DUFED, 10 (2) (2021) 163-184

Dicle Üniversitesi Fen Bilimleri Enstitüsü Dergisi

Dicle University Journal of the Institute of Natural and Applied Science

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

Araştırma Makalesi / Research Article

Quality Improvement in Routine Inspection and Control of Healthcare Products Using Statistical Intervention of Long-Term Data Trend

Uzun Vadeli Veri Trendinin İstatistiksel Müdahalesi Kullanılarak Sağlık Ürünlerinin Rutin Muayenesi ve Kontrolünde Kalite İyileştirme

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

[1] *Cairo University, Faculty of Pharmacy (independent Researcher Category), 11562, Cairo, Egypt*
[2] *National Centre for Radiation Research and Technology, 11787, Cairo, Egypt*
[3] *Royal Oldham Hospital, OL1 2JH, Oldham, United Kingdom*

ARTICLE INFO

Article History

Received, 26 February 2021
Revised, 18 July 2021
Accepted, 26 July 2021
Available Online, 01 October 2021

Keywords

Box plot, Histogram, Normality test, Matrix plot, Outlier, Principal component analysis.

ABSTRACT

Commercial products in markets must meet the regulatory and quality control criteria to be acceptable for the intended use. While it is mandatory that each product batch must meet specification limits, the stability and efficiency over the long term are underestimated. The present study was conducted on the chronological trend of a healthcare product using a statistical software package, including correlation matrix and multivariate analysis. The investigated quality characteristics were the average filling weight, relative density, pH and the relative potency of three active components, in addition to a chemical preservative. The database was processed in an Excel sheet and was subjected to descriptive statistical overview, histogram plot, box plot diagram, time series plot, correlation matrix table and Principal Component Analysis. The investigation showed that despite all batches passed quality control tests successfully yet there were signs of instability, fluctuations or oscillations and drifts in all quality metrics, with outlier values that could be observed. Non-parametric correlation demonstrated some level of association between some inspection characteristic indicators. PCA illustrated the major variability influencer and clustering tendency among studied quality markers that guided the grouping of the dataset. The study pinpointed the improvement needed to ensure product stability, efficiency and quality.

MAKALE BİLGİSİ

Makale Tarihi

Alınış, 26 Şubat 2021
Revize, 18 Temmuz 2021
Kabul, 26 Temmuz 2021
Online Yayınlama, 01 Ekim 2021

Anahtar Kelimeler

Kutu grafiği, Histogram, Normallik testi, Matris grafiği, Aykırı, Temel bileşenler analizi.

ÖZ

Piyasalardaki ticari ürünler, kullanım amacına uygun olarak kabul edilebilir olması için düzenleyici ve kalite kontrol kriterlerini karşılamalıdır. Her ürün serisinin spesifikasyon limitlerini karşılaması zorunlu olmakla birlikte, uzun vadede stabilite ve verimlilik hafife alınmaktadır. Bu çalışma, korelasyon matrisi ve çok değişkenli analiz içeren istatistiksel bir yazılım paketi kullanılarak bir sağlık ürününün

***Corresponding Author**

E-mail addresses: mostafaessameissa@yahoo.com (Mostafa Essam EISSA), engyrefaat@yahoo.com (Engy Refaat RASHED), daliaessameissa@yahoo.co.uk (Dalia Essam EISSA)

kronolojik eğilimi üzerinde yürütülmüştür. Araştırılan kalite özellikleri, bir kimyasal koruyucuya ek olarak, ortalama dolum ağırlığı, bağıl yoğunluk, pH ve üç aktif bileşenin bağıl gücüydü. Veritabanı bir Excel sayfasında işlendi ve tanımlayıcı istatistiksel genel bakışa, histogram grafiğine, kutu grafiğine, zaman serisi grafiğine, korelasyon matrisi tablosuna ve Temel Bileşen Analizine tabi tutuldu. Araştırma, tüm serilerin kalite kontrol testlerini başarıyla geçmesine rağmen, tüm kalite ölçümlerinde gözlemlenebilen aykırı değerlerle birlikte istikrarsızlık, dalgalanmalar veya salınımlar ve sapmalar olduğunu gösterdi. Parametrik olmayan korelasyon, bazı denetim karakteristik göstergeleri arasında bir düzeyde ilişki olduğunu göstermiştir. PCA, veri kümesinin gruplandırılmasına rehberlik eden incelenen kalite belirteçleri arasındaki ana değişkenlik etkileyicisini ve kümelenme eğilimini gösterdi. Çalışma, ürün istikrarını, verimliliğini ve kalitesini sağlamak için gereken iyileştirmeyi belirledi.

1. INTRODUCTION

In everyday life, millions of people are frequently get exposed to a fairly huge number of consumable products for application and/or ingestion [1]. A significant fraction of these goods includes healthcare items which must be tested routinely and rigorously for their indicator metrics to be accepted for release from the firm. Then, these products are distributed into the market for consumption by the customers or patients [2]. Safety, efficiency and quality are integral aspects that must be insured to be integrated into the final finished product for use [3]. In the modern business quality concept, acceptable goods that could be dispatched and distributed in the market after satisfying regulatory specifications must not be considered as the sole indicator for a solid long-lasting successful industry [4]. In the first place, it was the quality management concept which should be disseminated in the organization that would pool into the final product [5]. The benefits of ensuring good practices in a specific industry were reflected not only in its image and reputation but also in reproducible, stable and reliable outputs with predictable characteristics.

Every year, many recalls of products must be monitored, recorded and controlled worldwide with various degrees of severity and devastation. For example, Food and Drug Administration (FDA) had reported regularly an appreciable number of recalls for medicines and healthcare products for different causes such as lack of efficacy, safety and quality [6]. These events had happened despite many of these goods were not problematic before. In many cases, thorough investigations usually revealed a gap or problem associated with integrated quality management that is linked to the affected product [7]. Thus, it was understood that although products might meet the intended specification limits, yet challenging issues could be hidden when analysed quantitatively over the long-term which were affected by the processes and operations involved in the product creation life cycle [8]. These defects might eventually lead to unwanted events of out-of-specifications (OOSs), recalls and excursions campaigned with financial loss and distorted company image and reputation.

The present work aimed to establish a quality and safety monitoring system for the market products to provide an early warning system before aberrant goods and recall would occur reflecting hidden and/or minor defects that could spark major excursions in the future through a chronological shift in the inspection quality characteristics of the products. This would be important also for the manufacturer to investigate the unintended factors that contribute to the product manufacturing errors and defects. The study herein focused on long-run analysis and trending of a randomly selected commercial healthcare product in the market. Product analysis and inspection of the quality control characteristic metrics must be inspected for evaluation. Creation of long-term database was ensured, and the collected records should be arranged chronologically for further processing. Statistical investigation of the individual inspection characteristics to determine the quality, stability and dispersion pattern of the recorded observations. Degree of association between measured parameters should follow to assess the influence and the link between inspection properties datasets on data quality and pattern using correlation matrix and multivariate analysis. Identification of the quality of the inspection metrics quantitatively would be crucial to guide and manage the operational steps that lead to the final product excellence. Objective decision-making would be accomplished as either sustain or improve, when the examination procedure showed either efficiency and stability or deficiency and inadequate quality, respectively.

2. MATERIAL AND METHOD

The study subject comprised a chronological series of a liquid healthcare product in the market that was analysed physically and chemically in an independent drug research laboratory. Results were reported and recorded in an electronic storage system that could be converted into other software format for further processing and trending. Datasets of all inspection characteristics were created in form of an Excel sheet [9]. The examined properties were the average filling weight of the product in the primary packaging material, the relative density, pH value, the relative potencies of three active pharmaceutical ingredients (APIs) and a preservative that is used to preserve a multiuse product from the microbial spoilage [10-12]. A quantitative evaluation of the product quality was conducted using a statistical software platform as will be discussed in the following sections.

2.1 Graphical Analysis and Descriptive statistics of Long-Term Dataset of the Inspection Properties

The dispersion pattern of long-term datasets of the individual inspection characteristics columns was studied using both graphical presentation and summary of a statistical description to understand the distribution of the observations for a continuous data-type [13]. This type of analysis was used to study process stability, trend, pattern and clusters in the readings [14]. In the statistical program, the graphical Summary was used to summarize numeric data with a variety of statistics such as the sample size, mean, median, and standard deviation [15]. Data distribution could be described with graphs, through conducting an Anderson-Darling (AD) normality test and obtain Confidence Intervals (CI) for the mean, standard deviation, and median [15]. Complementary column statistics – including Coefficient of Variation (CV), geometric mean with upper and lower 95% CI, skewness and kurtosis determinations - was performed using GraphPad Prism v6.01 for Windows [16].

