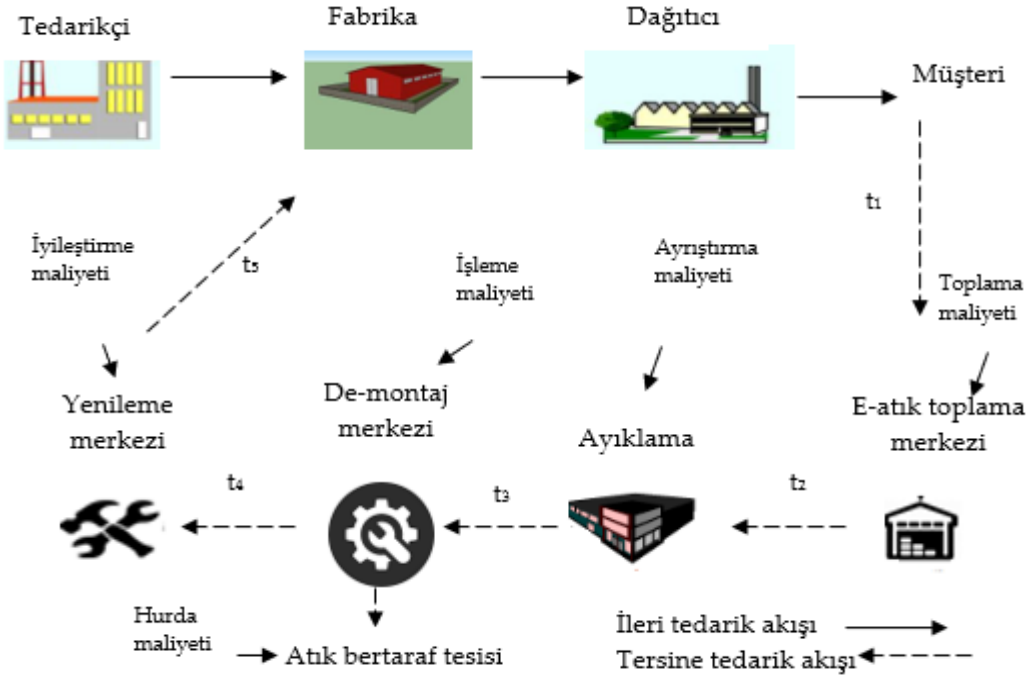
Bu çalışmada Kapalı Döngü Tedarik Zinciri Problemi için yeniden üretilebilir ve ortak parçalara sahip 3 farklı X, Y ve Z ürünleri ele alınmıştır. Yeniden üretilebilir ürünler 4 farklı parça, 2 farklı bileşen tipi ve 2 farklı alt parçadan oluşmaktadır. Yeniden üretilebilir X, Y ve Z ürünleri için kullanılması gereken parça ve bileşenler Şekil 2'de gösterilmiştir. Şekil 2'de belirtilen X, Y ve Z ürünlerine ait özellikler aşağıda mevcuttur.

- ✓ X, Y ve Z olmak üzere üç çeşit ürün modeli mevcuttur ve X, Y, Z ürününe ait parçalar X_{ijk}, Y_{ijk}, Z_{ijk} biçiminde tanımlanmıştır.
- ✓ X, Y, Z ürünleri için W ortak parça olarak tanımlanıp üç ürün için W_{ij} parçaları ortaktır.
- ✓ $i \in I$ ve $I = \{p, s, t, v\}$ olmak üzere i parça türünü ifade etmektedir.
- ✓ $j \in J$ ve $J = \{a, b\}$ olmak üzere j bileşen tipini ifade etmektedir.
- ✓ $k \in K$ ve $K = \{1, 2\}$ olmak üzere k (varsa) alt parça türünü ifade etmektedir.



Şekil 2. X, Y ve Z ürünleri için parça ve bileşenler

Tersine tedarik ağı müşteri- dağıtım merkezi arasında faaliyet gösterip ileri tedarik ağıyla bütünleşmesi sonucu KDTZ oluşmaktadır ve Şekil 3' de problemin ileri ve tersine tedarik akışları gösterilmektedir. Bu çalışmada örnek alınan 3 ürününün Kapalı Döngü Tedarik Zinciri'nde geri kazanım süresinin minimizasyonunun, toplam kârın ve geri kazanım oranının maksimizasyonunun sağlanmasıyla birlikte optimal ürün seçeneğinin belirlenmesi amaçlanmaktadır.



Şekil 3. KDTZ probleminin ileri ve tersine tedarik akışları

4.1. Çok Amaçlı Optimizasyon Modeli

Kapalı Döngü Tedarik Zinciri ağ tasarımı dikkate alındığında birimler arasında toplama, ayrıştırma, işleme, iyileştirme ve hurda maliyetleri olduğu, tersine tedarik akışında t_1, t_2, t_3, t_4 ve t_5 geri kazanım işlem sürelerinin bulunduğu görülmektedir. Model varsayımları kapsamında, müşterilerden geri dönen ürünlerin aynı kullanım süresinde ve aynı performansta olmadığı belirtilmiştir. Diğer varsayımlarla beraber de-montaj merkezinde farklı kullanım süresinde olan parçalar için kalite sınıfı (q) oluşturulmuştur. Kalite sınıfına göre q değeri 0 (düşük kalite) veya 1 (iyi kalite) değerlerini almıştır. De-montaj merkezine gelen ürünlerin yenileme merkezine gönderilen ürün oranı Q_1, de-montaj merkezinden atık merkezine gönderilerek bertaraf edilen ürün oranı ise $(1 - Q_1)$' dir. Diğer bir yandan müşterilerin iade ettiği X, Y ve Z ürünlerinin atık toplama merkezine ulaşması için gereken maliyet toplama maliyeti, farklı kullanım yüzdelerine göre sınıflama yapılması için gereken maliyet ayrıştırma maliyeti, de-montaj işlemiyle kalite durumunun değerlendirilmesi için gereken maliyet işleme maliyeti, geri kazandırılabilir ürünlerin yenileme merkezinde iyileştirilmesi için gereken maliyet iyileştirme maliyeti ve kalite düzeyi düşük parçaların çevreye zarar vermeden bertaraf edilme maliyeti hurda maliyeti olarak tanımlanmıştır. Bu bilgiler doğrultusunda KDTZ problemi için varsayımlar aşağıda belirtilmiştir.

- ✓ Müşterilerin iade ettiği X, Y ve Z ürünlerinin kullanım süreleri için 3 farklı dönem ele alınmıştır. Bu dönemler, 0-1 yıl, 1-3 yıl ve 3 yıldan daha uzun süre kullanımdır.
- ✓ Yeniden üretilen ürünlerin performansı fabrikadaki sıfır ürünlerin performansıyla aynıdır.
- ✓ Müşterilerin, geri kazandırılmış ürünlere olan talebinin pozitif olduğu dikkate alınmıştır.
- ✓ KDTZ merkezlerine gelen X, Y ve Z ürünleri için mevcut kapasitenin bulunduğu öngörülmüştür.
- ✓ Yenileme merkezinde iyileştirme işlemi yapılan parçalara ayrıca ek bileşen ihtiyacı duyulmadığı varsayılmıştır.

- ✓ Yok satma durumunun olmadığı varsayılarak elde bulundurmama maliyeti dikkate alınmamıştır.
- ✓ İade edilen ürünlerin geri kazandırılma süreci sonunda doğrudan satışa sunulduğu varsayılarak elde bulundurma maliyeti dikkate alınmamıştır.
- ✓ Satış geliri ve maliyetler dolar ($) olarak alınmıştır.
- ✓ KDTZ merkezleri arasındaki geri kazanım süreleri gün olarak alınmıştır.

KDTZ probleminde ele alınan X, Y ve Z ürünleri için çok amaçlı optimizasyon modeli önerilmiştir. Modelde kullanılan kısaltmalar ve indisler aşağıda verilmiştir:

Notasyonlar:

$R_{Xi,Yi,Zi}$: X, Y ve Z ürünlerinin i. dönemdeki iade edilen ürün miktarı $i \in \mathrm{I} = \{1, 2, 3\}$
$R_{T_{Xi,Yi,Zi}}$: X, Y ve Z ürünlerinin birim toplama maliyeti $i \in \mathrm{I} = \{1, 2, 3\}$
$R_{A_{Xi,Yi,Zi}}$: X, Y ve Z ürünlerinin birim ayrıştırma maliyeti $i \in \mathrm{I} = \{1, 2, 3\}$
$R_{K_{Xi,Yi,Zi}}$: X, Y ve Z ürünlerinin birim işleme maliyeti $i \in \mathrm{I} = \{1, 2, 3\}$
$R_{L_{Xi,Yi,Zi}}$: X, Y ve Z ürünlerinin birim iyileştirme maliyeti $i \in \mathrm{I} = \{1, 2, 3\}$
$R_{H_{Xi,Yi,Zi}}$: X, Y ve Z ürünlerinin birim hurda maliyeti $i \in \mathrm{I} = \{1, 2, 3\}$
$S_{Xi,Yi,Zi}$: X, Y ve Z ürünlerinin birim satış geliri $i \in \mathrm{I} = \{1, 2, 3\}$
$R_{R_{Xi,Yi,Zi}}$: X, Y ve Z ürünlerinin geri dönüşüm oranı $i \in \mathrm{I} = \{1, 2, 3\}$
$P_{Xi,Yi,Zi}(q)$: X, Y ve Z ürünlerinin q sınıfı için geri kazanım oranı
$t_{xj,yj,zj}$: X, Y ve Z ürünleri için müşteri- fabrika arasındaki geri kazanım süresi (j =1, 2, 3, 4, 5 süre(gün))
q: kalite sınıfı $q \in Q = \{0,1\}$
Q_1: de-montaj merkezinden yenileme merkezine gönderilen ürün oranı
$(1 - Q_1)$: de-montaj merkezinden atık merkezine gönderilen ürün oranı

Modelde kullanılan indisler:

i: Ürünün kullanım süresi indisi $i \in \mathrm{I} = \{1, \dots, N_i\}$
j: Geri kazanım süresi indisi $j \in \mathrm{J} = \{1, \dots, N_j\}$
Çok amaçlı optimizasyon modeli için hedefler aşağıda belirtilmiştir:

Max f_1: Toplam kâr = Toplam satış geliri − (Toplam toplama maliyeti + toplam ayrıştırma maliyeti + toplam işleme maliyeti + toplam iyileştirme maliyeti + toplam hurda maliyeti)

$$= \sum_{i=1}^{3} S_{Xi,Yi,Zi} \cdot R_{Xi,Yi,Zi} - (\sum_{i=1}^{3} R_{T_{Xi,Yi,Zi}} \cdot R_{Xi,Yi,Zi} + \sum_{i=1}^{3} R_{A_{Xi,Yi,Zi}} \cdot R_{Xi,Yi,Zi} +$$
$$\sum_{i=1}^{3} R_{K_{Xi,Yi,Zi}} \cdot R_{Xi,Yi,Zi} \cdot Q_1 \cdot R_{R_{Xi,Yi,Zi}} + \sum_{i=1}^{3} R_{Xi,Yi,Zi} \cdot R_{L_{Xi,Yi,Zi}} \cdot Q_1 \cdot R_{R_{Xi,Yi,Zi}} +$$
$$\sum_{i=1}^{3} R_{Xi,Yi,Zi} \cdot (1 - Q_1) \cdot R_{H_{Xi,Yi,Zi}}) \quad (13)$$

f_1 hedefi toplam kâr olarak tanımlanıp toplam satış gelirinden toplama, ayrıştırma, işleme, iyileştirme ve hurda maliyetlerinin toplam değerinden farkını ifade etmektedir. Ayrıca i indisi 0-1 yıl, 1-3 yıl, 3 ve daha fazla yıl kullanım süresi olarak tanımlanıp sırasıyla 1, 2 ve 3 değerlerini almaktadır.

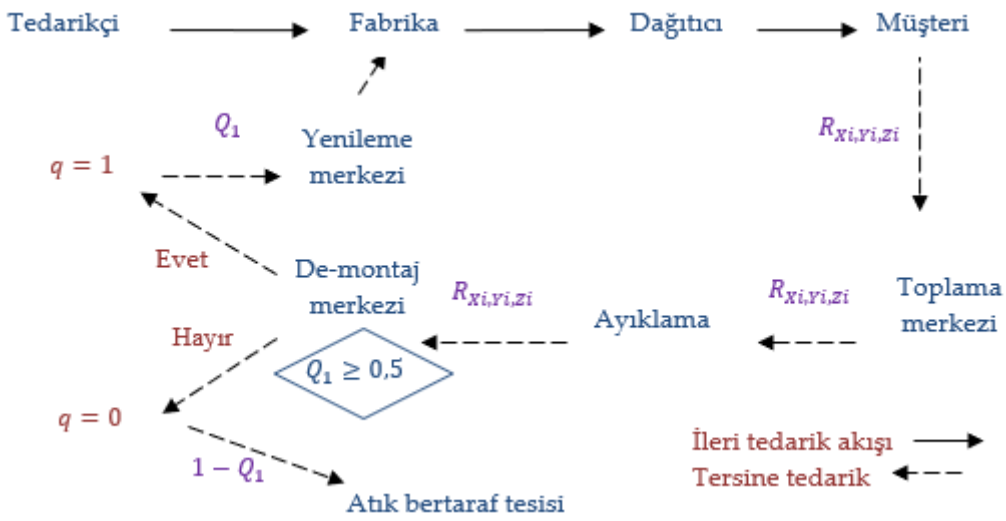
Min f_2: Toplam geri kazanım süresi = $\sum_{j=1}^{5} t_{xj,yj,zj}$ (14)

f_2 hedefi toplam geri kazanım süresi olarak tanımlanıp müşteri-toplama merkezi, toplama merkezi-ayıklama bölümü, ayıklama bölümü-de-montaj merkezi, de-montaj merkezi-yenileme merkezi, yenileme merkezi-fabrika arasında geçen sırasıyla t_1, t_2, t_3, t_4 ve t_5 sürelerinin toplamı olarak hesaplanmaktadır. t_j parametresindeki j indisi ise müşteri-fabrika arasındaki geri kazanım süresi olarak tanımlanıp sırasıyla 1, 2, 3, 4 ve 5 değerlerini almaktadır.

Max f_3: Geri kazanım oranı= $P_{Xi,Yi,Zi}(q)$ (15)

f_3 hedefi ise geri kazanım oranı olarak tanımlanıp X, Y ve Z ürünlerinin (q) kalite sınıfı için yenileme merkezine gönderilen ürün oranı Q_1, atık merkezine gönderilen ürün oranı $(1 - Q_1)$'na eşit veya 0,5'ten fazla olması durumunda q =1 (iyi kalite) ve olasılık değeri $P_{Xi,Yi,Zi}(q = 1)$, Q_1 yüzde oranının 0,5'ten küçük olması durumunda q =0 (düşük kalite) ve olasılık değeri $P_{Xi,Yi,Zi}(q = 0)$'dir. Şekil 4'te KDTZ

problemi için (q) kalite sınıfındaki ürün oranları gösterilmiştir.



Şekil 4. KDTZ problemi için (q) kalite sınıfındaki ürün oranları

Müşterilerden iade edilen X ürününün 0-1 yıl, 1-3 yıl, 3 ve daha fazla yıl kullanım süresindeki ürün miktarı sırasıyla R_{X1}, R_{X2} ve R_{X3}, Y ürününün 0-1 yıl, 1-3 yıl, 3 ve daha fazla yıl kullanım süresindeki ürün miktarı sırasıyla R_{Y1}, R_{Y2} ve R_{Y3}, Z ürününün 0-1 yıl, 1-3 yıl, 3 ve daha fazla yıl kullanım süresindeki ürün miktarı sırasıyla R_{Z1}, R_{Z2} ve R_{Z3} olup değişken olarak ifade edilmektedir.

Modele ilişkin denge kısıtları şu şekildedir:

$$Q_1 + (1 - Q_1) = 1 \quad (16)$$

$$q = 1 \quad |0,5 \leq Q_1 \leq 1 \quad (17)$$

$$q = 0 \quad |0 \leq Q_1 < 0,5 \quad (18)$$

$$t_{x1,y1,z1} + t_{x2,y2,z2} + t_{x3,y3,z3} + t_{x4,y4,z4} + t_{x5,y5,z5} = t_{xj,yj,zj} \quad (19)$$

Modele ilişkin negatif olmama kısıtı şu şekildedir:

$$\sum_{i=1}^{3} R_{T_{Xi,Yi,Zi}}, R_{A_{Xi,Yi,Zi}}, R_{K_{Xi,Yi,Zi}}, R_{L_{Xi,Yi,Zi}}, R_{H_{Xi,Yi,Zi}}, R_{R_{Xi,Yi,Zi}}, R_{Xi,Yi,Zi}, S_{Xi,Yi,Zi}, \sum_{j=1}^{5} t_{xj,yj,zj}, \sum_{i=1}^{3} P_{Xi,Yi,Zi}(q),$$
$$Q_1, (1 - Q_1), q \geq 0 \quad (20)$$

4.2. Sayısal Analiz

Bir önceki bölümde bahsedilen çok amaçlı optimizasyon modelini analiz edebilmek için sayısal bir örnek sunulmuştur.

Sayısal örnekle ilgili kısıtlar şunlardır:

$5000 \leq R_{X1} \leq 10500$ adet (iade edilen ürün sayısı) (21)

$5500 \leq R_{X2} \leq 16500$ adet (iade edilen ürün sayısı) (22)

$6000 \leq R_{X3} \leq 19500$ adet (iade edilen ürün sayısı) (23)

$4500 \leq R_{Y1} \leq 9850$ adet (iade edilen ürün sayısı) (24)

$5000 \leq R_{Y2} \leq 17750$ adet (iade edilen ürün sayısı) (25)

$5500 \leq R_{Y3} \leq 19200$ adet (iade edilen ürün sayısı) (26)

$4000 \leq R_{Z1} \leq 10250$ adet (iade edilen ürün sayısı) (27)

$5000 \leq R_{Z2} \leq 15800$ adet (iade edilen ürün sayısı) (28)

$6000 \leq R_{Z3} \leq 19750$ adet (iade edilen ürün sayısı) (29)

Geri kazanım olasılığını belirlemede kullanılacak Q_1 oranı, de-montaj merkezinden yenileme merkezine gönderilen ürün oranı ve $(1 - Q_1)$ oranı da de-montaj merkezinden atık merkezine gönderilen ürün oranı olarak tanımlanmıştır. Diğer bir taraftan yenileme merkezine gönderilecek ürünlerin geri dönüşüm oranları ($R_{R_{Xi}}$, $R_{R_{Yi}}$, $R_{R_{Zi}}$) ve q kalite sınıfı için geri kazanım oranları $P_{Xi}(q)$, $P_{Yi}(q)$, $P_{Zi}(q)$ olup X, Y ve Z ürünlerinin i. dönemdeki kullanım süresi için aldığı değerler Tablo 1'de gösterilmiştir.

Tablo 1. X, Y ve Z ürünlerinin i. dönemdeki geri dönüşüm ve geri kazanım oranları

Ürünler	**X**			**Y**			**Z**		
i. dönem	1	2	3	1	2	3	1	2	3
R_R	0,8	0,757	0,709	0,781	0,746	0,694	0,806	0,741	0,689
$P(q = 1)$	0,933			0,926			0,920		
$P(q = 0)$	0,067			0,074			0,08		

Sayısal örnek için çok amaçlı optimizasyon modelinde tanımlanan X, Y ve Z ürünlerinin toplama, ayrıştırma, işleme, iyileştirme ve hurda maliyetleri ile satış gelirlerinin alt ve üst sınır değerleri Tablo 2'de gösterilmiştir.

Tablo 2. X, Y ve Z ürünlerinin parametreler için alt-üst sınır değerleri

	X		**Y**		**Z**	
Parametreler ($)	**Alt Sınır**	**Üst Sınır**	**Alt Sınır**	**Üst Sınır**	**Alt Sınır**	**Üst Sınır**
Toplama maliyeti	60	-	60	-	60	-
Ayrıştırma maliyeti	50	85	48	80	55	75
İşleme maliyeti	200	250	215	260	230	270
İyileştirme maliyeti	210	245	220	246	205	247
Hurda maliyeti	50	67	54	65	58	69
Satış geliri	750	1000	800	1000	700	1000

X, Y ve Z ürünlerinin toplam geri kazanım süresinin hesaplanmasında kullanılan ve Şekil 3'teki KDTZ ağ tasarımında ifade edilen $t_{1,2,3,4,5}$ sürelerinin alt-üst sınır değerleri Tablo 3'te ifade edilmiştir. Ayrıca X, Y ve Z ürünleri için müşteri-fabrika arasındaki geri kazanım süresinin 90 günü geçmemesi ($t_{xj,yj,zj} \leq$ 90) istenmektedir.

Tablo 3. X, Y ve Z ürünlerinin $t_{1,2,3,4,5}$ sürelerinin alt-üst sınır değerleri

	X		Y		Z	
Geri kazanım süreleri (gün)	**Alt Sınır**	**Üst Sınır**	**Alt Sınır**	**Üst Sınır**	**Alt Sınır**	**Üst Sınır**
t_1	10	12	8	12	10	12
t_2	8	10	8	12	11	12
t_3	10	18	12	18	14	18
t_4	15	26	16	26	18	26
t_5	15	24	16	24	17	24

4.3. LPP Yaklaşımı ile Modelin Çözümü

Optimizasyon problemi için X, Y ve Z ürünleri arasından optimum seçeneğin belirlemesinde kullanılan kriterler aşağıda verilmiştir. Kriterler:

1. Toplam kâr
2. Toplam geri kazanım süresi

3. Geri kazanım oranı

Verilen kriterler çok amaçlı programlama modelinde ulaşılması istenen hedeflere göre belirlenmiş olup toplam kâr ($\boldsymbol{f_1}$) ve geri kazanım oranı ($\boldsymbol{f_3}$) Sınıf 2S (büyük olan daha iyi), toplam geri kazanım süresi ($\boldsymbol{f_2}$) ise Sınıf 1S (küçük olan daha iyi) fonksiyonunda yer almaktadır.
Çok amaçlı doğrusal programlama modeli için $\boldsymbol{f_1}$, $\boldsymbol{f_2}$ ve $\boldsymbol{f_3}$ kriter değerlerinin hesaplanmasında alt ve üst sınır değerleri verilen maliyet, satış geliri, iade edilen ürün sayıları ve geri kazanım sürelerinin beklenen değerleri göz önünde bulundurularak POM- QM ve Excel Solver programları kullanılmıştır. POM- QM ve Excel Solver programları kullanılarak hesaplanmış X, Y ve Z ürünleri için kriter değerleri Tablo 4'de mevcuttur.

Tablo 4. Önerilen modeldeki X, Y ve Z ürünleri için kriter değerleri

Ürünler	Toplam kâr ($)	Toplam geri kazanım süresi (gün)	Geri kazanım oranı (%)
X	17292898	75	93,3
Y	17728284	76	92,6
Z	15914634	82	92

Kriter ağırlıklarının hesaplanmasında kullanılacak olan tercih aralıkları Tablo 4'teki değerler ile belirlenmektedir. Toplam kâr için tercih aralıkları ve değerleri Tablo 5'te, toplam geri kazanım süresi (f_2) için tercih aralıkları ve değerleri Tablo 6'da, geri kazanım oranı (f_3) için tercih aralıkları ve değerleri Tablo 7'de gösterilmiştir.

Tablo 5. Kriter 1'e ilişkin tercih aralıkları ve sınır değerleri

Kriter 1(u=1): Toplam kâr			
Sınıf 2S			
Tercih	**Tercih aralığı**	**Limit**	**Limit değeri**
Kabul Edilemez	$\leq t_{u5}^{-}$		
Hiç istenmeyen	(t_{u5}^{-}, t_{u4}^{-})	t_{u5}^{-}	16000000
İstenmeyen	(t_{u4}^{-}, t_{u3}^{-})	t_{u4}^{-}	16500000
Tolere edilebilir	(t_{u3}^{-}, t_{u2}^{-})	t_{u3}^{-}	17000000
İstenen	(t_{u2}^{-}, t_{u1}^{-})	t_{u2}^{-}	17500000
İdeal	$\geq t_{u1}^{-}$	t_{u1}^{-}	18000000

Tablo 6. Kriter 2'ye ilişkin tercih aralıkları ve sınır değerleri

Kriter 2(u=2): Toplam geri kazanım süresi			
Sınıf 1S			
Tercih	**Tercih aralığı**	**Limit**	**Limit değeri**
Kabul edilemez	$\geq t_{u5}^{+}$		90
Hiç istenmeyen	(t_{u4}^{+}, t_{u5}^{+})	t_{u5}^{+}	85
İstenmeyen	(t_{u3}^{+}, t_{u4}^{+})	t_{u4}^{+}	80
Tolere edilebilir	(t_{u2}^{+}, t_{u3}^{+})	t_{u3}^{+}	75
İstenen	(t_{u1}^{+}, t_{u2}^{+})	t_{u2}^{+}	70
İdeal	$\leq t_{u1}^{+}$	t_{u1}^{+}	

Tablo 7. Kriter 3'e ilişkin tercih aralıkları ve sınır değerleri

Kriter 3(u=3): Geri kazanım oranı			
Sınıf 2S			
Tercih	**Tercih aralığı**	**Limit**	**Limit değeri**
Kabul Edilemez	$\leq t_{u5}^{-}$		
Hiç istenmeyen	(t_{u5}^{-}, t_{u4}^{-})	t_{u5}^{-}	91,5
İstenmeyen	(t_{u4}^{-}, t_{u3}^{-})	t_{u4}^{-}	92
Tolere edilebilir	(t_{u3}^{-}, t_{u2}^{-})	t_{u3}^{-}	92,5
İstenen	(t_{u2}^{-}, t_{u1}^{-})	t_{u2}^{-}	93
İdeal	$\geq t_{u1}^{-}$	t_{u1}^{-}	93,5

Toplam kâr (f_1), toplam geri kazanım süresi (f_2) ve geri kazanım oranı (f_3) kriterleri için sınıf fonksiyonları ve farklı derecelerdeki aralıkların hedef değerleri (t_u (x) değerleri) belirlendikten sonra LPP ağırlık algoritması kullanılarak ağırlıklar hesaplanmıştır.

