Asia Pacific Science Press

Journal of Advances in Engineering and Technology
ISSN(O):3005-7507
ISSN(P):3005-7493
Vol. 2 No.1 (2025)

Design and Application of a High-Dimensional Robust Control Chart for Joint Monitoring of Location and Scale Parameters

Meiling Lu*

School of Management, Xi'an Polytechnic University, Xi'an, Shaanxi, China

Corresponding author: Meiling Lu, lml118384@163.com

Abstract: In industrial production, statistical process control is a common method used to ensure process stability and product quality. With the development of production technology and the increasing complexity of products, the number of product index parameters that need to be monitored is also increasing, and the traditional control chart method often faces challenges in processing high-dimensional data. For example, the traditional control chart method is applied based on the assumption that the process data distribution is known, and the continuous data is usually assumed to be normally distributed, while many data in the actual process do not follow the normal distribution. Secondly, high-dimensional data often contains complex features, and there are often correlations between variables, which makes it difficult to describe the joint distribution of high-dimensional data. These problems will greatly affect the monitoring effect of the control chart. In view of the above problems and the characteristics of high-dimensional data, this paper first combines the score test statistics with the exponential weighted moving average (EWMA) method after mathematical transformation, and proposes a local statistic to monitor each one-dimensional data stream. Then the correlation between data streams is represented by the appropriate combination of marginal distribution functions, and the global statistics for monitoring high-dimensional data streams are constructed. The control chart proposed in this paper is different from the traditional control chart, it does not need to know the distribution of the process, and can monitor the position parameters and scale parameters simultaneously. The effectiveness and robustness of the control chart are verified by numerical simulation and example analysis.

Keywords: High Dimensional Data Flow; Robust Control Chart; Score Test; Statistical Process Control

Published: Feb 27, 2025

DOI: https://doi.org/10.62177/jaet.v2i1.175

1.Introduction

Statistical process control (SPC) is a process control tool based on statistical methods. At present, SPC has been widely used in the quality management process of manufacturing, retail, service and other industries[1] to achieve process stability by reducing variability. In the process of intelligent manufacturing and service driven by big data, data is usually presented with complex and multivariate characteristics[2], and the complexity of process variables is gradually increasing. In many cases, the data type is not a single continuous data. At present, most multivariate control charts are mainly applicable to the continuous data with known distribution, and there are certain limitations in application[3]. At the same time, in real scenarios, there are often correlations between multiple process variables, which may affect the performance of process monitoring and control charts. The traditional multivariate control chart often assumes that the variables are independent, thus ignoring this correlation, which may lead to misjudgment or missing judgment. In addition, because it is common to collect and analyze multiple related quality characteristics of a process at the same time, single-variable control charts may not be effective in detecting process changes.

Traditional control charts mostly assume that the parameters of the process data are known. However, the actual process parameters are largely unknown and need to be estimated using controlled (IC) data from phase I. The accuracy of parameter estimation requires enough samples, but in practice, there is often insufficient sample information to determine its distribution, and the assumed parameter distribution is rarely effective. In order to solve this important problem, many practitioners have designed non-parametric or distributorless control charts. Tercero-Gomez and Aguilar-Lleyda proposed a Lepage non-parametric CUSUM control chart based on Wilcoxon rank sum and Mood test[4]. However, non-parametric control charts designed based on rank method will lose some data information.

To sum up, with the progress of production technology and the increasing complexity of products, data often presents characteristics such as high dimension, diverse modes and complex correlation. Therefore, for the high-dimensional data flow with unknown distribution, this paper uses the appropriate transformation of score test statistics combined with the correlation between variables to build a control chart, monitor the position parameters and scale parameters of the high-dimensional data flow, find the problems in the process in time and take corresponding measures, so as to improve the stability of the process and product quality.

2.Monitoring Method

In the process of building the control chart, it is necessary to model each one-dimensional data flow separately to obtain the corresponding local statistics. Local statistics reflect the fluctuation of single dimension data and serve as an important basis for the subsequent construction of global control charts. However, considering the local statistics of each one-dimensional data stream alone is not sufficient to fully capture the interrelationships between the multidimensional data. Therefore, it is necessary to model the correlation between various dimensions to obtain a global statistic that can comprehensively reflect the information of all dimensions.

