Journal of Advanced Research in Applied Sciences and Engineering Technology 43, Issue 2 (2025) 203-219

SEMARAK ILMU
PUBLISHING
202103268166(003316878-P)

Journal of Advanced Research in Applied Sciences and Engineering Technology

Journal homepage:
https://semarakilmu.com.my/journals/index.php/applied_sciences_eng_tech/index
ISSN: 2462-1943

Robust Control Chart Application in Semiconductor Manufacturing Process

Sufinah Dahari[1,*], Muzalwana Abdul Talib[2], Adilah Abdul Ghapor[2]

[1] Department of Research and Development, Dominant Opto Technologies Sdn. Bhd., 75350 Melaka, Malaysia
[2] Department of Decision Science, Faculty of Business & Economics, Universiti Malaya, 50603 Kuala Lumpur, Malaysia

ARTICLE INFO

Article history:
Received 25 September 2023
Received in revised form 17 November 2023
Accepted 1 March 2024
Available online 17 April 2024

Keywords:
Robust statistics; Control limit; Attributes; Defectives

ABSTRACT

Statistical Process Control (SPC) charts are frequently used in the semiconductor manufacturing environment to monitor process quality and detect special-cause variations, hence, to take corrective actions when necessary. The important aspects of control charts to consider on production floors are identifying the primary objective of implementing control charts, the type of data to monitor and the most appropriate control limits to establish. When the quality data is a type of attribute data like the proportion of defectives from a production lot, a *p*-chart approach is most suitable. In *p*-chart applications, although the assumption of normally distributed process data is not mandatory, the widespread practice is to assume the normal distribution of process data when establishing the control limits. Yet again, the reality of industrial settings is that process data are most likely influenced by outliers, resulting in highly skewed distributions. This paper addresses these issues by proposing robust SPC charting techniques to detect special-cause variations in the semiconductor manufacturing processes. Here, we present a case study of a semiconductor company in Malaysia, Dominant Opto Technologies Sdn. Bhd. to propose three robust statistical approaches for monitoring the proportion of defectives in production lots. We apply M-estimates, median, and interquartile range to calculate the upper control limits (UCL) and found that robust estimators are more effective in detecting early process deterioration and capturing the out-of-control (OOC) conditions better than traditional control charts. By proposing robust methods, this study enlightens the practical aspects of process quality improvement for real-life manufacturing setups. Because a high OOC rate may impact manufacturing productivity, we recommend the decision-makers choose the types of control charts based on the implications of each robust approach toward quality and productivity. The significance of this study includes providing insights into setting up the appropriate attribute control charts for detecting defective proportions for professionals and SPC researchers working in these areas.

1. Introduction

The concept of quality has evolved throughout the history of manufacturing, starting with inspection during the Industrial Revolution in the 1700s-1800s, followed by quality control during

* *Corresponding author.*
E-mail address: sufinah.dahari@dominant-semi.com

https://doi.org/10.37934/araset.43.2.203219

mass production in the early 1920s, and statistical process control (SPC) in the 1930s-1940s [1,2]. The SPC approach uses statistical techniques such as a control chart to control a process by monitoring its variation [3]. Inspections ensure the manufactured item conforms to specification using visual inspection or go/no-go gauges; either by sampling inspections or by a 100% inspection [2].

Many visual inspections in the semiconductor manufacturing industry nowadays are performed by automated visual inspection (AVI) machines. This machine uses image-processing techniques to automatically detect defective items [4]. To ensure only good items are delivered to customers, the AVI machines are used to perform a 100% inspection to segregate the items that conform to the product specification (good) from the non-conformance ones (defectives) [4].

In recent years, with the advancement of information technology (IT), artificial intelligence, and connectivity, the data has usually been stored in the company database using the data warehouse concept [5]. May it be counting data or defective rates; control charts detect any special-cause variations to indicate early process deterioration. In the 100% inspection data, attribute data is commonly used. Attribute data are important when qualitative information is needed to assess the manufacturing quality performance. In the semiconductor manufacturing process, a severely skewed distribution of the inspection data is a typical problem faced due to extreme values such as outliers. This could lead to an ineffective control chart that fails to detect out-of-control (OOC) conditions associated with special-cause variations. In such cases of highly skewed data distribution, non-parametric or robust methods are necessary to estimate the control limits. The non-parametric control charts offer options to industrial practitioners; however, they are hard to implement in real-life practice due to a lack of knowledge and support in terms of computational tools [6].

Robust estimator-based control charts are particularly useful in processes dealing with extreme values or outliers, but their suitability for industrial applications may depend on specific process data or situations. Therefore, it is important to evaluate the performance of robust control charts in industrial settings with 100% inspection data in the presence of outliers. This is the motivation for this study. This paper is structured as follows: The next sub-sections examine the literature about SPC charts and robust statistics. The subsequent section presents the methodology used in this research. The next two sections present the results and discussions and end with the concluding remarks.

2. Literature Review

2.1 Statistical Process Control (SPC) Chart

Walter A. Shewhart first introduced control chart methodology to monitor common and special causes of variations [7]. Two types of Shewhart charts based on the type of data are the control chart for variables and the control chart for attributes. Control charts for variables handle continuous data, whereas attribute control charts cater for discrete or countable data. Proportion or defective rates are examples of attribute data. The SPC charts applied most in the manufacturing industry are run charts, individual *X*-charts for variables, and *p*-charts for attributes [8]. In run charts, defective rates are plotted against time to show changes over time and to assess the effects of new equipment or other factors on the process [8]. Variable *p*-charts are commonly used to monitor defective rates when it is not possible to maintain a constant sample size during the inspection process [8].

Generally, control charts can be applied in a two-phase procedure when the parameters of the process quality attribute are unknown. In phase 1, the individual *X*-chart of a historical data set will estimate the process parameters to establish the control limits. The established control limits will then be applied to a new set of process data in phase 2 to examine if the process data is stable where all data is within the control limits [9]. Theoretically, the Shewhart control chart consists of a centre line (CL), an upper control limit (UCL), and a lower control limit (LCL). The control limits are calculated based on the process mean forming the centre line plus or minus (+/-) the multiplier of sigma (scale or dispersion) [10]. When the control chart triggers an OOC due to some special causes, further investigations are needed to prevent poor-quality items from being produced [11]. The process's performance, however, depends on the stability of its parameters, such as the mean and dispersion of the parameters [12]. The following paragraphs discuss the characteristics of different types of control charts relevant to this research for the benefit of the readers, especially the industrial practitioners.

The AIAG SPC is a reference manual widely used by semiconductor suppliers in the automotive industry. These manual states if the data derived is of a discrete nature, like go/no-go or acceptable/not acceptable, an attribute control chart is the best option [13]. It also includes a chart based on count or percent data like a *p*-chart [13].

In Shewhart's *p*-chart for attributes, even though the underlying population distribution is binomial, parameters are estimated based on the normal distribution approximation by applying the central limit theorem [14]. The central limit theorem states that when all samples are of equal size, the distribution of a sample variable approximates a normal distribution and that the sample size grows regardless of the population's actual distribution shape [15].

