

5-1-2017

# Control Charts for Mean for Non-Normally Correlated Data

J. R. Singh

*Vikram University, Ujjain, India*

Ab Latif Dar

*School of Studies in Statistics, Vikram University, Ujjain, India, lateefdar.2007@rediffmail.com*

Follow this and additional works at: <http://digitalcommons.wayne.edu/jmasm>

 Part of the [Applied Statistics Commons](#), [Social and Behavioral Sciences Commons](#), and the [Statistical Theory Commons](#)

## Recommended Citation

Singh, J. R. & Dar, A. L. (2017). Control charts for mean for non-normally correlated data. *Journal of Modern Applied Statistical Methods*, 16(1), 452-460. doi: 10.22237/jmasm/1493598300

This Regular Article is brought to you for free and open access by the Open Access Journals at DigitalCommons@WayneState. It has been accepted for inclusion in *Journal of Modern Applied Statistical Methods* by an authorized editor of DigitalCommons@WayneState.

# Control Charts for Mean for Non-Normally Correlated Data

**J. R. Singh**  
Vikram University  
Ujjain, India

**Ab Latif Dar**  
Vikram University  
Ujjain, India

---

Traditionally, quality control methodology is based on the assumption that serially-generated data are independent and normally distributed. On the basis of these assumptions the operating characteristic (OC) function of the control chart is derived after setting the control limits. But in practice, many of the basic industrial variables do not satisfy both the assumptions and hence one may doubt the validity of the inferences drawn from the control charts. In this paper the power of the control chart for the mean is examined when both the assumptions of independence and normality are not tenable. The OC function is calculated and compared with the normal population.

*Keywords:* Control chart, correlation, Edgeworth Series, standardized cumulants

---

## Introduction

The quality control techniques currently used in industry are aimed at the detection of changes in the production process that result in quality defects. Quality control charts are currently the most widely-adopted control technique. Traditionally, quality control methodology is based on the assumption that serially-generated data are independent and normally distributed. Under these conditions, appropriate control limits for  $\bar{X}$  can be worked out from the tables available in standard textbooks on statistical quality control. But in practice, many of the basic industrial processes do not satisfy both the assumptions and hence one may doubt the validity of the inference drawn from the control charts.

Alwan (1992) studied the effect of auto-correlation on control chart performance. Maragah and Woodall (1992) studied the effect of auto-correlation on the retrospective  $X$ -chart. Alwan and Roberts (1995) conducted investigations of control charts when the assumptions of normality, independence, or both are violated. Dar and Singh (2015) studied the effect of correlation on the power of

---

*Ab Latif Dar is a Professor in the School of Studies in Statistics. Email them at: lateefdar.2007@rediffmail.com.*

the  $\bar{X}$  chart. The purpose of this study is to consider the power of the control chart and the effect of correlation on Type-I error and the OC function, and also to consider relaxing the assumption of normality and considering the production process to follow a non-normal distribution represented by the first four terms of an Edgeworth series.

### Effect of Correlation on OC Function for Normal Case

Suppose that the observations  $x_1, x_2, \dots, x_n$  have a multivariate normal distribution with  $E(x_i) = \mu$ ,  $V(x_i) = \sigma^2$  and  $\rho$  is the common correlation coefficient between any  $x_i$  and  $x_j$ ,  $i \neq j$ . Then

$$\begin{aligned} E(\bar{x}) &= \mu \\ \text{Var}(\bar{x}) &= \frac{\sigma^2}{n} [1 + (n-1)\rho] \\ &= \frac{\sigma^2}{n} T^2 \end{aligned} \quad (1)$$

where

$$T^2 = [1 + (n-1)\rho] \quad (2)$$

The power of the control chart is judged by its OC function. The control chart for the mean is set up by drawing the central line at the process average  $\theta$  and the control limits at

$$\theta \pm \frac{k\sigma}{\sqrt{n}}$$

where  $\sigma$  is the process standard deviation and  $n$  is the sample size. The OC function gives the probability that the control chart indicates the value  $\theta$  as the process average, when it is actually not  $\theta$ , but

$$\theta' = \theta + \frac{\gamma\sigma T}{\sqrt{n}}$$

## CONTROL CHARTS FOR MEAN

where  $T^2$  is as defined in equation (2).

The OC function is derived by integrating the distribution of the mean with  $\theta'$  as the process average between the limits of the control chart.