2.1 Control Chart Statistics

In the quality monitoring of a production process, it is assumed that there are p-dimensional data streams observed at time t, all data streams obey the same distribution, but the data streams are not independent, and the k-dimensional data streams are denoted as $\left\{X_{k,t}\right\}_1^\infty$， k=1， 2， ...， p.

First, given the complexity of high-dimensional data, no assumptions are made here about the distribution that the data flows follow. But suppose the mean of the data flow is μ_k, and the variance is σ_k^2. All variables in the k-dimensional data stream are standardized: $(X_{k,t} - \mu_k) / \sigma_k$, and the mean and standard deviation of the standardized variables are μ_{k0}=0 and σ_{k0}^2=1. Therefore, the cumulative distribution function of the k-dimensional data stream is denoted as:

$$G\left(X_{k,t};\mu_k,\sigma_k^2\right)=F\left(\left(X_{k,t}-\mu_k\right)/\sigma_k\right)$$

The probability density function corresponding to the k-dimensional data stream is

$$g\left(X_{k,t};\mu_k,\sigma_k^2\right)=\frac{1}{\sigma}f\left(\left(X_{k,t}-\mu_k\right)/\sigma_k\right)$$

Where is the standard distribution function.

The score test was proposed by Rao CRR, a famous statistician, in 1948[4]. Compared with the likelihood ratio test, the score test only needs to calculate the maximum likelihood estimate of the original hypothesis, and the steps are relatively easy, so it is widely used in statistical diagnosis. The Score Test, also known as the Lagrange Multiplier Test, is a hypothesis-testing method used to test the significance of parameters in a statistical model, often used in generalized linear models, regression analysis, and survival analysis. The main idea is to test the significance of parameters based on the Score of the likelihood function. At time t, the score test statistic of the observed value $X_{k,t}$ is:

$$s_{k,t}{}^T I_k^{-1} s_{k,t}$$

Where $S_{k,t} = [S_1 , S_2]^T$ is the score function vector and I_k is the information matrix.

$$s_1 = \frac{\partial \ln g_k\left(X_{k,t};\mu_k,\sigma_k^2\right)}{\partial \mu_k} = \frac{\partial \ln \frac{1}{\sigma_k} f_k\left(\left(X_{k,t}-\mu_k\right)/\sigma_k\right)}{\partial \mu_k} = -\frac{1}{\sigma_k}\frac{f'_k\left(\left(X_{k,t}-\mu_k\right)/\sigma_k\right)}{f_k\left(\left(X_{k,t}-\mu_k\right)/\sigma_k\right)}$$

$$s_2 = \frac{\partial \ln g_k\left(X_{k,t};\mu_k,\sigma_k^2\right)}{\partial \sigma_k} = \frac{\partial \ln \frac{1}{\sigma_k} f_k\left(\left(X_{k,t}-\mu_k\right)/\sigma_k\right)}{\partial \sigma_k} = -\frac{1}{\sigma_k}\left[1+\frac{\left(X_{k,t}-\mu_k\right)}{\sigma_k}\frac{f'_k\left(\left(X_{k,t}-\mu_k\right)/\sigma_k\right)}{f_k\left(\left(X_{k,t}-\mu_k\right)/\sigma_k\right)}\right]$$

Let $\frac{\left(X_{k,t}-\mu_k\right)}{\sigma_k} = y_{k,t}$, and the information matrix is:

$$I_{(g_k)} = \frac{1}{\sigma_k^2} I_{(f_k)} = \frac{1}{\sigma_k^2}\begin{pmatrix} I_{11} & I_{12} \\ I_{21} & I_{22} \end{pmatrix}$$

According to the definition of information matrix:

$$I_{11} = D\left(\frac{f'_k\left(y_{k,t}\right)}{f_k\left(y_{k,t}\right)}\right) = E\left[\left(\frac{f'_k\left(y_{k,t}\right)}{f_k\left(y_{k,t}\right)}\right)^2\right]$$

$$I_{12} = I_{21} = \mathrm{cov}\left(\frac{f'_k\left(y_{k,t}\right)}{f_k\left(y_{k,t}\right)}, 1+y_{kj}\frac{f'_k\left(y_{k,t}\right)}{f_k\left(y_{k,t}\right)}\right) = E\left[\left(\frac{f'_k\left(y_{k,t}\right)}{f_k\left(y_{k,t}\right)}\right)\left(1+y_{kj}\frac{f'_k\left(y_{k,t}\right)}{f_k\left(y_{k,t}\right)}\right)\right]$$

$$I_{22} = D\left(1+y_{k,t}\frac{f'_k\left(y_{k,t}\right)}{f_k\left(y_{k,t}\right)}\right) = E\left[\left(1+y_{kj}\frac{f'_k\left(y_{k,t}\right)}{f_k\left(y_{k,t}\right)}\right)^2\right]$$