However, should the data be contaminated by extreme observations or possible outliers that cause a highly skewed distribution, the *p*-charts may not be reliable for detecting special causes of variations. It is further assumed that the *p*-mean is constant over time [14]. In real-life applications, this assumption is not always valid but is noticeable when the subgroup sizes are large and vary. Because the overdispersion problem leads to a very sensitive chart, most points fall outside the control limits where there are more variations than expected from the theoretical limits [14,16,17]. The solution practiced in some industries is to treat the observations in the individual *X*-chart as variables [14]. However, there is an assertion that if the dispersion is measurable, why do we need to estimate the value? [18,19]. In the individual *X*-chart for this attributed data, the data is treated as continuous variables where control limits and sigma (estimated from range, sigma-r) are computed as follows [14,18,20]:

$$UCL_p = \bar{p} + k\sigma_p \tag{1}$$

$$CL_p = \bar{p} \tag{2}$$

$$LCL_p = \bar{p} - k\sigma_p \tag{3}$$

$$R_i = |p_i - p_{i-1}|,\ i = 2,3,\ldots m \tag{4}$$

$$\bar{R} = \frac{1}{m-1} * \sum_{i=2}^{m} R_i \tag{5}$$

$$\sigma_p = \frac{\bar{R}}{1.128} \tag{6}$$

Where;
i : number of subgroups (1,2,3….m)
p : defective proportion (p), which is percent of defective over-inspected quantities
$\bar{p}$ (or p-mean) : mean or average defective proportion (p), taken across subgroups
σ_p(or sigma-r) : estimated standard deviation of the sampling distribution

R_i : range between sample subgroups *i*
$\bar{R}$: average range
m : total number of subgroups

k is the multiplier of sigma (typically is 3) and 1.128 is the value of *d2* (*d2* is a control chart constant with subgroup size, *n* = 2 for ranges of two adjacent values). The criticism of this method is, it does not account for varying subgroup sizes and is only suitable if the subgroup size is constant [14,21].

Laney proposed an improved control chart for attributes (*p'*-chart) to cater for varying subgroup size concerns. In the Laney *p'*-chart, each *p* value is converted to a *z*-score (the number of sample standard deviations between that point and the overall mean), as given in the formula below.

$$Z_i = \frac{(p_i - \bar{p})}{\sigma_{p_i}} \tag{7}$$

Where;
Z : *z*-scores after transformation
σ_{p_i} : standard deviation of the sampling distribution

Then, the numbers of *z*-scores are plotted in an individual's chart with a mean equal to zero as the centre line of the chart. Laney justified that the theoretical mean of *z*-scores is zero and the standard deviation of *z* is "assumed to be unity", so the control limits can be set at +3 and -3 [14]. This results in flat control limits as follows:

$$UCL = +3 \tag{8}$$

$$CL = 0 \tag{9}$$

$$LCL = -3 \tag{10}$$

However, the assumption of "unit variance" is not valid because it relies only on the within-subgroup variation, although batch-to-batch variation is present [14]. Since batch-to-batch variation can be measured using range, Laney combined the z-chart with the individual *X*-chart concept to calculate the sigma-z using range, *R*, and average range, $\bar{R}$, of *z*-scores as described in the Eq. (11) and Eq. (12) below [14].

$$R_i = |z_i - z_{i-1}|,\ i = 2,3,\ldots m \tag{11}$$

$$\bar{R} = \frac{1}{m-1} * \sum_{i=2}^{m} R_i \tag{12}$$

The sigma-z is estimated by using the Eq. (13) below.

$$\sigma_z = \frac{\bar{R}}{1.128} \tag{13}$$

Transform the *z* back into the meaningful units of the *p* using the below Eq. (14).

$$p_i = \bar{p} + \sigma_z Z_i \tag{14}$$

The sigma of transformed *p* (sigma-*p'*) can be calculated using Eq. (15).

$$\sigma_{p\prime} = \sigma_{p\prime} * \sigma_z \tag{15}$$

The *p'*-chart control limits can be obtained by Eq. (16) to Eq. (18).

$$UCL = \bar{p} + k\sigma_{p\prime} \tag{16}$$

$$CL = \bar{p} \tag{17}$$

$$LCL = \bar{p} - k\sigma_{p\prime} \tag{18}$$

Laney suggested that the *p'*-chart should be adopted universally as the new standard for plotting attribute data to replace the *p*-chart. Mohammad *et al.,* confirmed that the *p'*-chart is the best way to deal with attribute data with large and varied subgroup sizes. However, the *p'*-chart could not detect outliers in small samples [22,23].

Most previous studies solely focused on attribute sampling rather than a 100% inspection. In attribute sampling, the number of defects or defective items is determined from a sample, subgroup batch, or lot inspected, and it is only a part of the total manufactured items. Based on the number of defects or defectives, a decision will be made on the entire lot to form a probability statement. In 100% inspection, on the other hand, the probability statement does not exist because the inspection result is a fact. Consider two scenarios below: scenario 1 represents attribute sampling inspection, and scenario 2 represents 100% attribute inspection.

i. Scenario 1: “In one production batch or lot, the total quantity in a lot is 1000 items. One hundred items were sampled for inspection and five items were found defective. The proportion of defective items or defective rate is 5/100, equivalent to 5%. The use of binomial distribution is appropriate because there are chances that the actual defective rate in a lot is not 5%”.
ii. Scenario 2: “In one production batch or lot, the total quantity in a lot is 1000 items. 100% or all the items are inspected and from the inspection, five items are found defective. The proportion of defective items or defective rate is 5/1000, equivalent to 0.5%”.

Attribute control charts that include the widely used *p*-charts [24], and recently the *p'*-chart to monitor defective rates as applied in scenario 1. In scenario 2, the most appropriate control chart is a run chart (treat one lot or one subgroup as one data point). In both scenarios, a question may arise on how to calculate the control limit for the run chart when there are outliers resulting in highly skewed data. We would like to address this issue by proposing robust statistics.

2.2 Robust Statistics

In most real datasets, some data may behave differently than most of the data owing to extraordinary events also known as outliers [25]. In manufacturing settings, these extraordinary events can be caused by machines or process breakdowns. It is estimated that between 5% and 10% of the total observations will be outliers [26]. The presence of outliers in a dataset would violate the normality assumption, rendering traditional statistical methods inappropriate for data analysis. A robust statistical approach is often used when the underlying normality assumption is violated [27].

Robust statistics are defined as having robustness or insensitivity to small deviations. These are primarily distributional robustness from the impact of skewed distribution [28]. Some examples of robust statistics in measures of locations (centre line) are median and robust mean M-estimates as alternatives to mean. The other examples of robust parameters for dispersion (sigma) are interquartile range (IQR), median absolute deviation (MAD), and robust standard deviation M-estimates as alternatives to standard deviation or range [28-32]. Below is the formula for the mentioned robust statistics [28,29,33]

Estimation for locations (centre line);
Median,

$$\mathrm{MD} = \{|\frac{m+1}{2}|\}\mathrm{th} \tag{19}$$

Robust mean M-estimates (Huber),

$$\widehat{\theta_m} = \min_t \sum_{i=1}^{m} \mathrm{rho}(p_i, \mathrm{t}), \qquad \text{where rho}(p_i,\text{t}) \text{ is an arbitrary function} \tag{20}$$

Estimation for dispersion (sigma);
Interquartile range,

$$\mathrm{IQR} = \{|\frac{3\,(m+1)}{4}|\}\mathrm{th} - \{|\frac{(m+1)}{4}|\}\mathrm{th} \tag{21}$$

Median-absolute-deviation,

$$\mathrm{MAD} = \mathrm{Median}\ \{|p_i - \mathrm{MD}|\}, \qquad \text{where MD is the median of } p_i \tag{22}$$

Robust standard deviation M-estimates (Huber),

$$\widehat{\sigma_m} = \mathrm{Q}(\theta, \sigma) = \sum_{i=1}^{N}[\mathrm{rho}(\frac{r_i}{\sigma}) + \mathrm{a}]\ \sigma,\ a > 0 \tag{23}$$

For details in computing Huber M-estimates for location, $\widehat{\theta_m}$ and dispersion, $\widehat{\sigma_m}$ refer to Huber or Chen [34].

