國 立 交 通 大 學

統計學研究所

碩 士 論 文

固定長度的眾數區間及其在品質管制上的應用

Fixed width mode interval and its application to quality control.

研 究 生：林怡均

指導教授：陳鄰安　教授

中 華 民 國 九 十 三 年 六 月

固定長度的眾數區間及其在品質管制上的應用

Fixed width mode interval and its application to quality control.

研 究 生：林怡均　　　　Student：Yi Jun Lin

指導教授：陳鄰安 博士　　　　Advisor：Dr. Lin An Chen

國 立 交 通 大 學

統計學研究所

碩 士 論 文

A Thesis

Submitted to Institute Statistics

College of Science

National Chiao Tung University

in partial Fulfillment of the Requirements

for the Degree of

Master

in

Statistics

June 2004

Hsin-chu, Taiwan, Republic of China

中華民國九十三年六月

固定長度眾數區間及其在品質管制上的應用

學生：林怡均　　　　　　　　　　　　　　　　　　　　　指導老師：陳鄰安 教授

國立交通大學統計學研究所 碩士班

摘　　　　要

通用的 Shewhart 管制圖是利用一個統計量 T 的平均數加減 3 倍標準差來做為管制上下限。這套方式應用於所有的分配，不管是對稱或不對稱，以及連續或不連續。我們把 Huang (2003)的眾數區間概念應用於固定長度，但具有最大覆蓋機率的管制圖。對此一區間我們討論了有母數及無母數估計並且做了資料分析。

Fixed width mode interval
and its application to quality control.

student：Yi Jun Lin Advisors：Dr. Lin An Chen

Institute of Statistics
National Chiao Tung University

ABSTRACT

The popular used Shewhart control chart is choosing a statistic T and setting its upper and lower control limits as T's mean plusing and minusing 3 times of T's standard deviation. This rule has been applied for variables with distributions continuous or discrete and symmetric or asymmetric. We extend the mode interval of Huang (2003) to define the fixed width mode interval which is one having largest coverage probability among the intervals with the same width. Estimation of this new chart has been discussed in parametric and nonparametric techniques. Moreover, a real data analysis has also been provided.

誌　　　　謝

時光飛似，轉眼間研究所求學階段即將結束，謝謝師長們的教導以及同學們的陪伴，讓我經歷了兩年愉快且充實的碩士生涯。

首先，我要感謝我的指導老師 陳鄰安教授，謝謝老師在忙碌之餘還能撥冗教導我，不厭其煩的指導我在研究時所遇到的問題，使我能如期完成論文。還要感謝已經畢業的博士班 黃景業學長，謝謝學長犧牲休假時間，回學校教導我、使我釐清學長論文中的觀念及問題。還要謝謝口試委員對我論文的指導及建議。當然還要感謝研究室的同學欣妤、巧慧、淑珍、政輝、志浩、忠庭、文祥、超毅、小強學長，謝謝他們陪我打球運動、度過學習中的低潮，讓我有健康的身心做研究。

此外，還要感謝我的家人以及阿昌給我的支持與鼓勵，讓我能毫無顧慮的求學。最後將此論文獻給我的家人、朋友，以及研究所的同學、學弟妹們。

林怡均　　謹致于

國立交通大學統計研究所

中華民國九十三年六月

Contents

Fixed Width Mode Interval and Its Application to Quality Control

SUMMARY

The popular used Shewhart control chart is choosing a statistic T and setting its upper and lower control limits as T's mean plusing and minusing 3 times of T's standard deviation. This rule has been applied for variables with distributions continuous or discrete and symmetric or asymmetric. We extend the mode interval of Huang (2003) to define the fixed width mode interval which is one having largest coverage probability among the intervals with the same width. Estimation of this new chart has been discussed in parametric and nonparametric techniques. Moreover, a real data analysis has also been provided.

1. Introduction

In statistical applications, we often face two problems of estimating an interval for a random variable or a statistic. In the first problem, we anticipate to find a random interval that covers (usually) a random variable with a given coverage probability. This problem usually is done by the so-called pivotal quantity method. In the second problem, we face the problem of estimating an interval that covers the random variable or statistic in some sense where the two ends of the interval are functions of unknown parameters. There are two main types of this unknown interval. One is setting covering the random variable or statistic with a fixed probability. In this problem, searching one among those with the same coverage probability with shortest width is, in general, a suitable solution. The other one is a certain interval with a fixed width. This interval of the type with fixed width is especially interesting in application in industrial statistics.

The popular way in setting an interval of fixed width is selecting T, a statistic T or random variable, with mean μ_t and standard deviation σ_t to form the symmetric interval $(\mu_t - k\sigma_t, \mu_t + k\sigma_t)$ of width fixed at $2k\sigma_t$ and centered at mean μ_t where the constant k popular is with value 3. We interpret this with two examples applying in statistical quality control. First, the general form of a Shewhart control chart considers the sample mean or sample range for statistic T and defines the two ends of the interval as upper control limit (UCL) and lower control limit (LCL) as

$$\begin{array}{r l} UCL = \mu_t + 3\sigma_t \\ CL = \mu_t \\ LCL = \mu_t - 3\sigma_t \end{array}.$$

where CL represents the central line. Since this control chart may be applied on the manufacturing process no matter what the distribution of the controlling variable is, this interval does not guarantee the coverage probability with a fixed value.

Second, using the width of this interval, process capability index is very popular representing the capability of a manufacturing process. For example, consider a random variable X and its standard deviation σ. The simplest version is defined as

$$C_p = \frac{USL - LSL}{6\sigma}$$

where USL and LSL, respectively, represent the upper and lower specification limits for the random variable that are determined by engineer. In this example, the index uses the 6σ of the interval $(\mu - 3\sigma, \mu + 3\sigma)$.

Consider the problem. The following

$$\{(a, a + 6\sigma_t) : a \in R\}, \tag{1.1}$$

provides the class of whole intervals with the same width $6\sigma_t$, why should we choose the symmetric one? Two criterions may be appropriate setting for making decision in selection. First, we may treat a fixed width interval as an extension of the traditional concept of location for a distribution of a random variable from a point to an interval. We then expect that it should fulfill several desired eqivariant properties for a location parameter. The traditional Shewhart control charts generally do not satisfy some expected equivariant properties where one is that the constructed intervals may be out of the support of the statistic T. For example, suppose that we have a sample mean $\bar{X}$ computed from a random sample $X_1, ..., X_n$ drawn from the distribution $Gamma(2, 3)$ where we have its mean $\mu = 6$ and standard deviation $\sigma = \sqrt{18}$. We then see that the lower control limit LCL is $6 - \frac{12.72}{\sqrt{n}}$ which is less than zero when $n \leq 4$ that makes LCL lie out side the support $(0, \infty)$. Since $n \leq 4$ is the very often case in quality control, we need to avoid this in-practical control limits. Second, for ensuring that the manufacturing process is running in appropriateness, a control chart should have control limits that work well in two aspects. 1. When the process is in control, we expect not to have data points lie outside the control limits which causes conclusion of possible distributional shift. 2. When the process is out of control, we expect to have more observations lie outside of the control limits that we can detect the fact of distributional shift. Searching a quantile interval that fulfill the two criterions above is our aim in this paper.