Üç kriter için hesaplanan $\widetilde{w}_{us}^{+}$ ve $\widetilde{w}_{us}^{-}$ ağırlıkları Tablo 8'de verilmiştir. Tablo 9'da X, Y ve Z ürünleri için toplam sapma değerleri verilmiş olup en küçük sapma değeri en iyi seçeneği belirlemektedir. Buna göre ürünleri sıralarsak birinci sırada Y, ikinci sırada X ve üçüncü sırada Z ürünü yer almıştır. Çok amaçlı optimizasyon modeli için normalleştirme işlemi yapıldığında X ürününün ağırlığı (w_1) 0,3663, Y ürününün ağırlığı (w_2) 0,3795 ve Z ürününün ağırlığı (w_3) 0,2537 olarak hesaplanmıştır. Bu durumda toplam sapma değeri en küçük olan Y ürünü optimal seçenek olarak belirlenmiştir.

Tablo 8. LPP ağırlık algoritmasıyla hesaplanan ağırlıklar

Kriterler	$\widetilde{w}_{u2}^{+}$	$\widetilde{w}_{u3}^{+}$	$\widetilde{w}_{u4}^{+}$	$\widetilde{w}_{u5}^{+}$	$\widetilde{w}_{u2}^{-}$	$\widetilde{w}_{u3}^{-}$	$\widetilde{w}_{u4}^{-}$	$\widetilde{w}_{u5}^{-}$
u=1					0.0529	0.1164	0.2560	0.5634
u=2	0.0530	0.1177	0.2591	0.5701				
u=3					0.0529	0.1164	0.2560	0.5634

Tablo 9. X, Y ve Z ürünleri için toplam sapma değerleri

Ürünler	**Toplam sapma değerleri**	**Sıralama**
X	902499	2
Y	869875	1
Z	1300209	3

Sonraki bölümde model için belirlenen parametrelerin hedef değerlerine olan etkisinin analiz edilebilmesi için üç farklı senaryo uygulanmış ve her bir senaryonun optimal çözüm değerleri hesaplanmıştır.

4.4. Senaryo 1 Uygulaması

Çok amaçlı optimizasyon modelinde senaryo 1 için X ve Z ürünlerinin iade ürün miktarı sabit tutulup Y ürününün iade edilen ürün miktarı (R_{Yi}) %10 azaldığında ve Y ürününün müşteri-toplama merkezi arası geçen geri kazanım süresi (t_1) sabit tutulup X ve Z ürünlerinin müşteri-toplama merkezi arası geçen geri kazanım süresi (t_1) iki kat artığında sonucun model çözümüne olan etkisi analiz edilmiştir. X, Y ve Z ürünlerinin değerlendirme sonucu için POM-QM ve Excel Solver programları kullanılarak hesaplanan X, Y ve Z ürünlerinin optimal değerleri Tablo 10'da mevcuttur.

Tablo 10'a bakıldığında Y ürününün iade edilen ürün miktarının azalması (R_{Yi}), geri kazanım süresine ve geri kazanım oranına etki göstermeyip toplam kâr miktarının düşmesine sebep olduğu söylenebilir. Y ürünü için toplam kâr miktarı 17.728.284$'dan 15.955.451$'a düşmüş olup toplam kâr için ürünler

arasında Tablo 4'e göre karşılaştırma yapıldığında Y ürünü ilk sırada iken ikinci sıraya geçmiş olup X ürünü toplam kâr durumu açısından optimal seçenek olarak belirlenmiştir. Ayrıca X ve Z ürünlerinin müşteri-toplama merkezi arası geçen geri kazanım süresinin (t_1) iki kat artması durumunda Z ürününün toplam geri kazanım süresi 93 gün olup hedef değerinde belirtilen 90 gün sınırını aşarak çok amaçlı programlama modelindeki ikinci hedefin yerine getirilememesine neden olmaktadır.

Tablo 10. Senaryo 1 için optimal çözüm değerleri

	X	**Y**	**Z**
f_1 **(\$)**	17292898	15955451	15914634
f_2**(gün)**	86	76	93
f_3**(%)**	93,3	92,6	92

4.5. Senaryo 2 Uygulaması

Çok amaçlı optimizasyon modelinde senaryo 2 için X ve Y ürünlerinin satış geliri sabit tutulup Z ürününün satış geliri (S_{Zi}) %10 arttığında ve X ve Z ürünlerinin iyi kaliteli ürün oranı sabit tutulup Y ürününün iyi kaliteli ürün oranı %1 arttığında sonucun model çözümüne olan etkisi analiz edilmiştir. X, Y ve Z ürünlerinin değerlendirme sonucu için POM-QM ve Excel Solver programları kullanılarak X, Y ve Z ürünlerinin optimal değerleri Tablo 11'te mevcuttur.

Tablo 11'e bakıldığında Z ürününün satış geliri (S_{Zi}) %10 arttırılması Z ürününün toplam kâr değerini yükselttiği ve ürünler arasında toplam kâr değerine göre kıyasla yapıldığında Y ürününün ilk sıradan ikinci sıraya gerilediği Z ürününün üçüncü sıradan ilk sıraya geçerek optimal seçenek olduğu belirlenmiştir. Ayrıca Y ürününün iyi kaliteli ürün oranı %1 artması sonucunda Tablo 4'e kıyasla ürünler arasındaki geri kazanım oranının büyükten küçüğe doğru sıralamasını Y, X ve Z olarak değiştirmiştir.

Tablo 11. Senaryo 2 için optimal çözüm değerleri

	X	**Y**	**Z**
f_1 **(\$)**	17292898	17728284	17827228
f_2**(gün)**	75	76	82
f_3**(%)**	93,3	93,6	92

4.6. Senaryo 3 Uygulaması

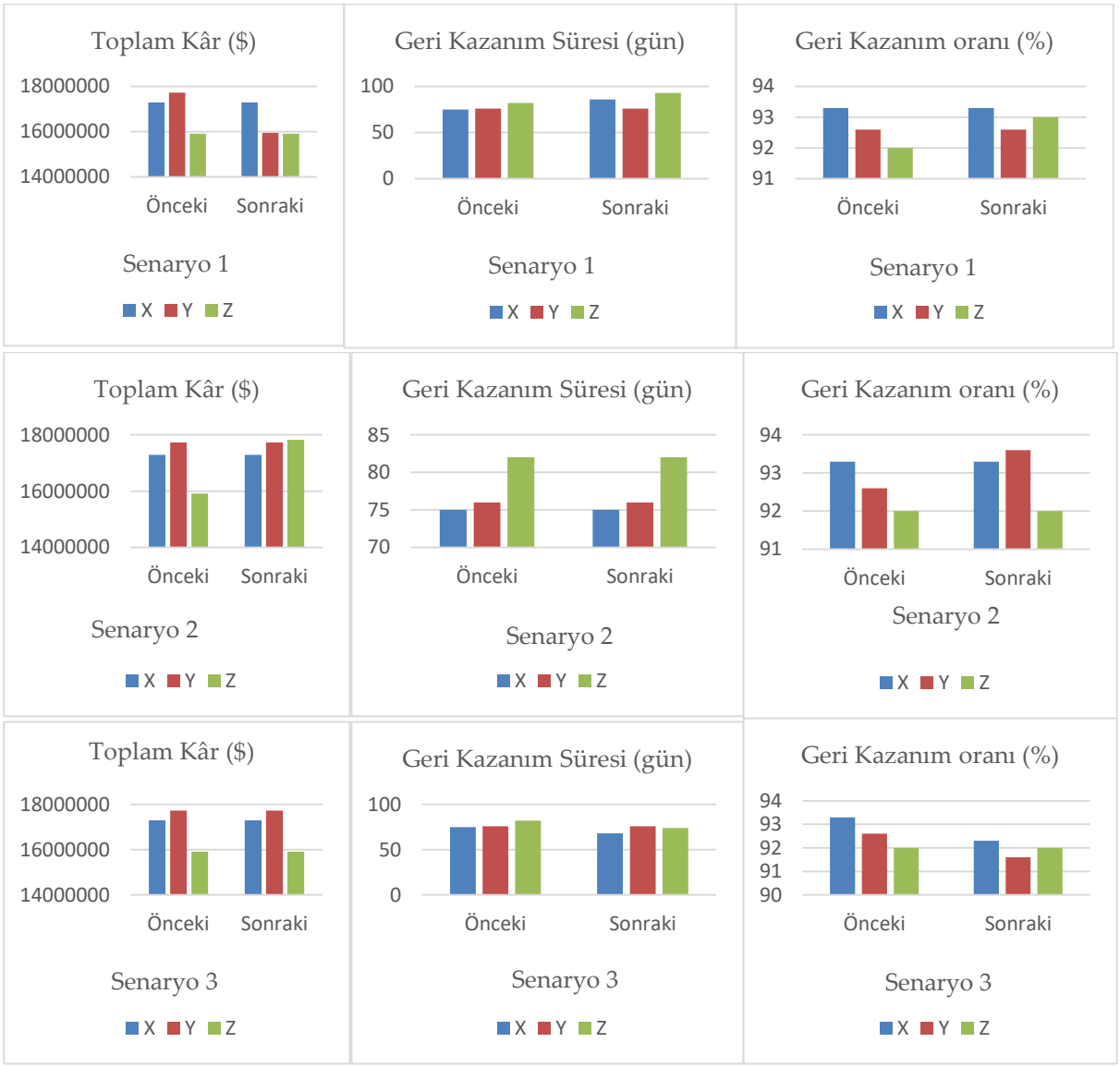
Çok amaçlı optimizasyon modelinde senaryo 3 için Z ürününün iyi kaliteli ürün oranı sabit tutulup X ve Y ürünlerinin iyi kaliteli ürün oranı %1 azaldığında ve Y ürününün ayıklama de-montaj merkezi arası geçen geri kazanım süresi sabit tutulup X ve Z ürünlerinin ayıklama de-montaj merkezi arası geçen geri kazanım süresi (t_3) %50 azaldığında sonucun model çözümüne olan etkisi analiz edilmiştir. X, Y ve Z ürünlerinin değerlendirme sonucu için POM-QM ve Excel Solver programları kullanılarak hesaplanan X, Y ve Z ürünlerinin optimal değerleri Tablo 12'de mevcuttur.

X ve Y ürünlerinin iyi kaliteli ürün oranlarının %1 azalması Tablo 4'e kıyasla ürünler arasındaki geri kazanım oranının büyükten küçüğe doğru sıralamasını X, Z ve Y olarak değiştirmiştir. Ayrıca X ve Z ürünlerinin ayıklama de-montaj merkezi arası geçen geri kazanım süresinin (t_3) %50 azalması sonucunda minimum geri kazanım süresine sahip X seçeneği optimal seçenek olarak belirlenirken Tablo 4'teki geri kazanım süresine göre ikinci sırada yer alan Y ürününün Senaryo 3 uygulamasında üçüncü sıraya geçtiği görülmektedir. Bu durum ile X ve Z ürünlerinin Y ürününe göre müşterilere daha kısa sürede ulaştığı ve zaman açısından tasarruf elde edildiği tespit edilmiştir.

Tablo 12. Senaryo 3 için optimal çözüm değerleri

	X	**Y**	**Z**
f_1 **(\$)**	17292898	17728284	15914634
f_2**(gün)**	68	76	74
f_3**(%)**	92,3	91,6	92

X, Y ve Z ürünleri için toplam kâr (f_1), geri kazanım süresi (f_2) ve geri kazanım oranının (f_3) çok amaçlı programlama modelindeki kriter değerleriyle Senaryo 1,2 ve 3 analizleri sonrası aldığı değerlerin grafiksel olarak gösterimi Şekil 5'te verilmiştir.



Şekil 5. X, Y ve Z ürünlerinin Senaryo 1,2 ve 3 analizlerinden önceki ve sonraki değerleri

5. Sonuçlar ve Öneriler

Kapalı Döngü Tedarik Zinciri için ileri ve tersine olmak üzere iki tip akış olduğu ve tersine akışta müşterilerden geri dönen ürünlerin miktarı, kalite düzeyleri, geri dönüşüm süresi ve toplam kazanç gibi belirsizliklerin yer aldığı belirtilmiştir. Kapalı Döngü Tedarik Zinciri'nde bu belirsizliklerden yola çıkarak toplam geri kazanım kâr miktarının maksimizasyonu ve toplam geri kazanım süresinin minimizasyonunu hedefleyen bir optimizasyon problemini ortaya çıkardığını ifade ederken problem çok amaçlı doğrusal programlama modeli olarak ele alınmıştır. Bu kapsamda sayısal bir örnekle uygulama için farklı ve ortak parçalara sahip 3 farklı ürün göz önüne alınmıştır. Modeldeki X, Y ve Z ürünlerinden optimum seçeneğin belirlenmesinde Doğrusal Fiziki Programlama yöntemi kullanılmış ve uygulama adımları açıklanmıştır. Doğrusal Fiziki Programlama yaklaşımı için toplam kâr, toplam geri kazanım süresi ve geri kazanım oranı kriterleri kullanılmıştır. Kriterlerin hedef değerleri ve tercih aralıkları kullanılarak üç ürün için toplam sapma miktarları hesaplanmıştır. Buna göre en küçük sapma değerine sahip Y ürünü optimal seçenek olarak belirlenmiştir. Bu durumda ileride tedarik edilecek stok miktarına karar verilirken Y ürün sayısının artırılmasının işletmeye stratejik açıdan fayda sağlayacağı

öngörülmektedir.
Çok amaçlı doğrusal programlama modeli için f_1, f_2 ve f_3 kriter değerlerinin hesaplanmasında alt ve üst sınır değerleri verilen maliyet, satış geliri, iade edilen ürün sayıları ve geri kazanım sürelerinin beklenen değerleri göz önünde bulundurularak POM- QM ve Excel Solver programları kullanılmıştır. Programdan elde edilen verilere bakıldığında X, Y ve Z ürünleri için maksimum kâr değerleri elde edilerek f_1 hedefi gerçekleştirilmiş, geri kazanım sürelerine bakıldığında optimal değerlerin 90 günü geçmemesi istendiğinden f_2 hedefine ulaşılmış ve maksimum geri kazanım oranı elde edilerek f_3 hedefinin de gerçekleştirilmesi sağlanmıştır. Çalışma iade edilen ürün sayısı, satış geliri, geri kazanım süresi ve geri kazanım oranındaki değişiklikler üzerinden üç farklı senaryo kullanılarak analiz edilmiştir. Senaryo 2'de Z ürünü için toplam kârın ve Y ürünü için geri kazanım oranının maksimum değer, Senaryo 3'te X ürünü için geri kazanım süresinin minimum değer aldığı belirlenmiştir.

Yazar Katkıları: Bu çalışmada G.Ö, literatür araştırması ve uygulama kısmının yapılması, H.K.A, çalışmanın yönetilmesi ve kontrol edilmesi; konularında katkı sağlamışlardır.

Finansman: Bu araştırma dışarıdan fon almadı.

Çıkar çatışmaları: Yazarlar çıkar çatışması beyan etmemektedir.

Kaynakça

[1] Amin, S. H., Zhang, G., "A three-stage model for closed-loop supply chain configuration under uncertainty," International Journal of Production Research, vol.51, no. 5, 2013, pp.1405-1425. Doi: https://doi.org/10.1080/00207543.2012.693643.

[2] Blog, L., "Tedarik zinciri nedir," URL:https://blog.logo.com.tr/tedarik-zinciri-ve-tedarik-zinciri-yonetimi-nedir-temel-evreleri-vefaydalari-nelerdir, visited on Jan.10,2020.

[3] Çalık, A., "Kapalı Döngü Tedarik Zinciri Optimizasyonu için Çok Kriterli Karar Verme Yöntemleri ile Yeni Etkileşimli Bulanık Programlama Yaklaşımları," Doktora tezi, Selçuk Üniversitesi Fen Bilimleri Enstitüsü, 2016.

[4] Gupta, S., Imtanavanich, P., "Linear Physical Programming Approach for a Disassembly-To Order System under Stochastic Yields and Product's Deterioration, Proceedings of the 2006 POMS Meeting", MA 2006b, 2006, pp. 004-0213.

[5] Fleischmann, M., Beullens, P., Bloemhof-Ruwaard, J. M., Wassenhove, L., "The Impact of Product Recovery on Logistics Network Design," Production Operations Management, vol.10, 2001, pp. 156–173.

[6] Hajipour, V., Tavana, M., Caprio, D. D., Akhgar, M., "An Optimization Model for Traceable Closed-Loop Supply Chain Networks," Applied Mathematical Modelling, vol. 71, 2019, pp 673-699. [7] Ilgın, M. A., Değirmenci, E., Demirtepe, S., "Personel Seçim Problemi için Doğrusal Fiziki Programlama Yaklaşımı," İstanbul Ticaret Üniversitesi Fen Bilimleri Dergisi, vol. 28, 2015, pp. 15-28. Doi: https://doi.org/10.1016/j.apm.2019.03.007.

[7] Ilgın, M. A., Değirmenci, E., Demirtepe, S., "Personel Seçim Problemi için Doğrusal Fiziki Programlama Yaklaşımı," İstanbul Ticaret Üniversitesi Fen Bilimleri Dergisi, vol. 28, 2015, pp. 15-28.

[8] Jayaraman, V., Guide, V., Srivastav, R., "A Closed Loop Logistics Model for Remanufacturing," Journal of the Operational Research Society, vol. 50, 1999, pp. 497-508. Doi: https://doi.org/10.2307/3009998.

[9] Kannan, G., Soleimani, H., Kannan, D., "Analysis of closed loop supply chain using genetic algorithm and particle swarm optimisation," International Journal of Production Research, vol. 47, 2009, pp. 1175-1200. Doi: http://dx.doi.org/10.1080/00207540701543585.

[10] Marin, A., Pelegrin, B., "The Return Plant Location Problem: Modelling and Resolution," European Journal of Operational Research, vol. 104, 1998, pp. 375-392.

[11] Pisvaee, M. S., Torabi, S. A., "A possibilistic programming approach for closed-loop supply chain network design under uncertainty," Fuzzy Sets and Systems, vol. 161, 2010, pp. 2668-2683. Doi: 10.1016/j.fss.2010.04.010.

[12] Subulan, K., Taşan, A. S., Baykasoğlu, A., "Fuzzy mixed integer programming model for medium-term planning in a closed-loop supply chain with remanufacturing option," Journal of Intelligent and Fuzzy Systems, vol. 23, no. 6, 2012, pp. 345-368. Doi: http://dx.doi.org/10.3233/IFS-2012-0525.

[13] Zhang, Z. H., Jiang, H., Pan, X., "A Lagrangian relaxation based approach for the capacitated lot sizing problem in closed-loop supply chain," International Journal of Production Economics, vol. 140, 2012, pp. 249-2. Doi: https://doi.org/10.1016/j.ijpe.2012.01.018

Journal of Statistics & Applied Science, Volume -2 Issue -2
ISSN 2718-0999

Article

Received: date: 08.06.2021
Accepted: date: 06.10.2021
Published: date: 28.11.2021

A New Kumaraswamy Class of Generalized Distributions with Applications to Exponential Model

Mohamed Ali AHMED[1,*]

[1]Department of Statistics, Mathematics and Insurance, Al Madina Higher Institute of Management and Technology, Giza, Egypt, mrmohamedali2005@yahoo.com
Orcid: 0000-0002-8320-6631[1]
*Correspondence Author, e-mail: mrmohamedali2005@yahoo.com

Abstract: In this paper, a new class of generalized distributions, so-called the alpha power Kumaraswamy (*AK*) class, is derived, three important classes of distributions are nested by the *AK* class. Some mathematical properties are studied and a parameters estimation method using maximum likelihood (*MLE*) is obtained. A simulation study using bootstrapping approach is applied to study the alpha power Kumaraswamy-exponential (*AKE*) distribution estimators' behavior. A real data set is used to investigate the *AKE* distribution flexibility.
Keywords: the alpha power Kumarasamy distribution; moments; order statistics; maximum likelihood estimation; bootstrapping approach.

1. Introduction

Adding parameters using alpha power transformation (*APT*) is a flexible method depending on the alpha power function, [1], having the following cumulative distribution function (*CDF)* and probability density function (*PDF)* for a continuous random variable *X*, respectively

$$F_{APT}(x)=\begin{cases}\dfrac{\alpha^{F(x)}-1}{\alpha-1};\alpha>0,\alpha\neq 1,\\ F(x);\alpha=1,\end{cases}$$

and

$$f_{APT}(x)=\begin{cases}\dfrac{\log\alpha}{\alpha-1}f(x)\alpha^{F(x)};\alpha>0,\alpha\neq 1,\\ f(x);\alpha=1.\end{cases}$$

Ahmed [2] presented for the first time the *AK* distribution having the following *CDF* and *PDF*, respectively,

$$F_{\alpha K}(x)=\frac{\alpha^{1-\left(1-x^{\beta}\right)^{\theta}}-1}{\alpha-1},0<x<1;\alpha,\beta,\theta>0;\alpha\neq 1, \quad (1)$$

and

 https://doi.org/10.52693/jsas.979832

$$f_{\alpha K}(x)=\frac{\beta\theta\log\alpha}{\alpha-1}\alpha^{1-\left(1-x^{\beta}\right)^{\theta}}x^{\beta-1}\left(1-x^{\beta}\right)^{\theta-1}. \quad (2)$$

The main object of this manuscript is to derive an extended class of generalized distributions naming the *AK* class of generalized distributions depending on the *AK* distribution, Ahmed [2], as a simple generator, also it aims to study some properties of the class and gives some applications to *AKE* distribution.

The rest of this paper is organized as follows: In section 2, the generalized class is presented. In section 3, some properties are derived. In section 4, the Hazard function is given. In section 5, order statistics are obtained. In Section 6, the *MLE* method is used. In section 7, a simulation study using bootstrapping is performed for the *AKE* distribution. Finally, in Section 8, an application is investigated, practically, for the *AKE* distribution.

2. The New Class of *AK* Distributions

Many classes of generalized distributions are derived based on generating method, Wahed [3], the Kumaraswamy (*KW*) class [4] and [5], the Kummer beta class [6], the McDonald class [7] and [8], the Kumaraswamy- Kumaraswamy (*KW-KW*) class [9] and [10]. Replacing x in (1) with the generalized parent $G(x;\Lambda)$ gives

$$F(x)=\frac{\alpha^{1-\left(1-G^{\beta}(x;\Lambda)\right)^{\theta}}-1}{\alpha-1};\alpha,\beta,\theta>0;\alpha\neq1, \quad (3)$$

differentiating (3) with respect to x yields

$$f(x)=\frac{\beta\theta}{\alpha-1}\left(\log\alpha\right)\alpha^{1-\left(1-G^{\beta}(x;\Lambda)\right)^{\theta}}g(x;\Lambda)\,G^{\beta-1}(x;\Lambda)\left(1-G^{\beta}(x;\Lambda)\right)^{\theta-1}, \quad (4)$$

where $G(x;\Lambda)$ and $g(x;\Lambda)$ are the *CDF* and *pdf* of the parent distribution, Λ is the parameter vector of the parent distribution. When $\alpha=1$, the *AK* class gives Kumaraswamy (*K*w) class [4] and [5], setting $\theta=1$ gives the alpha power (*AP*) class and setting $\alpha=1$, $\theta=1$ gives the power function (*P*) class [11]. Many distributions can be derived via the class of *AK* as the alpha power Kumaraswamy exponential (*AKE*) distribution, the *CDF* and *PDF* of the *AKE* distribution, respectively, can be given by

$$F(x)=\frac{\alpha^{1-\left[1-(1-e^{-\lambda x})^{\beta}\right]^{\theta}}-1}{\alpha-1};\alpha,\beta,\theta,\lambda>0;\alpha\neq1,$$

and

$$f(x)=\frac{\beta\theta\lambda}{\alpha-1}\left(\log\alpha\right)\alpha^{1-\left(1-(1-e^{-\lambda x})^{\beta}\right)^{\theta}}e^{-\lambda x}\,(1-e^{-\lambda x})^{\beta-1}\left(1-(1-e^{-\lambda x})^{\beta}\right)^{\theta-1},$$

some density function shapes for the *AKE* distribution are given in figure 1.