2.2 Correlation Matrix, Dendrogram and Multivariate Analysis of Product Quality Metrics

Numerical and graphical presentation of correlation between datasets - as a correlation matrix and matrix plot - were conducted using GraphPad Prism v6.01 and In Minitab® v17.1.0, respectively [17,18]. While multivariate study using Principal Component Analysis (PCA) was conducted using Minitab® v17.1.0. Matrix plots typically implicated the following elements: a matrix of scatterplots, rows and columns of the matrix, each of which represents a separate variable and an internal x- and y-axis scale for each scatterplot. In a matrix of plots, the lower-left graphs are mirror images of the upper right graphs. A matrix plot was used to assess the relationships among several pairs of variables at once. Displays of plots for judging the importance of the different principal components could be generated, along with the scores of the first two principal components to be investigated, and the outliers would be spotted.

The Scree figure (eigenvalue profile plot) included the eigenvalue associated with a principal component versus the number of the component. This plot was used to judge the relative magnitude of eigenvalues [19]. Biplot for the first two components was also selected to plot an overlay of the score and loading plots for the first two components. The PCA would be useful to help you to understand the underlying data structure and/or form a smaller number of uncorrelated variables (for instance, to avoid multicollinearity in regression) [20]. Points were the projected observations; vectors corresponded the projected variables. If the data were well-approximated by the first two principal components, a biplot could enable visualization of high-dimensional data by using a two-dimensional graph [21]. An outlier plot was also selected to plot the Mahalanobis distance for each data point. This diagram identified outliers in multivariate space [19,21]. A point above the Y reference line represented an unusual observation [22]. The dendrogram is a tree diagram that was used to display the groups that were formed by clustering observations at each step and their similarity levels. The similarity level was measured along the vertical axis (alternately, the distance level could be displayed), and the different observations were listed along the horizontal axis [23].

In the dialog box items for PCA, a type of matrix must be selected. The covariance would be chosen to calculate the principal components following covariance matrix selection. The covariance matrix was chosen taking into consideration that standardization was not needed for variables. Otherwise, the correlation must be selected to calculate the principal components using the correlation matrix. Since variables must be standardized to obtain a correct component score, a correlation type of matrix was selected [22,23]. The principal components included the linear combinations of the original variables that account for the variance in the data. The maximum number of components extracted always should equal the number of variables [24]. The eigenvectors, which were comprised of coefficients corresponding to each variable, were used to calculate the principal component scores. The coefficients were used to indicate the relative weight of each variable in the component [23,24]. The bigger the absolute value of the coefficient, the more important the corresponding variable would be in constructing the component.

3. RESULTS AND DISCUSSION

3.1 Graph and Summary Statistics of Data Distribution with Performance Analysis

Quality control (QC) inspection of the prepared medicinal products have become an indispensable task that provides oversight on the inspection characteristics of the final preparation in the compounding and formulation areas in a healthcare firm [25]. However, the reliance only on the batch-wise direct testing for just pass or fail criteria would not be not enough to judge the dissemination of the comprehensive quality management concept and system through the whole organization [26]. It is the statistical process trend analysis that would reveal the underneath deformities. The necessary adjustments in the system could be accomplished through the study of the inspection properties of the products. In turn, these characteristics were regarded as crucial aspects for the consistency and stability of the quality over fairly long runs [27,28]. All the monitored properties showed no Out-Of-Specification (OOS) with zero expected and the observed Defects-Per-Million-Opportunities (DPMO) or Part-Per-Million (PPM). There is no apparent defect in any of the 838 studied batches.

The tabulated statistics table of Figure 1-5 summarized information about each item and the total for all items: Total count which represented the number of observations, the mean (obtained from the sum of all observations divided by the total count) and the Standard Deviation (St.Dev.) which was used as a measure of dispersion analogous to the average distance (independent of direction) of each observation from the mean [29]. The collected sample data had more than 50 (N = 438 and 838 points for microbiological and non-microbiological analysis) observations, which would provide enough information to represent the dispersion pattern of the dataset. As the sample size would grow a more precise estimate of the process parameters, such as the mean and standard deviation could be attained [30]. The lower number of microbial enumeration database could be attributed to the applied skipping analysis plan schedule for the verified bioburden quality of the preparation [40]. Observations spreading were shown numerically in the column statistics in Figures 1 to 5, with mean, median and St.Dev. demonstrated at a 95% Confidence Interval (CI).

Outliers could strongly influence the results of any statistical analysis that is performed on the recorded observations [31]. A few of the dataset results were aberrant values, which were regarded as records that do not seem to be consistent with the remaining of the database. Since excursions might robustly lever the outcomes of any statistical study that was performed on the data, Identification of the cause of the unusual nature of these outliers should be investigated [32]. Verification of the sources of special-cause variation should be conducted and the data table in the program should be amended so that any entries in the cell(s) are corrected [33]. When there were no outliers - which are data points that do not appear to belong with the rest of the data – then all observations were considered homogenous and the noises in the time series plot could be attributed to the common-cause variations only which were considered normal statistical variation of the process.

The current study provided an integrative investigation of product goodness through the examination of the quality long-term database of the inspection characteristics of the medicinal dosage form. Graphical and descriptive summary of the statistical analysis of bioburden load (TVAC and TYMC), average filling weight, relative density, pH and the relative potencies of three APIs and antimicrobial preservative were shown in Figures 1 to 5. While data pattern and spreading tended to be simple and consistent in microbial enumeration test – with few aberrant values (indicated by asterisks in the box plot diagram, physical and chemical tests were more complex being two-sided properties around the mean values. The microbial score count dataset demonstrated a clear non-compliance to the bell-shaped of the Gaussian distribution due to limited count number variation which was an expected event in contrast to the continuous data type. This behaviour is common in microbiological count data and had been reported previously [34].

Nevertheless, the variable nature data type of both physical and chemical data properties (illustrated by histograms and box and whisker plot) showed dispersions that approached normality to a certain degree. This factual was observed despite failure of the Anderson-Darling (AD) test due to anomalies in the shape. Instances of unusual incidences included a flattened multimodal pattern in the average filling weight indicating multiple operations affecting one process. In addition, there were events that involved two isolated distributions as one major and one minor peaks with distant observations in the relative density measurement. Also, skewed distribution with distant outlier observations in pH measurements had been spotted. A possible two interfering distributions at the peak in the relative potency determination of API 1 was another observed phenomenon. The same - as in the average filling weight - could be observed with API 2 assay where an apparent multiple dispersion might interfere. Interestingly, it was the only parameter with no obvious outlier readings. The relative potency analysis dataset of API 3 was the closest in the spreading to the normal distribution despite the presence of some excursions in some readings. Again, the histogram of the relative potency of the preservative demonstrated a departure from normality due to excursions from high readings and single low excursion affecting the Gaussian distribution pattern.

Apart from the intermittent excursion of outlier observations in the time series plot, most inspection characteristics might show mixed and oscillating patterns of data trend. This fact could be attributed to intermittent shifts in the mean value along with some drifts and/or repeated behaviour. Thus, the out-of-control status of the means of inspected properties would be linked to the unstable variations between points in the moving range (MR) sequence. In turn, the instability of the operation could impact the precision of the evaluation of the process capabilities, despite a large number of observations reported [35]. All datasets failed to follow specific dispersion shapes even after transformation, except the average filling weight where it passed the normality test after Box-Cox transformation at λ = -5.00. The distorted distribution shape was affected by the heterogeneity of the reported outcome of the analysis which reflected inconsistent operations that influence the properties. In turn, the potential (within) process capability would be narrower and sharper in shape - if process shifts and drifts were eliminated – than the actual (overall) capability.

Table 1 showed other statistical tests that complement those in Figures 1 - 5 to be more informative about that quantitative nature of data. For example, skewness and kurtosis provided a numerical indication for the magnitude of deviation for the observations from the hypothetical distribution. Both geometric means and arithmetic means were almost identical. Coefficient of Variation (CV) expresses the spreading of data in the observation series around the average values [36]. In the current case, the lowest and the highest dispersions percent were demonstrated by the relative density and the relative potency of the preservative assay, respectively.