Extension of the location point of the mode, we define the fixed width interval that maximizes the corresponding coverage probability. From the expectation for being a location interval, we show that it satisfies several desired equivariant properties. Its estimation and application in constructing a new Shewhart $\bar{X}$ control chart are addressed. Finally, nonparametric estimation of this interval has also been discussed.

We define the mode type interval and show that it satisfies several equivariant properties in Section 2. Examples of mode type intervals for several distributions and their corresponding point estimations are displayed in Section 2. The application of this mode interval to the Shewhart $\bar{X}$ control chart is introduced in Section 3. Finally, we introduce a nonparametric estimation for the mode type interval and display several simulation results in Section 4.

2. Population Fixed Width Interval

Suppose that we have a random sample $X_1, ..., X_n$ drawn from distribution with p.d.f. $f(x, \theta)$. Let T be a statistic based on the random sample or simply the random variable X having p.d.f. f. Let σ^2 be the variance of T, which is usually dependent on θ, and we consider the maximum probability interval of width $k\sigma$.

Definition 2.1. A $k\sigma$ fixed width interval is $C(\theta) = (a^*(\theta), k\sigma + a^*(\theta))$ with

$$a^*(\theta) = \text{argsup}_{a \in R} P(a \leq T \leq a + k\sigma).$$

It is well known that the shortest confidence interval with a confidence coefficient may not exist. We then ask if the the $k\sigma$ fixed width interval which is one of shortest interval exist? The solution provides one reason that it is worthful to be proposed.

Theorem 2.2. If a random variable has finite variance σ^2, then, for $k > 0$, the $k\sigma$ fixed width interval exists.

Proof. Let's denote $P_a = P(a \leq T \leq a + k\sigma)$. Note that the set $\{P_a : a \in R\}$ is bounded so that its supremum, denotes it by P^*, exists. Let $\{a_n : n = 1, 2.3, ...\}$ be a set such that $p^* = \lim_{i\to\infty} p_{a_i}$. Since p.d.f. f satisfies $f(x) \to 0$ as $x \to \infty$ or $-\infty$, there exists a value a_0 such that $-x_0 \leq a_n \leq x_0$, for all n. Boundness of $\{a_n : n = 1, 2, 3, ...\}$ implies that there is a sequence n_i such that there is $a' = \lim_{i\to\infty} a_{n_i}$.

Suppose that the distribution is continuous. Then obviously a' is a solution of a^*. On the other hand, if the distribution is discrete. Then, from the fact that P_a is a step function in a, a^* has to be the limit of the sequence $\{a_{n_i}\}$.

Theorem 2.3. Suppose that the distribution F is symmetric at a value μ, then the $k\sigma$ fixed width interval is of the form

$$(\mu - \frac{k}{2}\sigma, \mu + \frac{k}{2}\sigma).$$

Proof. Consider only that the distribution F is also continuous. Then solution of a satisfies

$$0 = \frac{\partial}{\partial a}(F(a + k\sigma) - F(a)) = f(a + k\sigma) - f(a). \tag{2.1}$$

However, $f(a + k\sigma) = f(a)$ if and only if $a = \mu - \frac{k}{2}\sigma$ for symmetric p.d.f. f.

We extend the concept of measure of location point (see, for example, Staudte and Sheather (1990, p101)) to the location type interval.

Definition 2.4. A measure of fixed width coverage set for F is a set $D(X)$ that satisfies the following conditions:
(1). $D(X + b) = D(X) + b$ for $b \in R$.
(2). $D(aX) = aD(X)$ for $a \in R$.
(3). $X \geq 0$ implies $D(X) \geq 0$.

Not every type of parameterized interval fulfills the properties of measure of fixed width coverage set. Here we show that the $k\sigma$ fixed width interval does satisfies the conditions of a measure of coverage set.

Theorem 2.5. The $k\sigma$ fixed width interval is a measure of coverage set.
Proof. For convenience, redenote a^*, σ and $C(k\sigma)$ for random variable X, respectively, by $a^*(X)$, σ_X and $C(X)$. To show (1), for $c \in R$,

$$a^*(X + c) = \text{argsup}_{a\in R} P(a \leq X + c \leq a + k\sigma_{X+c})$$
$$= \text{argsup}_{a\in R} P(a - c \leq X \leq a + k\sigma_X - c)$$

where we use the fact that $\sigma_{X+c} = \sigma_X$. We have $a^*(X+c) - c = a^*(X)$ which implies $a^*(X+c) = a^*(X) + c$. Then

$$C(X+c) = (a^*(X+c), a^*(X+c)+k\sigma_{X+c}) = (a^*(X)+c, a^*(X)+c+k\sigma_X) = C(X)+c.$$

Consider (2). If $b > 0$,

$$\begin{aligned}a^*(bX) &= \text{argsup}_{a\in R} P(a \le bX \le a + k\sigma_{bX}) \\ &= \text{argsup}_{a\in R} P(\frac{a}{b} \le X \le \frac{a + kb\sigma_X}{b})\end{aligned}$$

where we use the fact that $\sigma_{bX} = b\sigma_X$. We have $\frac{a^*(bX)}{b} = a^*(X)$ which implies $a^*(bX) = ba^*(X)$. Then

$$C(bX) = (a^*(bX), a^*(bX) + k\sigma_{bX}) = b(a^*(X), a^*(X) + k\sigma_X) = bC(X).$$

If $b < 0$,

$$\begin{aligned}a^*(bX) &= \text{argsup}_{a\in R} P(a \le bX \le a + k\sigma_{bX}) \\ &= \text{argsup}_{a\in R} P(\frac{a - bk\sigma_X}{b} \le X \le \frac{a}{b})\end{aligned}$$

where we use the fact that $\sigma_{bX} = -b\sigma_X$. We have $a^*(X) = \frac{a^*(bX) - bk\sigma_X}{b}$ which implies $a^*(bX) = ba^*(X) + bk\sigma_X$. Then

$$\begin{aligned}C(bX) &= (a^*(bX), a^*(bX) + k\sigma_{bX}) \\ &= (ba^*(X) + bk\sigma_X, ba^*(X) + bk\sigma_X - bk\sigma_X) \\ &= b(a^*(X), a^*(X) + k\sigma_X) \\ &= bC(X).\end{aligned}$$

Condition (3) is induced by the fact that, for $X \ge 0$, we have $P(a_0 \le X \le a_0 + k\sigma) \le P(0 \le X \le k\sigma)$ if $a_0 < 0$.

