International Journal of Science and Research (IJSR)
ISSN: 2319-7064
SJIF (2022): 7.942

The Effect of Non-Normality and Measurement Error on the Economic Design of $\bar{X}$ Control Chart

J. R. Singh[1], N. Singh[2]

[1, 2]School of Studies in Statistics, Vikram University, Ujjain, M.P., India
[1]Email: *jrsinghstat[at]gmail.com*
[2]Email: *dr.niveditasingh114[at]gmail.com*

Abstract: ***An attempt is made to determine the effect of non-normality and measurement error on the economic design of $\bar{X}$ control chart. For non-normal population, we have considered the first four terms of an Edgeworth series. As one will be interested in having a suitable economic control chart under measurement error for non-normal variables, the optimum values of sample size n and sampling interval h are determined for different values of k.***

Keywords: Non-normality $\bar{X}$, control chart, Edgeworth series, Measurement error

1. Introduction

Traditionally, when designing control charts, one usually assumes that the measurements in the sample are normally distributed. However, this assumption may not be tenable. If the measurements really are normally distributed, then the statistic $\bar{x}$ is also normally distributed. If the measurements are asymmetrically distributed, then the statistic $\bar{x}$ will be approximately normally distributed only when the sample size n is sufficiently large (based on the central limit theorem). Unfortunately, when a control chart is applied to monitor the process, the sample size n is never sufficiently large due to the sampling cost. Therefore, if the measurements are not normally distributed, the traditional way of designing a control chart may reduce the ability of the control chart to detect the assignable causes. Yourstone and Zimmer (1992) used the Burr distribution to represent various non-normal distributions and, consequently, to statistically design the control limits of an control chart. However, they did not consider cost in the design of the chart. In designing a control chart, three parameters - the sample size n, time h between successive samples, and the number k of standard deviations away from the center line - must be determined. In economic-statistical design, the three parameters are chosen so that the expected cost per hour is minimized under constraints, e.g., minimum allowable values of Type I error probability (probability that point falls outside control limits while the process is in control) and Type II error probability (probability that point falls within control limits while out of control). Saniga (1989) first proposes the economic-statistical design. Al-Oraini and Rahim (2003) have shown that the statistical performance can be improved by the economic-statistical design significantly with only a slight increase in the cost. Kanazuka (1986) used to study the effect of measurement error on the performance of an $\bar{X}$-R chart. Mittag (1995) and Mittag and Stemann (1998) investigated how the measurement error affects the $\bar{X}$-S chart. Linna and Woodall (2001) assumed a linear relationship between the surrogate and the true quality characteristics to study the effect of measurement error on the performance of $\bar{X}$ and S^2 charts. However, the general practice to set symmetrical control limits for the mean to detect shifts in the process average when the process variation remain constant. The determination at k is mainly based upon the level of the control desired in a given situation depending upon the market price of defectives and effectives.

2. Mathematical Model for Cost Function

Duncan (1956) obtained an approximate function for the average net income per hour of using the control chart for mean of normal variables as :

$$I=[V_0]-\frac{\eta\mu B+(\alpha T/h)+\eta W}{1+\eta B}-\frac{b+cn}{h} \quad (2.1)$$

where
V_0= the average income per hour when the process is in control and the process average is μ,
V_1= the average income per hour when the process is not in control and the process average is $\mu'=\mu+\delta\sigma$,
$M=V_0-V_1$,
η = the average number of times the assignable causes occur within an interval of time,
$B=ah+Cn+D$,
$a=\frac{1}{P}-\frac{1}{2}+\frac{\eta h}{12}$,
h = interval between sampling in hours,
Cn = the time required to take and inspect a sample of size n,
D = average time taken to find the assignable cause after a point plotted on the chart falls out side the control limits,
P = probability of detecting an assignable cause when it exists;

$$=\int_{-\infty}^{\mu-k\sigma/\sqrt{n}} g(\bar{x}/\mu')d\bar{x}+\int_{\mu+k\sigma/\sqrt{n}}^{\infty} g(\bar{x}/\mu')d\bar{x}$$

$$\cong 1-\Phi(k-\delta\sqrt{n}) \quad for\ \delta>0,$$

where $g(\bar{x}/\mu)$ is the density function of $\bar{x}$ when the true mean is μ and $\Phi(x)$ is the normal probability integral,
α = probability of wrongly indicating the presence of assignable causes,

$$= \int_{\mu - k\sigma/\sqrt{n}}^{\mu + k\sigma/\sqrt{n}} g(\bar{x}/\mu) d\bar{x}$$

$$= 2\Phi(-k)$$

$$= \alpha_N \quad . \qquad (2.2)$$

T= the cost per occasion of looking for an assignable cause when no assignable cause exists,
W= the average cost per occasion of finding the assignable when it exists,
b= per sample cost of sampling and plotting, that is independent of sample size,
and *c*= the cost per unit of measuring an item in a sample.