For the normal population under correlated data,

$$f(X) = \frac{1}{\sigma} \left[ \phi \left( \frac{X - \theta}{\sigma} \right) \right] \quad (3)$$

The distribution of the sample mean is given by

$$g(\bar{X}) = \frac{\sqrt{n}}{\sigma T} \left[ \phi \left( \frac{\bar{X} - \theta / \frac{\sigma T}{\sqrt{n}}}{\frac{\sigma T}{\sqrt{n}}} \right) \right] \quad (4)$$

where

$$\phi(t) = \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{t^2}{2}\right) \quad \text{and} \quad \phi^r(t) = \frac{d^r}{dt^r} \phi(t)$$

The OC function is obtained after replacing  $\theta$  in (4) by  $\theta'$  and integrating it between the limits of the control chart as

$$L_N = \frac{\sqrt{n}}{\sigma T} \int_{\theta - \frac{k\sigma}{\sqrt{n}}}^{\theta + \frac{k\sigma}{\sqrt{n}}} \phi \left[ \frac{\bar{X} - \left( \theta + \frac{\gamma\sigma T}{\sqrt{n}} \right)}{\frac{\sigma T}{\sqrt{n}}} \right] d\bar{X} \quad (5)$$

$$L_N = \frac{\sqrt{n}}{\sigma T} \int_{\theta - \frac{k\sigma}{\sqrt{n}}}^{\theta + \frac{k\sigma}{\sqrt{n}}} \phi \left[ \frac{\bar{X} - \theta}{\frac{\sigma T}{\sqrt{n}}} + \gamma \right] d\bar{X} \quad (6)$$

Making the transformation

$$\frac{\bar{X} - \theta}{\frac{\sigma T}{\sqrt{n}}} = y$$

and  $y - \gamma = t$  sequentially, the above integral simplifies to

$$L_N = \Phi\left(\frac{k}{T} + \gamma\right) + \Phi\left(\frac{k}{T} - \gamma\right) - 1 \quad (7)$$

The error of Type I gives the probability of searching for assignable causes when in fact there are no such causes. It is given by

$$\alpha = 1 - \int_{\theta - \frac{k\sigma}{\sqrt{n}}}^{\theta + \frac{k\sigma}{\sqrt{n}}} g(\bar{X}) d\bar{X} \quad (8)$$

After integrating above as in the case of the OC function we will get

$$\alpha = 2\Phi\left(-\frac{k}{T}\right) \quad (9)$$

### The Effect of Non-Normally Correlated Data on OC Function

For non-normal populations represented by the first four terms of an Edgeworth series,

$$f(X) = \frac{1}{\sigma} \left[ \phi\left(\frac{\bar{X} - \theta}{\sigma}\right) - \frac{\lambda_3}{6} \phi^{(3)}\left(\frac{\bar{X} - \theta}{\sigma}\right) + \frac{\lambda_4}{24} \phi^{(4)}\left(\frac{\bar{X} - \theta}{\sigma}\right) + \frac{\lambda_3}{72} \phi^{(6)}\left(\frac{\bar{X} - \theta}{\sigma}\right) \right] \quad (10)$$

where  $\lambda_3 = \sqrt{\beta_1}$  and  $\lambda_4 = (\beta_2 - 3)$  are the standardized third and fourth cumulants, respectively.

The distribution of the sample mean for correlated data can be derived, by following Gayen (1952), as

## CONTROL CHARTS FOR MEAN

$$g(\bar{X}) = \frac{\sqrt{n}}{\sigma T} \left[ \begin{aligned} &\phi\left(\frac{\bar{X} - \theta}{\sigma T / \sqrt{n}}\right) - \frac{\lambda_3 T}{6\sqrt{n}} \phi^{(3)}\left(\frac{\bar{X} - \theta}{\sigma T / \sqrt{n}}\right) \\ &+ \frac{\lambda_4 T^2}{24n} \phi^{(4)}\left(\frac{\bar{X} - \theta}{\sigma T / \sqrt{n}}\right) + \frac{\lambda_3^2 T^2}{72n} \phi^{(6)}\left(\frac{\bar{X} - \theta}{\sigma T / \sqrt{n}}\right) \end{aligned} \right] \quad (11)$$

The OC function is obtained after replacing  $\theta$  in equation (11) by  $\theta'$  and integrating it between the limits of the control chart, i.e. between  $\theta \pm (k\sigma)/\sqrt{n}$ . Integrating in the similar way as for the normal case, we get

$$L' = L_N - L'_u + L'_b \quad (12)$$

where  $L_N$  is given by equation (7). The other two terms of the OC function are given by

$$L'_u = \frac{T^2}{72n} \left[ \frac{12\lambda_3\sqrt{n}}{T} \phi^{(2)}\left(\frac{k}{T} - \gamma\right) - 3\lambda_4 \phi^{(3)}\left(\frac{k}{T} - \gamma\right) - \lambda_3^2 \phi^{(5)}\left(\frac{k}{T} - \gamma\right) \right] \quad (13)$$

$$L'_b = \frac{T^2}{72n} \left[ \frac{12\lambda_3\sqrt{n}}{T} \phi^{(2)}\left(\frac{k}{T} + \gamma\right) + 3\lambda_4 \phi^{(3)}\left(\frac{k}{T} + \gamma\right) + \lambda_3^2 \phi^{(5)}\left(\frac{k}{T} + \gamma\right) \right] \quad (14)$$

