Implementation of Generalized Variance and T^2-Hotelling Control Charts for Monitoring the Water Quality Characteristics at the Perumda Air Minum Tugu Tirta Kota Malang

Karina Rahmah Dini[1*], Evelin Nuga Fitrah[2], Fachrur Rozi[3]
[1,2,3] Universitas Islam Negeri Maulana Malik Ibrahim Malang
karinarahmahdini165@gmail.com

***Abstract*- Water is one of the most important needs for humans, so it is necessary to ensure the quality of drinking water that is suitable for consumption according to the standards of the Ministry of Health. This study aims to analyze the quality of water produced by Perumda Air Minum Tugu Tirta Kota Malang using the Statistical Process Control (SPC) method, namely the Generalized Variance and T^2-Hotelling control charts. This quantitative research uses secondary data from September 2023 to June 2024 with turbidity, residual chlor, and pH parameters. This study analyzed in two phases, Phase I and Phase II. The normality assumption test and correlation between variables were carried out before making the control chart. The results showed that in Phase I, the Generalized Variance and T^2-Hotelling charts detected several observation points outside the control limits. However, after revision, the data showed a statistically controlled condition. In Phase II, the Generalized Variance graph showed in control and relatively stable process, while the T^2-Hotelling graph showed two points out of the control limits. The conclusion of this study is that quality control using the multivariate method is able to detect process uncontrollability effectively, and the quality of produced water generally satisfy the standards for consumption.**

Keywords—dringking water; generalized variance; multivariate statistics; quality control; T^2-Hotelling

I. INTRODUCTION

Competition among industries encourages innovation in product development. The more products circulate in the market, the tighter the competition among producers to gain consumer trust. One of the main factors in product selection is quality, making quality control an essential aspect. Quality control aims to measure and compare product characteristics with predetermined standards and take corrective actions when deviations are found. One of the methods used in quality control is Statistical Process Control (SPC). SPC can reduce variability by using control charts [5]. Control charts are used to monitor quality changes caused by common or special causes.

The concept of control charts was first introduced by Walter A. Shewhart in 1924. Based on the number of quality characteristics monitored, control charts are divided into univariate and multivariate charts. Shewhart charts are used for monitoring univariate quality processes. In multivariate cases, Harold Hotelling developed the T^2-Hotelling control chart in 1947 to monitor the process mean within control limits under the assumption of multivariate normal distribution. However, this chart focuses only on the meaning, making it less effective in monitoring process variability [5].

In 1932, Samuel S. Wilks introduced the concept of Generalized Variance (GV) as a measure of multivariate variability. GV is defined as the determinant of the covariance matrix and was developed into the Generalized Variance control chart, which functions to monitor the variability of multivariate processes. The main advantage of this chart is its ability to monitor overall process variability, not just individual means or variances. However, its implementation requires the assumption of multivariate normal distribution and involves complex mathematical calculations.

Several previous studies have demonstrated the effectiveness of this method. Putri Agustina Sri Rejeki and Naniek Utami Handayani (2024) used the T^2-Hotelling and Generalized Variance methods to monitor the quality of White Crystal Sugar production at PG Trangkil. Their analysis showed that although production generally met specifications, some observations still failed to meet the target due to factors such as human error, methods, machinery, raw materials, and environmental conditions [6]. Meanwhile, Arga Willy Widyasmara (2016) applied the same methods to newspaper production at PT Adiprima Suraprinta and found several out-of-control observations, although the process still met the expected capability standards [8].

Based on these studies, it can be concluded that the T^2-Hotelling and Generalized Variance control charts are effective in detecting data that fall outside the control limits. This study aims to apply both methods to assess the quality of drinking water characteristics produced from direct natural springs. According to the Regulation of the Minister of Health of the Republic of Indonesia Number 492/KEMENKES/PER/IV/2010, clean water must meet several criteria such as neutral pH, clarity, colorlessness, odorlessness, tastelessness, and freedom from toxins and microorganisms. However, the water consumed by the public often fails to meet these standards due to factors such as weather conditions, surrounding environmental activities, and human influence.

Perumda Air Minum Tugu Tirta Kota Malang is the regional-owned enterprise (BUMD) responsible for providing clean and safe drinking water for the people of Malang. In practice, quality control of water is essential to ensure that the final product complies with the established standards. This study analyzes water quality characteristics such as turbidity, chlorine residual, and pH, which are interrelated and mutually influential [2].