The average cost per hour involved for maintaining the control chart is $(b+cn)/h$. The average net income per hour of the process under the surveillance of the control chart for mean can be rewritten as,

$$I = V_0 - L,$$

where

$$L = \frac{\eta MB + (\alpha T/h) + \eta B}{1 + \eta B} + \frac{b + cn}{h} \qquad (2.3)$$

L can now be treated as the per hour cost due to the surveillance of the process under the control chart. The probability density function for non-normal population is represented by first four terms of Edgeworth series and $\dot{P} \; and \; \alpha'$ are determined from the sampling distribution of mean and are written as,

$$P' = 1 - \Phi(\xi) + \frac{\lambda_3}{6\sqrt{n}} \phi^{(2)}(\xi) - \frac{\lambda_4}{24n} \phi^{(3)}(\xi) - \frac{\lambda_3^2}{72n} \phi^{(5)}(\xi), \quad for \; \delta > 0 \qquad (2.4)$$

$$\alpha' = \alpha_N - \alpha_C \qquad (2.5)$$

Where $\xi = (k - \delta\sqrt{n})$

and $\alpha_C = \frac{\left[3\lambda_4 \phi^{(3)}(k) + \lambda_3^2 \phi^{(5)}(k)\right]}{36n}$

is the non-normality correction for α.

3. Optimum Value of Sample Size n and Sampling Interval h

One can determine the optimum value of sample size n_0 and sampling interval h_0 either by maximizing the gain function *I* or by minimizing the cost function *L* with respect to *n* and *h*. After solving for minimizing the cost function L, we find the following two equations

$$h = \left\{ \frac{\alpha' T + b + cn}{\eta M \left(\frac{1}{P'} - \frac{1}{2} \right)} \right\}^{\frac{1}{2}} . \qquad (3.1)$$

$$-\frac{\alpha' T + b + cn}{P'^2 \left(\frac{1}{P'} - \frac{1}{2} \right)} \cdot \frac{\partial P'}{\partial n} - \eta \alpha' T + \frac{T \alpha_C}{n} \left\{ \frac{\eta M \left(\frac{1}{P'} - \frac{1}{2} \right)}{\alpha' T + b + cn} \right\}^{\frac{1}{2}} + c = 0 . \; (3.2)$$

The value of *n* for which the equation (3.2) satisfy yield us the required optimum value of the sample size n_0. Substituting this value of *n* in equation (3.1), we find the optimum value of the sampling interval h_0.

4. Description for Optimum Value of Sample Size n and Sampling Interval h under Measurement Error

Assuming that the true measurement *x* and the random error of measurement *e* are additive, then

$$X = x + e. \qquad (4.1)$$

The mean and standard deviation of the observed measurement *X* can be written as

$E(X)=\mu, \quad V(X)=\sigma_X^2$. (say)

where μ is the mean of x and $e \sim N(0,\sigma_e^2)$,

The correlation coefficient ρ between the true and observed measurement is given by

$$\rho=\frac{r}{\sqrt{1+r^2}}, \qquad (4.2)$$

Where $r=\dfrac{\sigma_p}{\sigma_e}$.

Now since x and e are independent, the r^{th} cumulant of X is equal to the sum of the r^{th} cumulants of x and e. Further, since $e \sim N(0,\sigma_e^2)$, all the cumulants of e are zero except the second one which is σ_e^2 . Thus, if we denote by k_r and l_r the r^{th} cumulants of X and x respectively, we have

$$k_r = l_r, \qquad r \neq 2$$

Let v_r *and* λ_r $(r \neq 2)$ be the rth standardized cumulants of X and x respectively, then

$$v_r=\frac{k_r}{(k_r)^{r/2}}=\frac{l_r}{(\sigma_X)^r}=\frac{l_r}{(\sigma_p/\rho)^r}$$

or $$v_r=\rho^r\lambda_r. \qquad (4.3)$$

So the probability density function for non-normal population under non-normality and measurement error will be

$$P_e'=1-\Phi(\xi_e)+\frac{\rho^3\lambda_3}{6\sqrt{n}}\phi^{(2)}(\xi_e)-\frac{\rho^4\lambda_4}{24n}\phi^{(3)}(\xi_e)-\frac{\rho^6\lambda_3^2}{72n}\phi^{(5)}(\xi_e), \quad for\ \delta>0.$$

$$\alpha_e'=\alpha_{Ne}-\alpha_{Ce}$$

Where $\xi_e=\rho(k-\delta\sqrt{n})$,

$\alpha_{Ne}=2\Phi(-\rho k)$,

$$\alpha_{Ce}=\frac{\left[3\rho^4\lambda_4\phi^{(3)}(k)+\rho^6\lambda_3^2\phi^{(5)}(k)\right]}{36n}.$$