Table 1 Median and mode types interval for binomial distribution (b(12,0.3))

Length	π_{med}	C_{med}	π_{mod}	C_{mod}
0	0.2311	$\{4\}$	0.2397	$\{3\}$
1	0.4708	$\{3,4\}$	0.4708	$\{3,4\}$
2	0.6293	$\{3,4,5\}$	0.6386	$\{2,3,4\}$
3	0.7971	$\{2,3,4,5\}$	0.7971	$\{2,3,4,5\}$
4	0.8763	$\{2,...,6\}$	0.8763	$\{2,...,6\}$
5	0.9475	$\{1,...,6\}$	0.9475	$\{1,...,6\}$
6	0.9766	$\{1,...,7\}$	0.9766	$\{1,...,7\}$
7	0.9905	$\{0,...,7\}$	0.9905	$\{0,...,7\}$
8	0.9983	$\{0,...,8\}$	0.9983	$\{0,...,8\}$
9	0.9997	$\{0,...,9\}$	0.9997	$\{0,...,9\}$
10	0.9999	$\{0,...,10\}$	0.9999	$\{0,...,10\}$
11	0.99999	$\{0,...,11\}$	0.99999	$\{0,...,11\}$
12	1.0000	$\{0,...,12\}$	1.0000	$\{0,...,12\}$

We call an interval C_0 a highest density (HD) interval if

$$f(x) \geq f(x_1) \text{ for } x \in C_0, x_1 \notin C_0. \tag{2.2}$$

Theorem 2.6. Suppose that the underlying distribution is unimodal and continuous. Then a quantile interval $C(\theta)$ is a $k\sigma$ mode interval if and only if it is a width $k\sigma$ HD interval.

Proof. Let C_0 be a width $k\sigma$ HD interval. By the fact that $C = (C \cap C_0) \cup (C \cap C_0^c)$ and $C_0 = (C_0 \cap C) \cup (C_0 \cap C^c)$. Since C and C_0 are both width $k\sigma$ interval, we have

$$\text{width}(C \cap C_0^c) = \text{width}(C_0 \cap C^c).$$

From (2.2), we also have

$$\int_{C_0 \cap C^c} f(x)dx \geq \int_{C \cap C_0^c} f(x)dx.$$

Adding $\int_{C \cap C_0} f(x)dx$ to both sides, we further have

$$\int_{C_0} f(x)dx \geq \int_C f(x)dx. \tag{2.3}$$

However, from the definition of mode interval, strictly inequality in (2.3) can not hold. Then C_0 is a width $k\sigma$ mode interval.

On the other hand, let $C = (a, a + k\sigma)$ be a $k\sigma$ mode interval. From (2.1), we have

$$f(a) = f(a + k\sigma). \tag{2.4}$$

With (2.4) and the fact that the mode lying in the mode interval, interval C satisfies (2.2). Then C is a width $k\sigma$ HD interval.

3. Statistical Inferences for Fixed Width Mode Interval

Let a fixed width interval be of the form $(a_1(\theta), a_2(\theta))$. We also assume that there is a random sample $X_1, ..., X_N$ available. Consider $\hat{\theta}$ as an estimator of θ. In this section, we will develop statistical inference procedures for the fixed width interval when the interval is a function of unknown parameter θ and a random sample from the underlying distribution is available.

The simplicity of the expression of the fixed width mode interval determines the estimation technique. Suppose that a distribution F_θ makes the mode interval in the form

$$(a_1(\theta)c_0, a_1(\theta)c_0 + k\sigma) \tag{3.1}$$

where c_0 is the only fact that is determined by the maximization problem in Definition 2.1 in terms of a distribution F_0 which is free of parameter θ. Then as long as we have estimators $\hat{\theta}$ and $\hat{\sigma}$ respectively for θ and σ the estimator of the mode interval may be set as $\hat{C} = (a_1(\hat{\theta})c_0, a_1(\hat{\theta})c_0 + k\hat{\sigma})$. The distribution belonging to the family of location-scale family is the one with this advantage.

In the following, we presents the fixed width mode interval for several distributions that belong to the location-scale family.

Theorem 3.1. Let X be a random variable with distribution in the family of continuous location-scale distributions with p.d.f. of the form $f(x, \theta_1, \theta_2) = \frac{1}{\theta_2} f_0(\frac{x-\theta_1}{\theta_2})$ with parameter space $\theta_1 \in R$ and $\theta_2 > 0$ has $k\sigma$ mode interval of the type

$$(a^*, a^* + \frac{k\sigma}{\theta_2})$$

where

$$a^* = \text{argsup}_{a \in R} P(a \leq X_0 \leq a + \frac{k\sigma}{\theta_2}).$$

and $X_0 = \frac{X-\theta_1}{\theta_2}$.

Proof. The proof is obvious from the fact that

$$P(a \le X \le a + k\sigma) = \int_a^{a+k\sigma} \frac{1}{\theta_2} f_0(\frac{x-\theta_1}{\theta_2})dx$$
$$= \int_{\frac{a-\theta_1}{\theta_2}}^{\frac{a-\theta_1}{\theta_2}+\frac{k\sigma}{\theta_2}} f_0(y)dy.$$

The benifit of location-scale family is that the mode interval is explicitly displayed in terms of α^* and parameter θ and then we may easily develop the estimator of coverage interval through the existed theorems for statistical inferences for parameter θ.

