THE EFFECT OF CORRELATION ON $\overline{X}$ AND HOTELLING'S T^2 CONTROL CHARTS USING AVERAGE RUN LENGTH OF TYPE I ERROR

[1]Ali Adamu, [2]Sylvester Nwaosu, [3]Tela Adamu Bage

[1]Dept. of Mathematics, Federal College of Education (Tech), Gombe, Gombe State.
[2]Dept. of Mathematics, Statistics and Computer Science ,University of Agric., Makurdi, Benue State
[3]Dept. of Mathematics, Federal College of Education (Tech), Gombe, Gombe State.

Abstract

The effect of correlation on the performance of $\overline{X}$ and Hotelling's T^2 control charts were examined. Two random variables from normal distribution were generated at a specified level of correlations ranging from -0.9 to +0.9 for 5000 data points having repeated samples of size five each were simulated using R 10.2 software. Simulated data were analyzed using Shewart $\overline{X}$ -chart and Hotelling T^2 chart to investigate Type I error (α) average run lenght. It was observed that the average probability of Type I error for univariate increases as the level of correlation increases where as the average run length of Type I error for multivariate did not varied as a result of change in the level of correlation. It was observed also that multivariate chart yielded less error compared to univariate and therefore performed better. Furthermore,the average run length (ARL) of Type I error for multivariate control chart performed better than univariate $\overline{X}$ -control chart. Finally, it is recommended that level of correlation should be examined when subjecting several variable characteristics for quality testing.

1.0 Introduction

Conventional Shewart methods have successfully been the system of monitoring process performance over time (Shewart, 1931). They have been a foundation for maintaining and achieving new unprecedented levels of quality. However, these are generally categorized as univariate charts that can only be used to monitor a single variable of a stationary process. As a result of advancement in technology and customer expectations, the need has arisen to monitor more than one correlated variables simultaneously. The need to apply multivariate Statistical Process Control (SPC) became more desirable, especially with the complexity of processes and the dependency of product quality characteristics on each other (Prabhu and Runger,2007). In order for companies to remain competitive, they must achieve high levels of product quality, which is becoming challenging to achieve since quality characteristics are interrelated to each other as a result of technological advancement (Adamu,2011) . To achieve a state of statistical control, it requires a high level of knowledge regarding the process variables, the level of correlation among them and the accuracy at which they can be controlled (Neter, 2001) . The original work in multivariate quality control can be attributed to Hotelling (1947). His work led to a number of multivariate techniques to monitor quality.

When monitoring variables characteristics simultaneously, the value of correlation between the variables remained of paramount interest. The effect of correlation on statistical performance of univariate and multivariate control charts remained unanswered in the field. This paper will provide adequate investigation on the statistical performance of univariate and multivariate control charts considering the level of correlation. The univariate and multivariate control charts considered in this study are Shewart $\overline{X}$ -chart and Hotellings T^2 –chart.

2.0 Methodology

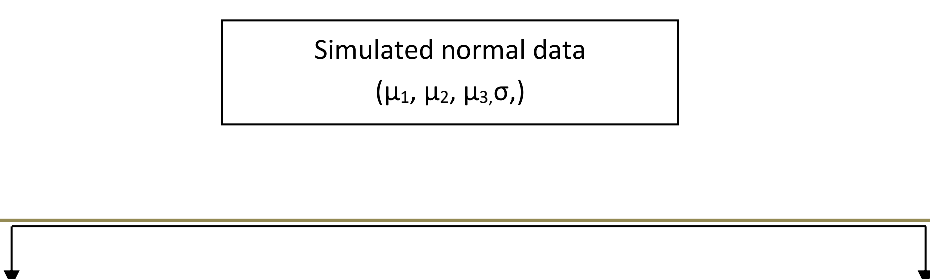
Random variables from bivariate normal distribution were generated at a specified level of correlation ranging from (-0.9) to +0.9 for 5000 data points having 5 sizes using R 10.2 software. The experiment was repeated 5 times for different values of correlation. Simulated data were analyzed by plotting the data on Shewart $\overline{X}$ -chart and Hotelling T^2 chart to account for Type I error probability (α). The values of α were obtained by recording the number of points falling outside the control limits .The objective of this simulation was to obtain the average error probability (α) based on simulated ARL. A sample of R-programme for only a case having vector 0 and covariance matrix of 0.5 is presented in Appendix 1. The chart indicating simulation procedure is presented in Fig. 1.The levels of correlation between the two variables were then varied from (-0.9) to (+0.9).