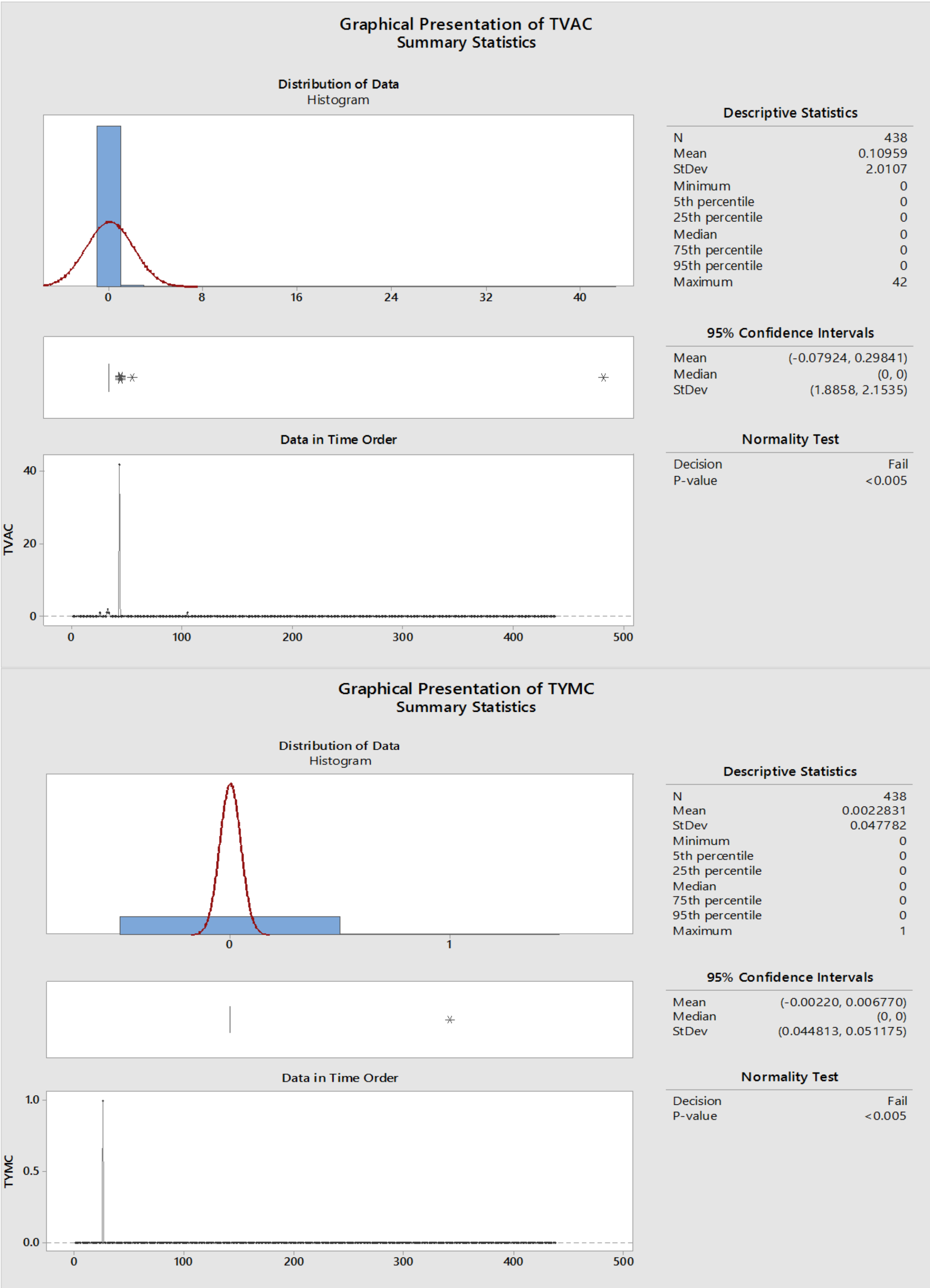

Figure 1. *Long-term statistical microbiological count assessment of TVAC and TYMC*

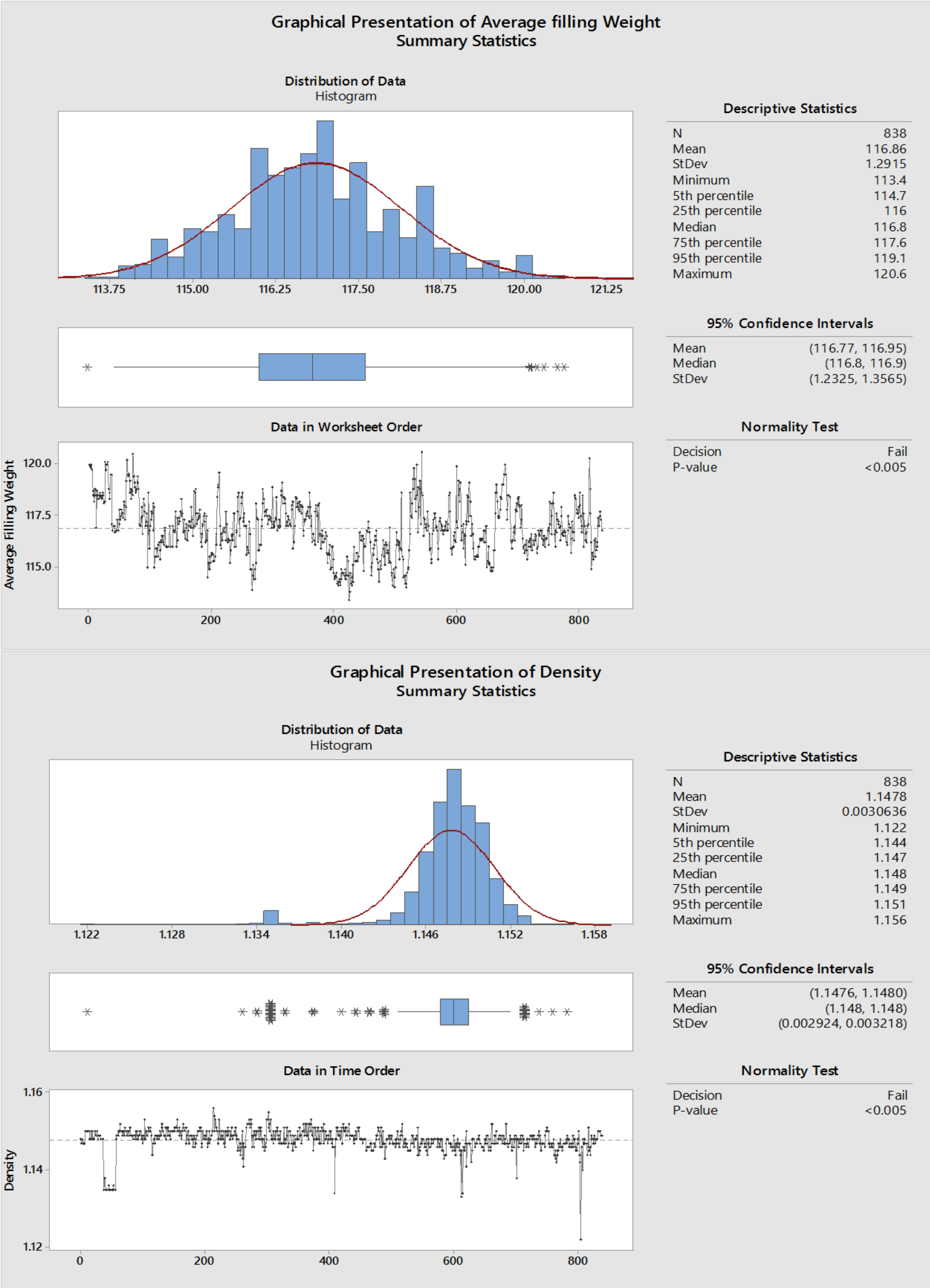

Figure 2. *Long-term statistical assessment of the average filling weight and relative density of the product*