2.1. The CDF Expansion

Applying the exponential expansion for (3) leads to

$$F(x)=\frac{1}{1-\alpha}\left\{1-\sum_{i=0}^{\infty}\frac{\left[1-\left(1-G^{\beta}(x;\Lambda)\right)^{\theta}\right]^{i}\left(\log\alpha\right)^{i}}{i!}\right\},\alpha,\beta,\theta>0;\alpha\neq1,$$

then, using binomial expansion gives

$$F(x)=\frac{1}{1-\alpha}\left\{1-\sum_{i=0}^{\infty}\frac{(\log\alpha)^i}{i!}\sum_{j=0}^{i}(-1)^j\binom{i}{j}\left(1-G^{\beta}(x;\Lambda)\right)^{\theta j}\right\},$$

replacing $\sum_{i=0}^{\infty}\sum_{j=0}^{i}$ with $\sum_{j=0}^{\infty}\sum_{i=j}^{\infty}$ yields

$$F(x)=\frac{1}{1-\alpha}\left\{1-\sum_{j=0}^{\infty}\sum_{i=j}^{\infty}\frac{(\log\alpha)^i}{i!}(-1)^j\binom{i}{j}\left(1-G^{\beta}(x;\Lambda)\right)^{\theta j}\right\},$$

using binomial expansion, again, leads to

$$F(x)=\frac{1}{1-\alpha}\left\{1-\sum_{j=0}^{\infty}\sum_{i=j}^{\infty}\sum_{k=0}^{\infty}(-1)^{j+k}\frac{(\log\alpha)^i}{i!}\binom{i}{j}\binom{\theta j}{k}G^{\beta k}(x;\Lambda)\right\},$$

where β is an integer, when β is real non integer yields

$$F(x)=\frac{1}{1-\alpha}\left\{1-\sum_{j=0}^{\infty}\sum_{i=j}^{\infty}\sum_{k=0}^{\infty}(-1)^{j+k}\frac{(\log\alpha)^i}{i!}\binom{i}{j}\binom{\theta j}{k}\left[1-\left(1-G(x;\Lambda)\right)\right]^{\beta k}\right\},$$

then, using binomial expansion two times gives

$$F(x)=\frac{1}{1-\alpha}\left\{1-\sum_{j=0}^{\infty}\sum_{i=j}^{\infty}\sum_{k=0}^{\infty}\sum_{p=0}^{\infty}\sum_{q=0}^{p}(-1)^{j+k+p+q}\frac{(\log\alpha)^i}{i!}\binom{i}{j}\binom{\theta j}{k}\binom{\beta k}{p}\binom{p}{q}G^{q}(x;\Lambda)\right\},$$

replacing $\sum_{p=0}^{\infty}\sum_{q=0}^{p}$ with $\sum_{q=0}^{\infty}\sum_{p=q}^{\infty}$ leads to

$$F(x)=\frac{1}{1-\alpha}\left\{1-\sum_{q=0}^{\infty}w_q\,G^{q}(x;\Lambda)\right\}, \tag{5}$$

where

$$w_q=\sum_{j=0}^{\infty}\sum_{i=j}^{\infty}\sum_{k=0}^{\infty}\sum_{p=q}^{\infty}(-1)^{j+k+p+q}\frac{(\log\alpha)^i}{i!}\binom{i}{j}\binom{\theta j}{k}\binom{\beta k}{p}\binom{p}{q}.$$

2.2. The PDF Expansion

Differentiating (5) with respect to x gives

$$f(x)=\frac{1}{1-\alpha}\left\{\sum_{q=1}^{\infty}w_q\,q\,G^{q-1}(x;\Lambda)\,g(x;\Lambda)\right\},$$

shifting q leads to

$$f(x)=\frac{1}{1-\alpha}\left\{\sum_{q=0}^{\infty}w_{q+1}(q+1)\,G^{q}(x;\Lambda)\,g(x;\Lambda)\right\},$$

then,

$$f(x)=\sum_{q=0}^{\infty} w_q^{*}\, G^{q}(x;\Lambda)\, g(x;\Lambda), \tag{6}$$

where

$$w_q^{*}=\frac{1}{1-\alpha}\, w_{q+1}(q+1). \tag{7}$$

The Condition for the *PDF* Expansion

since,

$$\sum_{q=0}^{\infty} w_q^{*} \int_{-\infty}^{\infty} G^{q}(x;\Lambda) g(x;\Lambda)\, dx = 1,$$

then,

$$\sum_{q=0}^{\infty} w_q^{*} \left[\frac{G^{q+1}(x;\Lambda)}{(q+1)} \right]_{-\infty}^{\infty} = 1,$$

hence,

$$\sum_{q=0}^{\infty} \frac{w_q^{*}}{q+1} = 1. \tag{8}$$

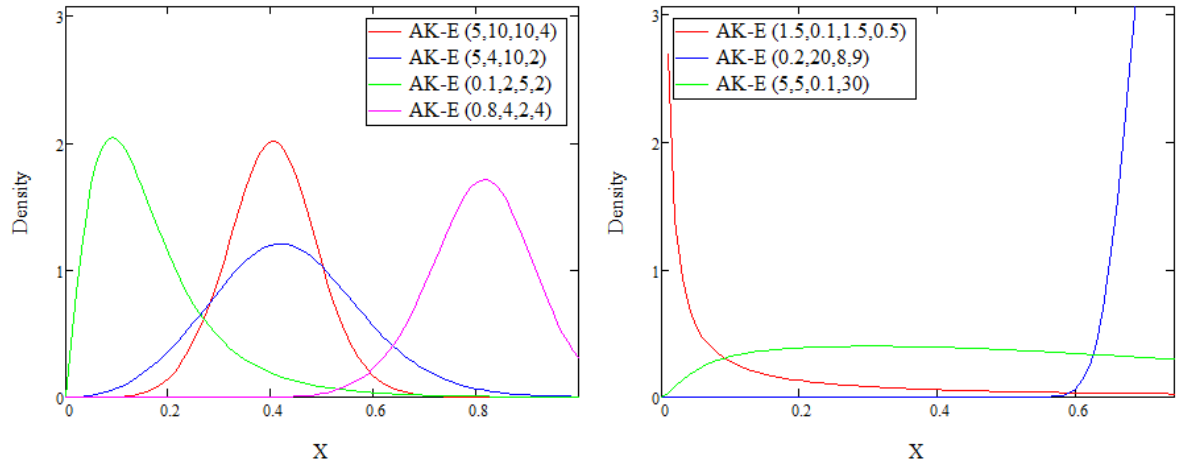


Figure 1: The *AKE* density functions

3. The *AK* Class of Distributions Properties

In this section some properties of the *AK* class of distributions will be given as follows:

3.1. The r-th Moment

Basically, a continuous random variable *X* has the following *r-th* moment [12]

$$E\left(X^{r}\right)=\int_{x} x^{r} f(x)\, dx,$$

substituting (6) into last equation yields

$$E\left(X^{r}\right)=\sum_{q=0}^{\infty}w_{q}^{*}\int_{-\infty}^{\infty}x^{r}g(x;\Lambda)G^{q}(x;\Lambda)dx, \tag{9}$$

then,

$$E\left(X^{r}\right)=\sum_{q=0}^{\infty}w_{q}^{*}\tau_{r,q,0}, \tag{10}$$

where τ is the probability weighted moment (*PWM*), Greenwood *et. al.* [13].
Obviously, setting *r=0* and using (9) leads to

$$E\left(X^{0}\right)=\sum_{q=o}^{\infty}w_{q}^{*}\int_{-\infty}^{\infty}g\left(x;\Lambda\right)G^{q}\left(x;\Lambda\right)dx,$$

then,

$$E\left(x^{0}\right)=\sum_{q=o}^{\infty}w_{q}^{*}\left[\frac{G^{q+1}\left(x;\Lambda\right)}{q+1}\right]_{-\infty}^{\infty},$$

substituting (8) into last equation gives

$$E\left(X^{0}\right)=1.$$

Using the Parent Quantile Function

Setting $G\left(x;\Lambda\right)=u$, $x=Q\left(u\right)$ and substituting into (9) gives

$$E\left(X^{r}\right)=\sum_{q=o}^{\infty}w_{q}^{*}\int_{0}^{1}Q^{r}\left(u\right)u^{q}du,$$

then,

$$E\left(X^{r}\right)=\sum_{q=o}^{\infty}w_{q}^{*}\tau_{r,q,o}.$$

3.2. The PWM

Basically, the *PWM* of a continuous random variable *X*, Greenwood *et. al.* [13], is given by

$$\tau_{r,s,0}=\int_{x}x^{r}f\left(x\right)F^{s}\left(x\right)dx,$$

substituting (5) and (6) into last equation leads to

$$\tau_{r,s,o}=\left(\frac{1}{1-\alpha}\right)^{s}\int_{-\infty}^{\infty}x^{r}g\left(x;\Lambda\right)\left[\sum_{q=o}^{\infty}w_{q}^{*}G^{q}\left(x;\Lambda\right)\right]\left[1-\sum_{q=0}^{\infty}w_{q}G^{q}(x;\Lambda)\right]^{s}dx,$$

using binomial expansion yields

$$\tau_{r,s,o}=\left(\frac{1}{1-\alpha}\right)^{s}\sum_{k=0}^{s}(-1)^{k}\binom{s}{k}\int_{-\infty}^{\infty}x^{r}g\left(x;\Lambda\right)\left[\sum_{q=o}^{\infty}w_{q}^{*}G^{q}\left(x;\Lambda\right)\right]\left[\sum_{q=0}^{\infty}w_{q}G^{q}(x;\Lambda)\right]^{k}dx,$$

since, $\left[\sum_{q=0}^{\infty}w_{q}G^{q}(x;\Lambda)\right]^{k}=\left[\sum_{q=0}^{\infty}c_{q}G^{q}(x;\Lambda)\right]$, Gradshteyn and Ryzhik [14], then,

$$\tau_{r,s,o}=\left(\frac{1}{1-\alpha}\right)^{s}\sum_{k=0}^{s}(-1)^{k}\binom{s}{k}\int_{-\infty}^{\infty}x^{r}g\left(x;\Lambda\right)\left[\sum_{q=o}^{\infty}w_{q}^{*}G^{q}\left(x;\Lambda\right)\right]\left[\sum_{q=0}^{\infty}c_{q}G^{q}(x;\Lambda)\right]dx,$$

where

$$c_{0}=w_{0}^{k},\ c_{m}=\frac{1}{m\,w_{0}}\sum_{q=1}^{m}\left(qk-m+q\right)w_{q}\,c_{m-q};m\geq 1,$$

since, $\left[\sum_{q=o}^{\infty}w_{q}^{*}G^{q}\left(x;\Lambda\right)\right]\left[\sum_{q=0}^{\infty}c_{q}G^{q}(x;\Lambda)\right]=\sum_{q=o}^{\infty}d_{q}G^{q}\left(x;\Lambda\right)$, Gradshteyn and Ryzhik [14],

hence,

$$\tau_{r,s,o}=\left(\frac{1}{1-\alpha}\right)^{s}\sum_{k=0}^{s}(-1)^{k}\binom{s}{k}\int_{-\infty}^{\infty}x^{r}g\left(x;\Lambda\right)\left[\sum_{q=o}^{\infty}d_{q}G^{q}\left(x;\Lambda\right)\right]dx,$$

where

$$d_{m}=\sum_{q=0}^{m}w_{q}^{*}c_{m-q},$$

since,

$$\tau_{r,s,o}=\sum_{q=o}^{\infty}h_{q}\int_{-\infty}^{\infty}x^{r}g\left(x;\Lambda\right)G^{q}\left(x;\Lambda\right)dx,\qquad(11)$$

where

$$h_{q}=\left(\frac{1}{1-\alpha}\right)^{s}\sum_{k=0}^{s}\sum_{q=o}^{\infty}d_{q}\left(-1\right)^{k}\binom{s}{k},$$

then,

$$\tau_{r,s,o}=\sum_{q=o}^{\infty}h_{q}\,\tau_{r,q,o}.$$

Using the Parent Quantile Function

Setting $G\left(x;\Lambda\right)=u$, $y=Q\left(u\right)$ and substituting into (11) yields

$$\tau_{r,s,o}=\sum_{q=o}^{\infty}h_{q}\int_{0}^{1}Q^{r}\left(u\right)u^{q}du,$$

then,

$$\tau_{r,s,o} = \sum_{q=o}^{\infty} h_q \, \tau_{r,q,o}.$$

3.3. The Moment Generating Function

A continuous random variable X moment generating function (*MGF*) can be written as

$$M_x(t) = E\left(e^{tx}\right) = \int_x e^{tx} f(x) dx, \tag{12}$$

applying the exponential expansion yields

$$E\left(e^{tx}\right) = E\left(\sum_{r=0}^{\infty} \frac{t^r x^r}{r!}\right),$$

then,

$$E\left(e^{tx}\right) = \sum_{r=0}^{\infty} \frac{t^r E\left(x^r\right)}{r!},$$

substituting (10) into last equation gives

$$E\left(e^{tx}\right) = \sum_{r=0}^{\infty}\sum_{q=0}^{\infty} \frac{t^r}{r!} w_q^* \tau_{r,q,0}.$$

Using the Parent Quantile Function

Substituting (6) into (12) leads to

$$E\left(e^{tx}\right) = \int_{-\infty}^{\infty} e^{tx} \sum_{q=0}^{\infty} w_q^* \, G^q(x;\Lambda) \, g(x;\Lambda) dx,$$

then,

$$E\left(e^{tx}\right) = \sum_{q=0}^{\infty} w_q^* \int_{-\infty}^{\infty} e^{tx} \, G^q(x;\Lambda) g(x;\Lambda) dx,$$

setting $G(x;\Lambda) = u$, $x = Q(u)$ and substituting into last equation yields

$$E\left(e^{tx}\right) = \sum_{q=o}^{\infty} w_q^* \int_0^1 e^{t\,Q(u)} u^q du,$$

using exponential expansion gives

$$E\left(e^{tx}\right) = \sum_{q=o}^{\infty} w_q^* \sum_{r=0}^{\infty} \frac{t^r}{r!} \int_0^1 Q^r(u) u^q du,$$

hence,

$$E\left(e^{tx}\right)=\sum_{q=o}^{\infty}w_q^{*}\sum_{r=0}^{\infty}\frac{t^r}{r!}\tau_{r,q,0}.$$

3.4. The Mean Deviation

Basically, a random variable X having the mean deviation about mean and median, respectively, can be written as

$$S_1(x)=\int_x|x-\mu|\,f(x)dx \quad \text{and} \quad S_2(x)=\int_x|x-M|f(x)dx,$$

which is given by, Ali Ahmed [15],

$$S_1(x)=2\mu F(\mu)-2t(\mu) \text{ and } S_2(x)=\mu-2t(M),$$

where $T(z)=\int_{-\infty}^{z}x\,f(x)dx$ is the linear incomplete moment.

Substituting (6) into *T(.)* yields

$$T(z)=\int_{-\infty}^{z}x\sum_{q=o}^{\infty}w_q^{*}\,g(x;\Lambda)G^q(x;\Lambda)dy,$$

then,

$$T(z)=\sum_{q=o}^{\infty}w_q^{*}\int_{-\infty}^{z}x\,g(x;\Lambda)G^q(x;\Lambda)dy.$$

Using the Parent Quantile Function

Setting $G(x;\Lambda)=u$, $x=Q(u)$ and substituting into last equation gives

$$T(z)=\sum_{q=o}^{\infty}w_p^{*}\int_0^{G(z)}Q(u)u^q du.$$

4. The Hazard Function of the *MLN* Class of Distributions

A random variable X survival function [16] can be written as

$$S(x)=1-F(x),$$

substituting (3) into last equation gives

$$S(x)=\frac{\alpha-\alpha^{1-\left(1-G^{\beta}(x;\Lambda)\right)^{\theta}}}{\alpha-1};\alpha,\beta,\theta>0;\alpha\neq1, \tag{13}$$

moreover, the Hazard function [16] can be written as

$$H(x)=\frac{f(x)}{S(x)},$$

substituting (4) and (13) into last equation yields

$$H(x)=\frac{\beta\theta(\log\alpha)\alpha^{1-\left(1-G^{\beta}(x;\Lambda)\right)^{\theta}}g(x;\Lambda)\,G^{\beta-1}(x;\Lambda)\left(1-G^{\beta}(x;\Lambda)\right)^{\theta-1}}{\alpha-\alpha^{1-\left(1-G^{\beta}(x;\Lambda)\right)^{\theta}}}.$$

The *AKE* Hazard function can be given by

$$H(x)=\frac{\beta\theta\lambda(\log\alpha)\alpha^{1-\left(1-(1-e^{-\lambda x})^{\beta}\right)^{\theta}}e^{-\lambda x}(1-e^{-\lambda x})^{\beta-1}\left(1-(1-e^{-\lambda x})^{\beta}\right)^{\theta-1}}{\alpha-\alpha^{1-\left(1-(1-e^{-\lambda x})^{\beta}\right)^{\theta}}},$$

some Hazard function shapes for the *AKE* distribution are given in figure 2.

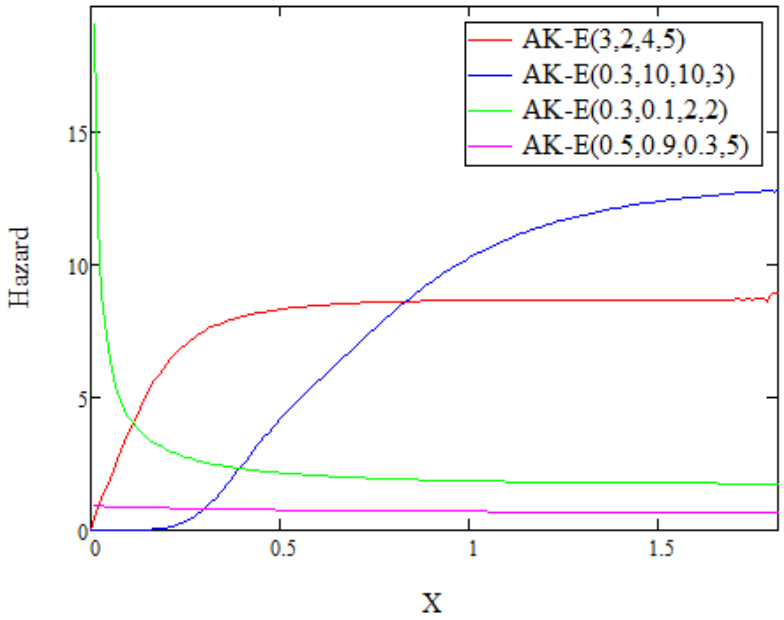


Figure 2: The *AKE* Hazard functions

In figure 2, Hazard functions curves four types for the *AKE* distribution are illustrated as follows: A decreasing then constant Hazard curve, a constant Hazard curve, an increasing then constant Hazard curve and a constant then increasing then constant Hazard curve.

5. Order Statistics of the *AK* Class of Distributions

The *u-th* order statistics density function $f(x_{u:v})$ for $u = 1,2,\ldots, v$ from *iid* random variables $X_1,X_2,\ldots, X_v$ following any *AK* generalized distribution [17] can be written as

$$f(x_{u:v})=\frac{f(x_u)}{B(u,v-u+1)}F^{u-1}(x_u)\{1-F(x_u)\}^{v-u},$$

applying binomial expansion for the last equation leads to

$$f\left(x_{uv}\right)=\sum_{p=0}^{v-u}\frac{(-1)^p\binom{v-u}{p}}{B\left(u,v-u+1\right)}f\left(x_u\right)F^{u+p-1}\left(x_u\right),$$

substituting (5) and (6) into last equation leads to

$$f\left(x_{uv}\right)=\sum_{p=0}^{v-u}\frac{(-1)^p\binom{v-u}{p}}{B\left(u,v-u+1\right)}\left[\sum_{q=0}^{\infty}w_q^{*}\,G^q(x_u;\Lambda)\,g(x_u;\Lambda)\right]\times\left[\frac{1}{1-\alpha}\left(1-\sum_{q=0}^{\infty}w_q\,G^q(x_u;\Lambda)\right)\right]^{u+p-1},$$

then,

$$f\left(x_{uv}\right)=\sum_{p=0}^{v-u}\frac{(-1)^p\binom{v-u}{p}\left(\frac{1}{1-\alpha}\right)^{u+p-1}}{B\left(u,v-u+1\right)}\left[\sum_{q=0}^{\infty}w_q^{*}\,G^q(x_u;\Lambda)\,g(x_u;\Lambda)\right]\left[1-\sum_{q=0}^{\infty}w_q\,G^q(x_u;\Lambda)\right]^{u+p-1},$$

using binomial expansion gives

$$f\left(x_{uv}\right)=\sum_{p=0}^{v-u}\frac{(-1)^p\binom{v-u}{p}\left(\frac{1}{1-\alpha}\right)^{u+p-1}}{B\left(u,v-u+1\right)}\left[\sum_{q=0}^{\infty}w_q^{*}\,G^q(x_u;\Lambda)\,g(x_u;\Lambda)\right]\times\sum_{s=0}^{u+p-1}(-1)^s\binom{u+p-1}{s}\left(\sum_{q=0}^{\infty}w_q\,G^q(x_u;\Lambda)\right)^s,$$

since, $\left(\sum_{q=0}^{\infty}w_q\,G^q(x_u;\Lambda)\right)^s=\left[\sum_{q=0}^{\infty}a_q\,G^q(x_u;\Lambda)\right]$, Gradshteyn and Ryzhik [14], then,

$$f\left(x_{uv}\right)=\sum_{p=0}^{v-u}\frac{(-1)^p\binom{v-u}{p}\left(\frac{1}{1-\alpha}\right)^{u+p-1}}{B\left(u,v-u+1\right)}\left[\sum_{q=0}^{\infty}w_q^{*}\,G^q(x_u;\Lambda)\,g(x_u;\Lambda)\right]\times\sum_{s=0}^{u+p-1}(-1)^s\binom{u+p-1}{s}\left(\sum_{q=0}^{\infty}a_q\,G^q(x_u;\Lambda)\right),$$

where

$$a_0=w_0^{s},\ a_m=\frac{1}{m\,w_0}\sum_{q=1}^{m}(qs-m+q)w_q\,a_{m-q};m\geq 1,$$

since, $\left(\sum_{q=0}^{\infty}w_q^{*}G^q(x_u;\Lambda)\right)\left(\sum_{q=0}^{\infty}a_qG^q(x_u;\Lambda)\right)=\left[\sum_{q=0}^{\infty}b_q\,G^q(x_u;\Lambda)\right]$, Gradshteyn and Ryzhik [14], then,

$$f\left(x_{uv}\right)=\sum_{p=0}^{v-u}\frac{(-1)^p\binom{v-u}{p}\left(\frac{1}{1-\alpha}\right)^{u+p-1}}{B\left(u,v-u+1\right)}\sum_{s=0}^{u+p-1}(-1)^s\binom{u+p-1}{s}g(x_u;\Lambda)\left[\sum_{q=0}^{\infty}b_q\,G^q(x_u;\Lambda)\right],$$

where

$$b_m = \sum_{q=0}^{m} w_q^* a_{m-q},$$

hence,

$$f\left(x_{u:v}\right) = \sum_{q=0}^{\infty} d_q \, g(x_u;\Lambda) G^q(x_u;\Lambda), \tag{14}$$

where

$$d_q = \sum_{p=0}^{v-u} \frac{(-1)^p \binom{v-u}{p} \left(\frac{1}{1-\alpha}\right)^{u+p-1}}{B(u, v-u+1)} \sum_{s=0}^{u+p-1} (-1)^s \binom{u+p-1}{s} b_q.$$

The order statistics *r-th* moment of the *AK* class of distributions can be written as

$$E\left(x_{u:v}^{r}\right) = \int_{x_u} x_u^r f\left(x_{u:v}\right) dx_u,$$

substituting (14) into last equation yields

$$E\left(X_{u:v}^{r}\right) = \sum_{q=o}^{\infty} d_q \int_{-\infty}^{\infty} X_u^r \, g(x_u;\Lambda) G^q(x_u;\Lambda) dx_u,$$

hence,

$$E\left(X_{u:v}^{r}\right) = \sum_{q=o}^{\infty} d_q \, \tau_{r,q,o}.$$

6. Estimation for Parameters of the *AK* Class of Distributions Using *MLE* Method

Let $X_1, X_2, \ldots, X_n$ be *iid* random variables following any *AK* generalized distribution $\left(x;\Delta\right)$ and the vector of parameter $\Delta = \left(\alpha, \beta, \theta, \Lambda\right)$ likelihood function [18] can be written as

$$L(x) = \left(\frac{\beta\theta}{\alpha-1}\right)^n (\log\alpha)^n \prod_{i=1}^{n} g(x_i;\Lambda) \prod_{i=1}^{n} G^{\beta-1}(x_i;\Lambda) \prod_{i=1}^{n} \left(1-G^{\beta}(x_i;\Lambda)\right)^{\theta-1} \prod_{i=1}^{n} \alpha^{1-\left(1-G^{\beta}(x_i;\Lambda)\right)^{\theta}},$$

the log likelihood function is given by

$$\ell(x) = n\log\frac{\beta\theta}{\alpha-1} + n\log(\log\alpha) + \sum_{i=1}^{n}\log g(x_i;\Lambda) + (\beta-1)\sum_{i=1}^{n}\log G(x_i;\Lambda)$$
$$+ (\theta-1)\sum_{i=1}^{n}\log\left(1-G^{\beta}(x_i;\Lambda)\right) + (\log\alpha)\sum_{i=1}^{n}\left[1-\left(1-G^{\beta}(x_i;\Lambda)\right)^{\theta}\right],$$

the score functions for the parameters α, β, θ and Λ can be obtained by

$$\frac{\partial \ell(x)}{\partial \alpha}=\frac{n}{1-\alpha}+\frac{n}{\alpha \log \alpha}+\frac{1}{\alpha}\sum_{i=1}^{n}\left[1-\left(1-G^{\beta}(x_i;\Lambda)\right)^{\theta}\right],$$

$$\frac{\partial \ell(x)}{\partial \beta}=\frac{n}{\beta}+(\log\alpha)\theta\sum_{i=1}^{n}G^{\beta}(x_i;\Lambda)\left[1-G^{\beta}(x_i;\Lambda)\right]^{\theta-1}\left[\log G(x_i;\Lambda)\right]+\sum_{i=1}^{n}\log G(x_i;\Lambda)$$
$$-(\theta-1)\sum_{i=1}^{n}\frac{\left[G^{\beta}(x_i;\Lambda)\log G(x_i;\Lambda)\right]}{1-G^{\beta}(x_i;\Lambda)},$$

$$\frac{\partial \ell(x)}{\partial \theta}=\frac{n}{\theta}-(\log\alpha)\sum_{i=1}^{n}\left[1-G^{\beta}(x_i;\Lambda)\right]^{\theta}\log\left[1-G^{\beta}(x_i;\Lambda)\right]+\sum_{i=1}^{n}\log\left[1-G^{\beta}(x_i;\Lambda)\right]$$

and

$$\frac{\partial \ell(x)}{\partial \Lambda}=(\log\alpha)\beta\theta\sum_{i=1}^{n}\left[1-G^{\beta}(x_i;\Lambda)\right]^{\theta-1}G^{\beta-1}(x_i;\Lambda)\frac{\partial G(x_i;\Lambda)}{\partial \Lambda}+\sum_{i=0}^{n}\frac{1}{g(x_i;\Lambda)}\frac{\partial g(x_i;\Lambda)}{\partial \Lambda}$$
$$+\sum_{i=0}^{n}\frac{(\beta-1)}{G(x_i;\Lambda)}\frac{\partial G(x_i;\Lambda)}{\partial \Lambda}-\sum_{i=0}^{n}\beta(\theta-1)\frac{G^{\beta-1}(x_i;\Lambda)}{\left(1-G^{\beta}(x_i;\Lambda)\right)}\frac{\partial G(x_i;\Lambda)}{\partial \Lambda}.$$

7. A Simulation Study

In this study, *MLEs* for parameters of the *AKE* distribution are obtained using random numbers to study the *MLEs* finite sample behavior via the bootstrapping resample approach. Obtaining parameters estimates algorithm is detailed in the following steps:

Step (1): Generating a random sample $X_1,X_2,\ldots,X_n$ of sizes *n=(5,15,30,50,100,300)* using the *AKE* distribution.

Step (2): Selecting parameters different set values as: set (1): *(α=2, β=0.5, θ=0.5, λ=0.2)*, set (2): *(α=2, β=0.5, θ=0.5, λ=0.5)* and set (3): *(α=2, β=0.5, θ=0.5, λ=2).*

Step (3): Solving normal equations of the *AKE* distribution via iteration to estimate distribution parameters.

Step (4): Replacing set (1), set(2) and set (3) with its estimators and repeating step (3) to compute, biases, *MLEs, RMSE* (the root of mean squared error) and the Pearson type [19] of parameters estimators of the *AKE* distribution.

Step (5): Repeating step (1) to step (4), 10000 times.

In this study, the conjugate gradient iteration method is performed in order to generate random numbers samples using Mathcad package v15. All results are included in tables and indicated in the appendix.

From study results, indicated in the appendix, one can see that: When sample size increases, biases, estimators, and *RMSEs* decrease. As sample size increases, the estimators can be consistent. $\hat{\alpha}$ and $\hat{\lambda}$ sampling distributions can be the Pearson type IV distribution in all times, $\hat{\beta}$ and $\hat{\theta}$ sampling

distributions differ according to sample size. When λ increases, for fixed values of α and β, the biases and *MSEs* of $\hat{\alpha}$ and $\hat{\beta}$ decrease.