Exponential distribution: Consider the case that the random sample is drawn from a right skewed exponential distribution with p.d.f.

$$f(x) = \frac{1}{\theta} e^{-\frac{x-\ell}{\theta}} I(\ell < x < \infty).$$

The $k\sigma$ fixed width interval is

$$C(\theta) = (\ell, \ell + k\theta).$$

Proof. For this distribution, we may see that standard deviation is $\sigma = \theta$ and $P(a < X < a + k\sigma) = e^{-\frac{a-\ell}{\theta}}(1 - e^{-k\sigma/\theta})$. Then the result is implied from the facts that $1 - e^{-k\sigma/\theta} > 0$ and $e^{-\frac{a-\ell}{\theta}}$ is a decreasing function of a on $[\ell, \infty)$. □

Table 2 Comparison of coverage probabilities of median and mode types intervals for the Exponential distribution with the coverage probabilities for normal distribution

k	π_{med}	π_{mod}	π_{med}	π_{mod}	π_{nor}
	$\lambda = 0.3$		$\lambda = 3$		Normal
1.0	0.419	0.644	0.417	0.643	0.382
2.0	0.830	0.865	0.830	0.865	0.682
3.0	0.924	0.954	0.923	0.954	0.866
4.0	0.956	0.988	0.956	0.988	0.954
5.0	0.977	0.999	0.977	0.999	0.987
6.0	0.989	1.000	0.989	1.000	0.997
7.0	0.996	1.000	0.996	1.000	0.999

For the point estimation, in case that we have a random sample $X_1, ..., X_n$ drawn from this exponential distribution, we may consider $\hat{C} = (\ell, \ell + k\bar{X})$ since $\bar{X}$ is the best estimator of θ.

On the other hand, one distribution highly asymmetric skewed to the left that has p.d.f. of the form

$$f(x) = \frac{1}{\theta} e^{\frac{x-\ell}{\theta}} I(-\infty < x < \ell).$$

We may also see that the $k\sigma$ fixed width interval is

$$C(\theta) = (\ell - k\theta, \ell)$$

where its estimator may be set as $\hat{C} = (\ell - k\bar{X}, \ell)$.

Gamma distribution: Suppose that X has distribution $Gamma(\frac{\ell}{2}, \theta)$. Since $E(X) = \frac{\ell\theta}{2}$ and $Var(X) = \frac{\ell\theta^2}{2}$, then the $k\sigma$ fixed width interval is

$$C_{med} = (\frac{\ell\theta}{2} - \frac{k}{2}\theta\sqrt{\frac{\ell}{2}}, \frac{\ell\theta}{2} + \frac{k}{2}\theta\sqrt{\frac{\ell}{2}})$$

and the mode type $2k\sigma$ fixed width interval is

$$C_{mod}(\theta) = (\frac{a_0^*}{2}\theta, \frac{a_0^*}{2}\theta + \frac{\theta k}{2}\sqrt{2\ell})$$

where a_0^* solves $\sup_{a \in R} P(a \leq \chi^2(\ell) \leq a + k\sqrt{2\ell})$.
Proof of mode type interval: Since $\sigma = \sqrt{\frac{\ell}{2}}\theta$, the $k\sigma$ fixed width interval is $C(\theta) = (a^*, a^* + k\sqrt{\frac{\ell}{2}}\theta)$ where a^* satisfies the followings

$$\begin{aligned} a^* &= \text{argsup}_{a \in R} P(a \leq X \leq a + k\sqrt{\frac{\ell}{2}}\theta) \\ &= \text{argsup}_{a \in R} P(\frac{2a}{\theta} \leq \frac{2X}{\theta} \leq \frac{2a}{\theta} + k\sqrt{2\ell}). \end{aligned}$$

Since $\frac{2X}{\theta}$ has χ^2 distribution with degrees of freedom ℓ, the theorem is followed. □

The coverage probabilities of the median type and mode type $2k\sigma$ fixed width interval are, respectively,

$$\begin{aligned} \pi_{med} &= P(\frac{\ell\theta}{2} - \frac{k}{2}\theta\sqrt{\frac{\ell}{2}} \leq X \leq \frac{\ell\theta}{2} + \frac{k}{2}\theta\sqrt{\frac{\ell}{2}}) \\ &= P(\ell - k\sqrt{\frac{\ell}{2}} \leq \chi^2(\ell) \leq \ell + k\sqrt{\frac{\ell}{2}}) \end{aligned}$$

and

$$\pi_{mod} = P(a_0^* \leq \chi^2(\ell) \leq a_0^* + k\sqrt{2\ell}).$$

Table 3 Comparison of coverage probabilities of median and mode types intervals for the Gamma distribution with the coverage probabilities for normal distribution

k	π_{med}	π_{mod}	π_{med}	π_{mod}	π_{nor}
	$\alpha=3$ $\beta=2$		$\alpha=9$ $\beta=2$		Normal
1.0	0.401	0.520	0.404	0.489	0.382
2.0	0.735	0.793	0.723	0.771	0.682
3.0	0.909	0.927	0.897	0.918	0.866
4.0	0.958	0.979	0.961	0.975	0.954
5.0	0.978	0.994	0.981	0.991	0.987
6.0	0.989	0.997	0.988	0.998	0.997
7.0	0.994	0.999	0.992	0.999	0.999

Let's turn to the situation that a distribution does not belong to the location-scale family. It is then quite often that the fixed $k\sigma$ mode interval may ne be formulated in the convenient form of (3.1). In this situation, we propose to estimate it as $\hat{C} = C(\hat{\theta})$ simply replacing the distribution F_θ of X by $F_{\hat{\theta}}$ where $\hat{\theta}$ is a suitable estimator of θ. In the following, we present a case of underlying Poisson distribution.

Poisson distribution: Let X be a random variable with Poisson distribution having p.d.f. of the form

$$f(x,\lambda) = \frac{\lambda^x e^{-\lambda}}{x!} I(x=0,1,2,...).$$

The variance of this distribution is λ so that the $k\sigma$ mode interval is $C(\lambda) = (a(\lambda), a(\lambda)+k\sqrt{\lambda})$ with

$$a(\lambda) = \text{argsup}_{a\geq 0} \sum_{x=a}^{a+k\sqrt{\lambda}} \frac{\lambda^x e^{-\lambda}}{x!}.$$

When λ is unknown, we may estimate $C(\lambda)$ by $\hat{C} = C(\bar{X})$ where $\bar{X}$ is the sample mean of the available random sample $X_1, ..., X_n$.