In presence of non-normality and measurement error, the equation (3.1) and (3.2) will reduce in following form

$$h_{e0}=h=\left\{\frac{\alpha_e'T+b+cn}{\eta M\left(\frac{1}{P_e'}-\frac{1}{2}\right)}\right\}^{\frac{1}{2}} \qquad (4.4)$$

and

$$-\frac{\alpha_e'T+b+cn}{P_e'^2\left(\frac{1}{P_e'}-\frac{1}{2}\right)}\cdot\frac{\partial P_e'}{\partial n}-\eta\alpha_e'T+\frac{T\alpha_{Ce}}{n}\left\{\frac{\eta M\left(\frac{1}{P_e'}-\frac{1}{2}\right)}{\alpha_e'T+b+cn}\right\}^{\frac{1}{2}}+c=0. \qquad (4.5)$$

Again the value of n for which the equation (4.5) satisfy give us the required optimum value of the sample size n_{e0}, and putting this value in equation (4.4), we find the optimum value of the sampling interval h_{e0} under non-normality and measurement error.

5. Numerical Illustration

For the purpose of numerical illustration, we take

$\lambda_3=-0.5,\ 0,\ 0.5, \quad \lambda_4=-0.5,\ 0,\ 1.0,\ 2.0, \quad k=2.0,\ 3.0, \delta=0.5,\ 1.0,\ 2.0,$

$\eta=0.01, \quad M=100, \quad W=25, \quad T=50, \quad C=0.05, \quad D=2, \quad b=0.5, \quad c=0.1,$

and $r=\infty, 2, 6$ and determine the optimum value of sample size and sampling interval. The values of n_0 and h_0 are presented in the Table-1,

Table 1: Values of the optimum sample size n and sampling interval h under measurement error for r =∞

δ	λ_3 → / λ_4 ↓	K=3						K=2					
		-0.5		0.0		0.5		-0.5		0.0		0.5	
		n	h	n	h	n	h	n	h	n	h	n	h
0.5	-0.5	66	3.2781	67	3.3254	68	3.3652	41	3.2800	42	3.3288	43	3.3897
	0.0	66	3.2990	67	3.3392	68	3.3782	42	3.3334	42	3.3568	43	3.3927
	1.0	67	3.3451	68	3.3894	68	3.4034	43	3.4173	44	3.4325	44	3.4440
	2.0	67	3.3576	68	3.4051	69	3.4483	44	3.4644	45	3.5057	45	3.5146
1.0	-0.5	17	1.8675	17	1.9237	18	1.9413	29	2.6164	29	2.6288	29	2.6395
	0.0	18	1.9911	18	2.0026	19	2.0605	31	2.7513	31	2.7709	31	2.7706
	1.0	21	2.2002	21	2.2088	21	2.2198	34	2.9405	34	2.9574	34	2.9575
	2.0	23	2.3422	23	2.3498	23	2.3590	37	3.1234	37	3.1076	37	3.1104
2.0	-0.5	3	1.0032	4	1.1150	4	1.1563	24	2.2733	25	2.3256	25	2.3526
	0.0	5	1.4340	5	1.4336	5	1.4543	26	2.4460	27	2.4898	27	2.4569
	1.0	13	1.9632	12	1.9192	12	1.9317	30	2.7167	34	2.9294	33	2.8916
	2.0	17	2.1800	17	2.1845	17	2.1931	34	2.9623	37	3.1026	37	3.0967

Table 2: Values of the optimum sample size n and sampling interval h under measurement error for r=2

δ	λ_3 → / λ_4 ↓	K=3						K=2					
		-0.5		0.0		0.5		-0.5		0.0		0.5	
		n	h	n	h	n	h	n	h	n	h	n	h
0.5	-0.5	95	4.3423	94	4.3129	95	4.3487	85	4.9819	85	4.9864	86	5.0030
	0.0	95	4.3548	95	4.3442	96	4.3796	86	4.9951	86	5.0154	87	5.0323
	1.0	96	4.3709	95	4.3600	97	4.4105	86	5.0141	87	5.0423	88	5.0671
	2.0	97	4.4017	96	4.3762	98	4.4409	87	5.0345	88	5.0665	89	5.0803
1.0	-0.5	112	4.8096	112	4.8137	113	4.8428	90	5.0842	91	5.1193	92	5.1429
	0.0	113	4.8347	114	4.8633	114	4.8675	90	5.0992	92	5.1410	94	5.1854
	1.0	114	4.8597	114	4.8756	115	4.8921	91	5.1166	93	5.1630	95	5.2068
	2.0	116	4.9085	116	4.9123	116	4.9164	92	5.1389	96	5.2262	96	5.2281
2.0	-0.5	130	5.2218	131	5.2455	131	5.2484	110	5.4975	112	5.5369	112	5.5376
	0.0	131	5.2430	132	5.2665	132	5.2693	112	5.5354	113	5.5558	112	5.5470
	1.0	132	5.2642	132	5.2770	134	5.3106	115	5.5916	114	5.5748	113	5.5570
	2.0	132	5.2747	134	5.3083	135	5.3312	116	5.6105	116	5.6121	117	5.6308