Moreover, maintaining the quality of water and protecting its natural sources is not just a technical or regulatory obligation, izt is also a moral and religious responsibility. As stated in the Qur'an Surah Al-A'raf (7:56):

"*And do not cause corruption upon the earth after its reformation. And invoke Him in fear and aspiration. Indeed, the mercy of Allah is near to the doers of good*."

This verse reminds us that humans are commanded to maintain the earth's natural order and refrain from actions that lead to environmental damage, including water pollution. Ensuring that products meet their expected quality standards aligns with the Qur'anic command to uphold fairness and justice in transactions, as stated in Surah Ash-Shu'ara (26:181–183):

"*Give full measure, and do not be of those who cause loss. And weigh with an even balance. And do not deprive people of their duty and do not commit abuse on earth, spreading corruption*."

These verses serve as a moral foundation for implementing quality control practices that ensure consumers receive products—including water—that are fair, accurate, and safe for consumption.

This study aims to describe and analyze the application of the Generalized Variance control chart in monitoring process variability and the T²-Hotelling control chart in monitoring the process mean of water production at Perumda Air Minum Tugu Tirta Kota Malang. The scope of the research is limited to evaluating water quality based on three main parameters: turbidity, chlorine residual, and pH. The study encompasses both Phase I and Phase II of statistical quality control implementation. The outcomes of this research are expected to provide practical insights for students in applying statistical quality control techniques, raise public awareness regarding the importance of clean water and the efforts to maintain its quality, and contribute to the body of applied statistical literature within the Mathematics Department, particularly in the field of quality control.

II. METHODOLOGY

A. Multivariate Normality Test

The multivariate normality assumption was verified to ensure that each variable follows a normal distribution. The hypotheses for testing multivariate normality are:

H_0 : The observation vectors follow a multivariate normal distribution.

H_1: The observation vectors do not follow a multivariate normal distribution.

The test was conducted using the Mahalanobis distance d_i^2, calculated by:

$$d_i^2 = (\boldsymbol{X_i} - \overline{\boldsymbol{X}})^T \boldsymbol{S}^{-1} (\boldsymbol{X_i} - \overline{\boldsymbol{X}}) \quad (1)$$

where $\boldsymbol{X_i}$ is the observation vector, $\overline{\boldsymbol{X}}$ is the mean vector, and $\boldsymbol{S}^{-1}$ is the inverse of the sample covariance matrix. The null hypothesis assumes that the observation vectors follow a multivariate normal distribution. The hypothesis is rejected if d_i^2 falls below the 50th percentile of the Chi-square distribution $\chi^2_{\alpha,df}$ [4].

B. Correlation Test

Bartlett's Test of Sphericity was used to determine whether the observed variables are significantly correlated. The hypotheses are:

H_0 : $\mathbf{R} = \mathbf{I}$ (No correlation among variables)

H_1 : $\mathbf{R} \neq \mathbf{I}$ (A Correlation exists among variables)

The test statistics are:

$$\chi^2_{hitung} = -\left(m - 1 - \frac{2p+5}{6}\right) \ln |\mathbf{R}| \quad (2)$$

where $\mathbf{R}$ is the correlation matrix, p is the number of variables, and m is the number of observations. The null hypothesis is rejected if the test statistic exceeds the critical value from the Chi-square distribution with $\frac{1}{2}p(p-1)$d egrees of freedom [1].

C. Generalized Variance Control Chart

The Generalized Variance control chart is used to monitor process variability in multivariate data. The variability is expressed using the covariance matrix Σ, with diagonal elements representing individual variances and off-diagonal elements representing covariances. The sample covariance matrix S is used to estimate Σ, and its determinant $|S_i|$ serves as the control statistic.

Control limits for the Generalized Variance chart are calculated using the formulas 3 below [2]:

$$UCL = |\bar{\mathbf{S}}| \left(\frac{b_1}{b_3} + 3\sqrt{\frac{b_2}{b_3^2 + b_4}} \right)$$
$$CL = |\bar{\mathbf{S}}| \frac{b_1}{b_3} \quad (3)$$
$$LCL = \max\left(0, |\bar{\mathbf{S}}| \left(\frac{b_1}{b_3} - 3\sqrt{\frac{b_2}{b_3^2 + b_4}} \right)\right)$$

where $\bar{\mathbf{s}}$ is the average of the sample covariance matrices. Constants b_1, b_2, b_3, and b_4 are calculated based on the subgroup size n, the number of characteristics p, and the number of subgroups m.