2.1 Average Run Length (ARL)

ARL can be used to evaluate the performance of control charts. ARL is the inverse probability of a single point falling outside the limits when the process is in-control (Caplen,2007). This can be calculated using the following formular.

$$ARL_0 = \frac{1}{\alpha} , \qquad (1)$$

Where α is the probability that any point will exceed the control limits.



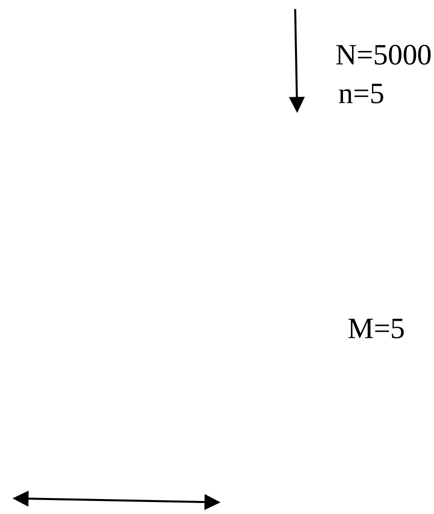


Fig.1 Simulated procedure (Type 1 Error probability)

3.0 DATA ANALYSIS

Investigation was carried out in order to account for Type 1 error probability (α). Results were obtained for Type I error probability for cases of univariates and multivariate SPCs and presented in Table 1. The SPC techniques used for the univariate and multivariate are Shewart $\overline{X}$ -chart and Hotellings T^2 chart respectively. The results were plotted on graphs shown in Figs.3 and 4. ARL of Type I error probability are computed using equation 1 above, are presented in Table 2

Table 1: Type I Error Probability

S/N	CORR (ρ)	UNI(A)	UNI(B)	AVG(UNIV)	MULTIV
1	-0.9	0.00332	0.0034	0.00336	0.00238
2	-0.7	0.0034	0.00332	0.00336	0.00224
3	-0.5	0.004	0.00404	0.00402	0.00224
4	-0.3	0.00412	0.00424	0.00418	0.00222
5	-0.1	0.0042	0.00428	0.00424	0.0022
6	0.0	0.0042	0.00448	0.00434	0.00227
7	0.1	0.00432	0.00436	0.00434	0.0022
8	0.3	0.0046	0.00456	0.00458	0.00222
9	0.5	0.00484	0.0048	0.00482	0.00222
10	0.7	0.00504	0.00524	0.00514	0.00222
11	0.9	0.00572	0.00632	0.00602	0.0022

Table 2: ARL of Type I Error for Univariate and Multivariate Cases

S/N	CORR(ρ)	ARL(B)	ARL (B)	ARL(AVG)	ARL(MULTIV)
1	-0.9	301.2048	294.1177	297.6191	420.1681
2	-0.7	294.1176	301.2048	297.6191	446.4286
3	-0.5	250	247.5248	248.7562	446.4286
4	-0.3	242.7184	235.8491	239.2345	450.4505
5	-0.1	238.0952	233.6449	235.8491	454.5455
6	0.0	238.0952	223.2143	230.4148	440.5286
7	0.1	231.4815	229.3578	230.4148	454.5455
8	0.3	217.3913	219.2983	218.3406	450.4505
9	0.5	206.6116	208.3333	207.4689	450.4505
10	0.7	198.4127	190.8397	194.5525	450.4505
11	0.9	174.8252	158.2279	166.1130	454.5455

4.0 Discussion of Results

It can be seen from table 1 that the average probability of Type I error for the two univariates SPC, increased as the level of correlation increases from (-0.9) to (0.9). Specifically, the average probability of Type I error for univariate (A) increased from 0.00332 to 0.00572 as the level of correlation increases. Also, the average probability of Type I error for univariate (B) increased from 0.0034 to 0.00632 as the level of correlation increases. Furthermore, the average of the two univariates revealed the same thing as the individual univariates. Type I error probability for the two univariates are plotted against the level of correlation and is presented on fig 2 and also the type I error probability of the average of the two univariates are plotted against the level of correlation and is presented on fig 3.

However, the probability of Type I error for the multivariate SPC revolves around 0.0022. This clearly shows that the level of correlation has no effect on Type I error of multivariate SPC. On the other hand, the level of correlation has effect on Type I error of Univariates SPC. It has generally observed from the fig 2 that multivariate SPC yielded less Type I error compared to univariate SPC. Therefore, multivariate SPC stand to perform better than the univariate SPC.