Table 3 : Values of the optimum sample size n and sampling interval h under measurement error for r=6

δ	$\lambda_3 \rightarrow$ $\lambda_4 \downarrow$	K=3						K=2					
		-0.5		0.0		0.5		-0.5		0.0		0.5	
		n	h	n	h	n	h	n	h	n	h	n	h
0.5	-0.5	87	4.1063	88	4.1453	88	4.1518	66	4.2658	66	4.2734	67	4.3059
	0.0	87	4.1230	89	4.1783	90	4.2169	67	4.2953	68	4.3310	68	4.3347
	1.0	88	4.1406	89	4.1949	91	4.2492	68	4.3249	68	4.3458	69	4.3637
	2.0	89	4.1739	90	4.2120	92	4.2810	68	4.3398	69	4.3611	69	4.3783
1.0	-0.5	104	4.6071	104	4.6125	105	4.6440	79	4.6105	80	4.6396	80	4.6418
	0.0	105	4.6333	105	4.6386	106	4.6697	81	4.6590	81	4.6638	82	4.6893
	1.0	105	4.6464	106	4.6646	107	4.6953	82	4.6834	82	4.6881	83	4.7133
	2.0	106	4.6599	108	4.7153	109	4.7451	84	4.7308	84	4.7351	85	4.7597
2.0	-0.5	119	4.9685	120	4.9943	120	4.9981	91	4.8850	91	4.8883	91	4.8894
	0.0	119	4.9796	121	5.0164	122	5.0416	92	4.9070	92	4.9101	93	4.9324
	1.0	120	4.9912	121	5.0273	123	5.0633	93	4.9292	94	4.9533	94	4.9543
	2.0	121	5.0135	123	5.0602	124	5.0848	94	4.9511	95	4.9750	95	4.9760

when $r = \infty$ (error free), it is evident from the table that for given *k* and δ the value of n_0 and h_0 increases with the increase in the value of λ_4 and practically remain the same with the increase in the value of λ_3. For large values of δ the effect of kurtosis seems to be more marked than that of skewness. As the size of error increases the value of n_0 and h_0 is seriously affected.

6. Conclusion

It may be inferred that measurement errors affects considerably the optimum value of the sample size and optimum sampling interval. It is necessary to point out that the measurement errors and non-normality of the population should be taken in to account while designing a control chart as the optimum values of the control chart parameters are affected by the measurement errors and non-normality of the population.

References

[1] ***Al-Oraini, H. A. and Rahim, M. A. (2003).*** "Economic Statistical Design of *X* Control Charts for Systems with gamma (λ, 2) In-control Times," *Journal of Applied Statistics 30*, 397-409.

[2] ***Duncan, A.J. (1956).*** "The Economic Design of X-bar Charts Used to Maintain Current Control of a Process," *American Statistical Association Journal*, 51, 228-242.

[3] ***Kanazuka, T. (1986).*** "The Effects of Measurement Errors on the Power of $\bar{X} - R$ Charts," *J. Quality Technology 18*, 91-95.

[4] ***Linna, K. W., and Woodall, W. H. (2001).*** "Effect of Measurement Error on Shewhart Control Charts," *Journal of Quality Technology 33*, 213-222

[5] ***Mittag, H. J. (1995).*** "Measurement Error Effects on Control Chart Performance," *ASQC 49th Annual Quality Congress Proceedings*, 66-73.

[6] ***Mittag, H. J., and Stemann, D. (1998).*** "Gauge Imprecision Effect on the Performance of the $\bar{X}$ −S Control Chart," *Journal of Applied Statistics. 25,* 307-317.

[7] ***Saniga, E. M. (1989)***. "Economic Statistical Control Chart Designs with an Application to $\bar{X}$ and *R* Charts," *Technometrics*, 31, 313-320.

[8] ***Yourstone, S. A. and Zimmer, W. J. (1992).*** "Non Normality and the Design of Control Charts for Averages," *Decision Science*, 23(5), 1099-1113.