$$
\begin{aligned}
b_1 &= \frac{1}{(n-1)^p}\prod_{i=1}^{p}(n-i) \\
b_2 &= \frac{1}{(n-1)^{2p}}\prod_{i=1}^{p}(n-i)\left[\prod_{j=1}^{p}(n-j+2)-\prod_{j=1}^{p}(n-j)\right] \\
b_3 &= \frac{1}{(m(n-1))^p}\prod_{k=1}^{p}(m(n-1)-k+1) \\
b_4 &= \frac{1}{(m(n-1))^{2p}}\prod_{i=1}^{p}(m(n-1)-i+1)\times\left[\prod_{j=1}^{p}(m(n-1)-j+3)-\prod_{j=1}^{p}(m(n-1)-j+1)\right]
\end{aligned} \tag{4}
$$

D. T^2-Hotelling Control Chart

The T^2-Hotelling control chart is applied to monitor the process meaning in multivariate data. It is an extension of the univariate t-distribution and is effective in detecting shifts in process mean. The test statistics are [5]:

$$T^2 = n\left(\bar{\boldsymbol{X}} - \bar{\bar{\boldsymbol{X}}}\right)' \bar{\mathbf{S}}^{-1}\left(\bar{\boldsymbol{X}} - \bar{\bar{\boldsymbol{X}}}\right) \tag{5}$$

where $\bar{\boldsymbol{X}}$ is the mean vector of a subgroup, $\bar{\bar{\boldsymbol{X}}}$ is the overall mean vector, n is the subgroup size, and $\bar{\boldsymbol{S}}^{-1}$ is the inverse of the pooled covariance matrix.

Control limits are established for two phases:

The control limits for the T^2-Hotelling control chart in Phase I are defined as follows:

$$
\begin{aligned}
UCL &= \frac{p(m-1)(n-1)}{mn-m-p+1} F_{\alpha,p,mn-m-p+1} \\
LCL &= 0
\end{aligned} \tag{6}
$$

And The control limits for the T^2-Hotelling control chart in Phase II are defined as follows:

$$
\begin{aligned}
UCL &= \frac{p(m+1)(n-1)}{mn-m-p+1} F_{\alpha,p,mn-m-p+1} \\
LCL &= 0
\end{aligned} \tag{7}
$$

E. Data and Data Sources

This study utilized secondary data on water quality parameters obtained from Perumda Air Minum Tugu Tirta Kota Malang. The data were sourced from laboratory records of water quality analysis results. Observations span the period from September 2023 to June 2024, consisting of a total of 43 observations.

For quality control purposes, the data were separated into two phases. Phase I consists of observations 1 to 33, which were used to establish baseline control limits and validate the necessary assumptions. Meanwhile, Phase II includes observations 34 to 43 and was used to assess the continued stability and control of the process based on the parameters determined in Phase I.

III. RESULTS AND DISCUSSION

A. Descriptive Statistical Analysis

Descriptive statistics were conducted on all 172 observations to provide an overview of the data. Table I presents summary statistics for three variables: Turbidity, Residual Chlorine, and pH, including their means, standard deviations, minimums, maximums, and relevant regulatory standards.

TABLE I
DESCRIPTIVE STATISTICS OF OVERALL DATA

Variable	Mean	Standard Deviation	Minimum	Maximum	Ministry of Health Standard	Unit
Turbidity	0.3328	0.2158	0	0.83	Max. 5	NTU
Residual Chlorine	0.3122	0.1234	0	0.7	0.2 – 0.6	mg/L
pH	7.0206	0.4115	6.5	8.3	6.5 – 8.5	-

Turbidity had a mean of 0.3328 NTU and a standard deviation of 0.2158, with all values well below the Ministry of Health's maximum limit of 5 NTU, indicating compliance with drinking water quality standards.

Residual chlorine showed a mean of 0.3122 mg/L and ranged from 0 to 0.7 mg/L. These values indicate that some observations fall below and above the recommended range of 0.2 – 0.6 mg/L. Chlorine levels below the minimum may pose microbiological risks, while levels above the maximum could lead to gastrointestinal issues due to excess chlorine.

The pH values ranged from 6.5 to 8.3, with a mean of 7.0206 and standard deviation of 0.4115, falling within the acceptable range of 6.5–8.5. However, values at the extremes (6.5 or 8.3) indicate relatively acidic or basic water, respectively, although still within the permissible limits.