8. Application

A real data set is given using the *MLE* method to study the new model performance by Mathematica package version 10, some distributions are used as: The *AKE* distribution, the gamma distribution (2 parameters), the normal distribution, the Weibull distribution, the Kumaraswamy exponential (*KE*) distribution, the alpha power – power exponential (*APE*) distribution and the power exponential *(PE)* distribution. The following data set represents the classic lamps lifetime (Hours) for 40 devices, the data are given from the *UK* National Physical Laboratory, http://www.npl.co.uk/,
1.104, 0.285, 0.627, 0.282, 0.542, 0.439, 0.884, 1.021, 0.281, 0.737, 0.504, 0.713, 1.186, 0.380, 0.583, 0.769, 0.334, 1.030, 0.242, 0.386, 0.936, 1.313, 0.263, 0.508, 0.654, 0.481, 0.353, 0.614, 0.736, 0.727, 0.264, 0.528, 0.588, 0.535, 0.590, 0.420, 0.458, 0.278, 0.405, 0.381
Some goodness of fit measures results are indicated in the table (1), the likelihood ratio tests results are indicated in the table (2), the figure (3) describes probability density functions for some distributions having skewness and kurtosis values similar to the *AKE* distribution (the gamma distribution with 2 parameters, the Normal distribution, the Weibull distribution), the figure (4) shows the empirical *CDF* compared to *CDFs* for some distributions (the gamma distribution with 2 parameters, the normal distribution, the Weibull distribution) and the figure (5) describes the probability density functions for special cases from the *AKE* distribution (the Kumaraswamy exponential (*KE*) distribution, the alpha power – power exponential (*APE*) distribution and the power exponential *(PE)* distribution).

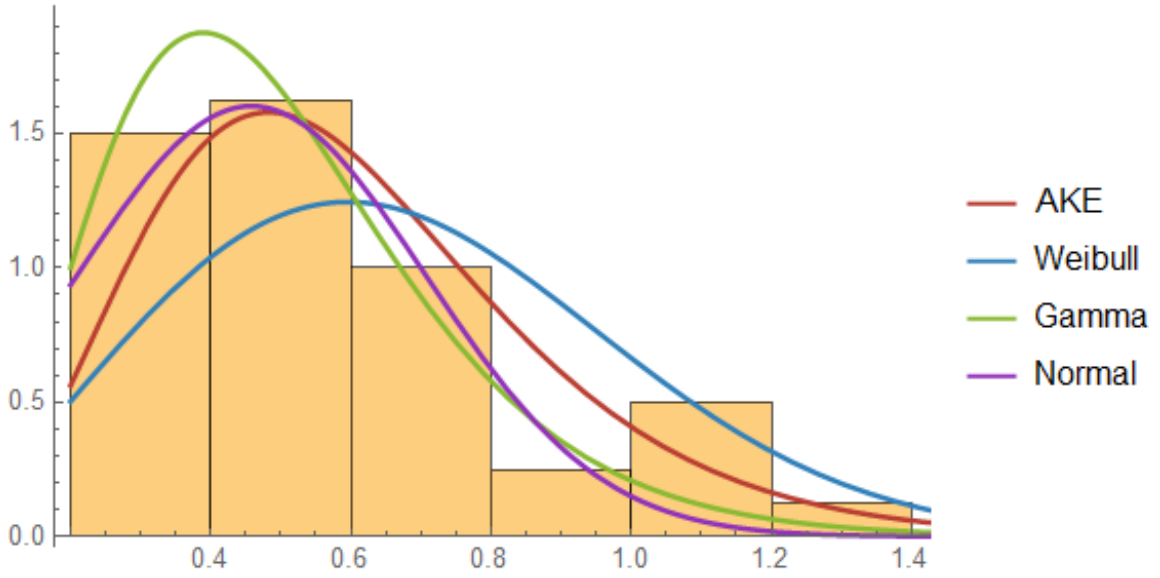


Figure 3: Probability density functions for some distributions having skewness and kurtosis Values similar to *AKE*

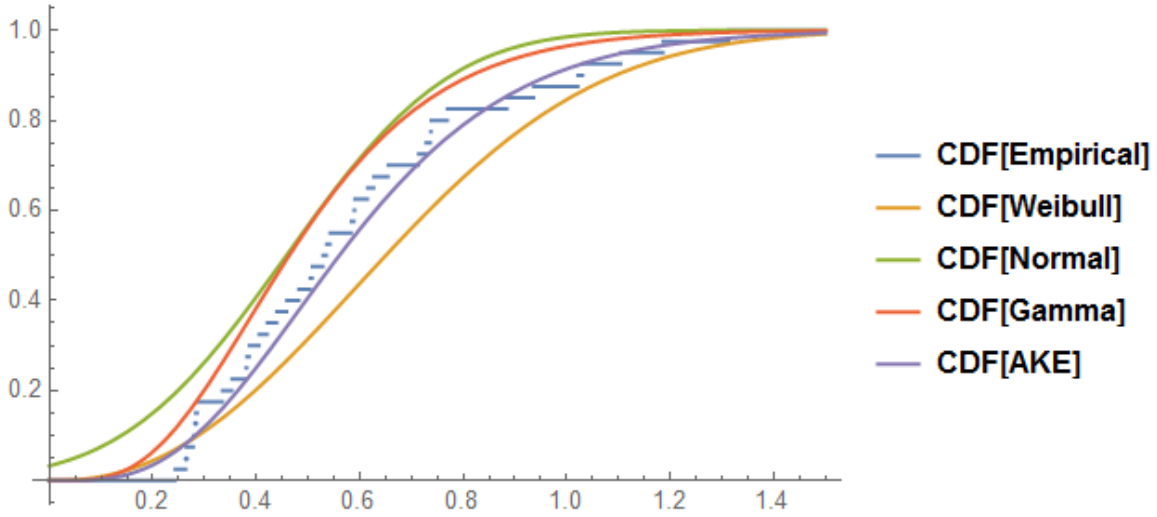


Figure 4: The empirical *CDF* compared to *CDF*s for some distributions

Table 1: The parameter(s) *MLE* and the associated *AIC* and *BIC* values.

Distribution	MLE_parameters				Skewness	Kurtosis	KS	P-value	Log Likel-ihood	AIC	BIC	CAIC
	α	β	θ	λ								
AK-E	2	4	3	2	0.058	2.403	0.08	0.938	-0.555	7.111	8.488	7.777
	(0.013)	(0.033)	(0.247)	(0.182)								
Weibull	2.307	0.762	—	—	0.451	2.961	0.203	0.062	-3.901	11.802	15.180	12.127
	(0.272)	(0.048)										
Gamma	4.503	0.111	—	—	0.942	4.332	0.144	0.338	-2.649	9.299	12.677	9.624
	(1.088)	(0.026)										
Normal	0.458	0.249	—	—	0	3	0.193	0.086	-10.05	24.115	27.492	24.439
	(0.042)	(0.0303)										

In table 1, the distributions parameters *MLE*s, parameters standard error (*SEs*), in parentheses, Kolmogorov-Smirnov (*KS*) test statistic, *AIC* (Akaike Information Criterion), *CAIC* (the consistent Akaike Information Criterion) and *BIC* (Bayesian information criterion), Merovcia and Puka [20], are computed for distributions having similar skewness and kurtosis values. Since, the *AKE* distribution has the smallest *KS*, *AIC*, *CAIC*, *BIC, SEs* and the largest p-value, hence the *AKE* distribution can be the best fitted distribution to the data compared with other distributions which have similar skewness and kurtosis values.

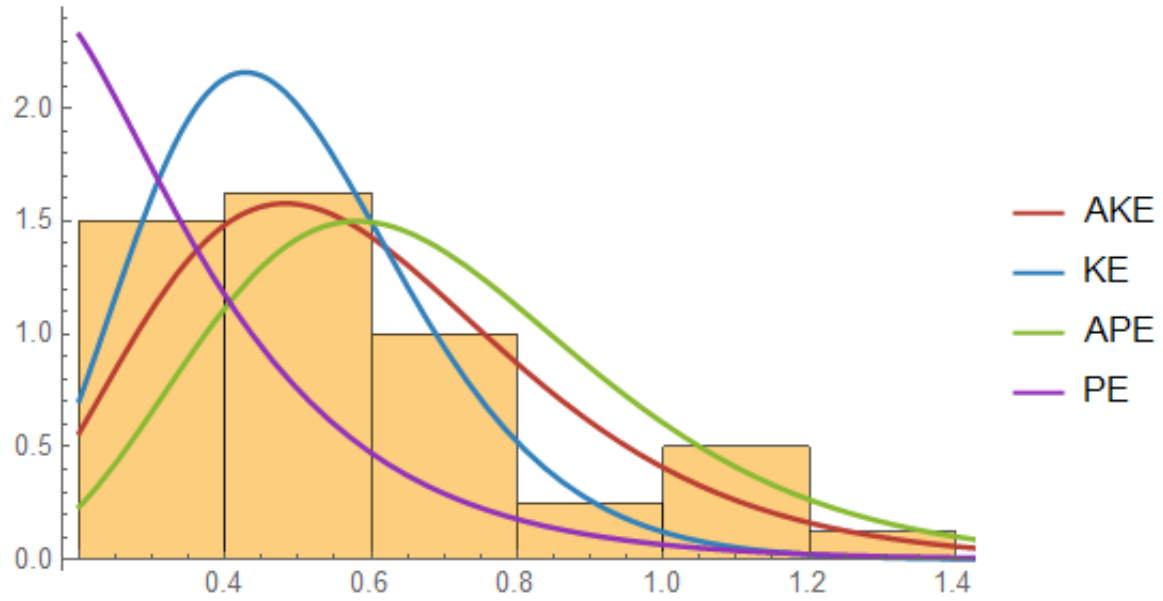


Figure 5: Probability density functions for special cases from the *AKE* distribution

Table 2: The log-likelihood function, the likelihood ratio tests statistic and p-values.

Distribution	Parameters				ℓ (log likelihood)	Λ (The likelihood ratio test statistic)	df (degrees of freedom)	p-value
	α	β	θ	λ				
KE	–	2.176	1.647	1.928	-8.214	15.318	1	9.085×10^{-5}
		(0.418)	(0.329)	(0.284)				
APE	1.275	3.217	–	1.099	-10.247	19.384	1	1.069×10^{-5}
	(0.214)	(1.517)		(0.314)				
PE	–	7.782	–	4.626	−14.869	28.628	2	6.074×10^{-7}
		(2.567)		(0.684)				

*Note that the *AKE* distribution log likelihood = - 0.555

In table 2, upon the likelihood ratio test, the null hypothesis is the data follow the nested model and the alternative is the data follow the full model, where the *KE* distribution, the *APE* distribution and the *PE* distribution are nested by *AKE* distribution, one can see that, all null hypotheses can be rejected at significance level is 0.05

9. Conclusion

The alpha power Kumaraswamy class has an explicit form gives more flexibility in mathematical properties and random number generating. The alpha power Kumaraswamy class generalizes some important classes of distributions as the Kumaraswamy class, the alpha power class and the power function class. The alpha power Kumaraswamy exponential distribution has wide applications in real data sets and in some cases it can be the best fitted distribution. Author encourages researchers to study more cases from that flexible class**.**

Competing interests: The author declare that he has no competing interests

Data and Material Availability: One can find the date set at: http://www.npl.co.uk/

Code availability: Not applicable

Acknowledgements: The author thanks anyone gave an advice or any comment for this manuscript.

References

[1] Mahdavi, A. & Kundu, D. (2017). A new method for generating distributions with an application to exponential distribution. *Communications in Statistics-Theory and Method, 46, 6543-6557* .

[2] Ahmed, M.A. (2020). On the alpha power Kumaraswamy distribution: Properties, simulation and application. Revista Colombiana de Estadística, 43, 285-313.

[3] Wahed, A.S. (2006). A general method of constructing extended families of distributions from an existing continuous class. Journal of Probability and Statistical Science, 4, 165-177.

[4] Kumaraswamy, P. (1980). A generalized probability density function for double- bounded random-processes. Journal of Hydrology, 46, 79-88.

[5] Cordeiro, G.M., & de Castro, M. (2011). A new family of generalized distributions. Journal of Statistical Computation & Simulation, 81, 883-898.

[6] Pescim, R.R., Cordeiro, G.M., Demétrio, C.G., Ortega, E.M. & Nadarajah, S. (2012). The new class of Kummer beta generalized distributions. SORT-Statistics and Operations Research Transactions, 153-180.

[7] McDonald, J.B. (1984). Some generalized functions for the size distribution of income. Econometrica, 52, 647-664.

[8] Alexander, C, Cordeiro, G.M., Ortega, E.M.M. & Sarabia J.M. (2012). Generalized beta-generated distributions. Comput Stat Data Anal., 56, 1880-1897.

[9] El-Sherpieny, E.A. & Ahmed, M.A. (2014). On the kumaraswamy Kumaraswamy distribution. International Journal of Basic and Applied Sciences, 3, 372-381.

[10] Mahmoud, M.R., El-Sherpieny, E.A. & Ahmed, M.A. (2015). The new Kumaraswamy Kumaraswamy family of generalized distributions with application. Pakistan Journal of Statistics and Operations Research, 11, 159-180.

[11] Meniconi, M. & Barry, D. (1996). The power function distribution: A useful and simple distribution to assess electrical component reliability. Microelectronics Reliability, 36, 1207-1212.

[12] Johnson, N.L., Kotz, S. & Balakrishnan, N. (1995). Continuous Univariate Distributions. John wiley and Sons, New York .

[13] Greenwood, J.A., Landwehr, J.M., Matalas, N.C. & Wallis, J.R. (1979). Probability weighted moments definition and relation to parameters of several distributions expressable in inverse form. Water Resources Research, 15, 1049-1054.

[14] Gradshteyn, I.S. & Ryzhik, I.M. (2000). Tables of Integrals, Series and Products.Academic Press, San Diego, CA.

[15] Ali Ahmed, M. (2021). The new form Libby-Novick distribution. Communications in Statistics-Theory and Methods, 50, 540-559.

[16] Meeker, W.Q. & Escobar, L.A. (1998). Statistical Methods for Reliability Data. John Wiley, New York.

[17] Arnold, B.C., Balakrishnan, N. & Nagaraja, H.N. (1992). A First Course in Order Statistics. John Wiley and Sons, Inc., New York.

[18] Garthwait, P.H., Jolliffe, I.P. & jones, B. (2002). Statistical Inference. Prentice HallInternational (UK) Limited, London.

[19] Pearson, K. (1895). Contributions to the mathematical theory of evolution. II. Skew variations in homogeneous material. Philosophical Transactions of the Royal Society of London, Series A, 186, 343-414.

[20] Merovcia, F. & Puka, L. (2014). Transmuted Pareto distribution. Prob Stat Forum, 7, 1-111.

Appendix

Results of the simulation study for different data sets:

Set(1):(α=2, β=0.5, θ =0.5, λ=0.2)

Sample Size	Parameters	Mean of Estimators	Biases	Total Bias	RMSE	Total RMSE	Pearson System Coefficients	Pearson Type
10	α=2	**2.262**	**0.262**	**1.474**	**1.258**	**2.185**	**0.202**	**IV**
	β=0.5	**0.759**	**0.259**		**0.75**		**0.27**	**IV**
	θ =0.5	**0.92**	**0.42**		**0.436**		**-0.27**	**I**
	λ=0.2	**1.565**	**1.365**		**1.563**		**0.576**	**IV**
20	α=2	**2.144**	**0.144**	**1.054**	**1.235**	**1.683**	**0.392**	**IV**
	β=0.5	**0.602**	**0.102**		**0.246**		**0.827**	**IV**
	θ =0.5	**0.844**	**0.344**		**0.36**		**0.054**	**IV**
	λ=0.2	**1.181**	**0.981**		**1.058**		**0.594**	**IV**
30	α=2	**2.14**	**0.14**	**0.898**	**1.103**	**1.462**	**0.461**	**IV**
	β=0.5	**0.568**	**0.068**		**0.173**		**-0.312**	**I**
	θ =0.5	**0.808**	**0.308**		**0.322**		**-0.496**	**I**
	λ=0.2	**1.03**	**0.83**		**0.888**		**0.14**	**IV**
50	α=2	**2.106**	**0.106**	**0.760**	**0.905**	**1.211**	**0.438**	**IV**
	β=0.5	**0.537**	**0.037**		**0.118**		**-2.505**	**I**
	θ =0.5	**0.776**	**0.276**		**0.289**		**0.021**	**IV**
	λ=0.2	**0.9**	**0.7**		**0.742**		**0.096**	**IV**
100	α=2	**2.073**	**0.073**	**0.224**	**0.368**	**0.391**	**0.44**	**IV**
	β=0.5	**0.511**	**0.011**		**0.085**		**-0.267**	**I**
	θ =0.5	**0.637**	**0.137**		**0.051**		**0.011**	**IV**
	λ=0.2	**0.362**	**0.162**		**0.091**		**0.666**	**IV**
300	α=2	**2.043**	**0.043**	**0.050**	**0.055**	**0.087**	**0.105**	**IV**
	β=0.5	**0.507**	**0.007**		**0.012**		**-0.043**	**I**
	θ =0.5	**0.506**	**0.006**		**0.018**		**-3.643**	**I**
	λ=0.2	**0.225**	**0.025**		**0.064**		**0.57**	**IV**

Set(2):(α=2, β=0.5, θ =0.5, λ=0.8)

Sample Size	Parameters	Mean of Estimators	Biases	Total Bias	RMSE	Total RMSE	Pearson System Coefficients	Pearson Type
10	α=2	2.239	0.239	3.506	1.234	6.406	0.203	IV
	β=0.5	0.743	0.243		0.460		0.268	IV
	θ =0.5	0.959	0.459		0.435		-0.372	I
	λ=0.8	4.26	3.46		6.255		0.583	IV
20	α=2	2.125	0.125	2.954	1.053	4.382	0.395	IV
	β=0.5	0.583	0.083		0.126		0.46	IV
	θ =0.5	0.894	0.394		0.391		0.062	IV
	λ=0.8	3.724	2.924		4.234		0.592	IV
30	α=2	2.117	0.117	2.352	1.002	3.709	0.461	IV
	β=0.5	0.549	0.049		0.101		-0.281	I
	θ =0.5	0.878	0.378		0.351		-0.221	I
	λ=0.8	3.118	2.318		3.553		0.692	IV
50	α=2	2.104	0.104	1.827	0.704	1.382	0.438	IV
	β=0.5	0.526	0.026		0.091		-4.266	I
	θ =0.5	0.796	0.296		0.20		0.018	IV
	λ=0.8	2.6	1.8		1.169		0.095	IV
100	α=2	2.053	0.053	1.25	0.171	0.470	0.439	IV
	β=0.5	0.509	0.009		0.064		-1.288	I
	θ =0.5	0.733	0.233		0.267		0.018	IV
	λ=0.8	2.036	1.236		0.341		0.046	IV
300	α=2	2.031	0.031	0.061	0.027	0.111	0.0018	IV
	β=0.5	0.504	0.004		0.009		-0.0021	I
	θ =0.5	0.521	0.021		0.037		0.00115	IV
	λ=0.8	0.848	0.048		0.101		0.529	IV

Set(3):(α=2, β=0.5, θ =0.5, λ=2)

Sample Size	Parameters	Mean of Estimators	Biases	Total Bias	RMSE	Total RMSE	Pearson System Coefficients	Pearson Type
10	α=2	2.218	0.218	13.667	1.202	15.709	0.2	IV
	β=0.5	0.727	0.227		0.384		0.279	IV
	θ =0.5	0.972	0.472		0.518		-0.231	I
	λ=2	15.656	13.656		15.65		0.561	IV
20	α=2	2.113	0.113	9.814	1.004	10.634	0.393	IV
	β=0.5	0.561	0.061		0.117		0.308	IV
	θ =0.5	0.897	0.397		0.438		0.06	IV
	λ=2	11.806	9.806		10.577		0.605	IV
30	α=2	2.095	0.095	8.307	0.809	8.934	0.454	IV

	$\beta=0.5$	0.524	0.024		0.096		-0.756	I
	$\theta=0.5$	0.885	0.385		0.472		-0.4	I
	$\lambda=2$	10.298	8.298		8.885		0.479	IV
50	$\alpha=2$	2.073	0.073	7.006	0.427	7.443	0.438	IV
	$\beta=0.5$	0.514	0.014		0.058		-4.187	I
	$\theta=0.5$	0.799	0.299		0.342		0.02	IV
	$\lambda=2$	9	7		7.423		0.097	IV
100	$\alpha=2$	2.029	0.029	5.519	0.062	5.756	0.975	IV
	$\beta=0.5$	0.508	0.0084		0.043		-1.329	I
	$\theta=0.5$	0.753	0.253		0.290		0.0079	IV
	$\lambda=2$	7.514	5.514		5.749		0.013	IV
300	$\alpha=2$	2.027	0.027	0.113	0.016	0.160	0.0026	IV
	$\beta=0.5$	0.5025	0.0025		0.007		-0.0031	I
	$\theta=0.5$	0.515	0.015		0.096		-0.0053	I
	$\lambda=2$	2.109	0.109		0.128		0.35	IV

Journal of Statistics & Applied Science, Volume -2 Issue -2
ISSN 2718-0999

Article

Received: date: 01.09.2021
Accepted: date: 24.10.2021
Published: date: 28.11.2021

Statistical Comparison of Parallel-Line Symmetrical Microbiological Models: Analysis of Agar Diffusion Assay in 8 x 8 Large Rectangular Plates

Dalia EISSA[1], Engy RASHED[2], Mostafa EISSA[3,*]

[1]Royal Oldham Hospital; Oldham; United Kingdom; eissade@yahoo.com
[2]National Centre for Radiation Research and Technology; Cairo; Egypt; engyrefaat@yahoo.com
[3]Independent Researcher; Pharmaceutical Research Facility; Cairo, Egypt; mostafaessameissa@yahoo.com
Orcid: 0000-0002-6340-8973[1] Orcid: 0000-0002-6593-378X[2] Orcid: 0000-0002-6340-8973[3]
*Correspondence Author, e-mail: mostafaessameissa@yahoo.com

Abstract: (1) Background: Microbiological assay of active medicinal compounds is superior to conventional chemical means in several circumstances to date. However, ensuring the validity and suitability of the assay design proposed for the intended purpose is crucial before deriving any records or conclusions from the results of the potency determination; (2) The present work represented statistical comparison between three design models for determination of the potency of Neomycin Sulfate antibiotic using agar diffusion technique for the same test material subject under identical conditions through the application of a combination of statistical software programs, including validated programmed Microsoft Excel Workbook for the statistical testing of each assay layout; (3) Results: raw data of the three assay designs were found to be reasonably valid for further analysis of the assay suitability. Examination of the sources of variations for each design demonstrated the validity of the conducted experimentation. Variation between the computed potencies from the three designs was lower than 5 μg/mg. However, there was significant variation between the confidence windows of each type; (4) Conclusions: 2 x 4 design had the narrowest confidence range. However, improving confidence would require investigation of the assay parameters, including the modification of the number of replicates per treatment.

Keywords: Agar Diffusion; ANOVA; Confidence Limit; Homoscedasticity; Neomycin Sulfate; Normality; Outlier; PLM; Potency; Two-Dose Symmetrical Assay

1. Introduction

Statistical evaluation of the suitability of assays for active medicinal materials is a pivotal task to be performed before conducting any further computations to estimate the biological activity of the subject [1]. Till nowadays, a microbiological assay of antimicrobials is still mandated by the official monographs for some antibiotics [2]. It is common for an antibiotic to be composed of a mixture of related active constituents or components [3]. It is a common practice to isolate and discover the most desired component from several compounds produced from the fermentation and industrial processes [3]. Also, finding the needed conditions to direct the biosynthesis or manufacturing of this active material of interest would be a desirable needed step.

However, there are many cases where the commercially available products in the market include raw materials that remain as a mixture of varying proportions of related substances such as members of the aminoglycoside antibiotics [4]. A prominent example, Neomycin represents a particularly interesting and challenging problem. It is composed of Neomycin B and Neomycin C, which are glycosides of the organic base Neamine (basically called Neomycin A) [5]. In commercially available Neomycin, factor B

is the major constituent and has greater antibiotic activity than factor C against a range of organisms. There is sometimes a small proportion of Neamine that has little activity [6,7]. Neomycin B and C are of the same molecular weight and they are different only in the aminoethyl group configuration and could not be easily separated by conventional analytical procedures [6].

The prime challenges in the microbiological antibiotic assay were discussed by a biological standard investigator in the early 1960s. One problem referred to the difficulty in obtaining agreement between successive assays in either the same or different laboratories [8]. The researcher had cited the reason for the lack of reproducibility as being the variable influence on the two active components of changes in pH, media composition, temperature, and the test organism culture age [8]. More deeply, A scientist and his coworkers in 1964 discussed the variable attitude of B and C fractions in the agar diffusion test depending on the constituents of the assay medium and especially its salt content [9]. They demonstrated that by choice of a medium of low ionic content and the right test organism [10]. An assay system design could be arranged in which components B and C showed an equivalent potency. However, unless it could be shown that Neomycin B and C were equipotent against a range of typical infecting organisms in vivo, then such an approach might conceal rather than solves the problem [6].

A study of commercially available Neomycin from several countries was made for the World Health Organization (WHO) 1970. It revealed that the then-current International Reference Preparation was unrepresentative of commercially available material about proportions of components B and C [11]. Wilson and colleagues in 1973 showed by gas liquid chromatography (GC) that in Neomycin products available on the Canadian market, the proportion of neomycin C variation ranged within 28.5 % [12]. Finally, Generally, Lightbown (1961) concluded the following: 1- For heterogeneous materials controlled biologically and assayed against a heterogeneous standard, it must be recognized that there is no true potency for any sample. 2- A sample will have a family of potencies depending on the conditions of the assay. These may be distributed about a mode, but the modal value has no intrinsic superiority over any individual value [9].

Recently, risk mitigation for the above problem could be achieved through standardization of the testing conditions according to the compendial chapters [13]. A further improvement might be reached by harmonization between the official reference monographs [14,15]. This embraced formulas for assay media which are available from pharmacopeias and various other publications [16]. Those for the assay of growth-inhibiting substances tend to be rather simple and based on natural nutrients. However, several assay designs have been devised for microbiological antibiotic potency determinations using zone inhibition [6]. Apart from the systematic validation of the assay, routine potency determination testing should be investigated for the suitability and quality of the experiment before calculating the actual activity of the antimicrobial substance.