The Type-I error for the non-normal population works out to be

$$\alpha' = 1 - \int_{\theta - \frac{k\sigma}{\sqrt{n}}}^{\theta + \frac{k\sigma}{\sqrt{n}}} g(\bar{X}) d\bar{X} = \alpha - c \quad (15)$$

where  $\alpha$  as defined by equation (9) is the Type-I error when the population is normal and dependent, and

$$c = \frac{T^2}{36n} \left[ 3\lambda_4 \phi^{(3)}\left(\frac{k}{T}\right) + \lambda_3^2 \phi^{(5)}\left(\frac{k}{T}\right) \right] \quad (16)$$

is the correction for non-normality and dependencies in Type-I error.

**Table 1.** Value of Type-I error for normally and correlated data

<i>n</i>	<i>K</i> = 2				<i>K</i> = 3			
	$\rho = 0.0$	$\rho = 0.2$	$\rho = 0.5$	$\rho = 0.8$	$\rho = 0.0$	$\rho = 0.2$	$\rho = 0.5$	$\rho = 0.8$
5	0.04550	0.13603	0.24821	0.32911	0.00269	0.02534	0.08326	0.14323
10	0.04550	0.23200	0.39377	0.48491	0.00270	0.07300	0.20083	0.29480
15	0.04550	0.30490	0.47950	0.56692	0.00270	0.12381	0.28884	0.39040

**Table 2.** Value of OC function for normally and correlated data

<i>n</i>	$\gamma$	<i>K</i> = 2				<i>K</i> = 3			
		$\rho = 0.0$	$\rho = 0.2$	$\rho = 0.5$	$\rho = 0.8$	$\rho = 0.0$	$\rho = 0.2$	$\rho = 0.5$	$\rho = 0.8$
5	-2	0.9932	0.3050	0.1981	0.1514	0.8413	0.5932	0.3942	0.2956
	-1	0.9997	0.6818	0.5458	0.4663	0.9772	0.8911	0.7647	0.6717
	0	0.9999	0.8639	0.7517	0.6708	0.9973	0.9746	0.9167	0.8567
	1	0.9997	0.6818	0.5458	0.4663	0.9772	0.8911	0.7647	0.6717
	2	0.9932	0.3050	0.1981	0.1514	0.8413	0.5932	0.3942	0.2956
10	-2	0.5000	0.2098	0.1235	0.0930	0.8413	0.4179	0.2350	0.1693
	-1	0.8400	0.5633	0.4095	0.3368	0.9772	0.7835	0.5986	0.4987
	0	0.9545	0.7680	0.6062	0.5151	0.9973	0.9270	0.7992	0.7052
	1	0.8400	0.5633	0.4095	0.3368	0.9772	0.7835	0.5986	0.4987
	2	0.5000	0.2098	0.1235	0.0930	0.8413	0.4179	0.2350	0.1693
15	-2	0.5000	0.1638	0.0946	0.0717	0.8413	0.3222	0.1727	0.1248
	-1	0.8400	0.4890	0.3409	0.2766	0.9772	0.6995	0.5045	0.4124
	0	0.9545	0.6951	0.5205	0.4331	0.9973	0.8762	0.7112	0.6096
	1	0.8400	0.4890	0.3409	0.2766	0.9772	0.6995	0.5045	0.4124
	2	0.5000	0.1638	0.0946	0.0717	0.8413	0.3222	0.1727	0.1248

## Results and Conclusion

For normal populations with correlation coefficient  $\rho = 0, 0.2, 0.5,$  and  $0.8,$  the values of Type-I error have been computed and given in Table 1 for  $k = 2, 3$  and  $n = 5, 10, 15.$  Table 1 clearly indicates that the effect of correlation on Type-I error is quite substantial as the error increases with the increase in  $\rho.$  For example, for  $n = 5, k = 2,$  and  $\rho = 0, 0.2, 0.5,$  and  $0.8,$  the corresponding values of Type-I error are  $0.04550, 0.13605, 0.24821,$  and  $0.32911.$  Though the effect goes on decreasing with increasing  $k,$  it still affects the value of Type-I error quite largely. For non-normal populations we have a similar result (Table 3) as the error goes on increasing with an increase in the value of  $\rho, \lambda_3,$  and  $\lambda_4.$  From Table 2, it is evident that the value of the OC are affected seriously as the correlation between the observations increases. For example, for  $\rho = 0, k = 2, n = 5,$  and  $\gamma = \pm 1,$  the value