B. Multivariate Normality Assumption

Before constructing the control chart, it is essential to verify several assumptions. The first assumption tested is multivariate normality, which aims to determine whether the data follows a multivariate normal distribution. The hypothesis tested is as follows:

H_0: The observation vectors follow a multivariate normal distribution

H_1: The observation vectors do not follow a multivariate normal distribution

The decision rule is to not reject H_0 if at least 50% of the calculated Mahalanobis distances (d_i^2) fall below the critical Chi-square value at a significance level of 0.05 and degrees of freedom equal to the number of variables.

For Phase I, Mahalanobis distances were computed based on the mean vector and the inverse of the covariance matrix. The results showed that 126 out of 132 calculated distances were below the critical Chi-square value of 7.815 (df = 3). Therefore, H_0 was not rejected, indicating that the data in Phase I follows a multivariate normal distribution.

The same procedure was applied to Phase II. The results indicated that all Mahalanobis distances were below the Chi-square critical value. Thus, H_0 was also not rejected for Phase II, confirming that the Phase II data also satisfies the assumption of multivariate normality.

C. Correlation Analysis

Following the multivariate normality test, Bartlett's Test of Sphericity was conducted to examine whether there is a significant correlation among variables. The hypotheses are:
H_0: $\mathbf{R} = \mathbf{I}$ (No correlation among variables)
H_1: $\mathbf{R} \neq \mathbf{I}$ (Correlation exists among variables)

For Phase I, the correlation matrix determinant was 0.9411, and the calculated χ^2 value was 7.8399. Since this value exceeds the critical value $\chi^2_{0.05;3} = 7.815$, H_0 is rejected, indicating significant correlations among the variables.

In Phase II, the determinant was 0.5523 and the χ^2 value was 22.0616, which is also greater than the critical value. Therefore, H_0 is again rejected, confirming that the quality characteristics are correlated in both phases.

D. Phase I Process Control

After all assumption tests were satisfied, the Generalized Variance control chart was constructed to monitor process stability based on multivariate variability during Phase I. The analysis used the first 33 subgroups, each consisting of 4 observations and 3 variables. Prior to constructing the chart, control limits were calculated using parameters derived from the sample covariance matrices.

The determinant of each subgroup's covariance matrix was computed, followed by the calculation of the pooled covariance matrix. Using this pooled determinant, the upper control limit (UCL), lower control limit (LCL), and central line (CL) were established. The resulting Phase I control chart is shown in Fig. 1.

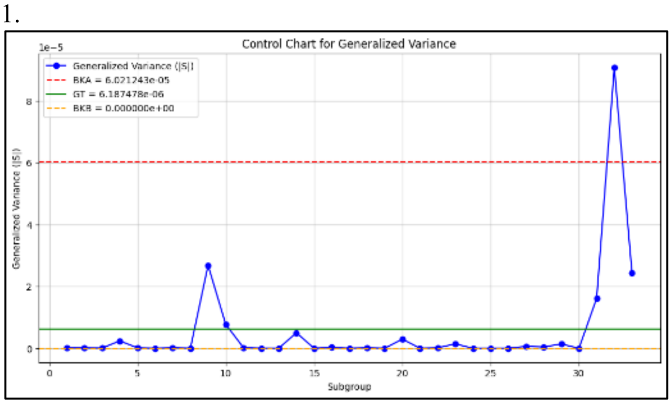

Fig. 1. Monitoring of Process Variability Using the Generalized Variance Control Chart in Phase I

Fig. 1 indicates that most of the data points lie within the control limits, suggesting stable variability. However, one subgroup (the 32nd) exceeded the upper control limit, indicating an out-of-control condition that may reflect unusual variation in the process. This anomaly suggests the need for further investigation to identify potential causes.

To ensure the accuracy and reliability of the control chart, the out-of-control point was removed. A revised chart was constructed using the remaining 32 subgroups, as presented in Fig. 2.

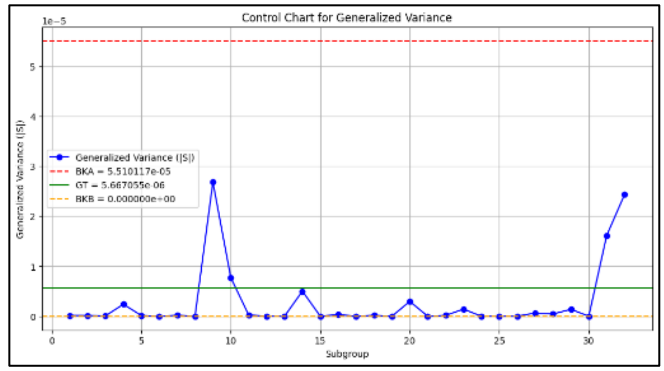

Fig. 2. Monitoring of Process Variability Using the Generalized Variance Control Chart in Phase I After Revision

Fig. 2 shows that all remaining points fall within the control limits, indicating the process is now in a stable and statistically controlled state. Consequently, the revised control limits — UCL = 5.101×10^{-5}, LCL = 0, and CL = 5.667×10^{-6} — will be used for Phase II monitoring.