4. Mode Interval Type Control Chart

Let T be a statistic based on a random sample from a distribution F with population mean μ_t and population variance σ_t^2. The general theory of Shewhart control chart is considering the mean μ_t as the central line and the two lines with distance $2k\sigma_t$ for some $k > 0$ as the upper and lower control limits as

$$UCL = \mu_t + k\sigma_t$$
$$CL = \mu_t$$
$$LCL = \mu_t - k\sigma_t$$

In practice, the most popular setting of value k is 3. Basically, the Shewhart control is an interval with fixed width $2k\sigma_t$. Inheriting the requirement of the fixed width $2k\sigma_t$, it is reasonable to consider the limits of the fixed width $6\sigma_t$ mode interval as the control limits, that is, we define the general form of mode interval type fixed width Shewhart control limits as

$$UCL_{mod} = 2k\sigma_t + a^*(\theta)$$
$$LCL_{mod} = a^*(\theta)$$

In practice of quality control, usually we assume that there is a history record of m samples $X_{j1}, X_{j2}, ..., X_{jn}, j = 1, ..., m$ drawn from a distribution F available to construct the control limits. Let $\hat{\mu}_{tj}, \hat{\sigma}_{tj}$ and $\hat{a}_j^*(\theta)$ are, based on jth random sample $X_{j1}, X_{j2}, ..., X_{jn}$, estimators of μ_t, σ_t and $a^*(\theta)$. By letting $\hat{\mu}_t = \frac{1}{m}\sum_{j=1}^m \hat{\mu}_{tj}, \hat{\sigma}_t = \frac{1}{m}\sum_{j=1}^m \hat{\sigma}_{tj}$ and $\hat{a}^*(\theta) = \frac{1}{m}\sum_{j=1}^m \hat{a}_j^*(\theta)$, the estimated Shewhart control chart and mode interval type Shewhart control chart are, respectively, with limits

$$UCL = \hat{\mu}_t + k\hat{\sigma}_t$$
$$CL = \hat{\mu}_t$$
$$LCL = \hat{\mu}_t - k\hat{\sigma}_t$$

and

$$UCL_{mod} = 2k\hat{\sigma}_t + \hat{a}^*(\theta)$$
$$LCL_{mod} = \hat{a}^*(\theta)$$

In application of this new control chart for on-line process surveillance, if the sample values of statistic T fall within the control limits LCL_{mod} and UCL_{mod} and do not exhibit any systematic pattern, we say the process is in control at the level indicated by the chart.

Exponential Distribution

Let $X_1, ..., X_n$ be a random sample drawn from the exponential distribution with p.d.f.

$$f(x,\theta) = \frac{1}{\theta}e^{-\frac{x}{\theta}}I(x>0).$$

Consider the $\bar{X}$-chart. The Shewhart control $\bar{X}$-chart is

$$\begin{aligned} UCL_{med} &= \theta + k\frac{\theta}{\sqrt{n}} \\ CL_{med} &= \theta \\ LCL_{med} &= \theta - k\frac{\theta}{\sqrt{n}} \end{aligned}$$

By letting $\bar{X}_j = \frac{1}{n}\sum_{i=1}^{n} X_{ji}$ and $\bar{\bar{X}} = \frac{1}{m}\sum_{j=1}^{m}\bar{X}_j$, the estimated Shewhart control $\bar{X}$-chart is

$$\begin{aligned} UCL_{med} &= \bar{\bar{X}} + k\frac{\bar{\bar{X}}}{\sqrt{n}} \\ CL_{med} &= \bar{\bar{X}} \\ LCL_{med} &= \bar{\bar{X}} - k\frac{\bar{\bar{X}}}{\sqrt{n}} \end{aligned}$$

On the other hand, the coverage probability of the Shewhart control $\bar{X}$-chart may be derived in the following

$$\begin{aligned} \pi_{med} &= P(\theta - k\frac{\theta}{\sqrt{n}} \le \bar{X} \le \theta + k\frac{\theta}{\sqrt{n}}) \\ &= P(n(\theta - k\frac{\theta}{\sqrt{n}}) \le \sum_{i=1}^{n} X_i \le n(\theta + k\frac{\theta}{\sqrt{n}})) \\ &= P(n(1 - \frac{k}{\sqrt{n}}) \le Y \le n(1 + \frac{k}{\sqrt{n}})) \end{aligned}$$

where $Y \sim Gamma(n,1)$.

The mode type Shewhart control $\bar{X}$-chart is

$$\begin{aligned} UCL_{mod} &= 2k\frac{\theta}{\sqrt{n}} + a^*(\theta) \\ LCL_{mod} &= a^*(\theta) \end{aligned}$$

where $a^*(\theta)$ may be derived in the following

$$a^*(\theta) = \text{argsup}_a P(a \le \bar{X} \le 2k\frac{\theta}{\sqrt{n}} + a)$$
$$= \text{argsup}_a P(n\frac{a}{\theta} \le \frac{\sum_{i=1}^n X_i}{\theta} \le n(\frac{2k}{\sqrt{n}} + \frac{a}{\theta})).$$

By letting $b^* = \text{argsup}_b P(nb \le Y \le n(\frac{2k}{\sqrt{n}} + b))$, we have $a^* = \theta b^*$. The alternative form of the mode type Shewhart control $\bar{X}$-chart is

$$UCL_{mod} = 2k\frac{\theta}{\sqrt{n}} + \theta b^*$$
$$LCL_{mod} = \theta b^*$$

Table 4 Comparison of coverage probabilities of median and mode types control limits for the Exponential distribution

k	π_{med}	π_{mod}	π_{med}	π_{mod}
	$n = 3$		$n = 5$	
1.0	0.382	0.441	0.382	0.415
2.0	0.715	0.751	0.700	0.722
3.0	0.909	0.909	0.892	0.893
4.0	0.955	0.972	0.958	0.966
5.0	0.976	0.992	0.980	0.990
6.0	0.988	0.998	0.990	0.997
7.0	0.994	0.999	0.995	0.999

The estimated mode type Shewhart control $\bar{X}$-chart is

$$UCL_{mod} = 2k\frac{\bar{\bar{X}}}{\sqrt{n}} + \bar{\bar{X}}\, b^*$$
$$LCL_{mod} = \bar{\bar{X}}\, b^*$$

The coverage probability of the mode type Shewhart control $\bar{X}$-chart is

$$\pi_{mod} = P(nb^* \le Y \le n(\frac{2k}{\sqrt{n}} + b^*)).$$

Gamma distribution Let $X_1, ..., X_n$ be a random sample drawn from the distribution $Gamma(\frac{\ell}{2}, \theta)$. Consider the Shewhart $\bar{X}$ control chart. Since $\bar{X}$ has mean $\frac{\ell\theta}{2}$ and variance $\frac{\ell\theta^2}{2n}$, the median type Shewhart $\bar{X}$ control chart is

$$UCL_{med} = \frac{\ell\theta}{2} + k\theta\sqrt{\frac{\ell}{2n}}$$
$$CL_{med} = \frac{\ell\theta}{2}$$
$$LCL_{med} = \frac{\ell\theta}{2} - k\theta\sqrt{\frac{\ell}{2n}}$$