It was observed from Table 2 that the average ARL of Type I error for univariates decreased approximately from 294 to 166 as the value of correlation increased from -0.9 to + 0.9. This means even when the process remains – in - control, an out – of – control signals will be generated every 294 to 164 samples depending on the value of correlation. On the other hand, the average ARL of Type I error for multivariate did not differ much as the value of correlation increases. The ARL of Type I error for multivariate rallied around 420 to 454. This means even if the process is in control, an out – of – control signal will be generated every 420 to 454 samples depending on the value of correlation. Based on that, multivariate performed better than univariate because the multivariate takes more samples compare to univariate before recording an out –of – control signal.

Graphs for ARL of Type I error of Univariates and Multivariate and that of

average of univariates and multivariate are presented in Fig. 4 and 5 respectively.

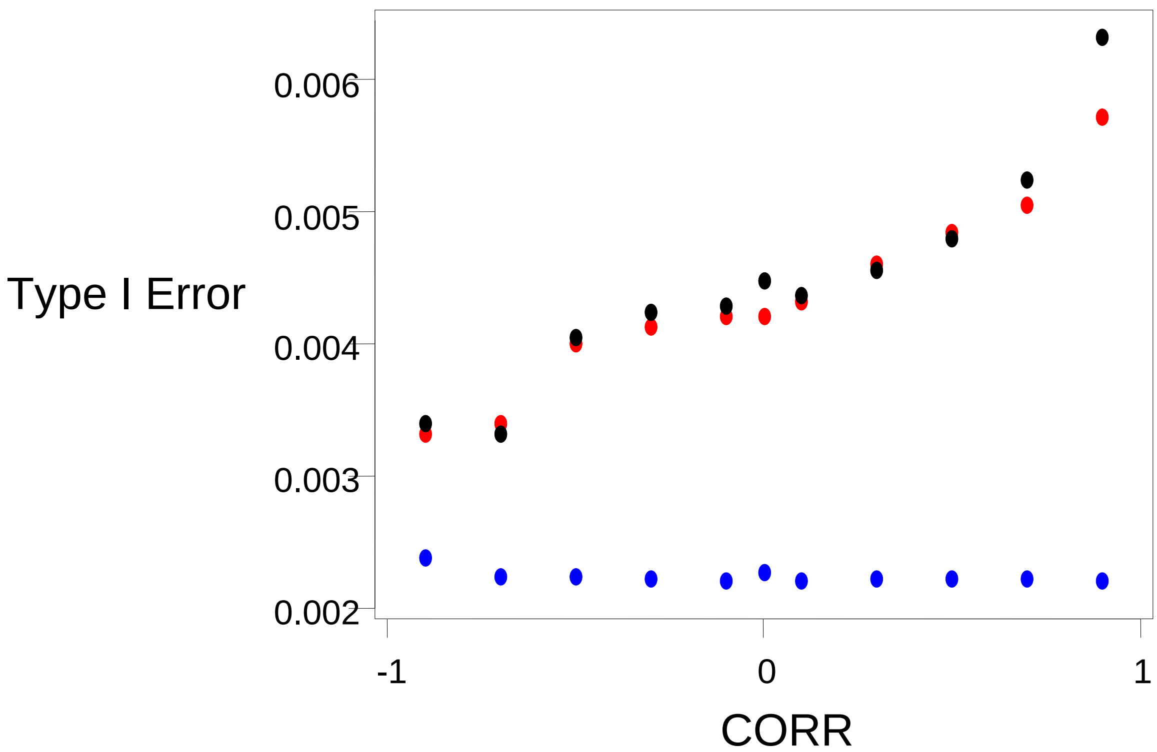


Fig 2 :Type I Error Probability of univariates and Multivariate

KEY

RED	Univariate (A)
BLACK	Univariate (B)
BLUE	Multivariate

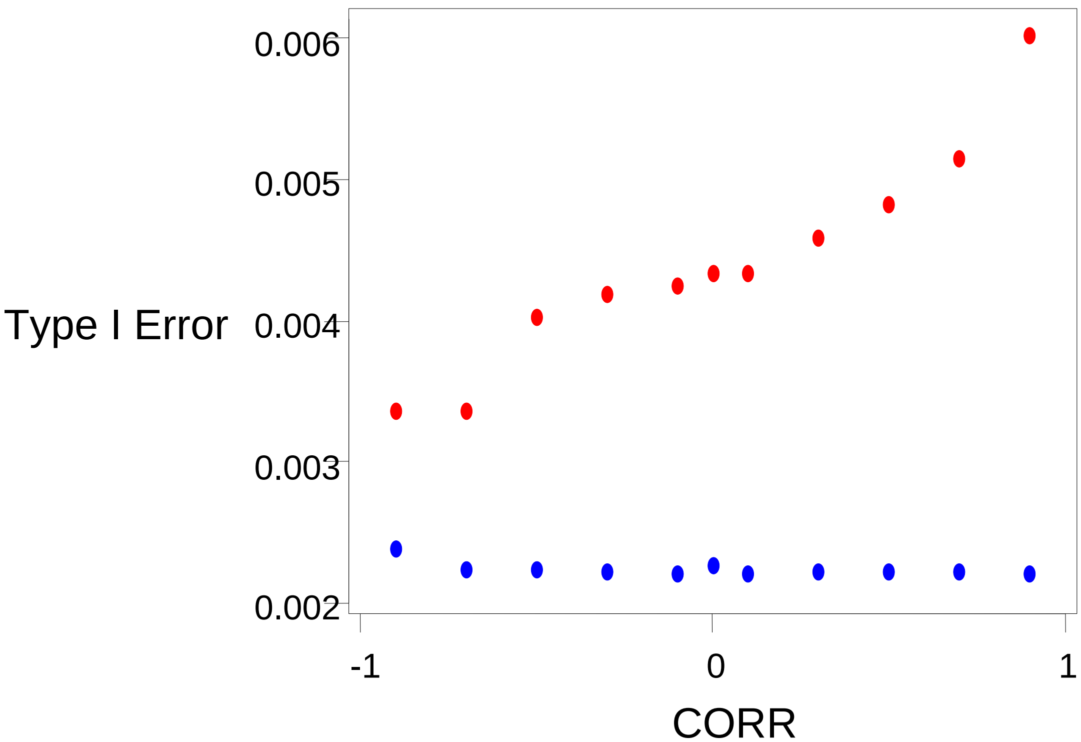


Fig 3: Type I Error Probability for average of Univariates and Multivariate

KEY

Red	Univariate
Blue	Multivariate