The initial step in constructing the Hotelling's T^2 control chart is calculating the pooled covariance matrix and its inverse. Using these parameters, the T^2 statistic was computed for each of the 32 subgroups. Control limits were then established, with the lower control limit set to zero and the upper control limit determined using the appropriate F-distribution.

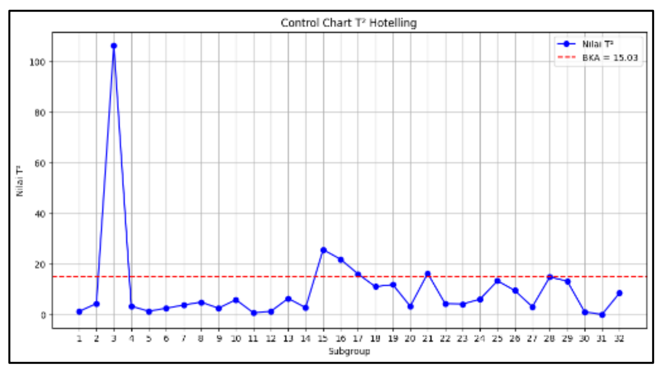

Fig. 3. Monitoring of Process Mean Using the T^2-Hotelling Control Chart in Phase I

The T^2 values were plotted on the control chart, as shown in Fig. 3. Several subgroups (3, 15, 16, 17, and 21) exceeded the upper control limit, indicating that the process was not fully in statistical control during Phase I. These out-of-control points suggest irregularities that violate the assumption of process stability.

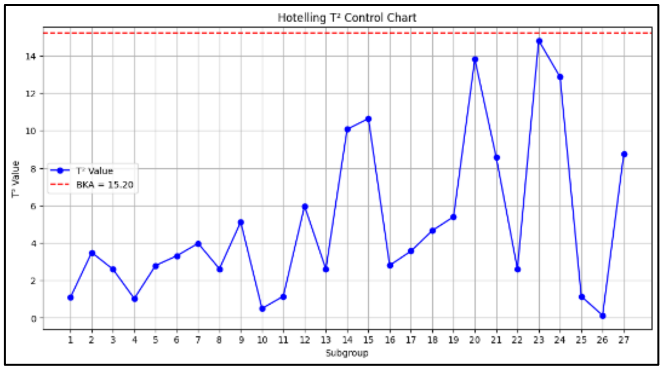

Fig. 4. Monitoring of Process Mean Using the T^2-Hotelling Control Chart in Phase I After Revision

To obtain a more accurate representation of the stable process, these subgroups were removed, leaving 27 subgroups for reanalysis. A revised control chart was then constructed, as shown in Fig. 4. All remaining observations fell within the control limits, indicating that the process mean was stable and in statistical control. The updated upper control limit of 15.20 was established as the new benchmark for monitoring process stability in subsequent phases.

E. Phase II Process Control

This section presents the control charts used to monitor the process during Phase II, including the Generalized Variance and Hotelling's T^2 control charts. The objective is to evaluate whether the process remains stable and within acceptable control limits over time.

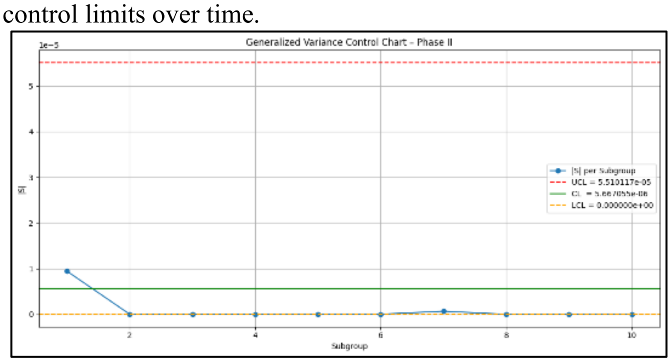

Fig. 5. Monitoring of Process Variability Using the Generalized Variance Control Chart in Phase II

The Generalized Variance control chart for Phase II was developed using the same procedures and control limits established in Phase I. As shown in Fig. 5, a noticeable decrease in the statistics occurs at the second observation. From the second to the tenth observation, the values remain relatively stable and close to zero. This indicates minimal variability between observations and a high level of consistency, suggesting that the process is well-controlled with no significant shifts detected.