The sample type Shewhart $\bar{X}$ control chart is

$$UCL_{med} = \bar{\bar{X}} + k\,\bar{\bar{X}}\sqrt{\frac{\ell}{2n}}$$
$$CL_{med} = \bar{\bar{X}}$$
$$LCL_{med} = \bar{\bar{X}} - k\,\bar{\bar{X}}\sqrt{\frac{\ell}{2n}}$$

The coverage probability of the median type Shewhart $\bar{X}$ control chart is

$$\pi_{med} = P(\frac{\ell\theta}{2} - k\theta\sqrt{\frac{\ell}{2n}} \le \bar{X} \le \frac{\ell\theta}{2} + k\theta\sqrt{\frac{\ell}{2n}})$$
$$= P(n(\frac{\ell}{2} - k\sqrt{\frac{\ell}{2n}}) \le Y \le n(\frac{\ell}{2} + k\sqrt{\frac{\ell}{2n}})).$$

For deriving the mode type $\bar{X}$ control chart, by letting $Y = \frac{\sum_{i=1}^{n} X_i}{\theta}$, and setting

$$a^*(\theta) = \text{argsup}_a P(a \le \bar{X} \le a + 2k\theta\sqrt{\frac{\ell}{2n}})$$
$$b^* = \text{argsup}_b P(b \le Y \le b + 2k\sqrt{\frac{\ell n}{2}}),$$

we have $a^* = \frac{\theta b^*}{n}$. The mode type Shewhart $\bar{X}$ control chart is

$$UCL_{mod} = k\theta\sqrt{\frac{2\ell}{n}} + \frac{\theta b^*}{n}$$
$$LCL_{mod} = \frac{\theta b^*}{n}.$$

Table 5 Comparison of coverage probabilities of median and mode types control limits for the Gamma distribution $Gamma(2, 2)$

k	π_{med}	π_{mod}	π_{med}	π_{mod}
	$n=3$		$n=5$	
1.0	0.382	0.409	0.382	0.398
2.0	0.697	0.715	0.691	0.702
3.0	0.888	0.888	0.879	0.880
4.0	0.959	0.964	0.958	0.960
5.0	0.981	0.990	0.983	0.989
6.0	0.991	0.997	0.993	0.997
7.0	0.996	0.999	0.997	0.999

The sample mode type Shewhart $\bar{X}$ control chart is

$$UCL_{mod} = 2k\bar{\bar{X}}\sqrt{\frac{2}{n\ell}} + \frac{2\bar{\bar{X}}b^*}{n\ell}$$

$$LCL_{mod} = \frac{2\bar{\bar{X}}b^*}{n\ell}.$$

The coverage probability of the mode type Shewhart $\bar{X}$ control chart is

$$\pi_{mod} = P(b^* \le Y \le b^* + 2k\sqrt{\frac{n\ell}{2}}).$$

Poisson distribution Let $X_1, ..., X_n$ be a random sample drawn from the distribution $Poisson(\lambda)$. Consider also the $\bar{X}$ control chart. Since $E(\bar{X}) = \lambda$ and $Var(\bar{X}) = \frac{\lambda}{n}$. the median type Shewhart control chart is

$$UCL_{med} = \lambda + k\sqrt{\frac{\lambda}{n}}$$

$$CL_{med} = \lambda$$

$$LCL_{med} = \lambda - k\sqrt{\frac{\lambda}{n}}$$

The sample type Shewhart control chart is

$$UCL_{med} = \bar{\bar{X}} + k\sqrt{\frac{\bar{\bar{X}}}{n}}$$

$$CL_{med} = \bar{\bar{X}}$$

$$LCL_{med} = \bar{\bar{X}} - k\sqrt{\frac{\bar{\bar{X}}}{n}}$$

The coverage probability of the median type Shewhart $\bar{X}$ control chart is

$$\pi_{med} = P(\lambda - k\sqrt{\frac{\lambda}{n}} \leq \bar{X} \leq \lambda + k\sqrt{\frac{\lambda}{n}})$$
$$= P(n(\lambda - k\sqrt{\frac{\lambda}{n}}) \leq Y \leq n(\lambda + k\sqrt{\frac{\lambda}{n}}))$$

where we let $Y = \sum_{i=1}^{n} X_i \sim Poisson(n\lambda)$.

The mode type Shewhart $\bar{X}$ control chart is

$$UCL_{mod} = a^* + 2k\sqrt{\frac{\lambda}{n}}$$
$$LCL_{mod} = a^*$$

where a^* satisfies

$$a^* = \text{argsup}_a P(a \leq \bar{X} \leq a + 2k\sqrt{\frac{\lambda}{n}}$$
$$= \text{argsup}_a P(na \leq Y \leq n(a + 2k\sqrt{\frac{\lambda}{n}})).$$

The sample type mode type Shewhart $\bar{X}$ control chart is

$$UCL_{mod} = a^*(\bar{\bar{X}}) + 2k\sqrt{\frac{\bar{\bar{X}}}{n}}$$
$$LCL_{mod} = a^*(\bar{\bar{X}})$$

where $a^*(\bar{\bar{X}})$ satisfies

$$a^*(\bar{\bar{X}}) = \text{argsup}_a P(na \leq Y \leq n(a + 2k\sqrt{\frac{\bar{\bar{X}}}{n}})).$$

We here display a comparison of coverage probabilities of median type and mode type $\bar{X}$ charts under the Poisson distribution with sample size $n = 3$

Table 6 Comparison of coverage probabilities of median and mode types control limits for distribution $Poisson(0.5)$ (n=3)

k	π_{med}	$(LCL, UCL)_{med}$	π_{mod}	$(LCL, UCL)_{mod}$
1.0	0.585	$(0.3, 0.704)$	0.585	$(0.3, 0.708)$
2.0	0.585	$(0.092, 0.908)$	0.808	$(0.0, 0.816)$
3.0	0.934	$(0.0, 1.112)$	0.934	$(0.0, 1.225)$
4.0	0.934	$(0.0, 1.316)$	0.981	$(0.0, 1.633)$
5.0	0.981	$(0.0, 1.521)$	0.999	$(0.0, 2.041)$
6.0	0.995	$(0.0, 1.725)$	0.999	$(0.0, 2.449)$
7.0	0.995	$(0.0, 1.929)$	0.999	$(0.0, 2.858)$