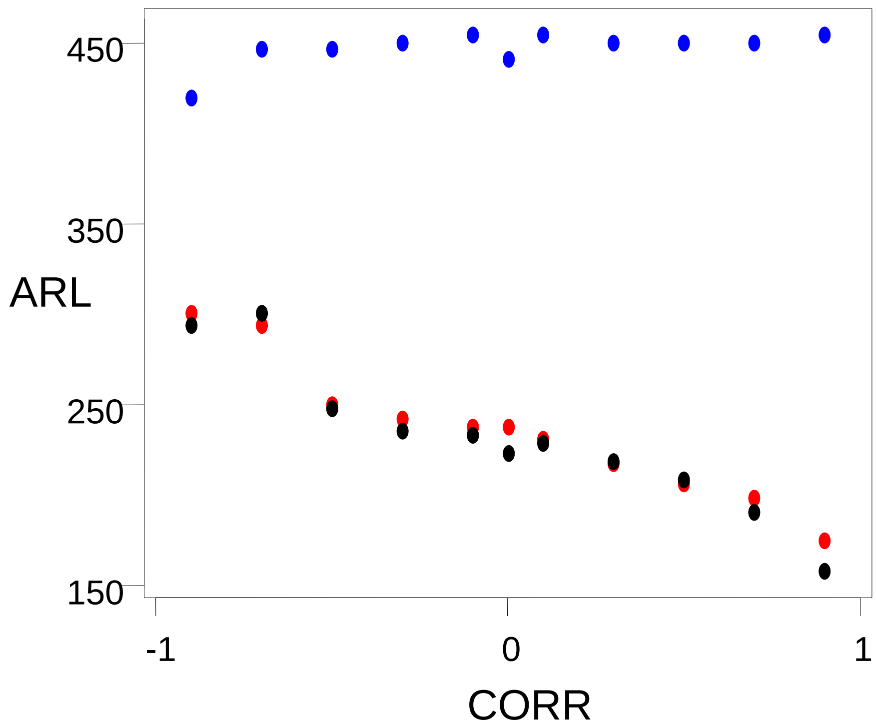


Fig 4:Average Run Length of Univariates and Multivariate

KEY

RED	ARL of Univariate (A)
BLACK	ARL of Univariate (B)
BLUE	ARL of Multivariate

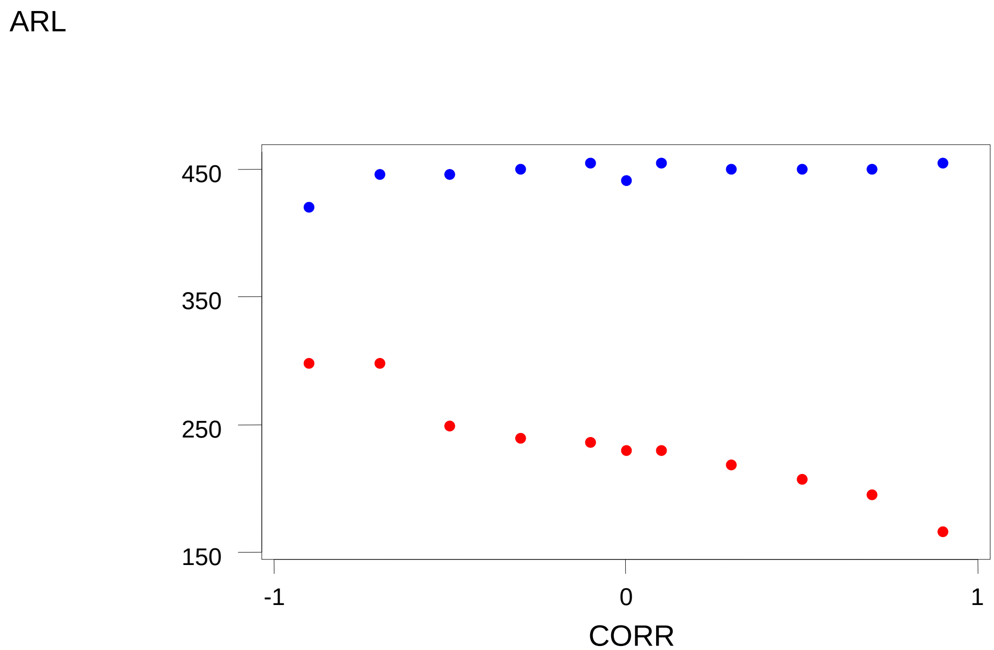


Fig .5 : ARL of Type I Error for Average of Univariates and Multivariate

KEY

Red	Univariate
Blue	Multivariate

4.0 CONCLUSION/RECOMMENDATION

From the investigations carried out, it has been observed that the average probability of Type I error for univariate vary as the level of correlation increases where as the average of Type I error for multivariate did not varied as a result of change in the level of correlation. It has been observed that multivariate chart yielded less error compared to univariate and therefore performed better.

It has been observed from average ARL of Type I error that multivariate performed better than univariate because the multivariate takes more samples compare to univariate before recording an out –of – control signal.

It is recommended that the level of correlation between the variables be examine when subjecting several variable characteristics to quality testing.

REFERENCE

Adamu,A. (2011). Statistical performance of Multivariate Charts. Mathematical and Computer Science series journal, Kolkata, India, 23,45-57.