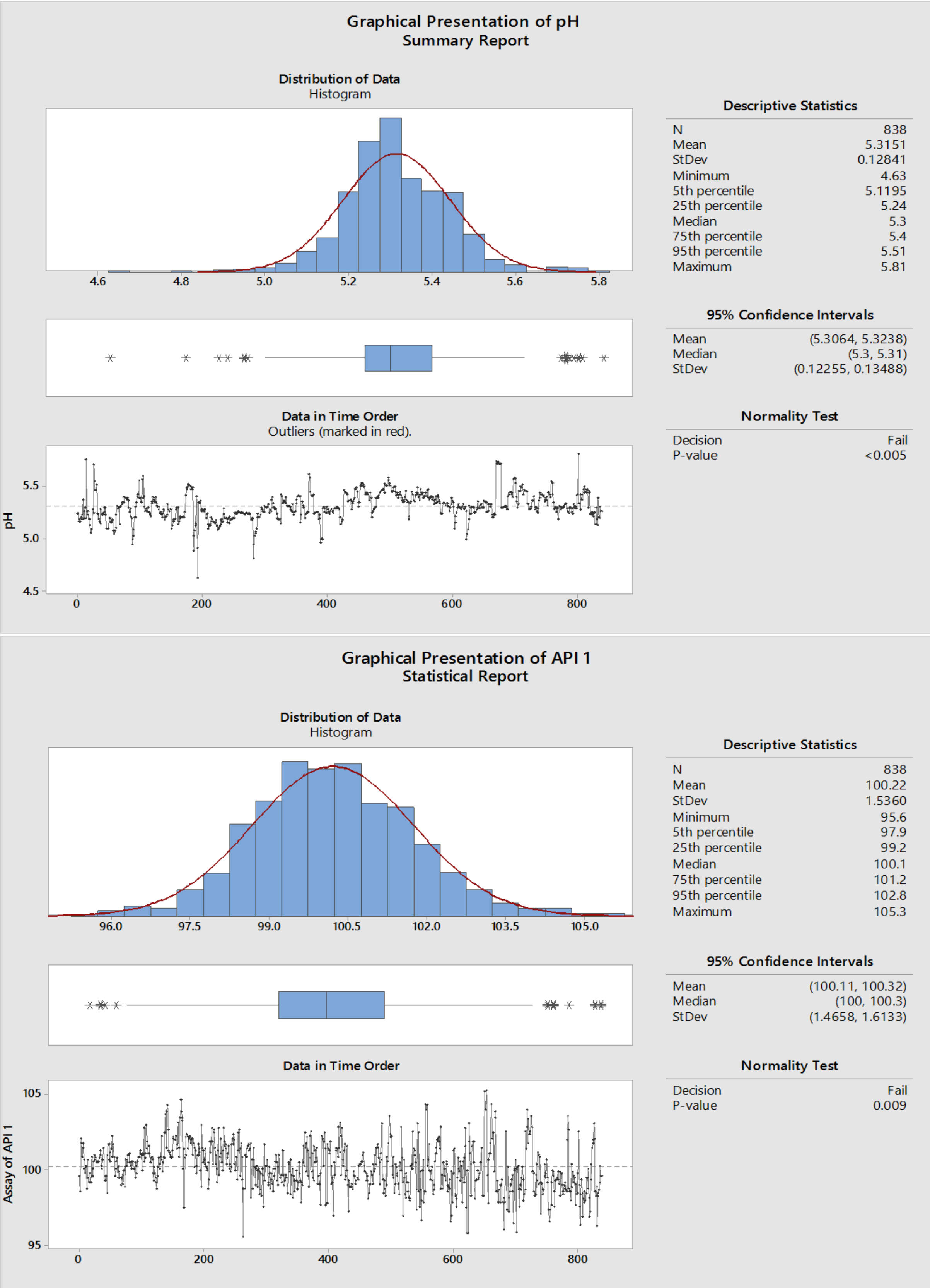

Figure 3. *Long-term statistical assessment of the pH and API 1 assay of the product*

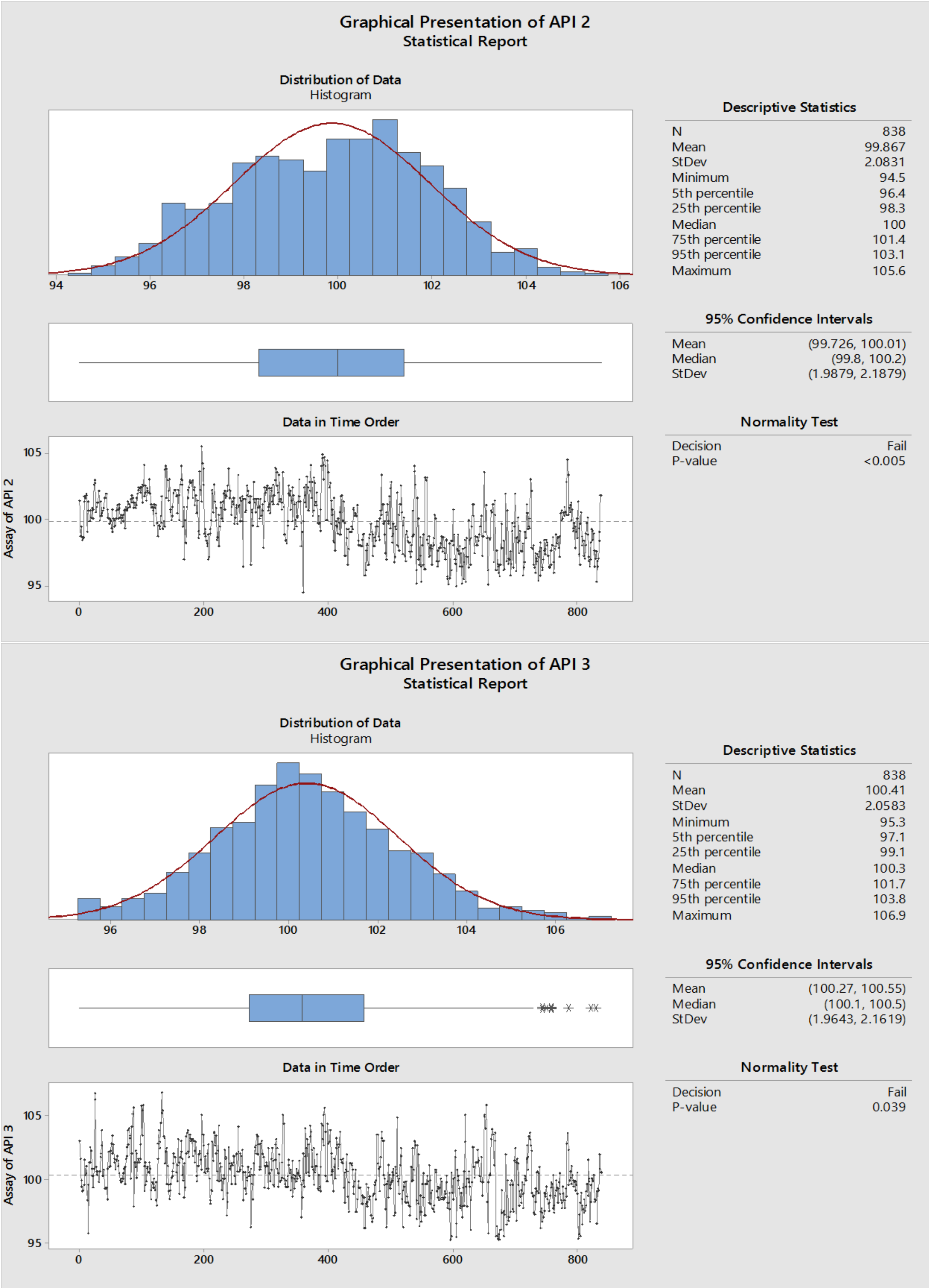

Figure 4. *Long-term statistical assessment of the API 2 and API 3 of the product*