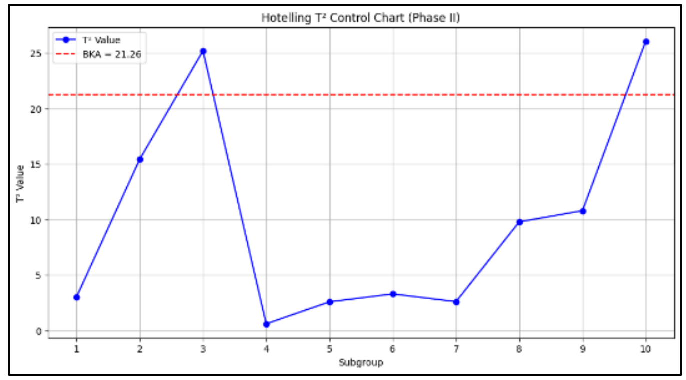

Fig. 6. Monitoring of Process Mean Using the T^2-Hotelling Control Chart in Phase II

The Hotelling's T^2 control chart for Phase II was constructed using the mean vector and inverse covariance matrix obtained from Phase I. The chart shows fluctuating process means, with two out-of-control signals identified at observations 3 and 10. This indicates potential instability in the process during these periods.

IV. CONCLUSION

Based on the analysis and discussion, it can be concluded that the Generalized Variance control chart in Phase I detected one out-of-control signal, indicating that the variability of the water production process was not statistically stable. After revising the data outside the control limits, all 32 subgroups were within the limits. These parameters from Phase I were then used to monitor Phase II, where the control chart showed statistical values that were predominantly close to zero. This indicates a low level of variability and suggests that the water production process at Perumda Air Minum Tugu Tirta Kota Malang is well-controlled in terms of variability.

In addition, the T^2 -Hotelling control chart in Phase I identified five out-of-control signals. After revising the data, 27 subgroups were within control limits. The process means vector and inverse covariance matrix obtained in Phase I were used in Phase II analysis. The results showed that two observations—specifically the 3rd and 10th—were outside the control limits, indicating that the process mean was not entirely stable. These deviations may be due to various factors such as weather conditions, infrastructure development, or other external influences.

Based on these findings, it is recommended to perform regular monitoring of the water production process to ensure consistent quality. Further diagnostic tests should be conducted to identify the root causes of any out-of-control signals. Additionally, it is suggested to carry out process capability analysis to determine whether the process meets the required quality standards.

REFERENCES

[1] Bartlett, M. S. (1951). The Effect of Standardization on a χ^2 Approximation in Factor Analysis. Biometrika, 38(3/4), 337. https://doi.org/10.2307/2332580.

.

[2] Djauhari. M. A. (2005). Improved monitoring of multivariate process variability. Journal of Quality Technology. 37(1). 32–39. https://doi.org/10.1080/00224065.2005.11980298.

[3] Jayanti. J. D.. & Wibawati. W. (2014). Penerapan Diagram Kontrol MEWMA dan MEWMV pada Pengendalian Kualitas Air Produksi di IPAM Ngagel I. Jurnal Sains Dan Seni ITS. 3(2). D314–D321.

[4] Johnson. R. A.. & Wichern. D. W. (2007). Applied Multivariate Statistical Analysis. Pearson Prentice Hall. In Pearson Prentice Hall.

[5] Montgomery. D. C. (2013). Introduction to Statistical Quality Control. Seventh Edition. United Stated of America: John Wiley & Sons. Inc.

[6] Rejeki. P. A. S.. & Handayani. N. U. (2024). Analisis Pengendalian Kualitas Dengan Statistik Multivariat Proses Produksi Gula Kristal Putih GKP. Industrial Engineering Online Journal. 13(1).

[7] Wardhani. R. N. (2017). Penerapan Grafik pengendali Multivariat Pada Pengendalian Kualitas Proses Produksi Air Di PDAM Surabaya.

[8] Widyasmara. A. W. (2016). Pengendalian Kualitas Proses Produksi Kertas Koran PT. Adiprima Suraprinta Menggunakan Multivariat Hotelling T2.

[9] Wilks. S. S. (1932). Certain Generalizations in the Analysis of Variance. Biometrika. 24(3–4). 471–494. https://doi.org/10.1093/biomet/24.3-4.471.