In order to make the score function and the information matrix have a more concise expression, let $u_{k,t} = G_k\left(X_{k,t};\mu_k,\sigma_k^2\right)$, and construct two functions

$$\begin{cases} \phi_1\left(u_{k,t}\right) = -\frac{f'_k\left(F_k^{-1}\left(u_{k,t}\right)\right)}{f_k\left(F_k^{-1}\left(u_{k,t}\right)\right)} \\ \phi_2\left(u_{k,t}\right) = -1-F_k^{-1}\left(u_{k,t}\right)\frac{f'_k\left(F_k^{-1}\left(u_{k,t}\right)\right)}{f_k\left(F_k^{-1}\left(u_{k,t}\right)\right)} \end{cases}$$

Therefore, the expression of the score function is

$$s_1 = \frac{1}{\sigma_k}\phi_1\left(u_{k,t}\right), s_2 = \frac{1}{\sigma_k}\phi_2\left(u_{k,t}\right)$$

The expression of the information matrix is:

$$I_{(g_k)} = \frac{1}{\sigma_k^2}\begin{pmatrix} E\left[\left(\phi_1\left(u_{k,t}\right)\right)^2\right] & E\left[\left(\phi_1\left(u_{k,t}\right)\right)\left(\phi_2\left(u_{k,t}\right)\right)\right] \\ E\left[\left(\phi_1\left(u_{k,t}\right)\right)\left(\phi_2\left(u_{k,t}\right)\right)\right] & E\left[\left(\phi_2\left(u_{k,t}\right)\right)^2\right] \end{pmatrix}$$

However, to compute the score function and the information matrix, you also need to know the probability density function expression of the variable, namely f(). In the selection of probability density function, the convenience of statistic calculation and the monitoring performance of control chart should be taken into account. Therefore, this paper selects the probability density function of the standard logistic distribution,

$$f(y)=\frac{e^{-y}}{\left(1+e^{-y}\right)^{2}}$$

At this time, the score function and information matrix have a more concise expression:

$$\phi_1\left(u_{k,t}\right)=2u_{k,t}-1,\phi_2\left(u_{k,t}\right)=\left(2u_{k,t}-1\right)\ln\frac{u_{k,t}}{1-u_{k,t}}-1$$

$$I_{(g_k)}=\frac{1}{\sigma_k^2}\begin{pmatrix}1/3 & 0\\ 0 & \left(\pi^2+3\right)/9\end{pmatrix}=\frac{1}{\sigma_k^2}I_0$$

Logistic distribution is selected in this paper for the following reasons. First, logistic distribution has a heavier tail than normal distribution, so it is more robust to outliers in the data and more effective in processing abnormal data; Secondly, the form of probability density function and cumulative distribution function of logistic distribution is simple, which is easy to calculate and implement. In this paper, score function vector and information matrix with simpler form are also obtained. Since the distribution function represents the relative position or order information of the observed values to some extent, the standardized rank form of the observed values is 2F(x_i)-1and the standard rank is robust to the distribution, independent of the distribution form of the original data.

Therefore, the k-dimensional statistic based on the score test can be expressed as:

$$Q_{k,t}=\Phi_{k,t}^{\mathrm{T}}I_0^{-1}\Phi_{k,t}$$

The statistics constructed in the previous article only take advantage of the current observations and completely ignore the influence of historical data. Specifically, let

$$\theta_{k,t}=(1-\lambda)\theta_{k,t-1}+\lambda\Phi_{k,t}$$

replaces the score test statistic $\Phi_{k,t}$ in the previous section. In this way, the new statistics are able to take into account both current observations and historical observations at several moments in the past to reflect the time series characteristics of the production process. Further, the expression of the statistic becomes:

$$R_{k,t}=\theta_{k,t}^{\mathrm{T}}I_0^{-1}\theta_{k,t}\quad 1\leq \mathrm{k}\leq \mathrm{p}$$