Caplen, H. (2007). On the study of quality of some variable characteristics, *Technometrics* 30(4), 11-21.

Hotelling, H. (1947) **Multivariate quality control**, Illustrated by the air testing of Sample bombsights, in Techniques of Statistical Analysis

Montgomery,D.C. (2008) **Introduction to Statistical Quality Control,** 6th edition, John Wiley, New York.

Neter (2001): Identification of out of control quality characteristics in a Multivariate manufacturing environment. *Commun. Statis.- Theory Methods 20*.

Prabhu, S.S. and Runger,G.C.(2007).Designing a EWMA Multivariate CUSUM Charts, *Journal of Quality Technology, 31, Pp. 66-67.*

Shewhart, W.A. (1931). **Economic Control of Quality of Manufacturing Product.** Princeton, N.J. Van Nostrand Reinhold.

APPENDIX 1

```
rm(list=ls())

library(MASS)

set.seed(1234)

muvec1=c(0,0)

sigma1=matrix(c(1,0.5,0.5,1),ncol=2,byrow=T)

muvec2=c(0,0)

sigma2= matrix(c(1,0.5,0.5,1),ncol=2,byrow=T)

muvec3=c(0,0)

sigma3= matrix(c(1,0.5,0.5,1),ncol=2,byrow=T)
```

```
muvec4=c(0,0)

sigma4= matrix(c(1,0.5,0.5,1),ncol=2,byrow=T)

muvec5=c(0,0)

sigma5= matrix(c(1,0.5,0.5,1),ncol=2,byrow=T)

nsim=5000

x=mvrnorm(5000,muvec1,sigma1)

y= mvrnorm(5000,muvec2,sigma2)

k= mvrnorm(5000,muvec3,sigma3)

q= mvrnorm(5000,muvec4,sigma4)

w= mvrnorm(5000,muvec5,sigma5)

z=rbind(x,y,k,q,w)

write.table(round(z,4),"c:\\ali\\newdata.txt",col.names=F,row.names=F)
```