5. Numerical Data Analysis

From the historical records (see the data in Besterfield (1990)), there is a data of 25 subgroups that gives the inspection results for the blower motor in an electric hair dryer for the motor department. In this case, we have a random sample $X_1, ..., X_m$ drawn from the binomial distribution $b(n, p)$ with unknown parameter p and $n = 300$ and $m = 25$. By letting $\bar{p} = \frac{1}{n}\sum_{j=1}^{m} \hat{p}_j$ with $\hat{p}_j = \frac{X_j}{n}, j = 1, ..., m$, the Shewhart X control chart is

$$
\begin{aligned}
UCL_{med} &= n\bar{p} + k\sqrt{n\bar{p}(1-\bar{p})} \\
CL_{med} &= n\bar{p} \\
LCL_{med} &= n\bar{p} - k\sqrt{n\bar{p}(1-\bar{p})}
\end{aligned}
$$

and by letting $\hat{X} \sim b(n, \bar{p})$ and

$$a^*(\bar{p}) = \text{argsup}_{a \geq 0} P(a \leq \hat{X} \leq a + 2k\sqrt{n\bar{p}(1-\bar{p})}),$$

the mode type Shewhart X control chart is

$$
\begin{aligned}
UCL_{mod} &= a^*(\bar{p}) + 2k\sqrt{n\bar{p}(1-\bar{p})} \\
LCL_{mod} &= a^*(\bar{p}).
\end{aligned}
$$

Table 7 Median and mode control chart for binomial distribution

k	$(LCL, UCL)_{med}$	π_{med}	$(LCL, UCL)_{mod}$	π_{mod}
2.0	$0.794, 10.00$	0.9742	$1.00, 10.0$	0.9742
2.2	$0.333, 10.46$	0.9742	$1.00, 11.0$	0.9967
2.4	$-0.126, 10.92$	0.9785	$1.00, 11.0$	0.9867
2.5	$-0.126, 10.92$	0.9910	$1.00, 12.0$	0.9922
2.6	$-0.587, 11.38$	0.9910	$1.00, 12.0$	0.9922
2.8	$-1.047, 11.84$	0.9910	$1.00, 12.0$	0.9922
3.0	$-1.508, 12.30$	0.9965	$0.00, 13.0$	0.9987
3.5	$-2.659, 13.45$	0.9987	$0.00, 16.0$	0.9999

We display a graph of the estimated control chart in Figure 1.

In statistical quality control, the number of defects, c, arises probabilitically according to the Poisson distribution. Suppose that we have a random sample $c_1, ..., c_m$

obeying this distribution. By denoting $\bar{c} = \frac{1}{m}\sum_{i=1}^{m} c_i$, the Shewhart c control chart has control limits

$$UCL_{med} = \bar{c} + k\sqrt{\bar{c}}$$
$$CL_{med} = \bar{c}$$
$$LCL_{med} = \bar{c} - k\sqrt{\bar{c}}.$$

On the other hand, the mode type Shewhart c control chart has control limits as

$$UCL_{mod} = a^*(\bar{c}) + 2k\sqrt{\bar{c}}$$
$$LCL_{mod} = a^*(\bar{c})$$

where $a^*(\bar{c}) = \text{argsup}_{a\geq 0} P(a \leq \hat{c} \leq a + 2k\sqrt{\bar{c}})$ where $\hat{c} \sim Poisson(\sqrt{\bar{c}})$.

Consider the process of the installation of front and rear bumpers on automobiles. In this automobile manufacturer, all autos were inspected for the bumper installation process. The numbers $c_1, ..., c_m$ of defects were recorded from shift to shift. For detail description of the data and defects, please see Devor, Chang, and Sutherland (1992). From the data of $m = 25$ samples, the mean of defects $\bar{c} = 16$. We list the two computed control limits in the following table.

Table 8 Median and mode control chart for bumper infects

k	$(LCL, UCL)_{med}$	π_{med}	$(LCL, UCL)_{mod}$	π_{mod}
2.0	$8.0, 24.0$	0.9677	$8.0, 24.0$	0.9677
2.2	$7.2, 24.8$	0.9677	$8.0, 25.0$	0.9769
2.4	$6.4, 25.6$	0.9829	$7.0, 26.0$	0.9885
2.5	$6.0, 26.0$	0.9912	$7.0, 27.0$	0.9919
2.6	$5.6, 26.4$	0.9912	$7.0, 27.0$	0.9919
2.8	$4.8, 27.2$	0.9955	$6.0, 28.0$	0.9964
3.0	$4.0, 28.0$	0.9977	$5.0, 29.0$	0.9985
3.5	$2.0, 30.0$	0.9994	$4.0, 32.0$	0.9998

We also display an estimated control chart in Figure 2.

6. Nonparametric Estimation

In the previous work in this paper, the observations were assumed to come from some underlying distribution, whose general form is assumed known. If these assumptions about the shape of the distribution are not made, then nonparametric methods to estimate the fixed width mode interval $C_{mod} = (a^*, k\sigma + a^*)$ must be used. Besides the nonparametric estimation of mode interval C_{mod}, we also simulate the efficiency of the mode interval C_{mod} comparing with median type interval. As we have defined earlier, the median type interval is $C_{med} = (\mu - \frac{k}{2}\sigma, \mu + \frac{k}{2}\sigma)$. For further study, we here also consider another type of median type interval as $C_{med2} = (F^{-1}(\alpha_0), k\sigma + F^{-1}(\alpha_0))$ with $F^{-1}(\alpha_0)$ satisfying $P(X \le F^{-1}(\alpha_0)) = P(X \ge k\sigma + F^{-1}(\alpha_0))$.

Let $X_1, ..., X_n$ be a random sample from a distribution F, we let $\bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$ and $S = (\frac{1}{n-1}\sum_{i=1}^{n}(X_i - \bar{X})^2)^{1/2}$, the, respectively, sample mean and sample standard deviation, and $X_{(1)}, X_{(2)}, ..., X_{(n)}$ be the corresponding order statistics. Let

$$n^* = \max\{n_i = \text{ number of observations in } [X_{(i)}, X_{(i)} + kS], i = 1, ..., n\}$$

and let i^* be the index i such that number of observations in $[X_{(i)}, X_{(i)} + kS]$ is equal to n^*. We then define the estimate of the $k\sigma$ mode interval by

$$\hat{C}_{mod} = [X_{(i^*)}, X_{(i^*)} + kS].$$

The corresponding coverage percentage estimate is $\hat{\pi}_{mod} = \frac{n^*}{n}$. The ordinary Shewhart control chart is $\hat{C}_{med} = (\hat{\mu} - \frac{k}{2}\hat{\sigma}, \hat{\mu} + \frac{k}{2}\hat{\sigma})$ where its corresponding coverage percentage is $\hat{\pi}_{med} = \frac{\text{number of observations in } \hat{C}_{med}}{n}$.