The advantage of using median, IQR, and MAD is that the calculation is simple and straightforward. Whereby, the calculation of robust mean M-estimates involves multiple iterations and requires statistical software assistance to perform the task.

Abu-Shawiesh proposed a robust control chart based on MAD as a robust scale estimator for process dispersion. The MAD's major characteristics are that it has a maximum 50% breakdown point that is twice as large as the IQR and is suitable for heavy-tailed distributions [27]. On the other hand, Schoonhoven and Does study alternative standard deviation estimators that are used as a basis to determine the *X*-bar control chart limit [35]. They proposed a robust mean control chart by constructing a phase 1 control chart derived from the trimmed mean of the sample IQR using variable data. Mehmood *et al.,* proposed skewed correction-based for various location control charts (i.e., median and Hodges-Lehmann) to monitor process characteristics with unknown probability [36]. Their study also found that various dispersion estimators (i.e., IQR and Gini mean difference) have high influences on the control charts [36].

Up until this section, we have reviewed the traditional SPC chart and deemed it not suitable for the highly skewed distribution data that is influenced by outliers, and robust statistics suggest promising results to overcome this issue. The next question is, how do we assess which robust control chart should be used for implementation?

In literature, the performance measure of control chart techniques is usually evaluated by the Average Run Length (ARL) [37,38]. ARL is defined as the average number of subgroups before an out-of-control (OOC) condition is identified on the control chart. However, this criterion is not suitable to be used when the distribution changes [39]. Lazariv and Schmid also showed that for certain scenarios, such as process deterioration with change points, ARL may not be an appropriate performance measure [40].

Hence, alternative measures should be considered for such cases. In our case of the semiconductor manufacturing process, we propose the use of an "out-of-control" rate (OOC) as the performance measure to help decision-makers choose the appropriate control charts for their application. This is because, when the control chart triggers OOC, the production lot normally needs to be held while pending investigation. Decision makers are primarily concerned with the potential number of OOCs because, to perform an investigation, they need to allocate resources such as engineering time, manpower, and machines for verification. As such, high OOC will lower down the productivity. Thus, the OOC rate should be calculated using phase 1 data to address such concerns and monitored in phase 2 when process deterioration may occur.

3. Methodology

3.1 Case Study Scenario

In this research, a case study from one of the semiconductor companies in Malaysia, Dominant Opto Technologies Sdn. Bhd., is presented. The company manufactures light-emitting diode (LED) products, and one of the products is coded as Product X. The product is processed in batches (also known as production lots) throughout all manufacturing processes in the company. Then, the product was subjected in batches to an automated visual inspection process.

The typical manufacturing input and output chain for Product X at assembly is shown in Figure 1, where the raw materials are subjected to the die-attached process (1), wire-attached process (2), and automated visual inspection (AVI) process (3) in sequence for all items in all production lots. In this case, process deterioration detected at the automated inspection process can be attributed to the die or wire attach process. Before the AVI machines are released for production, measurement system analysis (MSA) for attributes is done in accordance with the AIAG reference manual [41] to ensure the machines are able to detect and segregate defective items from good items.

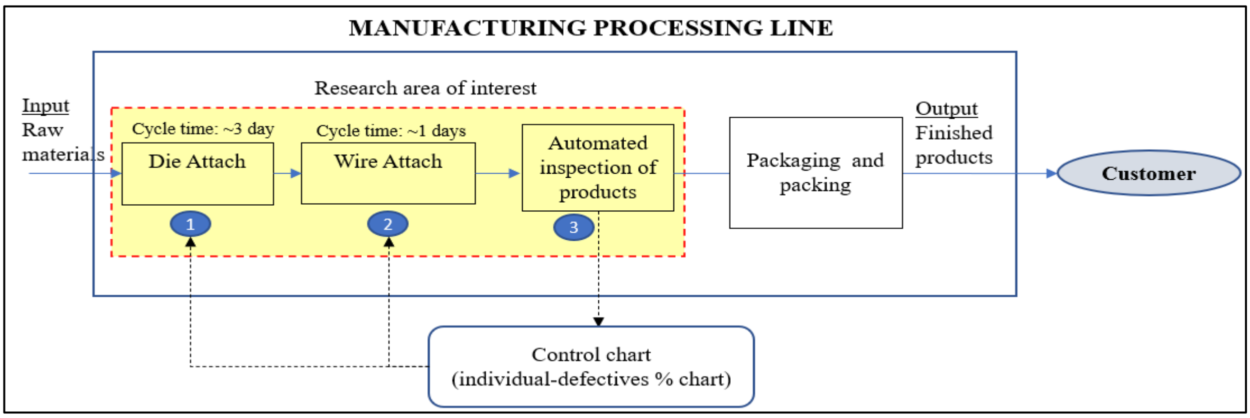


Fig. 1. Manufacturing processing line input and output chain

The AVI process inspects 100% of the product in batches. After each production lot completes the inspection process, the inspection data consists of the production lot ID, inspection date and time, quantity inspected, and number of defectives detected, which are automatically sent into the company IT database.

The Western Electric Handbook (1956) suggests four rules to indicate the presence of special causes in the SPC chart as below.

i. Rule 1: Any single data point that falls outside +3 sigma limit from the CL (i.e., any point that falls outside Zone A).
ii. Rule 2: Two out of three consecutive points fall beyond +2 sigma (i.e., beyond Zone A).
iii. Rule 3: Four out of five consecutive points fall beyond the +1 sigma (i.e., beyond Zone B).
iv. Rule 4: Eight consecutive points fall on the same side of the centerline (i.e., beyond Zone C).

However, in the case of this study, only Rule 1 (beyond limit rule) is applied to detect outliers or special-caused variations. Other rules are not applied in the SPC chart owing to the possibility that the AVI process contributes to the complex variations in the system. This is because there are three different processes that involve

i. die-attached machines
ii. wire-attached machines
iii. AVI machines

Complex and large variations could create more outliers in this case [41]; adding extra control chart rules like "four out of five points in Zone B or beyond", could not possibly contribute to the 'real' early signs of OOC because the control limits derived from the process data could potentially include many outliers.

The decision-makers are interested in knowing if the *p*-chart is still suitable. If not, what are the best alternative control charts to signal early process deterioration for process quality improvement? Considering the presence of outliers in phase 1 and the likelihood presence of more outliers in the preceding manufacturing stage (phase 2).

Typically, Average Run Length (ARL) is used to measure the performance of control charts; however, it is not used in this study because it is not applicable for the process change-point problem or process deterioration, as in the case of this study. Therefore, the performance of any control chart designs derived from different robust statistics proposed in this study will be assessed based on the OOC rates captured by the respective control limits.

3.2 Research Conceptual Framework

From the case study scenario, we attempt to establish a robust control charting technique as the SPC tool in the semiconductor process to detect any special-cause variations occurrence. Thus, the aims of this study are:

i. to review the suitability of the current control *p*-chart as a defective monitoring tool.
ii. to apply the three robust estimation techniques to the current control chart design.
iii. to explore the decision-making implications of robust estimation techniques on current control charts.

To review the suitability of current control chart applications, the distribution and summary statistics of the phase 1 data set are determined, and the total defective proportion (p) is verified to derive the upper control limits (UCL) and the respective OOC conditions for three different control limits derived from three robust methods (using Rule 1: beyond limit rule) shown by Eq. (24) to Eq. (27) below:

Let,

UCL = measure of location + k*measure of dispersion (24)

with k set to 3 (typical number of sigma multiplier).