The present work aimed to study three different designs for antibiotic potency determination using a balanced two-dose Parallel Line Model (PLM) in large rectangular 32 cm x 32 cm autoclavable agar plates using a case of Neomycin Sulfate raw material against working standard. The study covered statistical analysis of the validity of the examined designs before computing the potency and comparison of the quality of the outcome from the three experiment layouts. This statistical investigation would serve as the basis for the Quality Control (QC) analyst of stepwise zone inhibition dataset analysis from the regular laboratory activity in order to derive valid potency determination data.

2. Materials and Methods

2.1. Study subject

The powdered raw material of Neomycin Sulfate was assayed in large autoclavable 32 cm x 32 cm rectangular plates against a reference working standard of known potency (expressed as μg/mg) [17]. Three types of designs were used for potency determination under identical testing conditions as stated in the compendial method after the post-validation assessment [5-7]. Results were expressed as a zone of inhibition diameter (in mm) recorded to the nearest 0.01 mm at the edge boundaries [18]. This distinctive zone was a clear circular area of the agar medium surrounded by opaque growth space [19]. Raw data were stored in Excel Sheet files for further processing.

2.2. Assay design types

Balanced PLM using a two-dose level was used with randomization of the treatment groups across rows and columns. Three antimicrobial activity determinations were accessed based on the number of Preparations (P) and Treatments (T) per a single plate i.e. (P x T) [8]. These designs comprised two preparations and four treatments (2 x 4), two preparations and eight treatments (2 x 8) and four preparations and eight treatments (4 x 8). The first design included 16 replicates per one treatment. While the remaining treatments in the other layouts included groups of eight replicates [8].

2.3. Preliminary statistical analysis for data visualization

Initial evaluation of the scrambled and unscrambled datasets was conducted using Graphpad Prism V9 and Minitab V16 [20,21]. Data central tendency, pattern and spreading were examined through descriptive statistical analysis, outlier detection, normality testing and assessing the homogeneity of variances. If any true aberrant value was detected, it should be investigated for possible omission and replacement. Additional data visualization could be accomplished through screening different types of means with Confidence Limits (CL), percentiles, standard deviations (SD), Coefficient of Variations (CV), skewness and kurtosis. The graphical drawing was performed using Microsoft Excel 2016, GraphPad Prism and Minitab programs [20-23].

2.4. Analysis of the experimental sources of variation

Systematic analysis of the sources of variation in the assay was conducted using Analysis of Variance (ANOVA). Stepwise examination of datasets for each design was programmed in Excel Workbook according to Hewitt using provided model examples as a reference for validation. Empirically selected probability levels (P) for the variance criteria were selected as shown in Table 1 [6]. The potential sources of variations in the potency assay were the effect of preparation, row/column variation, the contrast of the standard and the unknown in the duplicate preparations, deviation from the parallelism and regression analysis.

Table 1. Selected probability limit for sources of variation in the experiment for the analysis of variance (ANOVA) [6].

Source of Variance	Arbitrary Probability Level, P
Preparation	0.050
Regression	0.001
Parallelism (deviation from)	0.050
Contrast of standard	0.200
Contrast of unknown	0.200
Contrast of standard slope	0.200
Contrast of unknown slope	0.200
Rows	0.050
Columns	0.050

2.5. Estimated activity determinations from valid assays

Excel sheets that were used in the previous analysis were extended to cover the computation of the potency of the unknown material using the known potency of the standard in the relation between the zone of inhibition (in mm) as a response in the y-axis and the logarithm of the potency to the base ten interpreted at x-axis [24]. The confidence intervals were calculated also in the program according to Hewitt methodology described in detail [25]. The final result of the potency estimate would be expressed in μg/mg units. All equations for the computation were included in the Excel worksheet and validated against detailed examples [26]. Graphical presentations were done using Both GraphPad Prism and Minitab software.

3. Results

Microbiological analysis using agar diffusion test - for the determination of the potency of the antibiotic Neomycin Sulfate – was assessed statistically using three types of PLM designs for the assay of one unknown against standard of known biological activity in large rectangular 32 cm x 32 cm plates as discussed in the following steps.

3.1. Preliminary statistical evaluation of the validity of the dataset

3.1.1. High and low doses assessment for treatment groups and the descriptive analysis

Preliminary visual examination of the dataset for each treatment in Figure 1 showed visually the parallelism of each dose pair (high (H) and low (L)) without any noticeable deviations. Maximum Standard Deviation of the Mean (S.D.) was found in the reference group low dose treatment (SBL) followed by U3L two preparations x eight treatments and four preparations x eight treatments designs, respectively. While the lowest value was found in the high dose of the unknown (UH) in two preparations x four treatments design. In general, all low doses of the standard and test preparations – except for SL – showed higher SEM than the corresponding high doses. The relative deviation percent of the means from medians relative to the average values for each treatment group were 0.00, 0.45, 0.16, 1.40, 0.45, 1.17, 0.16, -0.96, 0.25 and 0.50 for SAH, SAL, SBH, SBL, UAH, UAL, U8H, U8L, UH, UL, SH, SL, U1H, U1L, U2H, U2L, U3H, U3L, S8H and S8L, respectively (Table 2 and Figure 1). The coefficient of Variation (CV) range was relatively narrow ranging from 1.38 for U1H to 4.09 for SBL. Thus, data spreading around the center point would be minimal. It should be noted that Geometric Standard Deviation Factor (GSDF) followed the same order as CV% for the treatment groups. All values of GSDF were close to one.

Four types of means were used in the current analysis, namely: Arithmetic Mean (M), Geometric Mean (GM), Harmonic Mean (HM) and Quadratic Mean (QM) providing a measure of the central tendency. The deviation between these mean types did not exceed 4% between M and HM for SBL and U3L. The overall gap is slightly higher between M and HM than between HM and QM. In addition, 95% Confidence Interval (CI) for these means almost coincided with each other. Under ideal situations when all values per treatment group are the same, all means would yield the same value. Since the data were homogenous, the variations were very small. These observations were demonstrated in Table 2 numerically and complemented by Figure 2 visually. Skewness and kurtosis provided quantitative measures for the deviation of the datasets from the typical Gaussian pattern. In the present case, groups of 16 readings of 2 x 4 assay showed the minimum deviation from the expected bell-shaped, except SL. The highest distortion from the normal behavior could be spotted with U3L and UAL in the experiments involving eight replicates per the treatment group.

3.1.2. Normality of data distribution

QQ plot in Figure 3 showed the closeness of the record points to the straight line of the theoretical normality of the distribution. All points were reasonably close to the red dotted line signifying acceptable normality level. Datasets showed two groups of clusters in both graphs illustrating high and low doses of standard and test groups. Extreme results were shown in-between and at the edges of those treatment clusters. At α=0.05, the results of K^2 normality tests for S_{AH}, S_{AL}, S_{BH}, S_{BL}, U_{AH}, U_{AL}, U_{8H}, U_{8L}, U_H, U_L, S_H, S_L, U_{1H}, U_{1L}, U_{2H}, U_{2L}, U_{3H}, U_{3L}, S_{8H} and S_{8L} using D'Agostino & Pearson omnibus normality test were 0.7669, 3.272, 1.048, 2.442, 1.855, 6.133, 0.7377, 1.023, 5.590, 1.175, 0.1962, 5.130, 1.607, 1.961, 1.512, 0.1443, 0.7450, 5.661, 0.6315 and 3.455 with p-values of 0.6815, 0.1948, 0.5923, 0.2949, 0.3955, 0.0466, 0.6915, 0.5996, 0.0611, 0.5558, 0.9066, 0.0769, 0.4477, 0.3751, 0.4695, 0.9304, 0.6890, 0.0590, 0.7292 and 0.1777, respectively. Thus, all treatment groups passed the normality test – with a p-value that was not significant - except U_{AL} ($P \leq 0.05$).

Similar results were obtained using KS normality test with KS – in the same order - distance of 0.1965, 0.2500, 0.1929, 0.2148, 0.1942, 0.3100, 0.1638, 0.2478, 0.1639, 0.1658, 0.1761, 0.1754, 0.2399, 0.2495, 0.2433, 0.2437, 0.1856, 0.2406, 0.2033 and 0.2610 and p-values of 0.2000, 0.1599, 0.2000, 0.2000, 0.2000, 0.0228, 0.2000, 0.1699, 0.2000, 0.2000, 0.2000, 0.2000, 0.2000, 0.1623, 0.2000, 0.2000, 0.2000, 0.2000, 0.2000 and 0.1163. However, Shapiro-Wilk normality test showed another group (U_{1L}) that deviated from normal distribution ($P \leq 0.05$). W statistics were 0.9471, 0.9231, 0.9699, 0.8778, 0.8829, 0.8162, 0.9325, 0.9297,

0.9048, 0.9341, 0.9603, 0.9072, 0.8859, 0.7940, 0.8862, 0.8982, 0.9659, 0.8837, 0.9392 and 0.9192 with p-values 0.6821, 0.4553, 0.8970, 0.1793, 0.2008, 0.0425, 0.5389, 0.5130, 0.0958, 0.2832, 0.6664, 0.1051, 0.2142, 0.0247, 0.2158, 0.2786, 0.8643, 0.2041, 0.6031 and 0.4237, respectively. Nevertheless, the deviation in normality of U_{AL} and U_{1L} was not serious since both groups passed normality in all tests at α=0.01.

3.1.3. Analysis of outliers in the datasets of the experimental antibiotic designs

Implementation of the USP method for outlier detection using G value limit for upper and lower values showed only one outlier from the smallest figure of U_{AL} (0.789 which is slightly greater than the limiting G value of 0.780) in 2 x 8 experimental design. However, the absence of aberrant values was confirmed using the ROUT method at Q=1.0%. The result of the USP outlier test could be explained by that, there was a clustering tendency observed within the impacted group toward the upper values which might affect the outcome of the test. Nevertheless, this figure - by experience and trend observation - was found to be normal and not unusual. The outcome of the ROUT method was also confirmed with Grubbs' Test for outliers – at the significance level of α = 0.05 – where the null hypothesis assumption was that all data values have come from the same normal population.
On the other hand, the alternative hypothesis was based on that the smallest or largest data value was considered an outlier. In 4 x 8 design U_{1H}, U_{1L}, U_{2H}, U_{2L}, U_{3H}, U_{3L}, S_{8H} and S_{8L} treatments had (G, P) results of (1.57, 0.724), (1.57, 0.726), (1.40, 1.000), (1.64, 0.594), (1.81, 0.319), (2.10, 0.061), (1.58, 0.697) and (1.99, 0.129), respectively. The same was found in the 2 x 4 design with (G, P) statistics (1.47, 1000), (1.93, 0.654), (1.92, 0.686) and (2.30, 0.181) for U_H, U_L, S_H and S_L, respectively. For 2 x 8 design, S_{AH}, S_{AL}, S_{BH}, S_{BL}, U_{AH}, U_{AL}, U_{BH} and U_{BL}, (G, P) pair calculation was (1.54, 0.788), (1.98, 0.135), (1.82, 0.304), (1.87, 0.243), (1.50, 0.890), (2.09, 0.068), (1.46, 0.982), (1.45, 1.000), respectively. Accordingly, a decision has been made to continue the analysis without removing this suspected value.

3.1.4. Homogeneity of variances (homoscedasticity) within experimental designs

The assumption of the null hypothesis in the analysis of equal variances is that all variances are equal. On the other hand, the alternative hypothesis assumed that at least one variance is different at significance level α = 0.05. In 2 x 8 design, S_{AH}, S_{AL}, S_{BH}, S_{BL}, U_{AH}, U_{AL}, U_{BH} and U_{BL} showed 95% Bonferroni CI for S.D. of (0.307977 to 1.67726), (0.232627 to 4.25447), (0.307626 to 3.30808), (0.251840 to 4.91992), (0.166974 to 1.16824), (0.108259 to 4.16752), (0.202711 to 1.20854) and (0.347617 to 2.01593), respectively. For U_H, U_L, S_H and S_L in 2 x 4 design, CI was (0.321898 to 1.22862), (0.578193 to 2.71302), (0.922316 to 2.64605) and (0.673005 to 1.78064), respectively.
Complementarily, the 4 x 8 design of U_{1H}, U_{1L}, U_{2H}, U_{2L}, U_{3H}, U_{3L}, S_{8H} and S_{8L} showed CI (0.142608 to 1.21647), (0.187204 to 3.06863), (0.185335 to 1.39826), (0.189558 to 1.83306), (0.220910 to 2.18828), (0.244200 to 4.90146), (0.300196 to 1.70512), (0.229401 to 4.37330), respectively. The individual confidence level for the three assay designs was 99.38%, 98.75% and 99.38% for the three experimental designs, respectively. Levene test statistics for the three assays in the same experimental design order were 0.73, 2.12 and 0.55. In the summary plot of Figures 3 - 5, the p-values for all treatments in each experimental group are greater than the common significance level of 0.05. None of the differences between the groups are statistically significant, and all the comparison intervals overlap.

3.2. Statistical evaluation of the assay suitability system

Experimental sources of variation were investigated using analysis of variance (ANOVA) to exclude the inefficiency of the assay tests. In general, variance ratios showed satisfactory outcomes for different sources of variation from each test design as could be found in Tables 4, 5 and 6. Linear regression parameter should be fairly large above the critical limiting value. Deviation from parallelism was fairly below 4.05 and 2.83 limiting values. The calculated probability of linear regression should be significant ((<0.05) to compute the 95% confidence limit. The same also applied for non-parallelism but it should be not significant i.e. ≥0.05. Thus, the pharmacopeial criteria have been fulfilled. Other sources of variations that are non-official include the effect of preparation and the columns/rows matrix in large plates (such as 8 x 8 in this case) which were all within the acceptable limits. Effect of duplication in 4 x 8 assay has another special inspection characteristic which embraced the contrast of both the standard and the unknown. The variance ratios for these properties were well below the limiting value of 1.70.

Thus, it could be concluded that there are no signs for invalidating the assay and the process of potency determination could be conducted.

3.3. Potency determination of the antibiotic and establishing confidence limit for each design type

After confirmation of the validity and suitability of the assay design for the determination of Neomycin Sulfate antimicrobial potency, the microbiological activity for the active raw material was calculated for each design and the confidence range was established. The calculated antibiotic activity of the active medicinal material for the three designs was 745.87, 746.04 and 750.65 μg/mg for 2 x 8, 4 x 8 and 2 x 4 designs, respectively. The upper and lower confidence limits percent (at p=0.95) were as the following in the same order: (124.66%, 80.22%), (114.30%, 87.49%) and (109.47%, 86.95%) corresponding to (929.82 μg/mg, 598.31 μg/mg), (852.71 μg/mg, 652.72 μg/mg) and (821.77 μg/mg, 685.69 μg/mg), respectively. The confidence limit percent range was 18.1%, 26.8 and 44.4% for 2 x 4, 4 x 8 and 2 x 8 designs, respectively. These findings could be visualized in Figure 7 in the individual value plot and potency diagram with the corresponding confidence thresholds for each microbiological rectangular plate design.

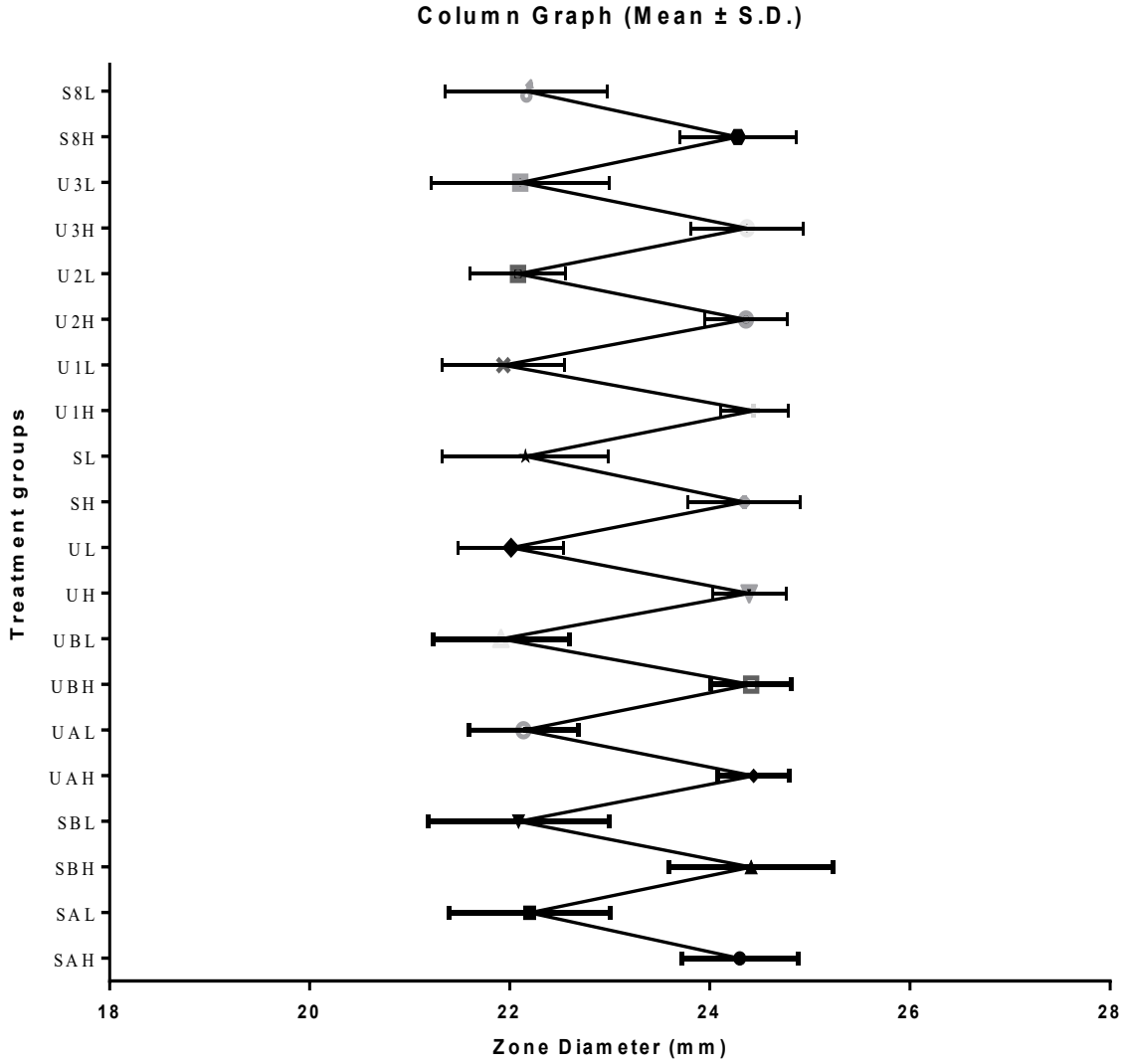


Figure 1. Means of treatment groups for three different designs ± standard deviations (S.D.) demonstrating parallelism between standards (S) and unknowns (U) Neomycin Sulfate preparations.

Table 2. Descriptive statistics showing comparison between treatment groups of three different designs expressed as zone of inhibition (mm) for the same experimental subject against the same control under identical conditions.

Design (P x T)	**2 x 8**								**2 x 4**						**4 x 8**					
CS	S_{AH}	S_{AL}	S_{BH}	S_{BL}	U_{AH}	U_{AL}	U_{BH}	U_{BL}	U_{H}	U_{L}	S_{H}	S_{L}	U_{1H}	U_{1L}	U_{2H}	U_{2L}	U_{3H}	U_{3L}	S_{8H}	S_{8L}
NoV	8	8	8	8	8	8	8	8	16	16	16	16	8	8	8	8	8	8	8	8
Minimum [3]	23.40	20.60	23.10	20.40	23.90	21.00	23.90	21.00	23.90	21.00	23.40	20.20	23.91	20.97	23.88	21.30	23.50	20.24	23.36	20.55
25% Percentile	23.85	21.75	23.88	21.33	24.05	21.80	24.03	21.38	24.03	21.75	24.00	21.85	24.08	21.23	24.03	21.87	23.99	21.86	23.87	21.72
Median [3]	24.30	22.30	24.45	22.40	24.55	22.40	24.45	21.70	24.45	22.05	24.45	22.25	24.57	22.24	24.25	22.00	24.43	22.08	24.28	22.28
75% Percentile	24.78	22.68	24.78	22.75	24.78	22.40	24.80	22.63	24.78	22.40	24.70	22.70	24.74	22.42	24.83	22.58	24.63	22.73	24.74	22.64
Maximum [3]	25.10	23.30	25.90	23.00	24.80	22.70	25.00	22.90	24.90	22.80	25.40	23.30	24.79	22.44	24.94	22.76	25.39	23.20	25.07	23.27
Range	1.70	2.70	2.80	2.60	0.90	1.70	1.10	1.90	1.00	1.80	2.00	3.10	0.88	1.47	1.06	1.46	1.89	2.96	1.71	2.72
10% Percentile	23.40	20.60	23.10	20.40	23.90	21.00	23.90	21.00	23.90	21.07	23.47	20.48	23.91	20.97	23.88	21.30	23.50	20.24	23.36	20.55
90% Percentile	25.10	23.30	25.90	23.00	24.80	22.70	25.00	22.90	24.90	22.73	25.19	23.23	24.79	22.44	24.94	22.76	25.39	23.20	25.07	23.27
95% CIoM [1]																				
ACL (%) [5]	99.22	99.22	99.22	99.22	99.22	99.22	99.22	99.22	97.87	97.87	97.87	97.87	99.22	99.22	99.22	99.22	99.22	99.22	99.22	99.22
LCL	23.40	20.60	23.10	20.40	23.90	21.00	23.90	21.00	24.00	21.70	24.00	21.80	23.91	20.97	23.88	21.30	23.50	20.24	23.36	20.55
UCL	25.10	23.30	25.90	23.00	24.80	22.70	25.00	22.90	24.80	22.40	24.70	22.70	24.79	22.44	24.94	22.76	25.39	23.20	25.07	23.27
Mean	24.30	22.20	24.41	22.09	24.44	22.14	24.41	21.91	24.39	22.01	24.34	22.16	24.44	21.94	24.36	22.08	24.37	22.11	24.28	22.17
Std. Deviation	0.583	0.807	0.818	0.903	0.358	0.545	0.402	0.679	0.366	0.529	0.561	0.831	0.338	0.615	0.413	0.478	0.564	0.888	0.580	0.813
SEM	0.206	0.285	0.289	0.319	0.126	0.192	0.142	0.240	0.091	0.132	0.140	0.207	0.119	0.217	0.146	0.169	0.199	0.313	0.205	0.287
LCIM 95%	23.81	21.53	23.73	21.33	24.14	21.68	24.08	21.34	24.20	21.73	24.04	21.71	24.16	21.42	24.01	21.68	23.90	21.36	23.79	21.49
UCIM 95%	24.79	22.87	25.10	22.84	24.74	22.59	24.75	22.48	24.59	22.29	24.64	22.60	24.72	22.45	24.71	22.48	24.84	22.85	24.77	22.85
CV (%) [6]	2.400	3.636	3.352	4.089	1.466	2.462	1.645	3.099	1.500	2.403	2.304	3.751	1.383	2.803	1.695	2.166	2.315	4.015	2.391	3.666
GM [7]	24.29	22.19	24.40	22.07	24.44	22.13	24.41	21.90	24.39	22.01	24.34	22.14	24.44	21.93	24.36	22.08	24.37	22.09	24.27	22.15
GSDF [8]	1.024	1.038	1.034	1.043	1.015	1.025	1.017	1.031	1.015	1.024	1.023	1.039	1.014	1.029	1.017	1.022	1.023	1.042	1.024	1.038
LCIGM 95%	23.81	21.51	23.73	21.32	24.14	21.67	24.08	21.34	24.20	21.72	24.04	21.69	24.16	21.42	24.02	21.68	23.90	21.34	23.79	21.48
UCIGM 95%	24.79	22.88	25.09	22.85	24.74	22.60	24.75	22.48	24.59	22.29	24.64	22.60	24.72	22.45	24.70	22.48	24.84	22.86	24.77	22.85
HM [4]	24.29	22.17	24.39	22.05	24.43	22.13	24.41	21.89	24.39	22.00	24.33	22.13	24.44	21.92	24.35	22.07	24.36	22.07	24.27	22.14
LCIHM 95%	23.81	21.50	23.73	21.30	24.14	21.67	24.08	21.34	24.20	21.72	24.04	21.68	24.16	21.41	24.02	21.68	23.90	21.33	23.79	21.46
UCIHM 95%	24.79	22.89	25.09	22.86	24.74	22.60	24.75	22.47	24.59	22.29	24.63	22.60	24.72	22.46	24.70	22.48	24.84	22.87	24.76	22.86
QM [2]	24.31	22.21	24.42	22.10	24.44	22.14	24.42	21.92	24.40	22.02	24.35	22.17	24.44	21.94	24.36	22.09	24.38	22.12	24.29	22.18
LCIQM 95%	23.81	21.54	23.73	21.35	24.14	21.69	24.08	21.34	24.20	21.74	24.05	21.73	24.16	21.43	24.01	21.68	23.90	21.38	23.80	21.50
UCIQM 95%	24.79	22.87	25.10	22.83	24.74	22.59	24.75	22.48	24.59	22.30	24.65	22.60	24.72	22.45	24.71	22.48	24.85	22.84	24.77	22.84
Skewness [9]	-0.173	-1.011	0.298	-1.132	-0.557	-1.536	0.249	0.321	0.020	-0.590	0.036	-1.084	-0.674	-0.938	0.523	0.170	0.347	-1.295	-0.217	-1.041
Kurtosis [10]	-1.221	1.754	1.247	0.331	-1.504	2.214	-1.166	-1.291	-1.599	-0.261	-0.585	1.285	-1.262	-0.983	-1.382	0.188	0.862	2.734	-1.112	1.816

[1]Confidence Interval = Sample mean ± t x Std. Deviation/Square root of sample size, with t-value for 95% confidence = 2.262.

[2] Calculate the square of every reading, then take the average of the squares and finally compute the square root of this mean.

[3] R = P x (n + 1)/100, where; P = Desired percentile and n = Number of values in the data set.

[4] Reciprocal of the mean of the reciprocal of the values in each treatment.

[5] Dependent on the precise values of numerator and denominator.

[6] Relative Variability = Std. Deviation/ Mean (expressed either as a fraction or a percent).

[7] Calculated by taking the average of the logarithms of the whole dataset, then calculating the antilog of the mean.

[8] The firsts step is to transform all the readings to logarithm values, calculate the sample Std. Deviation of these log sets, and then converting to the antilogarithm of those Std. Deviation.

[9]Quantification of the distribution symmetry. The ideal distribution which should be symmetrical possesses a skewness of zero.

[10]Measure how the tails of the dataset distribution are close to the perfect Gaussian spreading. That distribution has a kurtosis of zero.