By modeling the marginal distribution function of the variables, we get an expression that represents the correlation between the variables. This expression can be described in standard rank transform form. At time t, the correlation between P-dimensional data streams can be expressed as:

$$S_t=\sum_{i=1}^{p}\sum_{j=1}^{p}\left|\left(2u_{i,t}-1\right)\left(2u_{i,t}-1\right)\right|$$

On this basis, we further combine this correlation measure with traditional EWMA-type statistics to obtain a more robust monitoring method. Specifically, in order to make the monitoring process more stable and to reflect small changes in the high-dimensional data stream in a timely manner, we converted the correlation measure to EWMA form. Through this transformation, the statistics can better adapt to the long-term trend in the data stream, while effectively filtering out the impact of short-term fluctuations on the monitoring results, improving the overall stability and response speed. To make the monitoring process more stable, convert it to EWMA form:

$$\beta_t=(1-\lambda)S_{t-1}+\lambda S_t$$

Therefore, the statistics for monitoring high-dimensional data streams can be expressed as:

$$Z=\sum_{k=1}^{p}R_{k,t}+\beta_t$$

In summary, by combining the advantages of Lepage statistics and EWMA statistics, this paper proposes a new robust monitoring statistic, which not only makes no assumptions about the distribution of data, but also fully considers the correlation between high-dimensional data streams. This method provides a more flexible and effective means for monitoring changes in high dimensional data flow, and has high practical application value.

2.2 Determine the control limits

In this paper, is 500, and the dichotomy method is used to calculate the control limit CL. Dichotomy is a common numerical optimization method, especially suitable for solving the optimal solution by interval approximation. In statistical process control, dichotomy is applied to optimize the choice of control limits to ensure that normal and abnormal fluctuations can be distinguished effectively. By precisely calculating the control limits, the sensitivity and accuracy of the monitoring process can be improved, thereby identifying potential problems earlier and ensuring the stability of the production process.

3.Numerical simulation and performance evaluation

3.1 Simulation parameter setting

Since the proposed control chart is not affected by data distribution, three different distributions, symmetric, heavy-tail and skew, are selected to verify the effectiveness of the proposed control chart, which are as follows: (1) Multivariate normal distribution $N_p(0,\Omega)$; (2) The student distribution of degrees of freedom expressed by ξ : $t_p(\xi)$; (3) Gamma distribution with shape parameter ξ and scale parameter 1:$Ga_p(\xi,1)$. For the sake of generality, ξ is 5. The covariance matrix associated with these three distributions considers the covariance matrix of exponential decay, and the exponential decay structure represents the common data covariance structure in industrial production, expressed as:

$$\omega_{ij} = 0.5^{|i-j|} \quad i, j = 1, 2, ..., p$$

ARL is used to evaluate the performance of the control chart. In order to make the control chart more robust, according to the experience in the historical references, the ARL of this paper takes 500, and the false positive rate is 0.002. In the actual production process, the dimension of high dimensional data flow is uncertain. In this paper, dimension p is set to 200 only to verify the validity of the proposed control chart.

At time t, it is assumed that the dimension of the data flow in which the mean or variance shifts is $pa = p*\eta$, Where η is the proportion of the dimensions that drift, $\eta \in \{ 0.02, 0.05, 0.1, 0.25, 0.5, 0.8\}$, that is, $pa \in \{ 4, 10,20,50,100,160\}$. The drift of the mean is δ, where δ{0.1, 0.2, 0.5, 1,2}; The amount of drift of the variance is ζ, where ζ ∈{0.1, 0.2, 0.5, 1,2}. It is also assumed that when pa-dimensional data flows drift, the drift of process data mean or variance is the same.

In order to make the control chart more robust, the calculation of statistics should take full account of historical data, so the EWMA method is combined into the calculation of statistics in this paper. In this paper, 0.2 is chosen as the value of the smoothing parameter.