Then,

Robust method 1: (XMD-IQR) UCL_1 = pMD + 3*IQR (25)

Robust method 2: (XMD-MAD) UCL_2 = pMD + 3*MAD (26)

Robust method 3: (Huber M-estimate) UCL_3 = pRobustMean + 3*RobustStd (27)

Here, the performance of control limits is compared in terms of OOC rate as an implication towards decision makers consideration. OOC rate is defined as Eq. (28) below:

$$\mathrm{OOC.Rate} = \frac{number\ of\ OOC}{number\ of\ lots,\ m} \tag{28}$$

All OCC lots held are to be investigated to identify the root cause at the die or wire-attached process. No outliers have been removed to verify whether the calculated limits are robust based on actual production performance.

3.3 Data Collection and Procedure

Typically, in semiconductor manufacturing companies, one- or three-month data is used to setup control charts. In this case, two sets of AVI data for Product X are retrieved from the company IT database. One-month data is used to calculate the control limit (phase 1), and two months data is used to monitor the performance of the control limit (phase 2). Thus, the data used in this study is secondary data. The production lot size varies from lot to lot. Since 100% of the items in the production lot were inspected, each lot is one data point for defective proportion.

The company setup the p-chart using JMP Version 15 software and used phase 1 data to calculate the control limits. The p-chart captures the total defective proportion in the production lot (in percentage, %), with k set to 3 (the typical number of sigma multiplier) for the control limit computation. The chart is arranged in increased date order for trend monitoring at the AVI process and used as a tool to provide feedback on process deterioration to root-cause processes.

Using the similar data set in the case study scenario, we use JMP Version 15 software by SAS to generate the summary statistics such as mean, standard deviation, median, IQR, MAD, Huber M-estimates for location, and dispersion (robust mean and robust standard deviation). The JMP distribution platform uses ROBUSTREG procedure by SAS to compute the Huber M-estimates (robust mean and robust standard deviation) [34].

JMP hardcodes the default scale parameter, d for Huber-estimate, as 2.5 and the tuning constant for the weight function, c, as 1.345 [43]. Then, the UCL was calculated based on Eq. (25) for median and IQR (XMD-IQR), Eq. (26) for median and MAD (XMD-MAD), and Eq. (27) for Huber M-estimate by using Microsoft Excel 365. Afterwards, we calculate the OOC rate for phase 1 and phase 2 for all methods.

Figure 2 shows the data collection procedure flowchart for this case study research.

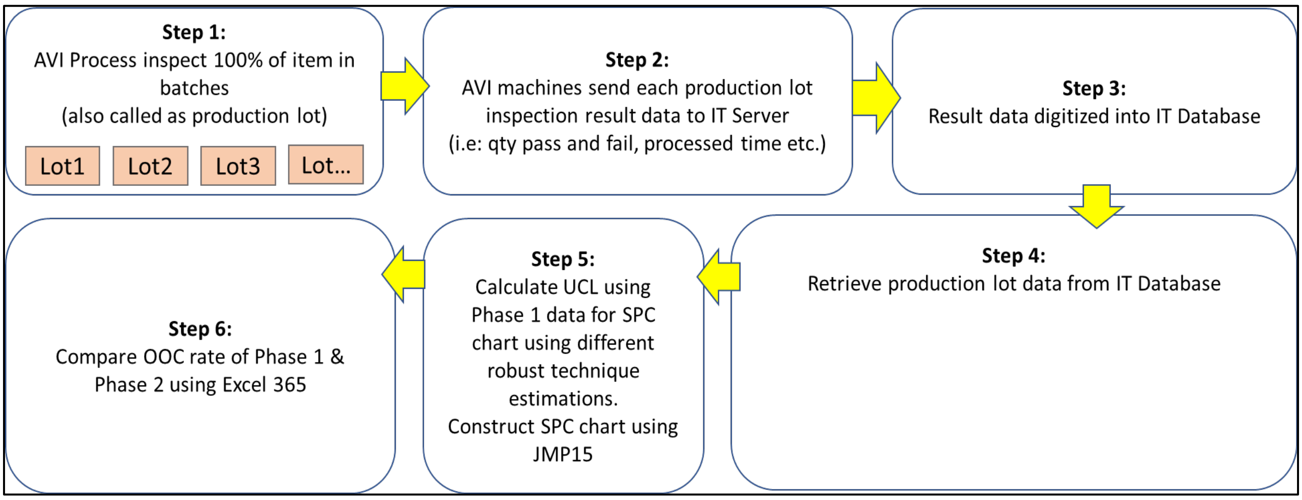


Fig. 2. Data collection procedure flowchart

Table 1 shows the summary of data collection for phase 1 and phase 2. The collected data from Lot 1 until Lot 115 and Lot 116 until Lot 825 are labelled as phase 1 and phase 2 data respectively.

Table 1
Summary of manufacturing attributes data of Product X

Location	Automated Inspection Process	
Type	Phase 1	Phase 2
Period	5th Oct 2022 to 5th Nov 2022	5th Nov 2022 to 19th Dec 2022
Lot No. / Lot ID	Lot 1 – Lot 115	Lot 116 – Lot 825
Number of lots, *m*	115	710
Total inspected quantity (in count per lot, *N*)	Varies between 1,199 to 12,721	Varies between 1,598 to 12, 114
Total defectives quantity (in count per lot, *D*)	Varies between 19 to 1,207	Varies between 25 to 6,047

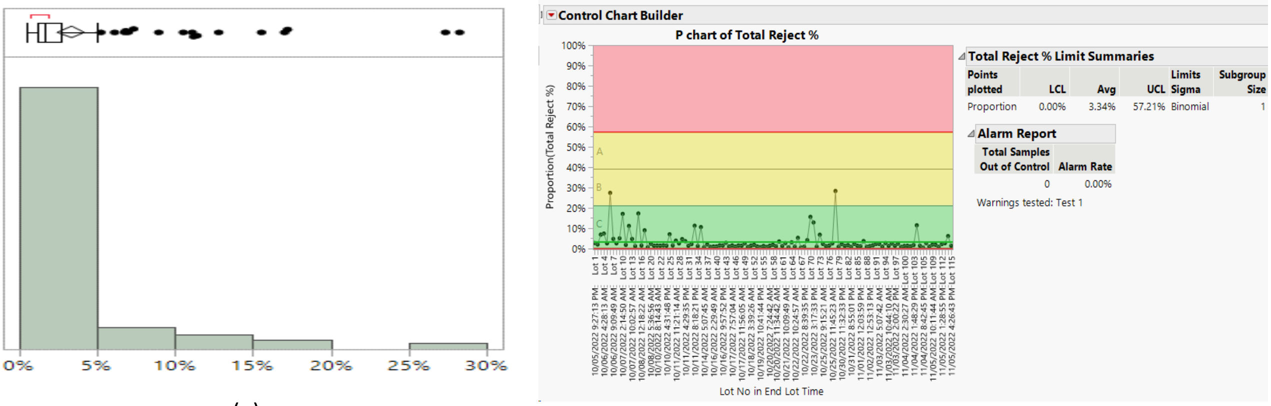
4. Results and Analysis

The summary of statistics and control limits calculated in phase 1 is shown in Table 2. The lower control limit (LCL) set in this control chart is zero because the proportion of defectives cannot be less than zero. The out-of-control (OOC) rule used in this *p*-chart is one point beyond the 3-sigma limit. Each lot that fails the *p*-chart upper control limit (57.21%) will be held for the process engineer to determine the cause of failure either at die-attached or wire-attached process.