ACL: Actual Confidence Level. CIoM: Confidence Interval of Median. CV: Coefficient of Variation. GM: Geometric Mean. GSDF: Geometric SD factor. HM: Harmonic mean. LCIGM 95%: Lower 95% CI of geo. Mean. LCIHM 95%: Lower 95% CI of harm. Mean. LCIM 95%: Lower 95% CI of mean. LCIQM 95%: Lower 95% CI of quad. Mean. LCL: Lower Confidence Limit. NOV: Number of Values. CS: Column Statistics. P x T: Preparations and Treatments. QM: Quadratic mean. SEM: Standard Error of Mean. UCIGM 95%: Upper 95% CI of geo. Mean. UCL: Upper Confidence Limit. UCIHM 95%: Upper 95% CI of harm. Mean. UCIM 95%: Upper 95% CI of mean. UCIQM 95%: Upper 95% CI of quad. Mean.

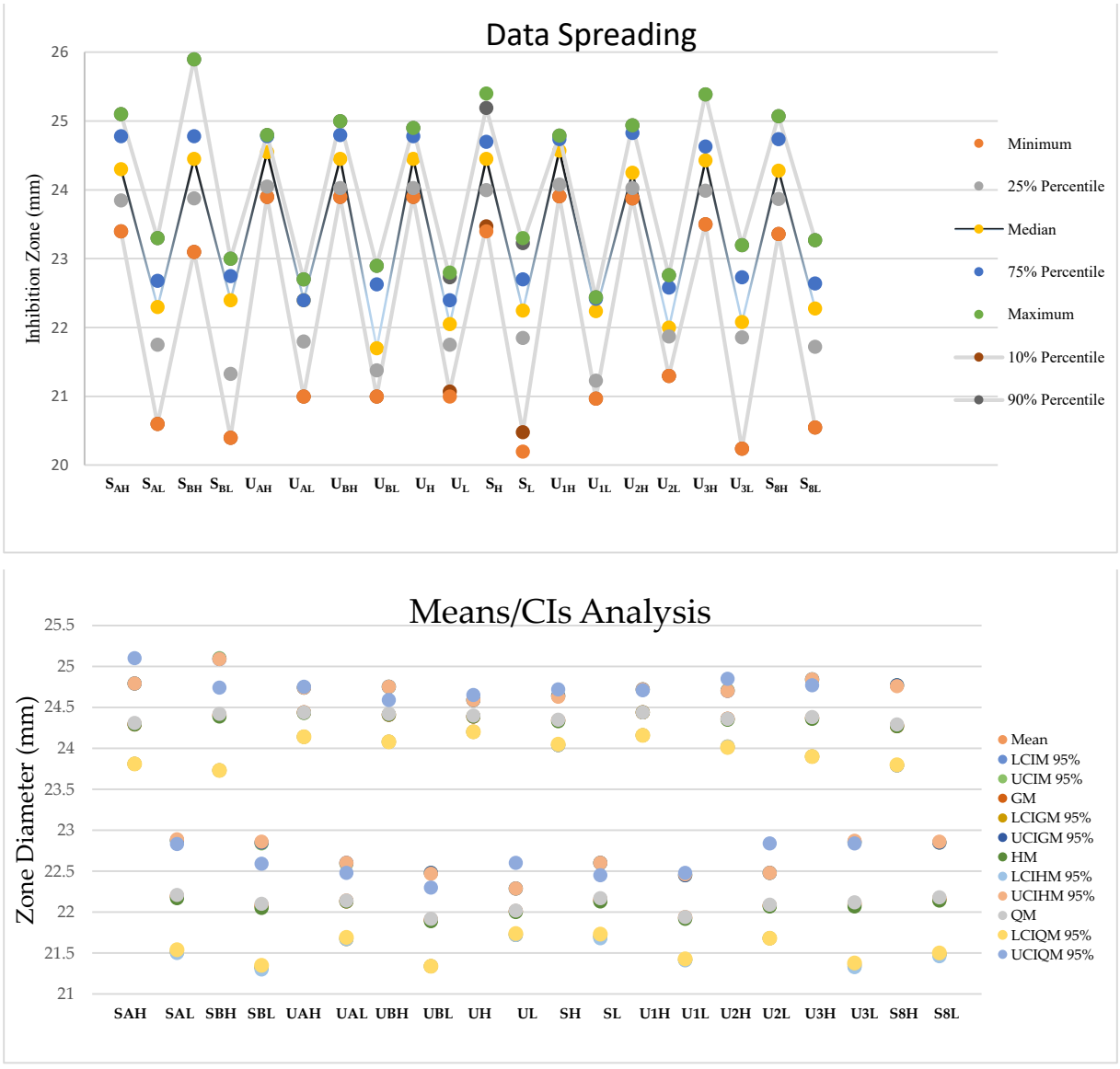


Figure 2. Visual descriptive analysis of data distribution for each treatment group of each antibiotic assay design showing dispersion pattern and the means with 95% CIs.

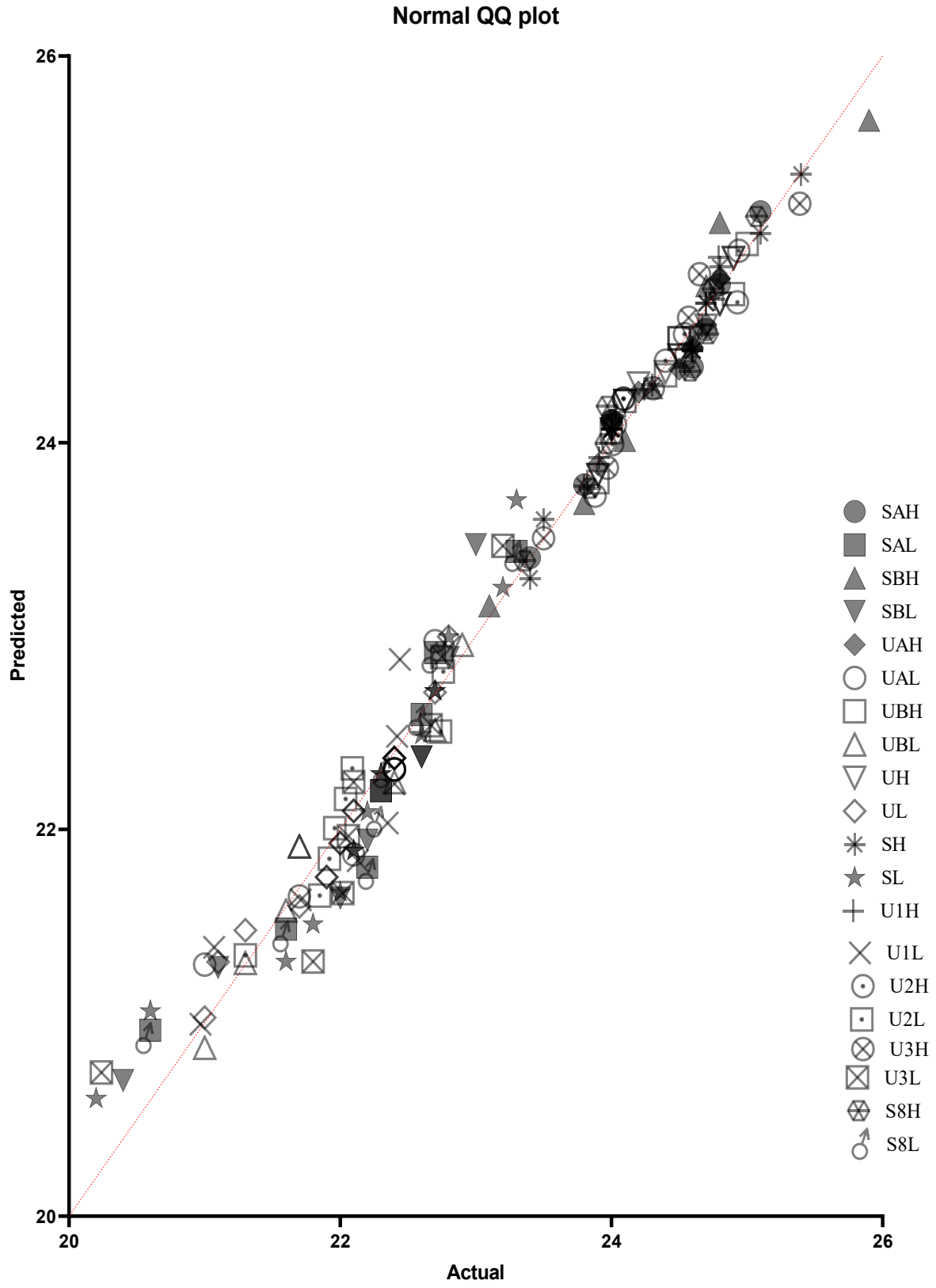


Figure 3. Scatter probability graph showing normality of the zone inhibition groups for assay designs of two preparations, (a) eight and (b) four treatments.

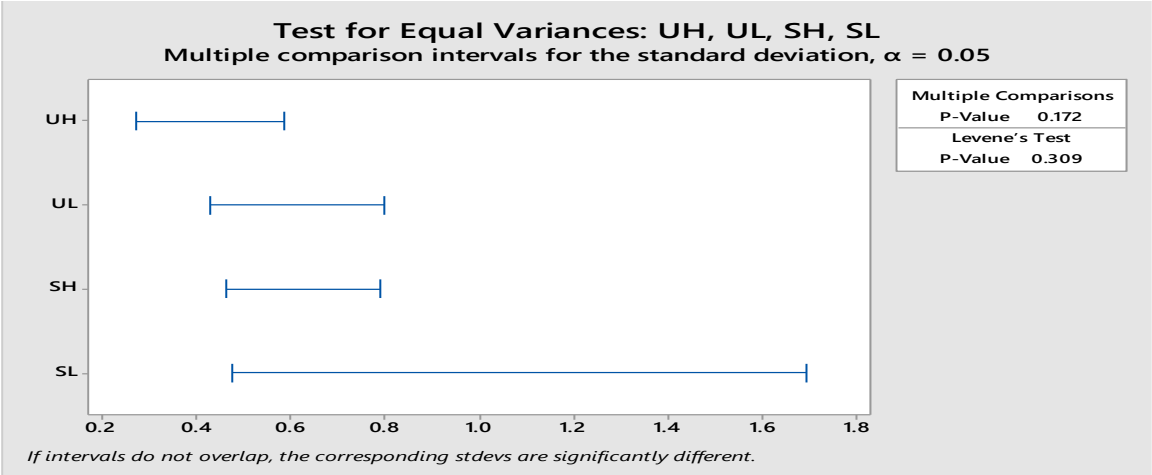


Figure 4. Homogeneity of variances test of two preparations x four treatments assay design, showing: equality of variances: significance of differences for standard deviations.

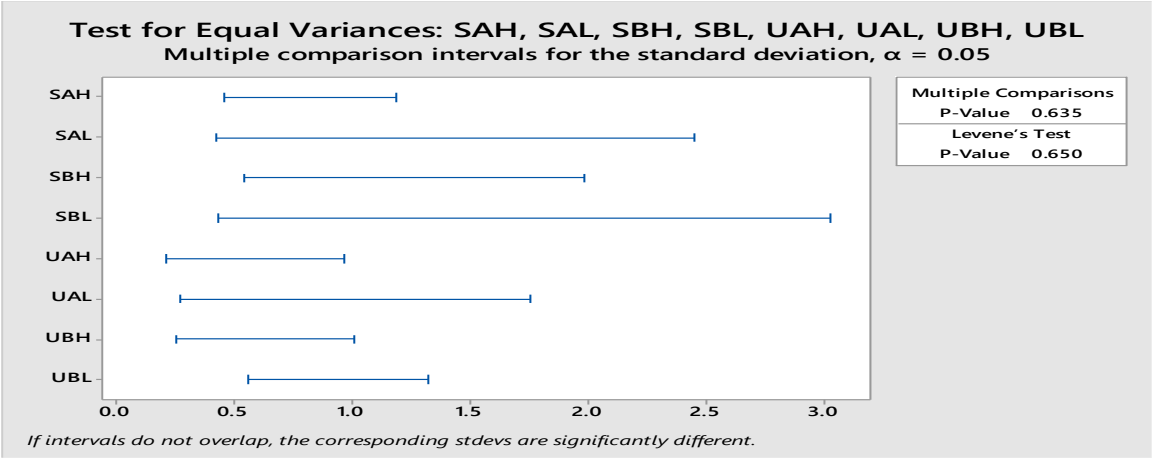


Figure 5. Homogeneity of variances test of two preparations x eight treatments assay design, showing: (equality of variances: significance of differences for standard deviations.

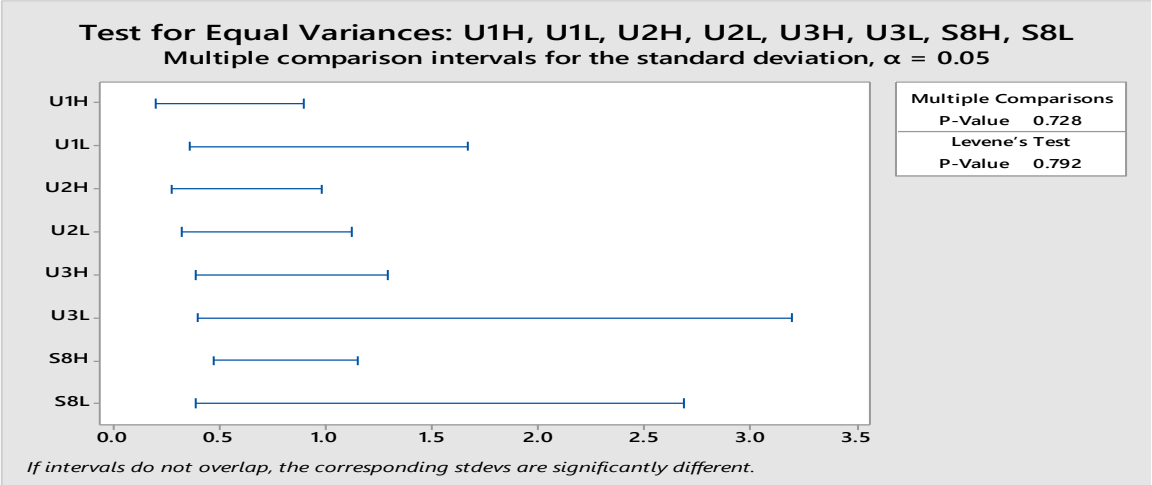


Figure 6. Homogeneity of variances test of two preparations x eight treatments assay design, showing: equality of variances: significance of differences for standard deviations.

Table 4. Analysis of variance report of two-dose level assay using a square plate and 8 x 8 Latin square 2 x 4 design.

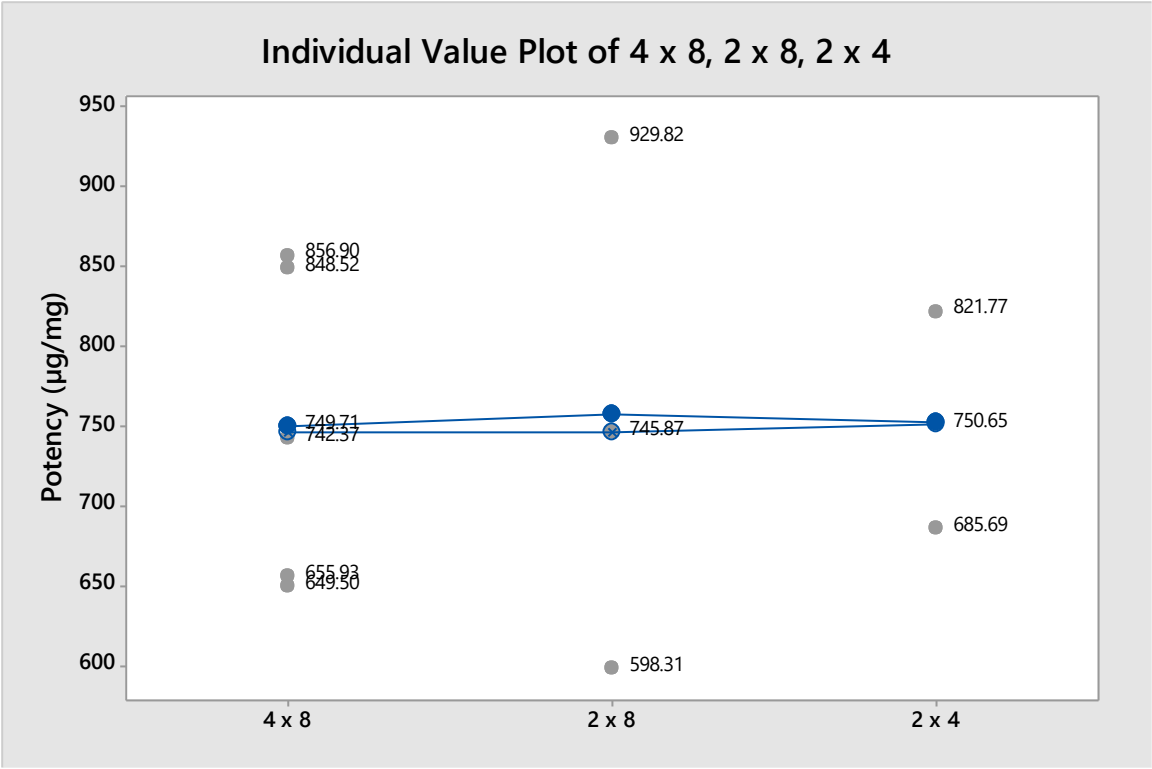
Source of Variance	d.f.	Sum of Squares	Mean Square	Variance Ratio (F-value)	Limiting Value	Calculated Probability
Derivative from Raw Total	7	2.79	0.40	1.129	<2.22	0.3618
Derivative from Column Total	7	2.09	0.30	0.844	<2.22	0.5573
Derivative from Treatment Total	3	84.10				
Derivative from Preparation Total	1	0.01	0.01	0.030	<4.05	0.8622
Regression Squares	1	83.93	83.93	237.624	>4.05	0.0000
Parallelism Squares	1	0.16	0.16	0.462	<4.05	0.5003
Residual Error (SSreg)	46	16.25	0.35	1.000	<1.63	
Total (SStot)	63	105.23				

Table 5. Analysis of variance report of two-dose level assay using a square plate and 4 x 8 Latin square 2 x 8 design.

Source of Variance	d.f.	Sum of Squares	Mean Square	Variance Ratio (F-value)	Limiting Value	Calculated Probability
Derivative from Raw Total	7	2.79	0.40	1.04	<2.24	0.6461
Derivative from Column Total	7	2.09	0.30	0.78	<2.24	0.7930
Derivative from Treatment Total	7	84.26				
Derivative from Preparation Total	3	0.02	0.01	0.018	<2.83	0.2761
Regression Squares	1	83.93	83.93	219.13	>12.52	0.0000
Parallelism Squares	3	0.31	0.16	0.27	<2.83	0.7138
Residual Error (SSreg)	42	16.09	0.35	1.00	<1.67	
Total (SStot)	63	105.23				

Table 6. Analysis of variance report of two-dose level assay using a square plate and 8 x 8 Latin square 2 x 8 design.

Source of Variance	d.f.	Sum of Squares	Mean Squares	Variance Ratio	Limiting Value	Calculated Probability
Preparation	1	0.16	0.16	0.42	<4.07	0.5229
Regression	1	84.94	84.94	217.09	>12.52	0.0000
Parallelism (deviation from)	1	0.14	0.14	0.35	<4.07	0.5587
Contrast of standard	1	0.00	0.00	0.00	<1.70	0.9485
Contrast of unknown	1	0.16	0.16	0.41	<1.70	0.5265
Contrast of standard slope	1	0.08	0.08	0.20	<1.70	0.6555
Contrast of unknown slope	1	0.08	0.08	0.20	<1.70	0.6535
Subtotal	7	85.40				
Treatments	7	85.40				
Rows	7	5.01	0.72	1.83	<2.24	0.1067
Columns	7	3.22	0.46	1.18	<2.24	0.3376
Error by difference	42	16.43	0.39	1.00		
Total	63	110.06				



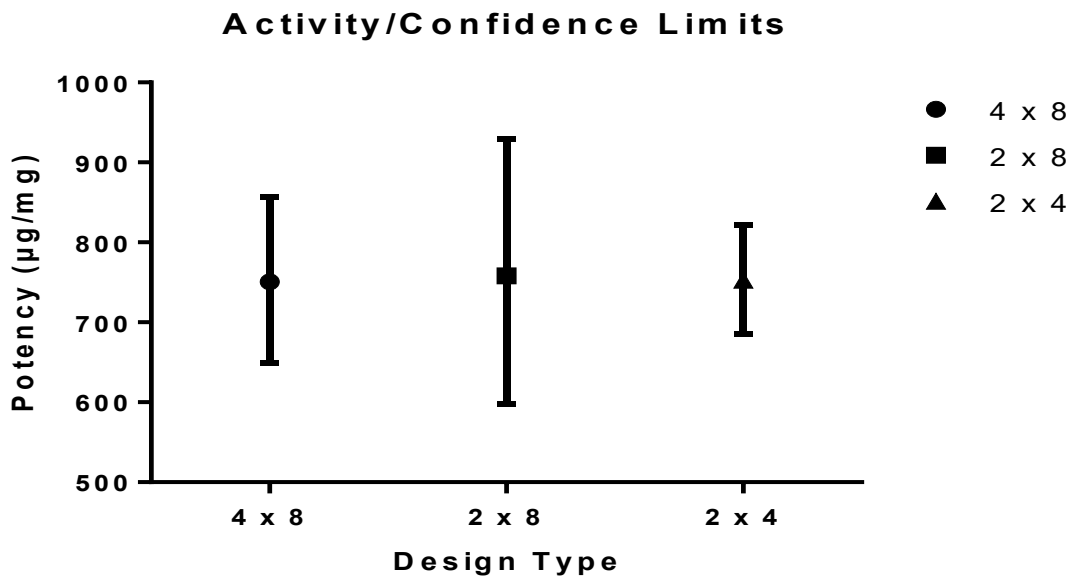


Figure 7. Mean and range graph showing the estimated confidence limits of the potency for three design types of microbiological assays presented as Individual and mean/range plots.

4. Discussion

Agar diffusion assay was known to be one of the most commonly used quality control (QC) testing for the determination of the antimicrobial potency of the antibiotic substance [27]. To date, the true antibiotic activity of Neomycin Sulfate must be officially determined using microbiological analysis [28]. This would be crucial as this antibiotic was added in "antibiotics and vitamins of the b group" from the WHO index of degradable substances [6]. This two-dose symmetrical PLM was statistically evaluated through four and eight treatments for two and four preparations. The stepwise statistical intervention

would serve as the basis for both the verification of the validity of the assay data – and hence the experimental model - for the regular testing of the drug potency and preferential comparison between the selected modules for the quality of the results obtained [8]. Ensuring the validity of the analysis system is a crucial step before drawing any results and/or conclusions from the test. Commercially available programs for the calculations should be assessed critically. The prepared Excel workbook designed for the potency assessment and determination was evaluated against already comprehensively expressed examples step-by-step to ensure the validity of software for computation.

4.1. Initial statistical investigation of the suitability of the assay output results

Initial examination of the raw data would be mandatory to exclude any unusual record from the dataset for this initial step. This could be followed by the investigation of the sources of the variation in the testing using ANOVA [29]. Statistical analysis of the raw data should reveal a dataset pattern that would be trustworthy to proceed in further computation to study the suitability of the assay design for potency calculation of the subject material [30]. The recorded datasets were the measured zone of inhibition diameter produced by the diffusion of the antibiotic through the agar matrix measured to the nearest 0.01 mm.

4.1.1. Overview of the pattern of treatment groups

The descriptive statistical study was essential in this study to detect and exclude any aberrant observations before further processing of the datasets which have been derived from the original raw data measured from the antibiotic plates [31]. While parallelism could be assessed statistically, it might be easily investigated visually using the graphical presentation of the successive high and low doses (Figure 1) [32]. In addition, a combination of tabulated and illustrative analysis of each treatment group might show the spreading, pattern and homogeneity of the records. At this stage, the clustering tendency of data could be detected which would explain an apparent outlier value as in Figure 2 [6]. While the general statistical description provided herein was comprehensive for clarity and demonstration, the actual routine preliminary analysis might make use of significantly reduced diversification in the examination of the datasets.

4.1.2. Normality of the treatment group datasets

While small deviation in the normality of data might not encounter a serious problem in the result interpretation [32]. It would be plausible to track - from time to time with reasonable frequency - the state of data distribution. An array of tests was used for a comparative study of the normality behavior. Yet, the Shapiro-Wilk normality test was considered as an official one for treatment groups with seven or more replicates [31]. Clustering patterns might show extreme values as excursions even if they were not true aberrant values (known by experience from the trending history of data). In turn, this behavior would influence the normality test results by affecting the expected shape of the standard expected bell-shaped of the Gaussian pattern of the distribution. Nevertheless, it should be understood that the actual distributions in real and practical experimentation would be expected to not follow the exact theoretically hypothesized dispersion and deviations that were common in study groups [33-35]. However, these deviations should be considered statistically within a reasonable range.

4.1.3. Outlier evaluation in each treatment group

As discussed earlier, outliers had an impact on data and their statistical evaluation. However, it should be highlighted that rejection and removal of the aberrant values blindly based solely on the outcome of the test would be a discouraging practice [36]. Each situation of detection and removal of the outlier should be evaluated case-by-case to ensure the avoidance of unintended bias that would lead to the omission of a truly valid result as it could be found in the current case [6,36]. The decision of canceling the rejection action - based on the previous experience - was supported by another battery of tests despite the fact that the compendial method was the original source of this marginal alarm for the excursion. Again, this might happen due to a possible clustering tendency in the datasets [6].

4.1.4. Analysis of homoscedasticity in the experimental designs

Homoscedasticity is another criterion that should be investigated between the treatment groups of the antibiotic potency assay experiment. Homogeneity of variances might be checked by either Bartlett's test or Cochran's test [31]. These appeared to be little used in the microbiological assay and so were not considered here. For further information, the referencing to the European Pharmacopoeia might be consulted [6,32]. However, it was noteworthy that Bartlett's test was criticized by Box (1953) as being not robust to non-normality. The author compared it with "putting to sea in a rowing boat to see if conditions were fit for an ocean liner to leave port" [6]. In the present work, a different test was used that was found convenient in terms of simplicity of implementation and ease of interpretation. Levene's test had the advantage of being less sensitive for the departure from normality than Bartlett's test [37].