3.2 Simulation results and performance comparison

When the variance of process data is unchanged, the mean value shifts from 0 to δ. The simulation results of the control chart proposed are shown in Table 1:

Table 1: ARL when the mean value shifts to different degrees under different distributions

δ	pa	$N_p(0, \Omega)$、	$t_p(5)$	$Ga_p(5, 1)$
0.1	4	290.5	350.5	394.2
	10	124.6	207.7	308.1
	20	35.2	79.5	259.4
	50	4.88	14.2	102.7
	100	2.03	3.56	5.55
	160	1.91	2.14	5.53
0.2	4	84.5	177.8	319.5
	10	11.6	33.72	164.6
	20	2.91	7.64	98.4
	50	1.91	2.07	5.96
	100	1.17	1.79	1.84
	160	1.0	1.15	1.83

δ	pa	$N_p(0, \Omega)$、	$t_p(5)$	$Ga_p(5, 1)$
0.5	4	6.35	20.9	230.9
	10	1.98	2.51	104.4
	20	1.78	1.98	6.16
	50	1.0	1.04	1.48
	100	1.0	1.0	1.0
	160	1.0	1.0	1.0
1	4	2.09	3.55	158.2
	10	1.78	1.92	6.69
	20	1.0	1.27	1.89
	50	1.0	1.0	1.0
	100	1.0	1.0	1.0
	160	1.0	1.0	1.0
2	4	1.98	2.01	25.17
	10	1.14	1.65	1.94
	20	1.0	1.0	1.0
	50	1.0	1.0	1.0
	100	1.0	1.0	1.0
	160	1.0	1.0	1.0

When the mean value of the process data is unchanged, the variance shifts from 0 to . The simulation results of the control chart proposed are shown in Table 2.

Table 2: ARL when the variance value shifts to different degrees under different distributions

ζ	pa	$N_p(0, \Omega)$、	$t_p(5)$	$Ga_p(5, 1)$
0.1	4	389.2	391.3	417.6
	10	312.2	342.6	398.1
	20	224.9	251.9	309.6
	50	93.9	104.8	156.5
	100	24.9	26.9	42.6
	160	7.75	9.18	15.3
0.2	4	297.4	312.8	341.6
	10	199.5	206.4	258.4
	20	88.7	88.1	105.7
	50	14.7	20.6	32.9
	100	3.33	12.9	21.4
	160	2.08	3.17	6.93

ζ	pa	$N_p(0, \Omega)$、	$t_p(5)$	$Ga_p(5, 1)$
0.5	4	272.6	277.3	310.2
	10	176.9	169.1	200.4
	20	64.8	68.1	90.3
	50	8.29	8.95	21.9
	100	2.63	3.51	14.0
	160	1.98	2.91	6.13
1	4	260.4	269.4	298.1
	10	154.8	156.9	168.2
	20	51.8	58.2	75.2
	50	6.88	7.02	15.8
	100	2.26	2.97	6.36
	160	1.95	2.21	4.56
2	4	256.3	251.2	270.3
	10	144.6	142.4	158.4
	20	46.5	51.7	63.5
	50	5.91	6.27	9.14
	100	2.14	2.78	3.62
	160	1.95	2.09	3.19

This can get the following conclusion:

The control chart has the ability to monitor the data of three different distributions, whether the mean or variance is shifted. In our simulation experiment, the control chart shows strong monitoring ability, which can not only effectively detect the change of the mean value, but also identify the drift of the variance. However, it is important to note that the sensitivity and response speed of the control chart may vary under complex non-normal distributions, especially in the case of smaller or low-dimensional drifts.

② The control chart scheme is more sensitive to the monitoring of the mean than the monitoring of the variance. The simulation results show that the control chart shows high sensitivity when monitoring the change of mean value. Especially in the case of multivariate normal distribution and multivariate student t distribution, a slight shift in the mean can trigger an alarm signal. For the monitoring of variance, the response of the control chart is relatively slow, especially when the variance is small, the control chart may take longer to detect the change. This may be related to the greater impact of mean change on the overall data distribution and the sensitivity of detection threshold setting. In the production process, the mean change is usually more significant, so the control chart can provide more timely and effective feedback when detecting mean drift.

③ When the three different distributions have a large drift or a large number of drifting dimensions, the control chart can immediately generate alarm signals, but when the amount of drift is small or the number of drifting dimensions is small, the control chart has the best monitoring effect on the normal distribution, followed by the t distribution. When the data distribution has a large mean or variance shift, whether it is in the normal, t or Gamma distribution, the control chart can issue an alarm signal in a very short time. This shows that the control chart is very agile in the face of significant drift and can catch abnormal changes in the process in time. However, the control chart behaves differently when the drift is small or occurs only in some dimensions.