## CONTROL CHARTS FOR MEAN

of the OC is 0.9997, while for  $\rho = 0.2, 0.5, 0.8, k = 2$ , and  $n = 5$ , the value reduces to 0.6818, 0.5458, 0.4663. For other values of  $n = 10, 15$ , we have a similar results. The values of the OC for non-normal populations with  $k = 2, n = 5$  and for different values of  $\rho = 0, 0.2, 0.5, 0.8$  are given in Table 4. For  $\rho = 0$  and  $(\lambda_3, \lambda_4) = (0, 0)$  we get tabulated values of Singh, Sankle, and Ahmad (2012), which are shown in Table 4. The effect of correlation on the OC function remains more or less of the same magnitude when we move from normal to non-normal populations. As is evident from the Table 4, for  $\rho = 0, \lambda_3 = 0, \lambda_4 = 0$ , and  $\gamma = \pm 1$ , the value of the OC function is 0.8400 while for  $\rho = 0.5, \lambda_3 = 0.5, \lambda_4 = 0.5$ , and  $\gamma = \pm 1$ , the corresponding value of the OC function is reduced to 0.3725. On changing  $\lambda_3$  (skewness),  $\lambda_4$  (kurtosis), or both at the same time, the value of the OC is affected. Therefore, it may be inferred that the violation in the assumptions of independence and normality have a serious effect on the control chart performance and it is advisable to take into account the dependence and non-normality of the parent population while designing control charts.

**Table 3.** Values of the Type-I error for non-normally correlated data

$\rho$	$n$	$\lambda_3$	$K = 2$				$K = 3$			
			$\lambda_4=0.0$	$\lambda_4=0.5$	$\lambda_4=1.0$	$\lambda_4=2.0$	$\lambda_4=0.0$	$\lambda_4=0.2$	$\lambda_4=0.5$	$\lambda_4=0.8$
0.0	5	0.0	0.0455	0.0464	0.0473	0.0491	0.0027	0.0034	0.0040	0.0054
		0.5	0.0442	0.0451	0.0459	0.0477	0.0028	0.0035	0.0041	0.0055
	10	0.0	0.0455	0.0459	0.0464	0.0473	0.0027	0.0030	0.0034	0.0040
		0.5	0.0448	0.0453	0.0457	0.0466	0.0028	0.0031	0.0034	0.0041
	15	0.0	0.0455	0.0458	0.0461	0.0467	0.0027	0.0029	0.0031	0.0036
		0.5	0.0451	0.0454	0.0456	0.0462	0.0027	0.0030	0.0032	0.0036
0.2	5	0.0	0.1360	0.1338	0.1315	0.1269	0.0253	0.0275	0.0297	0.0341
		0.5	0.1349	0.1326	0.1303	0.1258	0.0235	0.0257	0.0279	0.0323
	10	0.0	0.2320	0.2277	0.2234	0.2149	0.0730	0.0734	0.0737	0.0744
		0.5	0.2332	0.2290	0.2247	0.2161	0.0711	0.0715	0.0718	0.0725
	15	0.0	0.3049	0.2999	0.2950	0.2850	0.1238	0.1226	0.1213	0.1188
		0.5	0.3073	0.3023	0.2973	0.2874	0.1228	0.1215	0.1203	0.1178
0.5	5	0.0	0.2482	0.2384	0.2285	0.2088	0.0833	0.0833	0.0833	0.0833
		0.5	0.2516	0.2418	0.2319	0.2122	0.0794	0.0794	0.0794	0.0794
	10	0.0	0.3938	0.3814	0.3691	0.3445	0.2008	0.1938	0.1867	0.1727
		0.5	0.4012	0.3889	0.3766	0.3519	0.2020	0.1949	0.1879	0.1738
	15	0.0	0.4795	0.4673	0.4551	0.4307	0.2888	0.2788	0.2687	0.2487
		0.5	0.4878	0.4756	0.4634	0.4390	0.2933	0.2833	0.2732	0.2531



Table 3, continued.