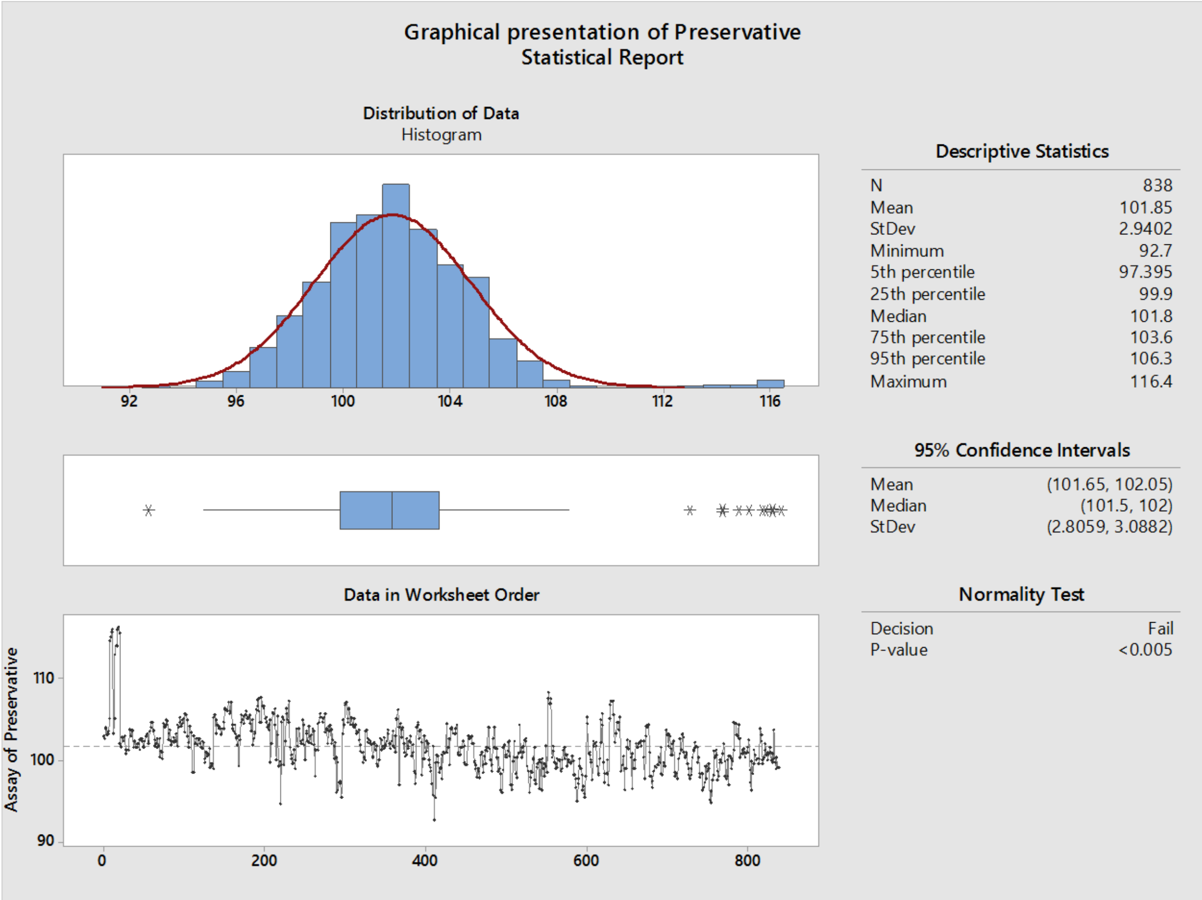

Figure 5. *Long-term statistical assessment of the preservative in the product*

Table 1. Column statistics of healthcare product inspection characteristics

Descriptive Statistics	Average Filling Weight	pH	Density	Assay of API 1	Assay of API 2	Assay of API 3	Assay of Preservative
Coefficient of variation (%) [a]	1.11	2.42	0.27	1.53	2.09	2.05	2.89
Geometric mean [b]	116.9	5.31	1.148	100.2	99.85	100.4	101.8
Lower 95% CI of geo. mean [c]	116.8	5.31	1.148	100.1	99.7	100.2	101.6
Upper 95% CI of geo. mean [c]	116.9	5.32	1.148	100.3	99.99	100.5	102.0
Skewness [d]	0.000	0.071	0.000	0.211	-0.088	0.144	0.966
Kurtosis [e]	-0.122	2.102	12.140	0.405	-0.614	0.077	3.807

[a] CV also is known as Relative Standard Deviation (RSD) which equals to ratio of the standard deviation to the mean.
[b] Geo. mean is an average value determined by x^{th} root of the product of multiplication of x numbers together.
[c] 95% confidence interval is a range of values that with 95% certainty would be expected to contain the true mean of the population.
[d] Quantification technique for the distribution degree of symmetry, where the ideal distribution has a value of zero.
[e] Quantification method of the tailing of certain distribution for matching to the Gaussian dispersion pattern.

3.2 Investigation of Degree of Association Between Inspection Properties of the Product

Table 2 showed the degree of correlation between different product markers numerically as coefficient and P values using non-parametric Spearman matrix at 95% confidence interval. The shape, pattern, predominant one-to-one relationship trend and cluster distribution were shown as matrix plot in Figure 6. A low-level correlation was observed between the relative density and the relative potency of the preservative with the assay of chemical components. In general, the other physical properties did not correlate appreciably either with each other or with the chemical potencies of the product. However, a higher level of data association was observed between the three APIs, notably, API 3 with API 1 and API 2. This overview might help in the decision on which variables to include in a model and how to model them when looking at the array of scatterplots to see which variables appeared correlated [37].

Table 2. Non-parametric two-tailed Spearman correlation matrix at 95% confidence interval for the indicators

Correlation Coefficient[¥] / P[©]	Average Filling Weight	pH	Density	Assay of API 1	Assay of API 2	Assay of API 3	Assay of Preservative
Average Filling Weight		-0.08	0.06	-0.08	0.06	0.03	0.14
pH	0.01566		-0.14	-0.17	-0.22	-0.25	-0.18
Density	0.10781	0.00008		0.42	0.47	0.40	0.40
Assay of API 1	0.02578	0.00000	0.00000		0.58	0.61	0.42
Assay of API 2	0.07439	0.00000	0.00000	0.00000		0.63	0.50
Assay of API 3	0.40487	0.00000	0.00000	0.00000	0.00000		0.41
Assay of Preservative	0.00005	0.00000	0.00000	0.00000	0.00000	0.00000	

[¥] Spearman r(rs) replaced Pearson correlation for non-Gaussian distribution.
[©] If the P value is diminishing, then rejection of the idea that the correlation is due to random sampling and vice versa.

If a predictor variable was not correlated with the response variable, it might be desirable to exclude the predictor from the model. If a relationship was curved, then it might be needed to include higher-order terms to accurately model the curvature (e.g. polynomial regression) [38]. Thus, the correlation matrix was used to assess the strength and direction of the relationship between two items or variables. Items that could measure the same construct should have high, positive correlation values. If the items were not highly correlated, then the items might be ambiguous, difficult to understand or measure different constructs [39]. Often, variables with correlation values greater than 0.7 were considered highly correlated. However, the correlation benchmark values would depend on subject area knowledge and the number of items in the examined data [40]. Nevertheless, the significance of the interconnection level between product properties was not high enough ($\geq$ 0.40 and $\leq$ 0.65) to justify modelling. The scattering patterns of the points around the positive relationship lines were not directive to a promising correlation between pairs of items to produce a satisfactory and reliable quantitative model.

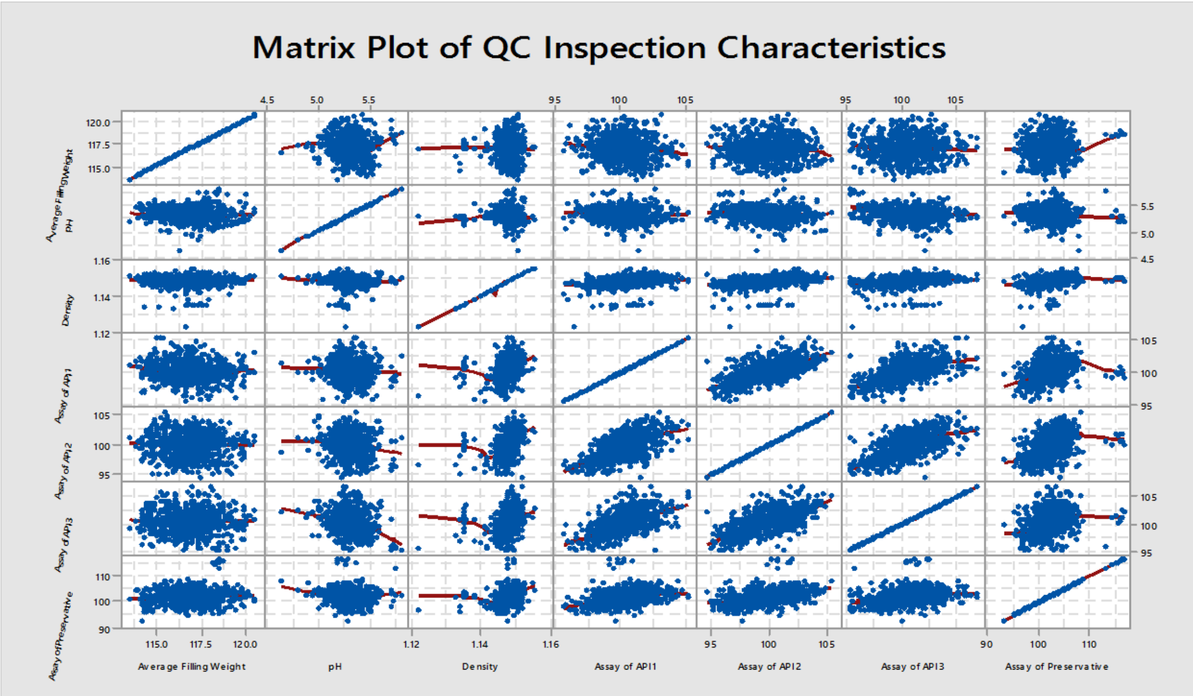

Figure 6. *Association pattern matrix between variables of the inspection quality characteristics*

3.3 Multivariate Analysis for Indicator Test of the Product

Survey analysis of continuous observations included studying the relations between the variables. PCA was used to form a smaller number of uncorrelated variables from a large set of data [41]. The goal of the PCA was to explain the maximum amount of variance with the fewest number of principal components. PCA could be used with large data sets in the study to determine the major metrics of the quality of a specific product. This multivariate methodology was commonly used as one step in a series of analyses [42]. Hence, the PCA was used to reduce the number of variables and avoid multicollinearity, or when too many predictors relative to the number of observations could be observed [43]. In the present case, a healthcare product was analysed for important quality aspects that were needed for its intended purpose. These inspection characteristics involved the average filling weight, relative density, pH and the assay of the three active components, in addition to the preserving compound. A PCA was conducted to see if they can form a smaller number of uncorrelated variables that were easier to interpret and analysed to be used as major indicators for the item quality.