Table 2
Summary of statistics for Product X (Phase 1)

Statistics	**Values**
Std Dev, $\sigma\hat{p}$	0.1796 (17.96%)
Mean, $\bar{p}$	0.0334 (3.34%)
$UCL\bar{p}$	0.5721 (57.21%)
$LCL\bar{p}$	0.00 (0.00%)
Number of lots, *m* (Phase 1)	115

Figure 3(a) shows that the distribution is highly skewed to the right. indicating the presence of outliers in the dataset. This is reasonable in the manufacturing industry because the ideal defective rate is always zero. Whereas Figure 3(b) depicts the SPC chart during phase 1.

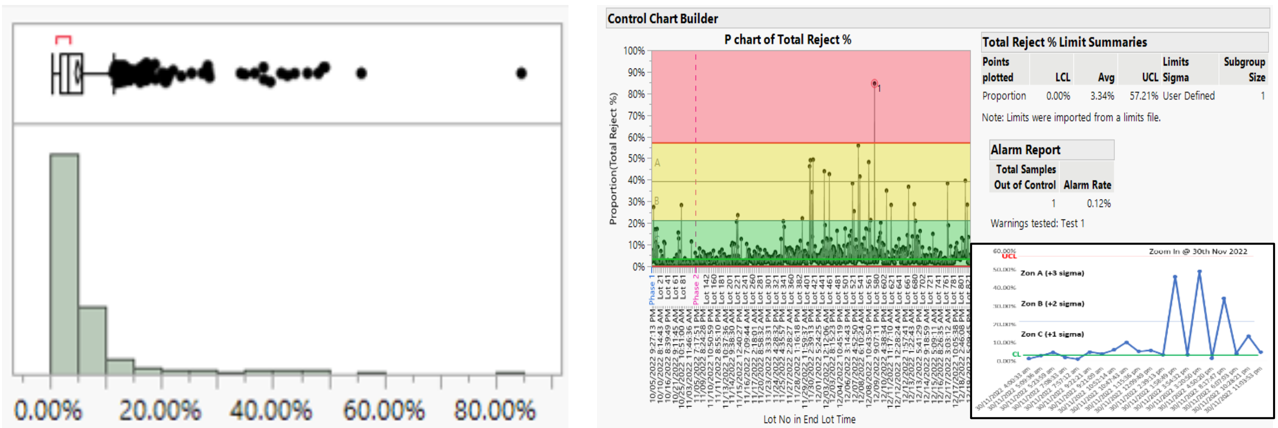


(a) (b)

Fig. 3. Distribution of phase 1 attribute data (a) and SPC chart with phase 1 control limit (b)

In this three-zone control chart [44,45], Zone C (in green area) indicates +1 sigma from CL; Zone B indicates +2 sigma from CL; and Zone A indicates +3 sigma from CL. So, the UCL when k sigma set to 3 is 57.21% (red line, beyond Zone A). The line that separates Zone C and B is 21.30% (+1 sigma from CL), while the line that separates Zone B and A is 39.25% (+2 sigma from CL). The OOC point is marked "1" in the SPC chart. During phase 1, no OOC should be detected with the control limits derived from the data set. Note that "total defective proportion" and "total reject %" are used interchangeably in the SPC chart.

Figure 4(a) shows the overall distribution of phase 1 and phase 2 data, and Figure 4(b) shows the p-chart of total defective proportion used at AVI process after the UCL is applied in phase 2.



(a) (b)

Fig. 4. Distribution of phase 1 and phase 2 attribute data (a) and SPC chart with phase 1 and phase 2 attribute data with zoom-in chart on November 30, 2022 (b)

From Figure 4, although major process deteriorations started to occur in the afternoon on November 30, 2022, where many points are observed in the +2-sigma zone (Zone B; refer to the zoomed-in diagram in Figure 4(b)), the OOC condition is only deemed to occur when the proportion hits above 57.21%, which is the UCL. Phase 1 depicted the condition of "stable process" as there are no OOC signals. This indicates there is no incentive or motivation for the process engineers to further improve the process quality. However, OOC are likely due to the potential occurrence of non-normality in the semiconductor manufacturing data distribution [46].

Table 3 shows the summary statistics used to calculate the UCL and its OOC rate for phase 1 and phase 2 data using different methods. Method 1 (UCL sigma-*p*) is used as the baseline. The *p*-chart puts UCL at 57.21%. The percentage of OOC is only 0.14% in phase 2 (only 1 lot over 710 lots) and zero in phase 1. On the other hand, the UCL for the robust techniques is 8.29% (Huber M-estimate method), 6.31% (XMD-IQR method), and 3.61% (XMD-MAD method). The difference in UCL between robust methods is only 4.68%; however, the OOC rate ranges from 9.56% (11 lots over 115 lots) to 21.74% (25 lots over 115 lots) in phase 1. More triggering happens in phase 2, where the OOC rate ranges from 13.10% (93 lots over 710 lots) to 47.26% (337 lots over 710 lots) since process deterioration is observed. These OOC rate differences indicate the opportunity for the engineers to take early corrective action in phase 1 to avoid further process deterioration in phase 2.

Table 3
Summary statistics to calculate UCL and its OOC for different methods

Method	Location of p	Dispersion (Sigma)	UCL	Phase 1: 115 lots		Phase 2: 710 lots	
				Count of OOC	Percentage of OOC	Count of OOC	Percentage of OOC
p-chart	Mean: 3.33%	17.96%	57.21%	0	0.00%	1	0.14%
XMD-IQR	Median: 1.58%	1.57%	6.31%	14	12.17%	157	22.11%
XMD-MAD	Median: 1.58%	0.68%	3.61%	25	21.74%	337	47.46%
Huber M-estimate	Robust Mean: 2.43%	1.95%	8.29%	11	9.56%	93	13.10%

Figure 5 visualizes the UCL and OOC locations of different robust methods and *p*-chart with respect to each production lot in the SPC chart. From Figure 5, we observed that robust statistic control charts are more effective in detecting outliers and special-caused variations (the count of outliers and special-caused variations can be referred to as the OOC in Table 3), particularly when there are change-point or process deterioration problems.