4.2. Statistical investigation of the sources of variations

After ensuring the quality and validity of the recorded results of the inhibition zone for further processing. It would be mandatory to examine the validity of the assay and to determine its suitability to calculate the potency of the sample under examination [36]. Otherwise, the estimated activity might be inaccurate. Identification of the sources of variations per an assay design was essential to control the possible sources of errors and investigate any abnormal result or outcome [8]. There were common criteria to be examined both official (regression and parallelism) and ono-official (preparation, row and column) [8]. There might be special criteria for a specific design such as that for 2 x 8 assay as there was contrast analysis for both duplicate standard and test with their slopes [6]. An acceptable ANOVA result would deliver a solid estimate for the potency of the examined antimicrobial material. These sources of variances were evident in Figures 4, 5 and 6.

4.3. Final result of antimicrobial activity evaluation for each design

Under identical experimental conditions, the results of potency determination for Neomycin Sulfate (expressed as μg/mg) were reasonably close within the three-experimental layout designs. Figure 7 demonstrated variable confidence intervals between them. When high confidence would be desirable, a modification in the assay test might be required [6,26]. To reach this goal. It would be necessary to align this criterion with the main purpose of the antibiotic assay [38]. Thus, a design that was aimed to screen the compounds for the antimicrobial activity should be different from that was designed for estimation of the activity of the product in the bulk or finished pharmaceutical preparations. Increasing the number of replicates for each treatment group must be considered as an important factor.

5. Conclusions

The current study provided an example for the quantitative assessment of the suitability of the testing system designs to determine the potency of the Active Pharmaceutical Ingredient (API) using PLM of 2 x 2 assay. In the present case of Neomycin Sulfate, the two-dose balanced experiment for three different designs in large rectangular antibiotic plates showed acceptable system suitability of 2 x 4, 2 x 8 and 4 x 8 for preparations x treatments. While the variation in the potency determination was <5.0%, the differences in the confidence limits were noticeable. Further study would be necessary to control experimental designs and conditions such as the number of replicates to bring the confidence range within the desired window depending on the main purpose and the target from the activity measurements. In the present situation, the 2 x 4 assay design showed a tighter confidence window in comparison to the 2 x 8 design. Long-term monitoring of the potency determination test using the current methodology might provide solid evidence for the adjustment and fine-tuning of the experiment layout design. Thus, the trending of data using control charts would assist the evaluation of the assay through a comparative study.

Author Contributions: Conceptualization, D.E and E.R; methodology, M.E; software, M.E; validation, E.R, D.E and M.E; formal analysis, M.E; investigation, E.R; resources, D.E; data curation, M.E; writing—original draft preparation, M.E; writing—review and editing, E.R; visualization, E.R; supervision, E.R; project administration, Dalia Eissa; funding acquisition, Dalia Eissa. All authors have read and agreed to the published version of the manuscript.

Funding: This work is supported by authors only without any external fund.

Acknowledgments: None to declare.

Conflicts of Interest: the authors declare that there is no competing conflict of interest.

References

[1] S. Gad, "Pharmaceutical manufacturing handbook: production and processes", Wiley-Interscience, 2008.

[2] R. Unissa, M. Sudhakar and M. Hadi, "Lab Manual for Pharmaceutical Microbiology", 1st ed. Ahmedabad: Nirav and Roopal Prakashan, 2011.

[3] V. Fedorenko et al., "Antibacterial discovery and development: from gene to product and back", BioMed Research International, vol. 2015, pp. 1-16, 2015. Available: 10.1155/2015/591349.

[4] Technical Guide for the elaboration of monographs, "European Pharmacopoeia", Edqm.eu, 2011. [Online]. Available: https://www.edqm.eu/medias/fichiers/technical_guide_for_the_elaboration_of_monographs_.pdf. [Accessed: 30- Aug- 2021].

[5] Health Ministers of the United Kingdom , "British Pharmacopoeia", 3rd ed. London: Medicines and Healthcare Products Regulatory Agency, 2021.

[6] W. Hewitt, "Microbiological assay for pharmaceutical analysis: a rational approach", Boca Raton, Fla: Interpharm/CRC, 2004.

[7] Pharmacopeial Forum, "The United States pharmacopoeia", 43rd ed. North Bethesda, Maryland , United States: United States Pharmacopeial Convention, Vol. 44(6), 2021.

[8] W. Hewitt, "Microbiological Assay", Saint Louis: Elsevier Science, 2014.

[9] J. Lightbown, "Biological standardisation and the Analyst. A review", The Analyst, vol. 86, no. 1021, p. 216, 1961. Available: 10.1039/an9618600216.

[10] W. Sokolski, C. Chidester, O. Carpenter and W. Kaneshiro, "Assay Methods for Total Neomycins B and C", Journal of Pharmaceutical Sciences, vol. 53, no. 7, pp. 826-828, 1964. Available: 10.1002/jps.2600530731.

[11] W. Köhler, "Who Expert Committee on Biological Standardization, 23. Report (WHO Techn. Rep. Ser. No. 463). 120. Genf 1970: World Health Organization", Zeitschrift für allgemeine Mikrobiologie, WHO/BS/70.1001., pp. 255-255, 1971. Available: 10.1002/jobm.19680080326.

[12] W. Wilson, G. Richard and D. Hughes, "Thin-layer chromatographic identification of the gentamicin complex", Journal of Chromatography A, vol. 78, no. 2, pp. 442-444, 1973. Available: 10.1016/s0021-9673(73)30103-0.

[13] L. Clontz," Microbial limit and bioburden tests", Boca Raton: CRC Press, 2009.

[14] WHO, "The International pharmacopoeia", Geneva: World Health Organization, 2006.

[15] C. Sheehan, "Overview of International Harmonization through the Pharmacopeial Discussion Group", Usp.org, 2014. [Online]. Available: https://www.usp.org/sites/default/files/usp/document/get-involved/stakeholder-forums/2b-excipients-and-harmonzation-overview-2014-02-19.pdf. [Accessed: 01- Sep- 2021].

[16] Technical Data, "Antibiotic Assay Medium F", Himedialabs.com, 2020. [Online]. Available: https://himedialabs.com/TD/M923.pdf. [Accessed: 01- Sep- 2021].

[17] Pharmaceutical Secondary Standard; Certified Reference Material, "Neomycin Sulfate Pharmaceutical Secondary Standard; Certified Reference Material | Sigma-Aldrich", Sigmaaldrich.com, 2021. [Online]. Available: https://www.sigmaaldrich.com/EG/en/product/sial/phr1491. [Accessed: 01- Sep- 2021].

[18] C. Valgas, S. Souza, E. Smânia and A. Smânia Jr., "Screening methods to determine antibacterial activity of natural products", Brazilian Journal of Microbiology, vol. 38, no. 2, pp. 369-380, 2007. Available: 10.1590/s1517-83822007000200034.

[19] J. Hudzicki, "Kirby-Bauer Disk Diffusion Susceptibility Test Protocol", Asm.org, 2009. [Online]. Available: https://asm.org/getattachment/2594ce26-bd44-47f6-8287-0657aa9185ad/Kirby-Bauer-Disk-Diffusion-Susceptibility-Test-Protocol-pdf.pdf. [Accessed: 01- Sep- 2021].

[20] GraphPad Software LLC, "GraphPad Prism 9 User Guide - Welcome to Prism 9 User Guide", Graphpad.com, 2021. [Online]. Available: https://www.graphpad.com/guides/prism/latest/user-guide/index.htm. [Accessed: 01- Sep- 2021].

[21] M. Evans, G. McCabe and D. Moore, “Minitab manual for Moore and McCabe's Introduction to the practice of statistics”, third edition. New York: W.H. Freeman, 1999.

[22] W. Winston, “Microsoft excel 2016”, Redmond, Washington: Microsoft, 2016.

[23] C. Carlberg, “Statistical analysis”, Indianapolis, Indiana: Que, 2018.

[24] T. Oppe, J. Menegola and E. Schapoval, "MICROBIOLOGICAL ASSAY FOR THE DETERMINATION OF CEFPIROME IN RAW MATERIAL AND INJECTABLE PREPARATION", Drug Analytical Research, vol. 2, no. 1, pp. 29-35, 2018. Available: 10.22456/2527-2616.84473.

[25] S. Nahar, M. Khatun and M. Kabir, "Application of microbiological assay to determine the potency of intravenous antibiotics", Stamford Journal of Microbiology, vol. 10, no. 1, pp. 25-29, 2020. Available: 10.3329/sjm.v10i1.50729.

[26] A. Zuluaga, M. Agudelo, C. Rodriguez and O. Vesga, "Application of microbiological assay to determine pharmaceutical equivalence of generic intravenous antibiotics", BMC Clinical Pharmacology, vol. 9, no. 1, 2009. Available: 10.1186/1472-6904-9-1.

[27] E. Cazedey and H. Salgado, "Development and Validation of a Microbiological Agar Assay for Determination of Orbifloxacin in Pharmaceutical Preparations", Pharmaceutics, vol. 3, no. 3, pp. 572-581, 2011. Available: 10.3390/pharmaceutics3030572.

[28] N. Dafale, U. Semwal, P. Agarwal, P. Sharma and G. Singh, "Development and validation of microbial bioassay for quantification of Levofloxacin in pharmaceutical preparations", Journal of Pharmaceutical Analysis, vol. 5, no. 1, pp. 18-26, 2015. Available: 10.1016/j.jpha.2014.07.007.

[29] F. Rebello Lourenco, M. Augusto Lyrio Traple, R. Takao Okamoto and T. de Jesus Andreoli Pinto, "Development and Validation of Microbiological Assay for Ceftriaxone and its Application in Photo-stability Study", Current Pharmaceutical Analysis, vol. 9, no. 1, pp. 77-81, 2013. Available: 10.2174/1573412911309010011.

[30] T. Mzolo, Statistical methods for the analysis of bioassay data. Technische Universiteit Eindhoven, 2016.

[31] F. Lourenço and T. Pinto, "Comparison of three experimental designs employed in gentamicin microbiological assay through agar diffusion", Brazilian Journal of Pharmaceutical Sciences, vol. 45, no. 3, pp. 559-566, 2009. Available: 10.1590/s1984-82502009000300022.

[32] Statistical analysis, "STATISTICAL ANALYSIS OF RESULTS OF BIOLOGICAL ASSAYS AND TESTS", Uspbpep.com, 2008. [Online]. Available: http://www.uspbpep.com/ep60/5.3.%20%20statistical%20analysis%20of%20results%20of%20biological%20assays%20and%20tests%2050300e.pdf. [Accessed: 01- Sep- 2021].

[33] M. Eissa, A. Mahmoud and A. Nouby, "Control Chart in Microbiological Cleaning Efficacy of Pharmaceutical Facility", Dhaka University Journal of Pharmaceutical Sciences, vol. 14, no. 2, pp. 133-138, 2016. Available: 10.3329/dujps.v14i2.28501.

[34] M. Eissa, "Application of Laney control chart in assessment of microbiological quality of oral pharmaceutical filterable products", Bangladesh Journal of Scientific and Industrial Research, vol. 52, no. 3, pp. 239-246, 2017. Available: 10.3329/bjsir.v52i3.34160.

[35] M. Essam Eissa, "Determination of the Microbiological Quality of Feed City Water to Pharmaceutical Facility: Distribution Study and Statistical Analysis", ATHENS JOURNAL OF SCIENCES, vol. 4, no. 2, pp. 143-160, 2017. Available: 10.30958/ajs.4-2-4.

[36] M. Essam Eissa, “Suitability system of microbiological method for nystatin potency determination in the routine analysis using agar diffusion method”, SciMedicine Journal (SMJ), in press.

[37] Engineering Statistics Handbook, "1.3.5.10. Levene Test for Equality of Variances", Itl.nist.gov, 2021. [Online]. Available: https://www.itl.nist.gov/div898/handbook/eda/section3/eda35a.htm. [Accessed: 01- Sep- 2021].

[38] M. Essam Eissa, “Validation of symmetrical two-dose parallel line assay model for nystatin potency determination in pharmaceutical product”, Journal of Advanced Pharmacy Research (JAPR), in press.

Journal of Statistics & Applied Science, Volume -2 Issue -2
ISSN 2718-0999

Review

Received: date: 28.09.2021
Accepted: date: 09.102021
Published: date: 28.11.2021

Structural Equation Modeling as a Marketing Research Tool: A Guideline for SEM Users About Critical Issues and Problematic Practices

Umut ÜNAL[1]

[1]Istanbul Medeniyet University, Department of Management; umutunalesp52@gmail.com
Orcid: 0000-0001-7723-6343[1]
*Correspondence: umutunalesp52@gmail.com

Abstract: Structural equation modeling (SEM) is a very powerful multivariate statistical technique that has increasingly been used in social sciences, particularly in marketing. As a consequence of the widespread use of this contemporary analysis method, several issues that SEM users face have become a matter of concern, which are discussed thoroughly in SEM literature. This paper aims to conduct an extensive review of these issues by benefitting from the previous review works, broaden the research criteria by bringing together the issues that are separately addressed in those previous studies, and make an empirical analysis to demonstrate how well these problems are dealt with. Along with the problematic practices identified, the solutions suggested in the literature are presented. By that, this study serves as a basic guideline for SEM users.
Keywords: Structural equation modeling, Critical review, Guideline, Mispractices

1. Introduction

In social and behavioral sciences, the theories become more and more complex, necessitating composite relationships between variables to be assessed in a way that the first-generation statistical tools cannot operate [1], as a result of which the use of multivariate statistical techniques has grown considerably over the last three decades [2]–[4].

Structural Equation Modeling, or *SEM*, which can be considered a combination of factor analysis and regression, is prevalent among such second-generation statistical methods [5]. It is a multivariate statistical technique with an unrivalled ability to simultaneously test complex webs of connections between variables [6]. SEM, unlike other statistic methods, takes potential measurement errors into consideration and makes the evaluation of the model accordingly, which makes it possible to eliminate the indicators having large error terms and/or low loadings and consequently improve the quality of the constructs that form the model [7].

The paths between theoretical constructs, which are represented by latent factors, form the basis of structural equation modeling while the latent constructs are measured via observed variables (i.e. indicators). This multi-layered nature of SEM gives it a unique power to incorporate second, even third-order factors, which provides a better understanding of relationships that may not be evident prima facie [8].

The growing use of SEM in social sciences is also evident in marketing. The increase in the complexity of newly proposed theories in this discipline is reflected onto the complexity of the causal structures that are based upon these theories [9] which led the marketing researchers to apply SEM much more frequently in their studies [10]. Structural equation modeling, which was started to be used by

marketing researchers in the mid-1970s, increasingly became a method of choice for theory development and expansion beginning from the early 80s [7], [11], [12]. The developments in SEM software also contributed to the rapid expansion of this technique among the researchers. When SEM was first introduced in the 70s, the statistics software required the structural model to be specified in terms of matrices to be able to create the path diagram. However, thanks to the recent developments in software, it has become possible to specify the path diagram directly without having to create the matrices first, which is a very tedious work, especially when working on models with many constructs [5]. Nevertheless, such user-friendly software does not magically make SEM a problem-free method. On the contrary, SEM is a complex statistical technique that provides excellent solutions to very complex problems but at the same time poses complex challenges for users [6].

Many reviews in the marketing literature address the problems that researchers have regarding the use of SEM. One of the, or maybe the, most renowned of such works is the paper by Baumgartner and Homburg [11] which reviews SEM applications in four major marketing journals (the Journal of Marketing, Journal of Marketing Research, International Journal of Research in Marketing, and the Journal of Consumer Research) covering an extended period of almost two decades (1977 and 1994). Martínez-López *et al.* [10] examines the SEM-based articles published between 1995 and 2007 in the same four journals to make an update of the paper by Baumgartner and Homburg [11] and at the same time draw a period-wise comparison to demonstrate how the issues identified in the early periods of SEM have been dealt with and what new problems have arisen. Even the sheer number of articles included in these two works, 149 in the former and 472 in the latter, is evidence of the striking increase in the popularity of SEM techniques in marketing literature.

Just as there are such works examining the use of SEM in marketing area covering a particular period or articles published in specific journals (e.g., [10], [11], [13]) there are bibliometric analyses reviewing the use of SEM in marketing studies on country basis as well (e.g., [7], [14], [15]). Dogan [14] examined conference proceedings of national marketing congresses in Turkey between 1999-2017 and reviewed 91 papers that used SEM as their analysis method as an attempt to fill the gap in Turkish marketing literature regarding the issues about the use of SEM. He addresses five common problems in his study (model fit issues, omission of CFA before model testing, misuse of covariance modification indices, insufficient reporting about the constructs' reliability and validity, and mischoice of PLS-SEM over CB-SEM). Even though these are critical problems to be investigated thoroughly, some other issues discussed extensively in the marketing literature are left untouched. Therefore, in this study, it is aimed to update and continue his work by analyzing a wider set of topics central to the use of SEM applications, which are determined based on the examination of previous SEM reviews as sample size adequacy, (non)application of CFA, measurement model specification, evaluation of model fit, model respecification, reliability, validity, common method bias, normality, and mediation.

In this meta-analysis, to demonstrate the most up-to-date picture regarding the issues pinpointed in the previous studies, it was decided to focus on the most recent works. Therefore, marketing-oriented Turkish articles published in 2020 were reviewed. 120 articles using the SEM technique were detected and analyzed based on the criteria mentioned above.

In the following section, these topics and issues related to them are explained one by one.

2. Methodological Issues in The Application of SEM

SEM is a complex statistical technique, the application of which requires special attention to many critical points. However, sometimes the users of SEM fail to take into consideration certain issues that they are supposed to pay attention. To highlight these problematic applications and provide solution suggestions, many researchers reviewed the use of SEM in their field [7], [10], [11], [13]–[15]. In this study, some of the problems that were identified to be frequently faced by SEM users in these previous studies are explained.

2.1. Sample Size Adequacy

One of the most significant limitations of SEM is about its sample size requirement. It is generally acknowledged that SEM is a large-sample technique [16]. Even relatively simple models require fairly large samples, not to mention the complex structures [17]. An inadequate sample size directly affects the reliability of the results, power of analysis and generalizability of research findings [10], [13]. Therefore, sample size is one of the most crucial criteria that SEM modellers should pay attention when applying this technique.

However, even though there is a general agreement on the large sample requirement of SEM, there is no simple rule of thumb answering the question of what is a "large enough" sample size in SEM.

There are two main schools of thought that can be referred to. The first one considers the sample size in isolated terms and makes recommendations of minimum sample size independent of the number of indicators included in the model. For instance, Ding *et al.* [18] suggest a minimum of 100-150 cases. Likewise, Anderson and Gerbing [19] claim that a sample size of 150 is usually enough to obtain sufficiently convergent and proper results. Loehlin [20] recommends an absolute minimum sample size of 200 observations. The second school of thought, on the other hand, recommends determining the sample size through the number of parameters to be estimated. By many researchers this stream of though is considered more reasonable because the varying degrees of complexities of models and extra factors such as missing data make it very difficult to identify a "one-size-fits-all" sample size that will work across all SEM models [16]. In their review, Baumgartner and Homburg [11] also recommend the second school indicating that there cannot be one size that is suitable to all SEM applications. In this sense, the recommendation of *N:q* ratio by Bentler and Chou [21], which considers the ratio of number of observations to the number of free parameters as the criteria of sample size assessment and take 1:5 as acceptable and 1:10 as recommended ratio, is frequently referred to by researchers.

2.2. Measurement Model Specification

In SEM, most of the time the prime interest is on the connections between theoretical constructs which are measured by the observed variables. Therefore, how well the observed variables represent their respective unobserved (latent) variables is very crucial to the main objective of SEM analysis [5]. The substructure of the model, that is, the number of observed variables (i.e. indicators) per construct, profoundly influences the extent to which the structural model is well identified [10]. Out of the two types of constructs, i.e. single-item and multi-item constructs, Baumgartner and Homburg [11] were highly critical of the former and discouraged their use. As for the number of indicators in multi-item constructs, at least 3 indicators per construct is recommended to obtain reliable results [22].

2.3. (Non)application of Confirmatory Factor Analysis

In SEM, prior to the application of structural model by which the research hypotheses are tested, the measurement model needs to be created, and the structure of constructs and their measurement items should be identified [1]. To do this, confirmatory factor analysis (CFA) is applied by which it is tested how well the measured variables represent their respective theoretical latent constructs and how well the prespecified measurement theory matches reality as captured by data [17].

Even though CFA is a prerequisite to the hypothesis testing in SEM, Dogan [14] indicated that almost one quarter of the papers reviewed in his study lacks a CFA and directly skip to the application of structural model. Therefore, even though such an issue was not identified in other SEM reviews (e.g., [10], [11], [13]), in this paper it is included among the subjects to be examined.

2.4. Assessment of Model Fit

Goodness-of-fit indices are the indicators to assess how well the prespecified theorical model explains the data collected. Based on the model fit values, it is decided whether to accept or reject the measurement model. In case that the overall model fit is not acceptable, all the loadings, parameters and estimations within the model become null and void [17]. Therefore, the correct assessment of model fit is very crucial in SEM analysis.

Even though chi-square test is a well-known statistical method to test goodness of fit, it has a downside that it is very sensitive to sample size [1]. If the sample size is very large, which is often the case in SEM especially when working with complex models, the statistical test will almost certainly be significant and even if it actually describes the data very well, the model will always be rejected. Likewise, when the sample is small, the model will always be accepted, even if in reality it fits poorly [5]. As a solution to this problem, researchers have proposed a lot of different alternative fit indices over time, which provided SEM an "arsenal" of goodness-of-fit indices [10, pp. 129]. Among them, the most used ones are GFI, AGFI, CFI, TLI, RMSEA and SRMR. The acceptable thresholds of these indices are as follows: GFI ≥ .90, AGFI ≥ .90, CFI ≥ .90, TLI ≥ .90, SRMR ≤ .08, RMSEA ≤ .06 [1], [17], [23].

2.5. Model Respecification

When an adequate level of fit is not achieved, it is common practice to modify the model, by deleting nonsignificant parameters that are upsetting the model fit, and/or by adding new parameters, which would improve the fit if specified [5]. To make this process easier, SEM provides the user a diagnostic output, modification indices, which calculates the prospective impact of every possible covariance that is not estimated in the model on the overall fit. Adding the new covariance path suggested in the modification indices to the model reduces the overall $\chi 2$ value, thus improves the model fit [17]. However, the misuse of these indices is a serious problem in SEM applications [10]. When performing model respecification by utilizing modification indices, researchers should consider the theorical basis of the changes they make [17], [19]; otherwise the modifications would amount to baseless data-driven adjustments that lack validity [24].

2.6. Reliability

Reliability refers to the assessment on degree of consistency among a set of items that measure a construct. For a scale to be reliable, its items are expected to be reflective of the same underlying latent construct and to be highly intercorrelated. In simpler terms, they are expected to be closely related as a group and vary together [17]. Establishment of construct reliability is very crucial in SEM, because only then the results achieved can be treated with confidence [6]. In marketing literature, researchers customarily apply the Cronbach's alpha to evaluate the internal consistency of measures [10]. However, this coefficient suffers from several weaknesses. First, the number of items in the test influences the accuracy of the reliability estimation [25]. Second, Cronbach's alpha operates under the assumption of equally weighted indicator loadings [22]. Third, it does not weight each individual item in the calculations [17]. Due to these limitations, Cronbach's alpha is not considered a true index for unidimensionality assessment [10]. In response to these weaknesses, more accurate construct reliability coefficients have been proposed, among which composite reliability (CR) is the most commonly applied measure. Unlike Cronbach's alpha, it is not influenced by the number of items in the scale [26], does not assume indicators to be equally loaded and weight the individual indicators based on their loadings [17]. The acceptable threshold for CR is 0.70, with each item having a minimum factor loading of 0.707 [26].

2.7. Validity

In SEM, two types of validity have to be established in order to obtain statistically admissible results. These are convergent validity and discriminant validity. Convergent validity is a measure of how closely the indicators of a scale converge, or 'load together,' on a single latent construct. In other words, it tests if the measures of a construct which are theoretically related are really related or not [17]. Convergent validity, or communality, is assessed through the calculation of average variance extracted (AVE) across all indicators linked with a particular construct. The rule of thumb for an acceptable AVE, which is the average of the squared loadings of all indicators under a particular construct, is 0.50 or higher [26].

Discriminant validity, on the other hand, measures if the indicators that theoretically should not be related are indeed not related. In other words, discriminant validity is established unless the items belonging to different constructs move very closely as if they were the members of the same group [17].

Typically, CB-SEM uses Fornell–Larcker criterion, which verifies discriminant validity if the AVE of a particular construct is higher than the variance that construct shares with each of other constructs [26]. An alternative method for assessing discriminant validity is heterotrait-monotrait ratio of correlations (HTMT), which was recently proposed by Henseler *et al.* [27] . HTMT is considered is a more precise contemporary measure of assessment. Even though it was originally recommended to be used in PLS-SEM, HTMT can be used in any SEM model regardless of the estimation method used [28]. An HTMT value below 0.90 suggests the evidence of discriminant validity [27].

2.8. Common Method Bias

Common method bias (CMB), or common method variance, is the spurious variance that can be attributed to the measurement instrument rather than to the constructs that the measures are assumed to represent [29]. Among possible causes of CMB, social factors such as implicit social desirability and acquiescence are the possibilities that are frequently mentioned in the literature [6], [30], [31]. Also the research instrument can cause CMB; for instance, the instructions on the top of a survey may lead the responses of the participants to a certain direction. As a result, the constructs end up sharing a certain amount of common variation which is not actually related to the network of causal relationships in the model tested [30]. CMB is a very serious phenomenon that needs to be checked because it can bias the reliability and validity of measures, which will cast suspicion on the correctness of the analysis results obtained, and can lead to misjudgements about the hypotheses tested in a research model because it may inflate or deflate the estimates between two constructs [31]. Nevertheless, despite the widespread acknowledgement about the problems that it poses, a surprising volume of studies do not check the existence of CMB in their data [30], [31]. To make sure if the data is contaminated with CMB or not, the most frequently used method is to apply Harman's single factor test, which measures if a single factor accounts for the majority of the covariance among the constructs. If the cumulative variance extracted by one factor is lower than the threshold of 50 per cent it can be concluded that CMB is not present in the data [32]. Some researchers show an inclination to test the reliability and validity of the constructs but skip measuring CMB. This is an important problem to be addressed because models may have acceptable convergent and discriminant validity values but still be contaminated by common method bias [30].