The results suggested the following patterns: Relative density and assay of APIs 1, 2, 3 and the preservative could be categorized under a "product efficiency" component. Average filling weight forms a "product amount" component. The level of pH value forms a "stability/solubility" component. Figure 7 showed the grouping of the QC examination items for the product. Interpretation of dendrogram illustrated how the clusters were formed at each step and to assess the similarity (or distance) levels of the clusters that were formed. Based on the similarity (or distance) levels, the pattern of how similarity or distance values change from step to step would help to choose the final grouping for the database. The step where the values changed abruptly might identify a good point to define the final grouping [44]. In the present situation the similarity value changed from 36.51 for the overall observations to 48.30 in the left branch, then 63.26, 66.71, 79.72 and 81.82 on the right side. The right branch made the most sense of data grouping. Thus, the product efficiency component as in the biplot was sought as a convenient cluster. This final decision was called “cutting the dendrogram” [45]. the angles between the vectors represented a visual indication to the degree by which characteristics correlate with one another. When two or more vectors were close, forming a small angle, the variables under investigation were positively correlated. If the lines were perpendicular, they were not likely to be correlated. When they diverged and formed a large angle (close to 180°), they are negatively correlated [23].

The biplot combined the score and loading graphs [46]. The biplot was used to examine the data structure and loadings on one figure. The second principal component scores were plotted versus the first principal component scores [22]. The loadings for these two principal components were plotted on the same graph. This biplot demonstrated the following: the left and bottom axes were for the PCA plot which was used to read PCA scores of the observations (points). The top and right axes belonged to the loading plot which was used to read how strongly each quality property (vector) influences the principal components [23]. Relative potency determinations for the four chemical entities had large positive loadings on component one. Therefore, this component focused on the product's long-term quality stability. On the other hand, pH had large negative loadings on component two. Therefore, this component focused on the product's bygone quality trend.

The outlier graph in Figure 8 demonstrated outlier points beyond the reference line. Any point that was above the reference line was considered an outlier. Outliers could significantly affect the pattern and stability of the trend under examination [22,45]. Therefore, if an outlier was identified in the dataset, Investigation of the aberrant observation should be executed to understand why it is unusual to improve process or characteristic stability. Thus, the test pinpoints to the excursions in the pattern to correct any measurement or data entry errors [23,45]. Accordingly, removing data that were associated with special causes and repeating the analysis should be considered. Scree plot displayed the eigenvalues associated with a component or factor in descending order versus the number of the components or factors. It was used in PCA to visually assess which components accounted for most of the variability in the data [47].

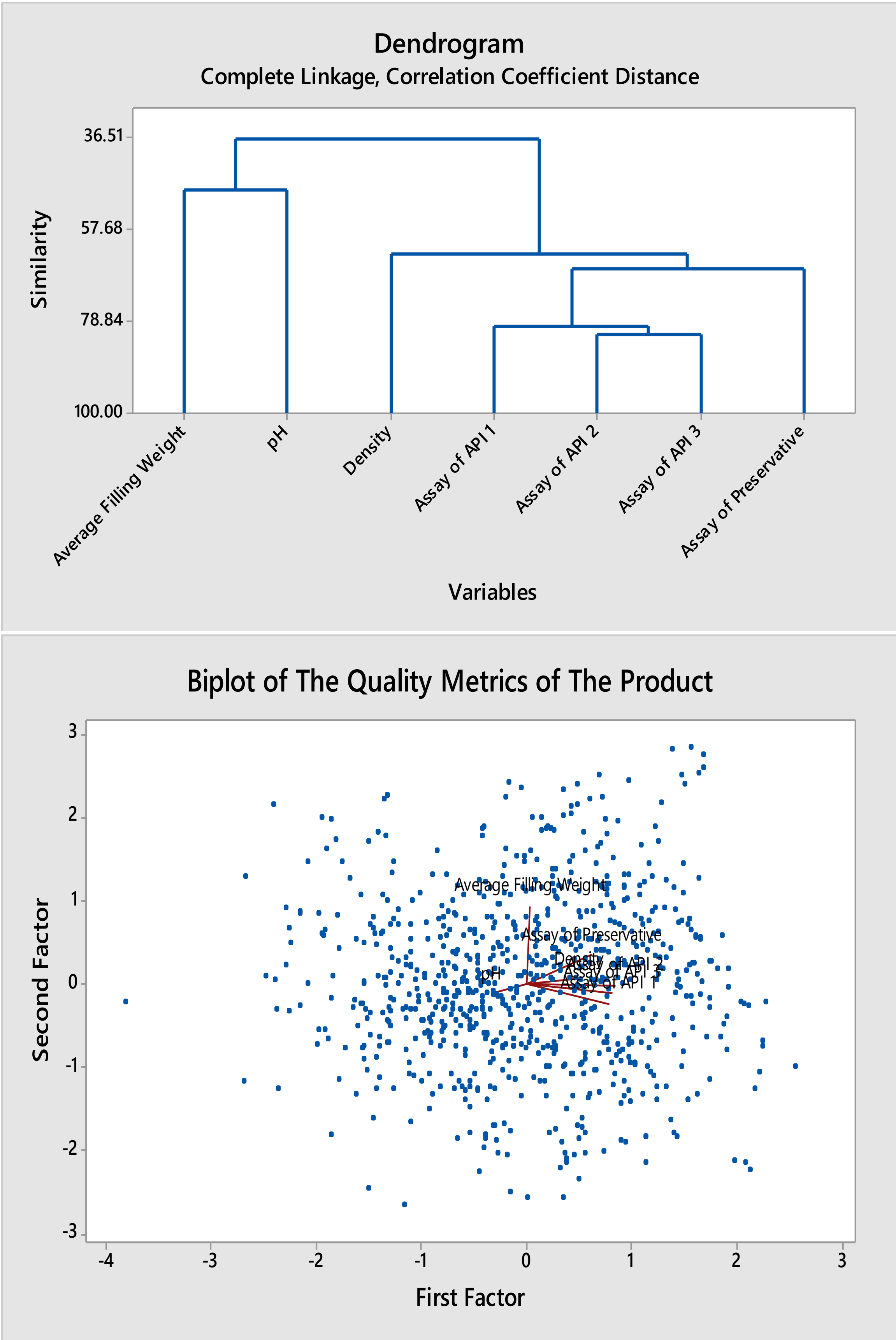

Figure 7: *Major components categorization of the product under examination*

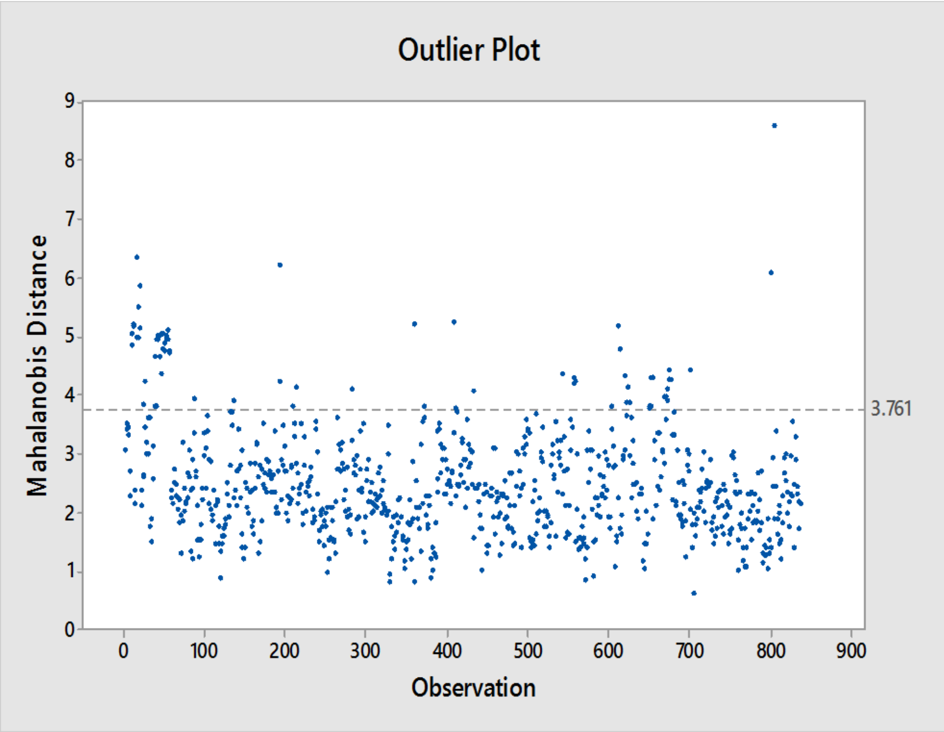

Figure 8. *Outlier plots of Principal Component Multivariate analysis of the product*