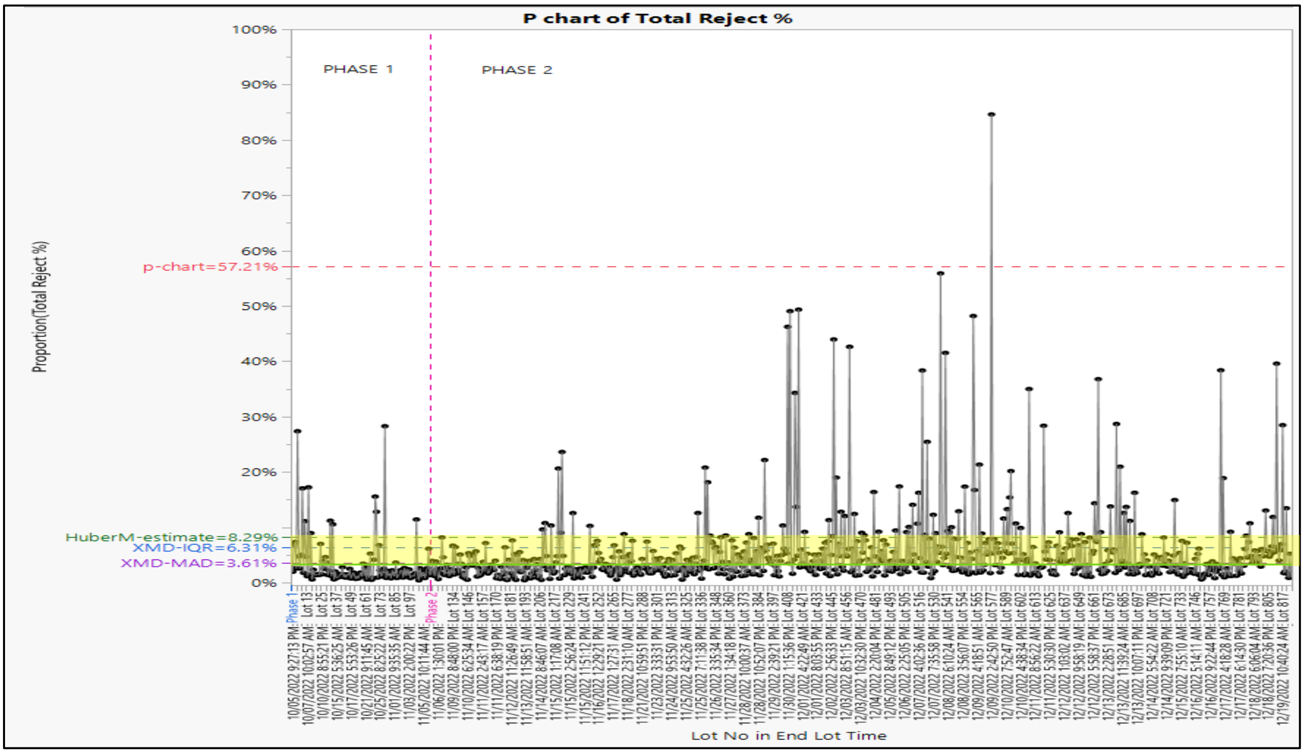


Fig. 5. Upper control limit using various methods in control charts

To select the best robust method for the SPC chart in this case, we compare the OOC rate between phase 1 and phase 2. Figure 6 presents the results, where XMD-MAD has the highest OOC rate.

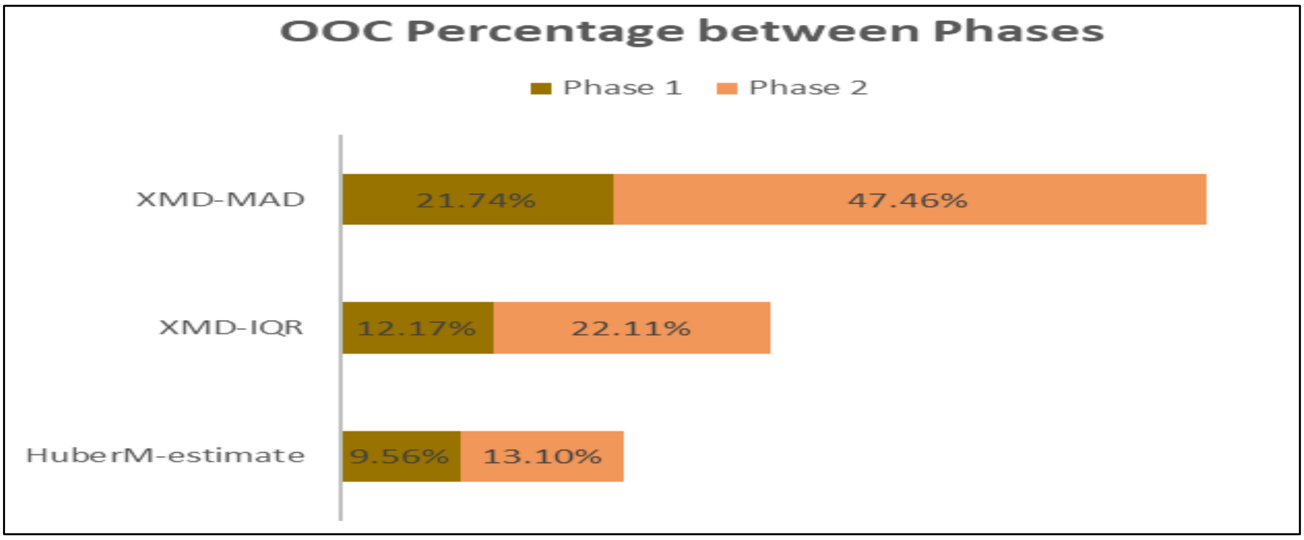


Fig. 6. OOC rate (in percentage) between phases

5. Discussion

In the previous section, we observed that the *p*-chart control limits are not robust against the presence of outliers, as shown by the highly skewed distribution. It has hampered early signals of process deterioration. Too wide control limits imply no special causes of variation in the process, whereas outliers are known as the cause of special or assignable process variations. Theoretically, *p*-

chart control limits are unreliable when dealing with highly skewed data because they are based on the binomial distribution assumption [47]. However, we found evidence that robust statistics can increase the performance of attribute control charts, i.e., the *p*-charts.

We also opined that if the outliers in phase 1 are detected as OOC, special causes of variation can be investigated early to identify the root cause at the die or wire-attached process and prevent further process deterioration. As highlighted by Walfish and Durivage, in the actual manufacturing processes, any potential outliers should be recorded for future evaluation when more data is available [42,48].

In real-life practice, decisions on the control chart limits should be made during phase 1. The information on different OOC rates observed during phase 1 could better assist in preparing for any occurrences of OOC signalled by the control limits set for phase 2 operation. In our case, all the proposed robust methods, Method 1 – 3 have proven their capability to produce early signals of process deterioration. Our findings reveal Method 2 (XMD-MAD) has the tightest control limit at 3.61%, which captures the 21.74% OOC rate in phase 1. The chart is also able to detect more shifts of defective proportions during phase 2 (47.46%). As such, Method 2 also has a higher OOC rate compared to Method 1 (XMD-IQR) by 44% to 53% (about 50%). This could be due to MAD breakdown points being 50%, twice as much as the IQR. The effectiveness of control chart designs with a MAD-robust approach was also proven by Raji *et al.*, [49]. However, several decision-makers do not favour this because it is too stringent. For this study, we assessed the presence of outliers using Mathieu (2021) and actual datasets, and we found 5% and 10% outliers, as demonstrated in Method 3 (Huber M-estimates), which indicates a 9.56% OOC rate in phase 1.

This study enlightens the practical aspects of process quality improvement in real-life manufacturing setups. Despite stimulating findings revealed by this research study, to improve the quality of the manufacturing process, the decision-makers or practitioners need to balance out the number of resources (such as labour and machine capacity) allocated to investigate any occurrence of OOC. Similar to the number of OOC triggers. In Method 3, we observed that the increase in the OOC rate in phase 2 (at 13.1%) is less than 5% higher than in phase 1 (at 9.56%). It could be that the same number of resources allocated during phase 1 would be sufficient to deal with the increase. Because of the trade-offs between quality improvement based on OOC rate and productivity, in this study, the decision-makers should review the chart in Figure 5 and the OOC rate in Figure 6. Depending on special-caused variations, the Huber M-estimates approach can be the best alternative method compared to the others.

Unlike other studies that utilized simulated data, we significantly demonstrate that robust statistics promote effective *p*-charts when applied to actual semiconductor manufacturing process data. The technique also delivers early signals of process deterioration. We also provide practical recommendations for decision-makers to choose the best estimator and the consequences of each robust strategy.