2.9. *Normality*

Normality is another concern involving SEM analysis. Most estimation methods in SEM, including the most popular ones (maximum likelihood estimation and general least squares), operates under the assumption of normal distribution of the data. Violation of this rule may lead to distorted goodness-of-fit measures and undervalued standard error terms [33]. Therefore, researchers need to check the multivariate normality of their data and, if necessary, apply remedies to account for non-normality [6]. To test the normality of data, skewness and kurtosis values need to be calculated. While skewness is used to describe the balance of the distribution; i.e. if the data are distributed symmetrically or accumulated on side (left or right), kurtosis is the measure of peakedness or flatness of distribution [17]. In the case that the data show non-normal distribution, bootstrapping is a method that can be resorted to. Bootstrapping, which refers to the regeneration of sampling distributions via resampling without replacing the original data, can be used when fitting covariance structures to data with non-normal distribution [34]. By using bootstrapped samples, researchers can reach estimations with accurate significance levels and appropriate standard errors from non-normal data [35]. Another way to account for data non-normality is to use alternative non-conventional estimations techniques such as weighted least squares (WLS), general weighted least squares (GWLS), and elliptical reweighted least squares (ERLS), which are, unlike maximum likelihood (ML), the most frequently used estimation technique, have less stringent requirements regarding data distribution. However, this method is discouraged because changing the estimation technique may violate the theoretical logic underpinning the original dataset [6].

2.10. Mediation

In SEM, researchers can make many complex and comprehensive analyses which are very difficult to do with first generation statistical methods. One of such calculations is mediation analysis [7], which refers to the testing of a hypothesized causal chain in which one variable affects a second variable which, in turn, affects a third variable. The variable in the middle of this causal chain, that is, the mediator, intervenes the relationship between the independent and the dependent variables [17]. The result of the mediation analysis shows the indirect effect of the independent varible, i.e predictor, on the dependent variable, i.e outcome. When testing this indirect effect, researchers resort to two general methods, which are bootstrapping and Monte Carlo method [36].

3. Descriptive Results and Findings

3.1. Sample Size

The articles reviewed in this study show that generally researchers are aware of the large sample size requirement of SEM analysis. In these 120 articles, only 3 of them have a sample size smaller than 200. Nevertheless, as explained in the previous section, it is widely recommended to assess the adequacy of sample size not in terms of absolute numbers, rather based on the complexity of the model, that is, the number of parameters to be estimated. Therefore, the sample sizes of these articles were assessed based on the of N:q ratio by proposed by Bentler and Chou [21], taking 1:5 as the threshold for an acceptable sample size.

In order to calculate the free parameters in a model, in addition to the structure among the latent variables, the number of observed variables need to be known as well. However, 6 articles (5%) share only the path diagrams between the latent constructs without any information regarding the number of observed variables, therefore no assessment could be done whether the sample size is large enough or not. This is a critical issue that needs special attention. Considering the widely acknowledged sample size requirements of SEM in the literature, researchers are recommended to report at least the number, if not the full detail, of items of each construct.

When the remaining 114 articles are analyzed, it is seen that 21 of them (18%) have insufficient sample size. Actually, the median sample size of those articles is 309. So, most of them seemingly have quite large sample sizes but the complexity of their models makes these samples "large but not large enough". The rest, 93 articles (77%), are assessed to have sufficient sample sizes.

Table 1. Sample Size Adequacy

	Frequency	Mean	Percentage
Articles with adequate sample size	93	408	77%
Articles with insufficient sample size	21	309	18%
Unsuitable for interpretation	6	455	5%
Total	120	403	100%

During the sample size assessments, another issue that is not identified in previous works came into sight, which is the misuse of Kaiser-Meyer-Olkin (KMO) and Bartlett's tests. KMO and Barlett's test of sphericity are used to determine whether the sample is suited for factor analysis by measuring the proportion of variance and correlations among variables, that is, if the data is suitable to form factors, these factors can explain each other and related to each other [17]. However, it is detected that in some articles these values are used to check the sample size adequacy. This is a critical misapplication because these tests should be applied to see if the quality, not quantity, of the data is adequate for the factor analysis.

Out of 120 articles review in this study, 69 did not use KMO and Bartlett's tests. 10 articles of the remaining 51 misused these tests and take the KMO and Bartlett's measures to conclude if the sample size of their study is adequate. This frequency corresponds to almost 20% which is not a negligible ratio.

Therefore, the researchers are advised to pay attention to this issue and verify the sample size of their study is large enough by using the criteria acknowledged in the literature.

3.2. (Non)application of CFA

The proper application of SEM technique requires a confirmatory factor analysis (CFA) prior to the hypothesis testing. In other words, first a measurement model is created to see how well the measured variables represent the number of constructs specified in the model [17]. After it is verified that factor structure fits the data, then the structural model is finalized. Even though such a problem is not identified in other leading SEM review studies (see [7], [10], [11], [13], [15]) Dogan [14] indicated that in some works researchers skip CFA and proceed directly to the hypothesis testing by building a structural model. Out of 120 articles analyzed in this study, in 25 of them (21%) CFA is not performed. This is a clear violation of the universally acknowledged two-step approach of SEM [19] that the researchers are strongly advised to follow.

3.3. Measurement Model Specification

The extent how well the observed variables represent their respective unobserved variables directly influences the quality of model specification [10]. Thus, in the previous SEM reviews, the ratio between observed/latent variables was examined to see if each construct has ideally sufficient indicators under them (see [10], [11], [13]). The median ratio of observed/latent variables for the articles reviewed in this study is 4.1, which is more than enough based on the recommendation of "at least three indicators per construct" by Baumgartner and Homburg [11].

Secondly, the use of single and multi-item constructs is analyzed. Measurement of constructs by only one item is strictly discouraged and instead the use of multi-item measures is conventionally recommended in the literature [11], [37]. Excluding the 6 articles that didn't explicitly give information regarding the number of measurement items used in the analysis, most of the remaining articles (112) used multi-item constructs. Only 2 articles used single item constructs, which corresponds to less than 2% of the works reviewed in this study. However, even though it was generally refrained from using single-item constructs, the use of constructs measured by only two indicators is not equally uncommon. Despite the use of only 2 items to measure a construct is generally opposed [11], [22], 29 articles (24%) under the scope of this review include two-item constructs in their models.

Table 2. Construct types

	Frequency	Percentage
Constructs with one item	2	2%
Constructs with two items	29	24%
Constructs with three or more items	83	69%
Non-specified	6	5%
Total	120	100%

3.4. Evaluation of Model Fit

In SEM, reporting the model fit results is extremely crucial because through the model fit indices it can be understood how well the prespecified theoretical model fits the data collected. If the model does not fit, in other words, does not explain the data, then further analysis is meaningless [5], [22]. To test the fit of the model, there are several goodness of fit indices available to researchers. The most widely used of these indices are chi-square test, incremental fit index (IFI), comperative fit index (CFI), normed fit index (NFI), relative fit index (RFI) and root mean square error of approximation (RMSEA) [10]. The indices used in the articles reviewed and their frequency are as follows:

Table 3. Use of Model Fit Indices

Model Fit Index	Frequency	Percentage
χ2/df	93	87%
CFI	92	86%
RMSEA	90	84%
GFI	77	72%
NFI	64	60%
AGFI	46	43%
SRMR	43	40%
IFI	33	31%
TLI	20	19%
RFI	9	8%

As it can be seen in the Table 3, chi-square test, CFI, RMSEA, CFI, GFI and NFI are used most of the articles.

Out of 120 articles reviewed, 13 of them (11%) do not share any information regarding the model fit. This is a misapplication that should definitely be abstained from because if it is not known whether the model fit is adequate or not, the analysis results lose their credibility and validity.

During the examination of the use of model fit indices, another problem that has not been explicitly addressed in the previous works is identified. Even though 107 articles reported the fit values of their models, 44 of them (36%) does not indicate the goodness of fit values of both the measurement and the structural models. The proper application of SEM requires the researchers to measure the fit of both models [38] but in these 44 articles the model fit values of either the measurement or the structural model are calculated. SEM users should avoid this kind of a mispractice and report the fit values of both models.

Table 4. Reporting of Model Fit

	Frequency	Percentage
Both measurement and structural models	63	53%
Only structural model	28	23%
Only measurement model	16	13%
Not used	13	11%

3.5. *Model Respecification*

In SEM, once the measurement model is built and the model fit is calculated, adjustments on the model can be made in order to improve model fit. Modification indices that are suggested by the SEM software make this respecification process very easy. However, previous studies indicate that sometimes the modifications made by the researchers may lack theoretical plausibility even if technically possible [10], [14]. Out of 120 articles review in this study, 40 of them (33%) report that they made use of modification indices in order to improve the fit of the model. In these 40, most of them (31) utilized the modification indices properly. For 4 articles, it cannot be interpreted whether the use of modification indices is proper or not because in these studies, even though the authors indicate that they used modification indices, no further detail is presented.

For the remaining 5 articles, modification indices are assessed to be employed in an inappropriate way. The reasons leading to this assessment can be divided into two: misuse and overuse.

It is needless to say that SEM software produce the modification indices with a data driven approach. It is researchers' responsibility to judge if these suggestions are rational and well-reasoned on a theorical basis [17], [19], [24]. In 2 articles, the researchers draw correlations between the error terms belonging to different constructs. Even though such a modification may have a favorable impact on the overall fit of the model, there cannot be any theoretical foundation to correlate two error terms under different latent factors [39].

In 3 articles, there is an overuse of modification indices. Many authors advise the researchers to be cautious in using modification indices and recommend that such modifications should be kept at a minimum level [33], [40], [41]. Despite this call for caution, in these 3 articles correlations are drawn between almost all the error terms under the same construct. Such an application may improve the model fit substantially, but it is deemed problematic in the literature [39].

Table 5. Model Respecification

	Frequency
No model respecification reported	80
Model respecification reported	40
Appropriate respecification	31
Inappropriate respecification	5
Misuse of modification indices	2
Overuse of modification indices	3
Non-specified	4

3.6. Reliability

As explained in the previous section, to test reliability of the constructs in SEM, two main measures are used: Cronbach's alpha and composite reliability. In this review, it is examined in how many articles, and with which measure the reliability test is employed. Out of 120 studies, only two does not report any reliability result at all, which means that more than 98% of the articles conducted reliability test. While in 60 articles (50%) only Cronbach's alpha is used, only 5 articles (4%) use CR alone. The remaining 53 articles (44%) include both alpha and CR values in their analyses.

Table 6. Reliability measures

	Frequency
Reliability not measured	2
Reliability measured	118
Cronbach's alpha alone	60
Composite reliability (CR) alone	5
Cronbach's alpha and CR together	53

Cronbach's alpha is considered to be most commonly used measure of scale reliability in SEM applications [6] and this study points to the same finding. More than 94% of the articles reviewed use Cronbach's alpha either alone or together with CR. However, due to the limitations of alpha value explained in the previous section, it is not recommended to use only this measure. Even Cronbach himself expresses that it would be safer to use alpha coefficient along with other successor procedures rather than alone [42]. Therefore, the researchers are recommended to use both Cronbach's alpha and CR to determine the reliability of their constructs.

3.7. Validity

In this review, the articles are examined if they include proper controls of convergent and discriminant validity of the constructs. Out of these two validity types, convergent validity is much more commonly

measured. 74 articles (62%) in 120 test the convergence of their indicators. When it comes to discriminant validity, it is measured in much fewer studies. Only in 45 articles (38%) discriminant validity is analyzed (33 follow only Fornell and Larcker procedure; 11 employ both Fornell & Larcker procedure and the HTMT method; and only 1 use HTMT method alone). In 46 articles (38%), neither convergent nor discriminant validity is measured. This result indicates an alarming situation because just like reliability, validity is also a prerequisite in SEM application. Therefore, it is surprising to see that construct validity is ignored by such a big portion of these studies even though reliability is addressed in almost all the articles reviewed.

Table 7. Validity measures

	Frequency
Validity not measured	46
Validity measured	74
Convergent validity	74
Discriminant validity	45

3.8. Common Method Bias

Despite its critical importance, in several previous SEM review works CMB has not been scrutinized (see [10], [11], [13]–[15]). To fill this gap in the literature and set an example for the future SEM review studies, the articles are examined to see if CMB is addressed or not. Out of 120 articles, only 9 of them (8%) ascertain that their data is not contaminated with CMB. All these 9 articles use Harman's single factor test as the measurement method.

SEM researchers collect their data very frequently through surveys, which are always under a certain risk of being affected by CMB [30], [31]. Therefore, to establish credibility of their analyses, it needs to become a common practice for SEM user to ensure that CMB is not a major issue in their dataset and take preventive countermeasures against CMB such as guaranteed anonymity and confidentiality, reverse coding of construct items, pretesting for item wording refinement and improvement [43].

3.9. Normality

Even though normal data distribution is one of the assumptions of SEM, previous review studies show that users of SEM fail to report the normality results of their data more often than not [10], [13]. Therefore, in this study reporting of data distribution is examined. In 23 articles (19%) the partial least squares estimation technique (PLS-SEM) was used. Out of these 23 articles, only 5 report normality measures (4 of them report kurtosis and skewness measures while 1 conducted Kolmogorov-Smirnov test). In the literature, there are two views regarding the normality requirement of PLS. The traditional view asserts that PLS-SEM, in contrast with CB-SEM, does not need the data to be distributed normally. Therefore, PLS-SEM based studies do not need to conduct a normality test [8]. The modern view, on the other hand, claims that PLS-SEM does not differ from CB-SEM regarding normality assumption. The non-parametric technique, bootstrapping, that PLS-SEM applies can equally be applied by other SEM techniques [44]. Therefore, since there is no unity in the literature agreeing that PLS-SEM based studies do not need to report the distribution results of their data, it can be seen a safer way to perform the normality test regardless of the estimation method and refer to bootstrapping in case of non-normal distribution as a remedy. 66 articles (56%) of the remaining 97 do not examine the normality of their data. Only in 31 articles (26%) measures of normality are reported (27 used kurtosis and skewness values; 7 conducted Kolmonogov-Smirnov test; and 1 employed Shapiro–Wilk test). In all the studies that report normality results, there is no non-normal distribution issue expect for 3 articles, all of which utilize bootstrapping as a solution.

Table 8. Normality measures

Frequency	Estimation Method

		CB-SEM	PLS-SEM
Normality not measured	84	66	18
Normality measured	36	31	5

3.10. Mediation

Researchers widely use SEM applications when conducting mediation analyses. In literature, there are many studies recommending the use of bootstrapping to define the confidence intervals for mediation effects [45], [46]. Therefore, in this study, it is examined how many articles test the mediating effect of a construct and how many of these articles make use of bootstrapping.

Out of 120 articles reviewed, 29 of them (24%) have analyses involving a mediating effect. In these 29 studies, 10 use bootstrapping technique. Also in one article Sobel test was applied. Therefore, only 11 articles (38%) out of 29 employed a statistical method designed to be used in mediation analyses.

4. Conclusion and Implications

In this study, 10 topics central to the use of SEM are examined by reviewing 120 SEM-based articles which are all published in 2020. The selection of these recently published articles ensures that all the issues detected in this review are current, not outdated, problems in SEM applications. The issues detected in previous SEM reviews form the backdrop to this study. Also some critical problems discussed in many works in the literature but were not included in those leading SEM reviews are included within the scope of this study.

SEM is a very effective statistical method frequently used for theory testing and development. As a tool that is used by researchers very commonly, it has a substantial role in the generation of new knowledge in many disciplines. However, only when it is applied properly, the information obtained through SEM analysis can be treated with confidence. Therefore, researchers should pay the utmost attention to the problematic practices, misuses and critiques regarding SEM applications when employing this technique.

Even though the descriptive statistics explain the practices regarding the use of SEM in Turkish marketing literature, the issues identified in this study are not limited to one discipline or country. Therefore, this study serves as a guideline to all SEM users about the most common problems faced in SEM applications, wrongful practices that should be avoided, exercises that needs to be executed and remedies that may be applied.

Funding: This research received no external funding.

Conflicts of Interest: The authors declare no conflict of interest.

References

[1] R. P. Bagozzi and Y. Yi, "Specification, evaluation, and interpretation of structural equation models," *Journal of the Academy of Marketing Science*, vol. 40, no. 1, Jan. 2012, doi: 10.1007/s11747-011-0278-x.

[2] W. W. Chin, R. A. Peterson, and S. P. Brown, "Structural Equation Modeling in Marketing: Some Practical Reminders," *Journal of Marketing Theory and Practice*, vol. 16, no. 4, Sep. 2008, doi: 10.2753/MTP1069-6679160402.

[3] M. Sarstedt, C. M. Ringle, and J. F. Hair, "PLS-SEM: Looking Back and Moving Forward," *Long Range Planning*, vol. 47, no. 3, pp. 132–137, Jun. 2014, doi: 10.1016/J.LRP.2014.02.008.

[4] R. Weston and P. A. Gore, "A Brief Guide to Structural Equation Modeling," *The Counseling Psychologist*, vol. 34, no. 5, Sep. 2006, doi: 10.1177/0011000006286345.

[5] J. J. Hox and T. M. Bechger, "Introduction Structural Equation Modeling An Introduction to Structural Equation Modeling 1 What is Structural Equation Modeling?," 1998.

[6] G. T. M. Hult *et al.*, "An Assessment of the Use of Structural Equation Modeling in International Business Research," *Research Methodology in Strategy and Management*, vol. 3. pp. 385–415, 2006. doi: 10.1016/S1479-8387(06)03012-8.

[7] J. F. Hair Jr., M. L. D. da S. Gabriel, and V. K. Patel, "AMOS Covariance-Based Structural Equation Modeling (CB-SEM): Guidelines on Its Application as a Marketing Research Tool," *Brazilian Journal of Marketing*, vol. 13, no. 2, pp. 44–55, May 2014, doi: 10.5585/remark.v13i2.2718.

[8] C. B. Astrachan, V. K. Patel, and G. Wanzenried, "A comparative study of CB-SEM and PLS-SEM for theory development in family firm research," *Journal of Family Business Strategy*, vol. 5, no. 1, pp. 116–128, Mar. 2014, doi: 10.1016/J.JFBS.2013.12.002.

[9] C. L. Shook, D. J. Ketchen, G. T. M. Hult, and K. M. Kacmar, "An assessment of the use of structural equation modeling in strategic management research," *Strategic Management Journal*, vol. 25, no. 4, Apr. 2004, doi: 10.1002/smj.385.

[10] F. J. Martínez-López, J. C. Gázquez-Abad, and C. M. P. Sousa, "Structural equation modelling in marketing and business research: Critical issues and practical recommendations," *European Journal of Marketing*, vol. 47, no. 1, pp. 115–152, Feb. 2013, doi: 10.1108/03090561311285484.

[11] H. Baumgartner and C. Homburg, "Applications of structural equation modeling in marketing and consumer research: A review," 1996.

[12] P. M. Bentler, "Multivariate Analysis with Latent Variables: Causal Modeling," *Annual Review of Psychology*, vol. 31, no. 1, Jan. 1980, doi: 10.1146/annurev.ps.31.020180.002223.

[13] Y. Koubaa, R. S. Tabbane, and R. C. Jallouli, "On the use of structural equation modeling in marketing image research," *Asia Pacific Journal of Marketing and Logistics*, vol. 26, no. 2, pp. 315–338, 2014, doi: 10.1108/APJML-10-2013-0113.

[14] V. Doğan, "PAZARLAMA ARAŞTIRMACILARININ YAPISAL EŞİTLİK MODELİ ANALİZİ UYGULAMALARI: SORUNLAR VE ÖNERİLER," *Journal of Administrative Sciences*, vol. 16, no. 32, pp. 201–230, 2018.

[15] D. Frías-Navarro and M. P. Soler, "Exploratory factor analysis (EFA) in consumer behavior and marketing research," *Suma Psicológica*, vol. 19, pp. 47–58, 2012.

[16] Kline Rex B., *Principles and Practice of Structural Equation Modeling*, 4th ed. New York: The Guilford Press, 2016.

[17] J. F. Hair, W. C. Black, B. J. Babin, and R. E. Anderson, *Multivariate Data Analysis*, 8th ed. Hampshire, UK: Cengage Learning EMEA, 2019. [Online]. Available: www.cengage.com/highered

[18] L. Ding, W. F. Velicer, and L. L. Harlow, "Effects of estimation methods, number of indicators per factor, and improper solutions on structural equation modeling fit indices," *Structural Equation Modeling: A Multidisciplinary Journal*, vol. 2, no. 2, Jan. 1995, doi: 10.1080/10705519509540000.

[19] J. C. Anderson and D. W. Gerbing, "Structural Equation Modeling in Practice: A Review and Recommended Two-Step Approach," 1988.

[20] J. C. Loehlin, *Latent Variable Models: An Introduction to Factor, Path, and Structural Analysis*. Mahwah, NJ: Lawrence Erlbaum Associates, 1998.

[21] P. M. BENTLER and C.-P. CHOU, "Practical Issues in Structural Modeling," *Sociological Methods & Research*, vol. 16, no. 1, Aug. 1987, doi: 10.1177/0049124187016001004.

[22] K. A. Bollen, *Structural Equations with Latent Variables*. New York, NY: Wiley Interscience, 1989.

[23] L. Hu and P. M. Bentler, "Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives," *Structural Equation Modeling: A Multidisciplinary Journal*, vol. 6, no. 1, Jan. 1999, doi: 10.1080/10705519909540118.

[24] R. MacCallum, "Specification searches in covariance structure modeling.," *Psychological Bulletin*, vol. 100, no. 1, 1986, doi: 10.1037/0033-2909.100.1.107.

[25] M. S. Garver and J. T. Mentzer, "Logistics research methods: Employing structural equation modeling to test for construct validity," *Journal of Business Logistics*, vol. 20, no. 1, pp. 33–57, 1999.

[26] C. Fornell and D. F. Larcker, "Evaluating Structural Equation Models with Unobservable Variables and Measurement Error," *Journal of Marketing Research*, vol. 18, no. 1, Feb. 1981, doi: 10.1177/002224378101800104.

[27] J. Henseler, C. M. Ringle, and M. Sarstedt, "A new criterion for assessing discriminant validity in variance-based structural equation modeling," *Journal of the Academy of Marketing Science*, vol. 43, no. 1, Jan. 2015, doi: 10.1007/s11747-014-0403-8.

[28] J. F. Hair Jr., L. M. Matthews, R. L. Matthews, and M. Sarstedt, "PLS-SEM or CB-SEM: updated guidelines on which method to use," *International Journal of Multivariate Data Analysis*, vol. 1, no. 2, 2017, doi: 10.1504/IJMDA.2017.087624.

[29] P. M. Podsakoff, S. B. MacKenzie, J. Y. Lee, and N. P. Podsakoff, "Common Method Biases in Behavioral Research: A Critical Review of the Literature and Recommended Remedies," *Journal of Applied Psychology*, vol. 88, no. 5. pp. 879–903, Oct. 2003. doi: 10.1037/0021-9010.88.5.879.

[30] N. Kock, "Common method bias in PLS-SEM: A full collinearity assessment approach," 2015.

[31] P. J. Jordan and A. C. Troth, "Common method bias in applied settings: The dilemma of researching in organizations," *Australian Journal of Management*, vol. 45, no. 1, pp. 3–14, Feb. 2020, doi: 10.1177/0312896219871976.

[32] P. M. Podsakoff and D. W. Organ, "Self-Reports in Organizational Research: Problems and Prospects," *Journal of Management*, vol. 12, no. 4, Dec. 1986, doi: 10.1177/014920638601200408.

[33] R. C. MacCallum, M. Roznowski, and L. B. Necowitz, "Model modifications in covariance structure analysis: The problem of capitalization on chance.," *Psychological Bulletin*, vol. 111, no. 3, 1992, doi: 10.1037/0033-2909.111.3.490.

[34] K. A. Bollen and R. A. Stine, "Bootstrapping Goodness-of-Fit Measures in Structural Equation Models," *Sociological Methods & Research*, vol. 21, no. 2, Nov. 1992, doi: 10.1177/0049124192021002004.

[35] A. J. Tomarken and N. G. Waller, "Structural equation modeling: Strengths, limitations, and misconceptions," *Annual Review of Clinical Psychology*, vol. 1. pp. 31–65, 2005. doi: 10.1146/annurev.clinpsy.1.102803.144239.

[36] D. Tofighi and D. P. MacKinnon, "Monte Carlo Confidence Intervals for Complex Functions of Indirect Effects," *Structural Equation Modeling: A Multidisciplinary Journal*, vol. 23, no. 2, Mar. 2016, doi: 10.1080/10705511.2015.1057284.

[37] J.-B. E. M. Steenkamp and H. Baumgartner, "On the use of structural equation models for marketing modeling," *International Journal of Research in Marketing*, vol. 17, no. 2–3, Sep. 2000, doi: 10.1016/S0167-8116(00)00016-1.

[38] J. B. Schreiber, A. Nora, F. K. Stage, E. A. Barlow, and J. King, "Reporting Structural Equation Modeling and Confirmatory Factor Analysis Results: A Review," *The Journal of Educational Research*, vol. 99, no. 6, Jul. 2006, doi: 10.3200/JOER.99.6.323-338.

[39] R. Hermida, "The problem of allowing correlated errors in structural equation modeling: concerns and considerations," *Computational Methods in Social Sciences*, vol. 3, no. 1, pp. 5–17, 2015.

[40] A. J. Tomarken and N. G. Waller, "Potential problems with 'well fitting' models.," *Journal of Abnormal Psychology*, vol. 112, no. 4, 2003, doi: 10.1037/0021-843X.112.4.578.

[41] R. Shah and S. M. Goldstein, "Use of structural equation modeling in operations management research: Looking back and forward," *Journal of Operations Management*, vol. 24, no. 2, pp. 148–169, Jan. 2006, doi: 10.1016/J.JOM.2005.05.001.

[42] L. J. Cronbach and R. J. Shavelson, "My Current Thoughts on Coefficient Alpha and Successor Procedures," *Educational and Psychological Measurement*, vol. 64, no. 3, Jun. 2004, doi: 10.1177/0013164404266386.

[43] I. Rodríguez-Ardura and A. Meseguer-Artola, "Editorial: How to Prevent, Detect and Control Common Method Variance in Electronic Commerce Research," *Journal of theoretical and applied electronic commerce research*, vol. 15, no. 2, 2020, doi: 10.4067/S0718-18762020000200101.

[44] J. Henseler, G. Hubona, and P. A. Ray, "Using PLS path modeling in new technology research: updated guidelines," *Industrial Management & Data Systems*, vol. 116, no. 1, Feb. 2016, doi: 10.1108/IMDS-09-2015-0382.

[45] G. W. Cheung and R. S. Lau, "Testing Mediation and Suppression Effects of Latent Variables," *Organizational Research Methods*, vol. 11, no. 2, Apr. 2008, doi: 10.1177/1094428107300343.

[46] D. P. MacKinnon, C. M. Lockwood, and J. Williams, "Confidence Limits for the Indirect Effect: Distribution of the Product and Resampling Methods," *Multivariate Behavioral Research*, vol. 39, no. 1, Jan. 2004, doi: 10.1207/s15327906mbr3901_4.