It can be seen from the above results that the sensitivity of the control chart is different in the face of different types of distributions. Because of its symmetry and central tendency, normal distribution is more direct and accurate in the drift detection of mean and variance. The thick tail characteristic of the student t distribution makes the control chart may encounter challenges in detecting variance drift, especially in the case of small drift or interference with extreme values, which may lead to false positives or missed positives. Because of the skewness characteristic of the Gamma distribution, the control chart is relatively slow to react, especially when small drifts occur, and more data points may be needed to confirm the exception.

4.Case Analysis

The data comes from the semiconductor manufacturing process of Secom, which is the product high-dimensional data collected by real-time monitoring sensors during the semiconductor manufacturing process [5]. Data dimension p=591, sample size is 1567. Firstly, the data is preprocessed. There are missing values in the data set. From the 590 dimensional variables, 218 variables containing only constant values or too many missing values are removed, and then the remaining p=445 dimensional variables are analyzed. In the remaining data, the column mean interpolation method is used to fill in each missing value of the corresponding column vector. Then, 1463 groups of controlled observation vectors with 445 dimensional variables are tested through normal Q-Q graphs. It can be judged that almost all variables do not obey normal distribution, which also shows the complexity of high-dimensional data.

In order to illustrate the monitoring effect of the proposed control chart in practical application, 1463 controlled samples were sampled in the first stage. Then 104 out-of-control samples were used as online test samples in phase II.In order to make the results more accurate on the premise of continuous distribution, we modify the empirical distribution function by using （−0.5）/1463 as the distribution function for the k-dimensional data stream , t=1,...,1463. Use the sample data to calculate the control chart statistics, the specific process is: first calculate the experience distribution function of the controlled sample data and store it in the new table, and then calculate the control chart statistics, determine the control limit of the control chart according to ARL=500, and then apply the calculated control limit to the monitoring in phase II. The statistics of the out-of-control sample data are compared with the control limit obtained in stage I. If the statistics exceed the control limit, it means that the control chart detects the process anomaly. After python calculation, the results in Table 3 are obtained.

Table 3 High-dimensional Robust control charts monitor the ARL of out-of-control data flow in semiconductor manufacturing processes

λ	Control limit	RL
0.2	462.3	10

Therefore, the effectiveness and practicability of the high-dimensional robust control chart method are verified by the detailed analysis of the example data.

5. Conclusion

Based on the characteristics of high-dimensional data, a new high-dimensional robust control chart method based on score test statistic transformation is proposed in this paper. In order to verify the practical effect of the proposed method, this paper uses Monte Carlo simulation method to generate high-dimensional data samples with different distributions, and uses Python for simulation analysis. The experimental results show that the control chart can effectively monitor the drift of position parameters and scale parameters, especially in the monitoring of position parameter drift shows better sensitivity. As for the monitoring of scale parameters, although the performance is good in most cases, the monitoring effect is slightly inferior to that of the normal distribution in the case of the scale parameter drift of the Gamma distribution. This shows that there are some differences in the performance of the control chart in the face of different distribution characteristics, and further improvement is needed to improve the universality of the control chart in various distributions.

Funding

no

Conflict of Interests

The author(s)declare(s) that there is no conflict of interest regarding the publication of this paper.

References

[1] Ruoyu L ,Xin L ,Jiayin W , et al. SCR-CUSUM: An illness-death semi-Markov model-based risk-adjusted CUSUM for semi-competing risk data monitoring[J]. Computers &amp; Industrial Engineering,2023,184.

[2] Liu M ,Lv J ,Du S , et al. Multi-resource constrained flexible job shop scheduling problem with fixture-pallet combinatorial optimisation [J]. Computers & Industrial Engineering, 2024, 188 109903-.

[3] Alevizakos V ,Koukouvinos C ,Chatterjee K . A nonparametric double generally weighted moving average signed-rank control chart for monitoring process location[J]. Quality and Reliability Engineering International,2020,36(7).

[4] Tercero-Gómez V, Aguilar-Lleyda V, Cordero-Franco A, et al. A distribution-free CUSUM chart for joint monitoring of location and scale based on the combination of Wilcoxon and Mood statistics[J]. Quality and Reliability Engineering International, 2020, 36(4).

[5] Sen P K .Introduction to Rao (1948) Large Sample Tests of Statistical Hypotheses Concerning Several Parameters with Applications to Problems of Estimation[J].Springer New York, 1997.

[6] http://archive.ics.uci.edu/ml/machine-learning-databases/secom/[DB/OL]