6. Conclusions

The findings of this case study add to the empirical literature on *p*-charts. There are claims that *p*-charts are not robust to the presence of outliers and highly skewed data, preventing early detection of special causes of variation. This present study proposes robust control charts, which are more effective in detecting process deterioration. In this case, the OOC rate is utilized as the new control chart performance measure. But to select the best control chart for production floor usage, decision-makers must also review the OOC rate versus the potential impact on productivity. This study provides insights and guidance for professionals and researchers working in the SPC field and semiconductor manufacturing to set up appropriate attribute control charts for defective proportions. However, this has its limitations.

This case study only applies to a specific semiconductor manufacturing process currently employing 100% inspection using AVI machines. To address these limitations, we suggest future researchers investigate the effectiveness of other statistical process control methods in detecting process deterioration in the context of different industries or with various types of data. Moreover, our model is limited to a typical *k*-sigma multiplier of 3, to calculate the control limits. Although the technique is common in practice, we are concerned if this limitation would hinder the industrial practitioner from using other sigma multipliers; like 1 or 2, or even 4 or 6, to control their process. The 3-sigma multiplier was proposed by Shewhart in the 1920s for economic reasons. But, as Woodall and Faltin argued, what was economical in the 1920s may not be economical in the 21st century [50].

Since the original SPC theory dates to the 1920s in the era of Industry 2.0 [51], more studies are needed to investigate the modern way we can determine control limits. As the semiconductor manufacturing industry moves into Industry 4.0, the data involved can become far too complex for typical statistical process monitoring tools to handle [51]. One of the main roles of applied statistics is to advise the decision-makers on how to refine their decisions in the face of uncertainty [52]. This case study could assist decision-makers in developing a robust SPC framework for their organization. The development of framework can help organizations be more resilient in the presence of a technological revolution [53]. Future studies should also benchmark the cognitive skills related to brain-based abilities (i.e., information processing, knowledge acquisition and reasoning) [54] with artificial intuition for automated decision-making methods [55]. This is essential for the innovation of adaptable, reliable artificial intelligence quality monitoring systems that can withstand various Industry 4.0 decision-making challenges including new process development with limited data.

Acknowledgement

This publication is partially funded by the Faculty of Business and Economics, Universiti Malaya Special Publication fund. The authors would like to express gratitude to Dominant Opto Technologies Sdn. Bhd., for consenting to the use of the case study, raw data, and statistical software, JMP Version 15 for this research. Special acknowledgement given to the technical managers and process engineers that provide industrial input in research concepts and framework. The authors would also like to thank an anonymous reviewer for proofreading and suggestions throughout the early writing stage, which we incorporated to improve the reporting of literature review and methods. The authors would like to express gratitude to the anonymous referees and audience who provided thoughtful questions, comments, and suggestions during the conference proceeding at the 4th International Conference on Applied & Industrial Mathematics and Statistics 2023 (ICoAIMS 2023). Special thanks to the editorial for further suggestions to improve this article.

References

[1] Nicholas, John M. *Instructor's Manual to Accompany" Competitive Manufacturing Management: Continuous Improvement, Lean Production, and Customer-focused Quality"*. Irwin/McGraw-Hill, 1998.

[2] Scallan, Peter. *Process planning: the design/manufacture interface*. Elsevier, 2003. https://doi.org/10.1016/B978-075065129-5/50003-9

[3] ASQ. "Quality-resources/Statistical-process-control". https://asq.org/quality-resources/statistical-process-control#History

[4] Huang, Szu-Hao, and Ying-Cheng Pan. "Automated visual inspection in the semiconductor industry: A survey." *Computers in industry* 66 (2015): 1-10. https://doi.org/10.1016/j.compind.2014.10.006

[5] Jones-Farmer, L. Allison, Jeremy D. Ezell, and Benjamin T. Hazen. "Applying control chart methods to enhance data quality." *Technometrics* 56, no. 1 (2014): 29-41. https://doi.org/10.1080/00401706.2013.804437

[6] Smajdorová, Tereza, and Darja Noskievičová. "Analysis and application of selected control charts suitable for smart manufacturing processes." *Applied Sciences* 12, no. 11 (2022): 5410. https://doi.org/10.3390/app12115410
[7] Shewhart, Walter Andrew. "Control of quality of manufactured product." (1929).
[8] Oakland, John, and John S. Oakland. *Statistical process control*. Routledge, 2007. https://doi.org/10.4324/9780080551739
[9] Vining, Geoff. "Technical Advice: Phase I and phase II control charts." *Quality Engineering* 21, no. 4 (2009): 478-479. https://doi.org/10.1080/08982110903185736
[10] Abu-Shawiesh, Moustafa O., and Mokhtar B. Abdullah. "New robust statistical process control chart for location." *Quality Engineering* 12, no. 2 (1999): 149-159. https://doi.org/10.1080/08982119908962572
[11] Huat, Ng Kooi, and Habshah Midi. "Robust individuals control chart for shifts in process mean and variance." In *Proceedings of the 9th WSEAS international conference on Applications of computer engineering*, pp. 270-275. 2010.
[12] Schoonhoven, Marit, Muhammad Riaz, and Ronald JMM Does. "Design and analysis of control charts for standard deviation with estimated parameters." *Journal of Quality Technology* 43, no. 4 (2011): 307-333. https://doi.org/10.1080/00224065.2011.11917867
[13] Chrysler Group LLC, Ford Motor Company, General Motors Corporation. (2005). STATISTICAL PROCESS CONTROL (SPC) Reference Manual (Vol. 2nd Edition). Automotive Industry Action Group (AIAG).
[14] Laney, David B. "Improved control charts for attributes." *Quality Engineering* 14, no. 4 (2002): 531-537. https://doi.org/10.1081/QEN-120003555
[15] Ahad, Nor Aishah, Che Rohani Yaacob, and Abdul Rahman Othman. "Central Limit Theorem in a Skewed Leptokurtic Distribution." *Journal of Science and Mathematics Letters* 3, no. 1 (2011): 64-71.
[16] Montgomery, Douglas C. *Introduction to statistical quality control*. John wiley & sons, 2019.
[17] Goedhart, Rob, and William H. Woodall. "Monitoring proportions with two components of common cause variation." *Journal of Quality Technology* 54, no. 3 (2022): 324-337. https://doi.org/10.1080/00224065.2021.1903823
[18] Wheeler, Donald J. *Advanced topics in statistical process control*. Vol. 470. Knoxville, TN: SPC press, 1995.
[19] Woodall, William H. "Controversies and contradictions in statistical process control." *Journal of quality technology* 32, no. 4 (2000): 341-350. https://doi.org/10.1080/00224065.2000.11980013
[20] Western, E. "Statistical quality control handbook. Western Electric Co." (1956): 25-8.
[21] Mohammad, Rozhan Kamal, Aras, Jalal, and Saeed, Azad. "An Improvement in the use of Control Chart (p'-chart)." In *Proceedings of the 5th International Scientific Conference to Arab Statistician Union*. 2016.
[22] Vidmar, Gaj, and Rok Blagus. "Outlier detection for healthcare quality monitoring–a comparison of four approaches to over-dispersed proportions." *Quality and Reliability Engineering International* 30, no. 3 (2014): 347-362. https://doi.org/10.1002/qre.1581
[23] Ahsan, Muhammad, Muhammad Mashuri, and Hidayatul Khusna. "Evaluation of Laney p'chart performance." *International Journal of Applied Engineering Research* 12, no. 24 (2017): 14208-14217.
[24] Abbas, Zameer, Hafiz Zafar Nazir, Noureen Akhtar, Muhammad Abid, and Muhammad Riaz. "On designing an efficient control chart to monitor fraction nonconforming." *Quality and Reliability Engineering International* 36, no. 2 (2020): 547-564. https://doi.org/10.1002/qre.2590
[25] Rousseeuw, Peter J., and Mia Hubert. "Anomaly detection by robust statistics." *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery* 8, no. 2 (2018): e1236. https://doi.org/10.1002/widm.1236
[26] Mathieu, Timothée. "M-estimation and Median of Means applied to statistical learning." PhD diss., Université Paris-Saclay, 2021.
[27] Abu-Shawiesh, Moustafa OA. "A simple robust control chart based on MAD." *Journal of Mathematics and Statistics* 4, no. 2 (2008): 102. https://doi.org/10.3844/jmssp.2008.102.107
[28] Huber, Peter J. *Robust statistics*. Vol. 523. John Wiley & Sons, 2004.
[29] Huber, Peter J. *Robust statistical procedures*. Society for Industrial and Applied Mathematics, 1996.
[30] U.S. Department of Commerce. https://www.itl.nist.gov/div898/software/dataplot/refman2/auxillar/qn_scale.htm
[31] Croux, Christophe, and Catherine Dehon. "Robust estimation of location and scale." *Wiley StatsRef: Statistics Reference Online* (2014). https://doi.org/10.1002/9781118445112.stat07416
[32] Manikandan, S. "Measures of dispersion." *Journal of Pharmacology and Pharmacotherapeutics* 2, no. 4 (2011): 315. https://doi.org/10.4103/0976-500X.85931
[33] JMP. https://www.jmp.com/support/help/en/17.0/#page/jmp/the-summary-statistics-report.shtml#ww311398
[34] Chen, Colin. "Paper 265-27 Robust regression and outlier detection with the ROBUSTREG procedure." In *Proceedings of the Proceedings of the Twenty-Seventh Annual SAS Users Group International Conference*. 2002.