$\rho$	$n$	$\lambda_3$	$K=2$				$K=3$			
			$\lambda_4=0.0$	$\lambda_4=0.5$	$\lambda_4=1.0$	$\lambda_4=2.0$	$\lambda_4=0.0$	$\lambda_4=0.2$	$\lambda_4=0.5$	$\lambda_4=0.8$
0.8	5	0.0	0.3291	0.3118	0.2944	0.2598	0.1432	0.1372	0.1312	0.1192
		0.5	0.3381	0.3208	0.3034	0.2688	0.1411	0.1351	0.1291	0.1171
	10	0.0	0.4849	0.4662	0.4474	0.4099	0.2948	0.2791	0.2634	0.2320
		0.5	0.4978	0.4790	0.4603	0.4228	0.3020	0.2863	0.2706	0.2392
	15	0.0	0.5669	0.5494	0.5318	0.4967	0.3904	0.3722	0.3541	0.3177
		0.5	0.5799	0.5623	0.5447	0.5096	0.4013	0.3832	0.3650	0.3286

Table 4. Values of OC function for non-normally correlated data

$\rho$	$\gamma$	$(\lambda_3, \lambda_4)$							
		(0.0,0.0)	(0.0,0.5)	(0.0,1.0)	(0.0,2.0)	(0.5,0.0)	(0.5,0.5)	(0.5,1.0)	(0.5,2.0)
0.0	-2	0.5000	0.5000	0.4999	0.4999	0.4850	0.4850	0.4849	0.4849
	-1	0.8400	0.8408	0.8417	0.8434	0.8396	0.8413	0.8430	0.8464
	0	0.9545	0.9540	0.9536	0.9527	0.9545	0.9536	0.9527	0.9509
	1	0.8400	0.8408	0.8417	0.8434	0.8404	0.8420	0.8437	0.8471
	2	0.5000	0.5000	0.4999	0.4999	0.5150	0.5149	0.5149	0.5148
0.2	-2	0.2098	0.2063	0.2028	0.1957	0.2022	0.1986	0.1951	0.1881
	-1	0.5633	0.5638	0.5643	0.5653	0.5633	0.5430	0.5435	0.5445
	0	0.7680	0.7723	0.7766	0.7851	0.7680	0.7723	0.7766	0.7851
	1	0.5633	0.5638	0.5643	0.5653	0.5633	0.5846	0.5851	0.5861
	2	0.2098	0.2063	0.2028	0.1957	0.1229	0.2139	0.2104	0.2033
0.5	-2	0.1235	0.1178	0.1120	0.1006	0.1229	0.1172	0.1115	0.1000
	-1	0.4095	0.4069	0.4042	0.3990	0.3752	0.3725	0.3699	0.3646
	0	0.6062	0.6186	0.6309	0.6555	0.6062	0.6186	0.6309	0.6555
	1	0.4095	0.4069	0.4042	0.3990	0.4439	0.4413	0.4386	0.4333
	2	0.1235	0.1178	0.1120	0.1006	0.1241	0.1183	0.1126	0.1012
0.8	-2	0.0930	0.0860	0.0790	0.0649	0.0968	0.0898	0.0828	0.0687
	-1	0.3368	0.3314	0.3260	0.3152	0.3314	0.3314	0.3260	0.3152
	0	0.5151	0.5338	0.5526	0.5901	0.5338	0.5338	0.5526	0.5901
	1	0.3368	0.3314	0.3260	0.3152	0.3314	0.3314	0.3260	0.3152
	2	0.0930	0.0860	0.0790	0.0649	0.0860	0.0860	0.0790	0.0649

## References

Alwan, L. C. (1992). Effects of autocorrelation on control chart performance. *Communications in Statistics – Theory and Methods*, 21(4), 1025-1049. doi: 10.1080/03610929208830829

## CONTROL CHARTS FOR MEAN

Alwan, L. C., & Roberts, H. V. (1995). The problem of misplaced control limits. *Journal of the Royal Statistical Society. Series C (Applied Statistics)*, 44(3), 269-278. doi: 10.2307/2986036

Dar, A. L., & Singh, J. R. (2015). The power of  $\bar{X}$ -chart in presence of data correlation. *Journal of Reliability and Statistical Studies*, 8(1), 25-30. Retrieved from <http://www.jrssi.in.net/assets/8103.pdf>

Gayen, A. K. (1952). On setting up control charts for non-normal samples. *Indian Society for Quality Control Bulletin*, 53, 43-47.

Maragah, H. D, & Woodall, W. H. (1992). The effect of autocorrelation on the retrospective  $X$ -chart. *Journal of Statistical Computation and Simulation*, 40(1-2), 29-42. doi: 10.1080/00949659208811363

Singh, J. R., Sankle, R., & Ahmad, M. (2012). Control charts for mean under correlated data. *Journal of Rajasthan Statistical Association*, 1(1), 21-30.