The ideal pattern in a scree plot would be observed as a steep curve, followed by a bend and then a flat or horizontal line. Those components or factors in the steep curve - before the first point that started the flat line trend – should be retained [20,48]. Using knowledge of the data and the results from the other methods of selecting components or factors would help to decide the number of important components or factors [20,48]. In these results of Figure 9, the first three principal components had eigenvalues greater than one. These three components explained most of the variation in the data. The scree plot showed that the eigenvalues started to form a straight line after the third principal component. Therefore, the remaining principal components accounted for a very small proportion of the variability (close to zero) and were probably unimportant. Eigen analysis of the Correlation Matrix was as the following: Eigenvalue (2.7696, 1.0836, 1.0160, 0.7398, 0.6572, 0.3906 and 0.3433), as a proportion (0.396, 0.155, 0.145, 0.106, 0.094, 0.056 and 0.049) and as a cumulative (0.396, 0.550, 0.696, 0.801, 0.895, 0.951 and 1.000). For the variables that showed Eigenvalue greater than one from the observation of variables viz (average filling weight, pH, density, relative potency of API 1, API 2, API 3 and preservative), (PC1, PC2, PC3) were as the following: (0.024, 0.897, -0.026), (-0.196, -0.116, -0.826), (0.303, 0.055, -0.524), (0.476, -0.241, -0.060), (0.509, -0.029, -0.002), (0.492, -0.111, 0.152) and (0.377, 0.328, -0.126), respectively.

Commercial products that were released in the market for consumption by the customers were normally expected to meet the rigorous official and regulatory specifications as they may impact their health and even life. Yet every year, a growing list of the recalled products could be observed due to various issues affecting products' usability. This work highlights the potential risk of the failure of the goods due to drifts in the inspection characteristics of the product that would not be detected using the classical reliance on the specification release criteria. The present study demonstrated the detection of the stability of each quality inspection criteria over the long term. The insight into the product quality could be interpreted easily through dimensionality reduction of the multiple inspection tests. It was evident from this research that the investigated inspection characteristics of the studied product needed further improvement from the manufacturer side in terms of the stability and consistency of the quality criteria.

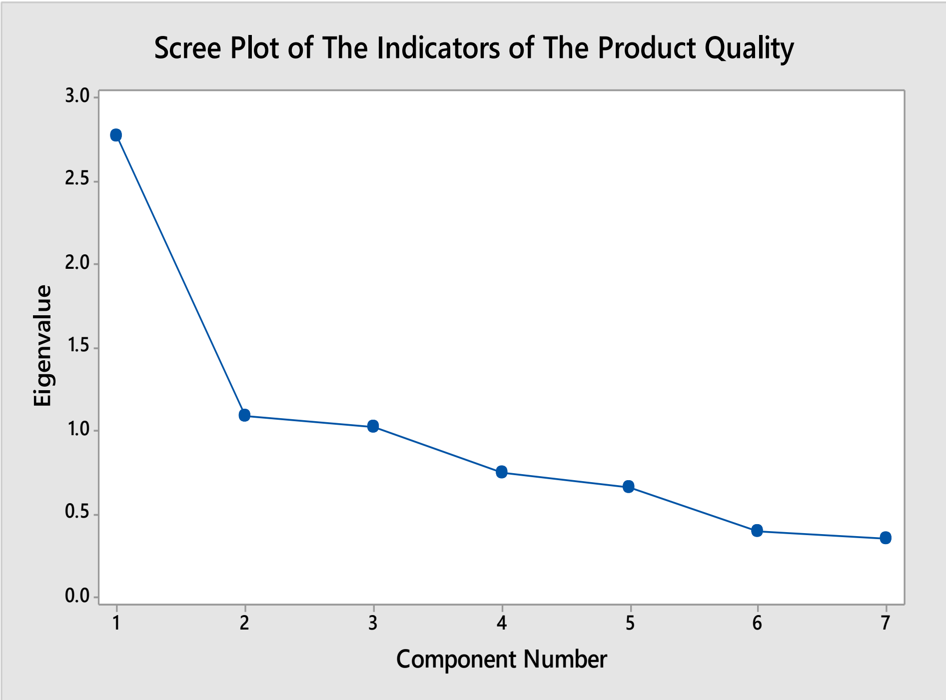

Figure 9. *Scee plots of Principal Component Multivariate analysis of the product*

4. CONCLUSION

The dissemination of quality concepts through the implementation of good practices in various fields (GxP) in any industry is crucial for a long-lasting successful business. Long-term monitoring of the key quality metrics of any market goods reveals the stability and consistency of the operations and processes performed on the created product. The studied inspection characteristics should be evaluated in terms of correlation, variability and impact on the value of the delivered items to the customers. The applied statistical tools in the present case were simple, effective and time-saving for large database analysis and investigation where safety, efficacy and quality of healthcare products are crucial for the health of the consumers in the challenging world of the ever-increasing ailment conditions that affect the human being and impact healthcare sector. Also, the study sounds an alarm clearly for the potential risk of the presence of a fairly large number of the products in the market that meet specification limits although they are not showing stable characteristics patterns of long-run operations.

CONFLICTS OF INTEREST

No conflict of interest was declared by the authors.

AUTHORS' CONTRIBUTIONS

Mostafa Essam EISSA: Designing and planning of the study, analysis, data collection, statistical interpretation and investigation. Engy Refaat RASHED: Topic suggestion, analysis, data collection, writing- original draft preparation. Dalia Essam EISSA: Topic suggestion, designing and planning of the study, writing- editing.

REFERENCES

[1] Y. Christopher, *Inadvertent ingestion exposure in the workplace*. Edimbourg, Royaume-Uni: Health and safety executive, 2007.

[2] OECD, *Improving Healthcare Quality in Europe: Characteristics, Effectiveness and Implementation of Different Strategies*. OECD Publishing, 2019.

[3] World Health Organization (WHO), *Quality Assurance of Pharmaceuticals 2016*. Geneva: World Health Organization, 2016.

[4] C. Yang, *The Evolution of Quality Concepts and the Related Quality Management*, Quality Control and Assurance - An Ancient Greek Term Re-Mastered. London, UK: Intech Open, 2017.

[5] M. Nguyen, A. Phan and Y. Matsui, "Contribution of Quality Management Practices to Sustainability Performance of Vietnamese Firms", *Sustainability*, vol. 10, no. 2, pp. 375-385, 2018, doi: 10.3390/su10020375.

[6] M. Eissa, "Rare event control charts in drug recall monitoring and trend analysis of data record: A multidimensional study", *Global Journal on Quality and Safety in Healthcare*, vol. 2, no. 2, pp. 34-39, 2019. Available: 10.4103/jqsh.jqsh_3_19.

[7] Food & Drug Administration, *Guidance for Industry Q9 Quality Risk Management*. (2006). Accessed: Feb. 21, 2021. [Online]. Available: https://www.fda.gov/media/71543/download.

[8] G. Sonnemann and M. Margni, *Life Cycle Management*. LCA Compendium – The Complete World of Life Cycle Assessment, (2015). Accessed: Feb. 22, 2021. [Online]. Available: 10.1007/978-94-017-7221-1.

[9] J. Nielsen, *Microsoft Official Academic Course Microsoft Excel 2016*. (2016). Accessed: Feb. 20, 2021. [Online]. Available: https://www.dit.ie/media/ittraining/msoffice/MOAC_Excel_2016_Core.pdf.

[10] J. Kim, K. Lee, U. Jerng and G. Choi, "Global Comparison of Stability Testing Parameters and Testing Methods for Finished Herbal Products", *Evidence-Based Complementary and Alternative Medicine*, vol. 2019, pp. 1-14, 2019. Available: 10.1155/2019/7348929 [Accessed:22 February 2021].

[11] SCCS Members & External experts, *The Sccs Notes Of Guidance For The Testing Of Cosmetic Ingredients And Their Safety Evaluation.* (2015). Accessed: Feb. 19, 2021. [Online]. Available: https://ec.europa.eu/health/scientific_committees/consumer_safety/docs/sccs_o_190.pdf.