[35] Schoonhoven, Marit, and Ronald JMM Does. "A robust control chart." *Quality and Reliability Engineering International* 29, no. 7 (2013): 951-970. https://doi.org/10.1002/qre.1447
[36] Mehmood, Rashid, Muhammad Hisyam Lee, Muhammad Rizwan Iqbal, and Sajdah Hassan. "New robust location control charts for unknown process distribution with practical significance." *Soft Computing* 26 (2022): 137-153. https://doi.org/10.1007/s00500-021-06497-4
[37] Jamali, Abdul Sattar, Jinlin Li, M. Usman Keerio, and Altaf Rajpar. "Performance measure of shewhart $\bar{X}$ control chart for armature resistance process." *Information Technology Journal* 6, no. 4 (2007): 573-578. https://doi.org/10.3923/itj.2007.573.578
[38] Tran, Phuong Hanh, Adel Ahmadi Nadi, Thi Hien Nguyen, Kim Duc Tran, and Kim Phuc Tran. "Application of machine learning in statistical process control charts: A survey and perspective." In *Control charts and machine learning for anomaly detection in manufacturing*, pp. 7-42. Springer, Cham, 2022. https://doi.org/10.1007/978-3-030-83819-5_2
[39] Tepat, Berdasarkan Taburan Panjang Larian, Jk Chong, Wl Teoh, Michael BC Khoo, Zl Chong, and Sy Teh. "Performance Comparison of the Exact Run-Length Distribution Between the Run Sum X and Ewma X Charts." *Journal of Quality Measurement and Analysis JQMA* 12, no. 1-2 (2016): 11-26.
[40] Lazariv, Taras, and Wolfgang Schmid. "Surveillance of non-stationary processes." *AStA Advances in Statistical Analysis* 103 (2019): 305-331. https://doi.org/10.1007/s10182-018-00330-4
[41] Chrysler, L. L. C. "Ford motor company, general motors corporation." *Potential Failure Mode and Effect Analysis (FMEA)* (2008): 67-112.
[42] Durivage, Mark. "Identifying Outliers in Process Data Using Visual and Analytical Techniques." (2018). https://www.pharmaceuticalonline.com/doc/identifying-outliers-in-process-data-using-visual-and-analytical-techniques-0001
[43] Inc, S. I. "SAS/STAT 14.2 User's Guide." *Cary, NC, USA: SAS Institute Inc* (2016).
[44] SAS, I. "JMP®, Version Pro 15.0." *SAS Institute Inc., Cary, NC* 2019 (1989): 2019.
[45] Nelson, Lloyd S. "The Shewhart control chart—tests for special causes." *Journal of quality technology* 16, no. 4 (1984): 237-239. https://doi.org/10.1080/00224065.1984.11978921
[46] Kubo, Tomoaki, Tomomi Ino, Kazuhiro Minami, Masateru Minami, and Tetsuya Homma. "A statistical process control method for semiconductor manufacturing." *SICE Journal of Control, Measurement, and System Integration* 2, no. 4 (2009): 246-254. https://doi.org/10.9746/jcmsi.2.246
[47] Wang, Hsiuying. "Comparison of p control charts for low defective rate." *Computational statistics & data analysis* 53, no. 12 (2009): 4210-4220. https://doi.org/10.1016/j.csda.2009.05.024
[48] Walfish, Steven. "A review of statistical outlier methods." *Pharmaceutical technology* 30, no. 11 (2006): 82.
[49] Raji, Ishaq Adeyanju, Muhammad Hisyam Lee, Muhammad Riaz, Mu'azu Ramat Abujiya, and Nasir Abbas. "Outliers detection models in shewhart control charts; An application in photolithography: A semiconductor manufacturing industry." *Mathematics* 8, no. 5 (2020): 857. https://doi.org/10.3390/math8050857
[50] Woodall, William H., and Frederick W. Faltin. "Rethinking control chart design and evaluation." *Quality Engineering* 31, no. 4 (2019): 596-605. https://doi.org/10.1080/08982112.2019.1582779
[51] ProcessMiner, Inc. "The Evolution and Required Transformation of Statistical Process Monitoring". https://processminer.com/the-evolution-and-required-transformation-of-statistical-process-monitoring
[52] Stefan, Angelika M. "Statistics for Making Decisions, by Nicholas T. Longford. Boca Raton, FL: Chapman & Hall/CRC Press, 2021, xv+ 292 pp., $120.00 (H), ISBN: 978-0-36-734267-8." (2022): 87-88. https://doi.org/10.1080/00031305.2021.2020003
[53] Othman, Zakirah, Mohamad Faizal Ahmad Zaidi, and Wan Ahmad Jaafar Wan Yahaya. "Adoption Strategy for Electrical and Electronics (E&E) Small and Medium-Sized Enterprises (SMEs): Malaysia IR4. 0 Perspective." *Journal of Advanced Research in Applied Sciences and Engineering Technology* 28, no. 3 (2022): 27-38. https://doi.org/10.37934/araset.28.3.2738
[54] Ibrahim, Muhammad Shafiq, Seri Rahayu Kamat, and Syamimi Shamsuddin. "The Role of Brain Wave Activity by Electroencephalogram (EEG) in Assessing Cognitive Skills as an Indicator for Driving Fatigue: A Review." *Malaysian Journal on Composites Science & Manufacturing* 11, no. 1 (2023): 19-31.
[55] Trovati, Marcello, Khalid Teli, Nikolaos Polatidis, Ufuk Alpsahin Cullen, and Simon Bolton. "Artificial Intuition for Automated Decision-Making." *Applied Artificial Intelligence* 37, no. 1 (2023): 2230749. https://doi.org/10.1080/08839514.2023.2230749
[56] Dahari, Sufinah (2023). G002_SupplementaryData. figshare. Dataset.