We consider a simulation with replication $m = 10,000$. For each replication, we draw a random sample $X_1, ..., X_n$ of sample size $n = 50$ from a distribution F. Besides the estimated coverage probabilities, we also setting the following vector type mean squares errors (MSE):

$$MSE_{med} = \{\frac{1}{m}\sum_{j=1}^{m}(\hat{\mu}_j - \frac{k}{2}\hat{\sigma}_j - (\mu - \frac{k}{2}\sigma))^2, \frac{1}{m}\sum_{j=1}^{m}(\hat{\mu}_j + \frac{k}{2}\hat{\sigma}_j - (\mu + \frac{k}{2}\sigma))^2\}$$

$$MSE_{mod} = \{\frac{1}{m}\sum_{j=1}^{m}(X_{(i^*)j} - a^*)^2, \frac{1}{m}\sum_{j=1}^{m}(X_{(i^*)j} + kS_j - (k\sigma + a^*))^2\}.$$

The first we consider the exponential distribution with pdf $f(x, \lambda) = \lambda e^{-\lambda x}, x > 0$ and we display both the MSE's and the estimated coverage probabilities.

Table 9 MSE's and coverage percentages for Exponential distribution

λ	MSE_{med}	$\hat{\pi}_{med}$	MSE_{mod}	$\hat{\pi}_{mod}$
0.1	$0.0799, 87.755$	0.8955	$0.0799, 48.384$	0.9544
0.3	$0.0132, 10.594$	0.8957	$0.0088, 5.3760$	0.9542
0.5	$0.0047, 3.8140$	0.8950	$0.0031, 1.9353$	0.9543
0.7	$0.0024, 1.9461$	0.8949	$0.0016, 0.9874$	0.9542
0.9	$0.0014, 1.1771$	0.8956	$0.0009, 0.5973$	0.9544

We have two conclusions:

(a). The MSE's of C_{mod} for estimation of its two ends are uniformly smaller or equal to the corresponding MSE's of C_{med}. This indicates that the location of mode interval is relatively easy to estimate than the location of median interval.

(b). The estimated coverage probabilities based on C_{mod} are also significantly larger than those based on C_{med}.

Table 10 Coverage percentages for distribution $Beta(\alpha, \beta)$

β	$\hat{\pi}_{med}$	$\hat{\pi}_{mod}$	$\hat{\pi}_{med}$	$\hat{\pi}_{mod}$
	$\alpha = 3$		$\alpha = 10$	
1	0.9032	0.9309	0.9099	0.9419
3	0.8642	0.8942	0.8875	0.9104
5	0.8709	0.8986	0.8728	0.8997
7	0.8791	0.9047	0.8682	0.8956
9	0.8855	0.9092	0.8669	0.8949

References

Besterfield, D. H. (1990). *Quality Control.* Prentice Hall: New Jersey.

DeVor, R. E., Chang, T.-H. and Sutherland, J. W. (1992). *Statistical Quality Design and Control.* Prentice Hall: New Jersey.

Huang, J.-Y. (2003). Mode interval and its application to construct a new Shewhart control chart. Ph.D. thesis, National Chiao Tung University.

Staudte, R. G. and Sheather, S. J. (1990). *Robust Estimation and Testing*, Wiley: New York.

Figure.1 np control chart for blower motor data

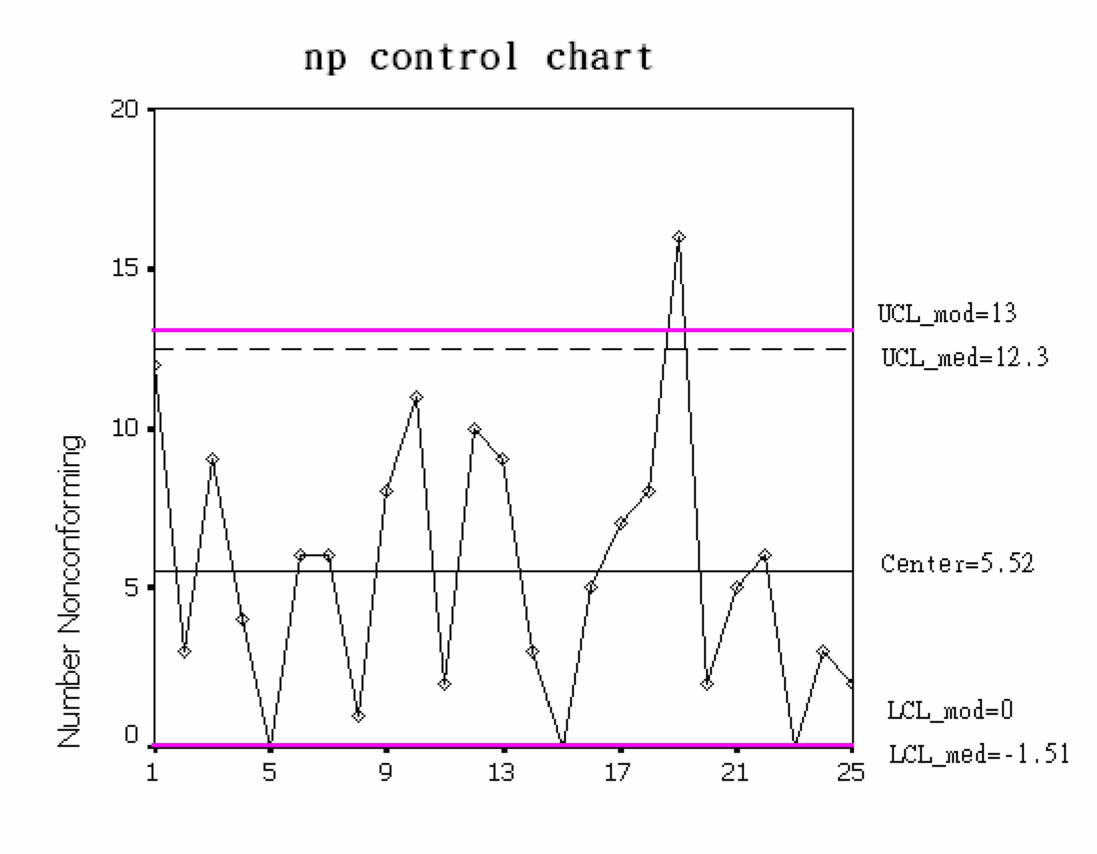

Figure.2 C control chart for rear bumper data