[12] S. Gad, *Pharmaceutical manufacturing handbook: Production & Processes*. Hoboken, N.J.: Wiley-Interscience, 2008.

[13] Natural Resources Biometrics, *Chapter 1: Descriptive Statistics and the Normal Distribution | Natural Resources Biometrics*. (2021). Accessed: Feb. 18, 2021. [Online]. Available: https://courses.lumenlearning.com/suny-natural-resources-biometrics/chapter/chapter-1-descriptive-statistics-and-the-normal-distribution/.

[14] A. Sleeper, *Minitab Demystified*. New York: McGraw-Hill Education, 2011.

[15] C. Steele, *An Easy Data Set to Summarize with Minitab's Assistant*. (2015). Accessed: Feb. 16, 2021. [Online]. Available: https://blog.minitab.com/en/statistics-and-quality-improvement/an-easy-data-set-to-summarize-with-minitabs-assistant.

[16] H. Motulsky, *GraphPad Prism*. San Diego, Calif.: GraphPad Software, 2003.

[17] H. Motulsky, *Analyzing data with graphPad prism*. San Diego: GraphPad Software Inc., 1999.

[18] H. Pham, *Springer Handbook of Engineering Statistics*. New York: Springer, 2007.

[19] L. Ngo, *How to read PCA biplots and scree plots - BioTuring's Blog*. (2018). Accessed: Feb. 15, 2021. [Online]. Available: https://blog.bioturing.com/2018/06/18/how-to-read-pca-biplots-and-scree-plots/.

[20] R. Wicklin, What are biplots?. (2019). Accessed: Feb. 16, 2021. [Online]. Available: https://blogs.sas.com/content/iml/2019/11/06/what-are-biplots.html.

[21] J. Fish, *Outliers Episode 3: Detecting outliers using the Mahalanobis distance (and T2)*. (2021). Accessed: Feb. 14, 2021. [Online]. Available: https://community.jmp.com/t5/JMP-Blog/Outliers-Episode-3-Detecting-outliers-using-the-Mahalanobis/ba-p/351183.

[22] P. Python, *PCA: Practical Guide to Principal Component Analysis in R & Python*. (2016). Accessed: Feb. 1, 2021. [Online]. Available: https://www.analyticsvidhya.com/blog/2016/03/pca-practical-guide-principal-component-analysis-python/.

[23] Kassambara, *PCA - Principal Component Analysis Essentials - Articles – STHDA*. (2021). Accessed: Feb. 2, 2021. [Online]. Available: http://www.sthda.com/english/articles/31-principal-component-methods-in-r-practical-guide/112-pca-principal-component-analysis-essentials/.

[24] WHO Technical Report Series, No. 957, *WHO good practices for pharmaceutical quality control laboratories*. (2010). Accessed: Feb. 5, 2021. [Online]. Available: https://www.who.int/medicines/areas/quality_safety/quality_assurance/Goodpractices PharmaceuticalQualityControlLaboratoriesTRS957Annex1.pdf.

[25] S. Kulkarni, *Total Quality Management*. (2021). Accessed: Feb. 6, 2021. [Online]. Available: http://hsit.ac.in/e-learning/mechanical%20engineering/vi%20semester/total%20quality%20management(17me664)/tqm%20 notes.pdf.

[26] C. Berardinelli and L. Yerian, *Run Charts: A Simple and Powerful Tool for Process Improvement*. (2021). Accessed: Feb. 7, 2021. [Online]. Available: https://www.isixsigma.com/tools-templates/control-charts/run-charts-a-simple-and-powerful-tool-for-process-improvement/.

[27] M. Eissa and A. Abid, "Application of statistical process control for spotting compliance to good pharmaceutical practice", *Brazilian Journal of Pharmaceutical Sciences*, vol. 54, no. 2, 2018. Available: 10.1590/s2175-97902018000217499.

[28] J. Bland and D. Altman, "Statistics notes: Measurement error", *BMJ*, vol. 312, no. 7047, pp. 1654-1654, 1996. Available: 10.1136/bmj.312.7047.1654 [Accessed: 25 Feb. 2021].

[29] A. Bhattacherjee, *Social science research*. Tampa, FL: University of South Florida, 2012.

[30] M. Essam Eissa, "Investigation of Microbiological Quality of Water from the Feed Source to the Terminal Application in the Healthcare Facility: A Case Study", *Health Research*, vol. 2, no. 1, pp. 16, 2018. Available: 10.31058/j.hr.2018.21002 [Accessed: 25 Feb.2021].

[31] N. Balakrishnan, V. Barnett and T. Lewis, "Outliers in Statistical Data.", *Biometrics*, vol. 51, no. 1, pp. 381, 1995. Available: 10.2307/2533352.

[32] J. Frost, *Guidelines for Removing and Handling Outliers in Data - Statistics By Jim*. (2021). Accessed: Feb. 24, 2021. [Online]. Available: https://statisticsbyjim.com/basics/remove-outliers/.

[33] A. Monnappa, *Common Cause Variation vs Special Cause Variation*. (2021). Accessed: Feb. 8, (2021). [Online]. Available: https://www.simplilearn.com/common-vs-special-cause-of-variance-article.

[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 [Accessed: 25 February 2021].

[35] S. Wachs, *Process Stability vs. Capability Explained | WinSPC.com*. (2021). Accessed: Feb. 9, 2021. [Online]. Available: https://www.winspc.com/what-is-the-relationship-between-process-stability-and-process-capability/.

[36] İ. K. Keser, İ. D. Kocakoç and A. Şehirlioğlu, "A New Descriptive Statistic for Functional Data: Functional Coefficient of Variation", *Alphanumeric Journal*, vol. 4, no. 2, 2016. Available: 10.17093/aj.2016.4.2.5000185408 [Accessed: 25 February 2021].

[37] B. Illowsky and S. Dean, *Outliers*. (2021). Accessed: Feb. 10, 2021. [Online]. Available: https://opentextbc.ca/introstatopenstax/chapter/outliers/.

[38] Department of Statistics Online Programs, *7.7 - Polynomial Regression | STAT 462*. (2021). Accessed: Feb. 11, 2021. [Online]. Available: https://online.stat.psu.edu/stat462/node/158/.

[39] S. Nickolas, *What Do Correlation Coefficients Positive, Negative, and Zero Mean?*. (2021). Accessed: Feb. 12, 2021. [Online]. Available: https://www.investopedia.com/ask/answers/032515/what-does-it-mean-if-correlation-coefficient-positive-negative-or-zero.asp.

[40] B. C. Cronk, *How to Use SPSS, Fifth Edition*. (2021). Accessed: Jan. 31, 2021. [Online]. Available: http://www.xysfxy.cn/wcm.files/upload/null/201604/2016041309440 20 .pdf.

[41] D. Bartholomew, *International Encyclopedia of Education | ScienceDirect*. (2010). Accessed: Jan. 30, 2021. [Online]. Available: https://www.sciencedirect.com/referencework/9780080448947/international-encyclopedia-of-education.

[42] I. Jolliffe and J. Cadima, "Principal component analysis: a review and recent developments", *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*, vol. 374, no. 2065, 2016. Available: 10.1098/rsta.2015.0202 [Accessed: 25 February 2021].

[43] M. Patel, *PCA*. (2016). Accessed: Jan. 29, 2021. [Online]. Available: https://rstudio-pubs-static.s3.amazonaws.com/231900_91fafc24d70943f19b1c016987c886d9.html.

[44] P. Tan, M. Steinbach and V. Kumar, I*ntroduction to Data Mining* - *Instructor's Solution Manual*. (2006). Accessed: Jan. 28, 2021. [Online]. Available: https://www-users.cs.umn.edu/~kumar001/dmbook/sol.pdf.

[45] K. Hartmann, J. Krois and B. Waske, *SOGA*. (2018). Accessed: Jan. 26, 2021. [Online]. Available: https://www.geo.fu-berlin.de/en/v/soga/index.html.

[46] S. Serneels and T. Verdonck, "Principal component analysis for data containing outliers and missing elements", *Computational Statistics & Data Analysis*, vol. 52, no. 3, pp. 1712-1727, 2008. Available: 10.1016/j.csda.2007.05.024.

[47] C. Rao, J. Miller and D. Rao, *Essential statistical methods for medical statistics*. Amsterdam: Elsevier, 2011.

[48] *screeplot — Scree plot of eigenvalues*. (2021). Accessed: Jan. 25, 2021. [Online]. Available: https://www.stata.com/manuals/mvscreeplot.pdf.