In manufacturing industry, data collection is expensive and time consuming. Therefore, to monitor the quality of production process, the use of small sample is strongly required. Unfortunately, the current techniques of multivariate process variability monitoring for small sample are developed based on heuristic approach. As a consequence, their reliability is unknown. This book introduces a reliable version of GV-Chart (the most adopted technique) and a reliable version of VV-Chart. Since both charts are complementary to each other, their simultaneous use is recommended. Furthermore, for practical purposes, the book is completed with statistical tables of reliability constant for some selected sample sizes and selected number of quality characteristics. Real industrial applications are also provided to illustrate the advantages of the two reliable charts.

Maman A. Djauhari was born in Dayeuh Garut, Tanah Pasundan, Indonesia, in 1948. He got his Doctorandus in Mathematics from Institut Teknologi Bandung (ITB), Indonesia, in 1974. His Doctorate in Pure and Applied Mathematics was obtained from Université Montpellier 2, France, in 1979. He was Dean and Chairman of Professors Council at ITB.

978-613-9-58609-7

Djauhari, Zolkeply, Herwindiati

Maman Djauhari
Zunna'aim Zolkeply
Dyah Herwindiati

Reliable Shewhart-type Control Charts for Multivariate Process Variability

LAP LAMBERT Academic Publishing

Maman Djauhari
Zunna'aim Zolkeply
Dyah Herwindiati

Reliable Shewhart-type Control Charts for Multivariate Process Variability

Maman Djauhari
Zunna'aim Zolkeply
Dyah Herwindiati

Reliable Shewhart-type Control Charts for Multivariate Process Variability

LAP LAMBERT Academic Publishing

Imprint
Any brand names and product names mentioned in this book are subject to trademark, brand or patent protection and are trademarks or registered trademarks of their respective holders. The use of brand names, product names, common names, trade names, product descriptions etc. even without a particular marking in this work is in no way to be construed to mean that such names may be regarded as unrestricted in respect of trademark and brand protection legislation and could thus be used by anyone.

Cover image: www.ingimage.com

Publisher:
LAP LAMBERT Academic Publishing
is a trademark of
International Book Market Service Ltd., member of OmniScriptum Publishing Group
17 Meldrum Street, Beau Bassin 71504, Mauritius

Printed at: see last page
ISBN: 978-613-9-58609-7

RELIABLE SHEWHART-TYPE CONTROL CHARTS FOR MULTIVARIATE PROCESS VARIABILITY

By:
Maman Abdurachman Djauhari,
Zunna'aim Zolkepley and Dyah Erny Herwindiati

FOREWORD FROM THE PRESIDENT OF MALAYSIAN ACADEMY OF MATHEMATICAL SCIENTISTS

The experience of advanced nations shows that the industrial sector especially in manufacturing is one of the important factors to generate economic growth of a nation. The role of the industrial sector to provide opportunities for employment, inventing new technologies and sustain economic growth is of great importance. The driving force behind the progress of the industrial sector needless to say is the knowledge-workers. A nation needs to develop such human capital in order for it to continue progress and advance to more elevated level with sound economic activities that will lead to generation of higher income among its populace. It is one of the pillars that will support and maintain social stability of a nation. It is due to this that the industrial sector ought to remain vibrant and sustain its strength for as long as it is possible.

One of the factors that are essential in maintaining strength and always ready to face up to any challenges as time changes that the main players in industry should not neglect is the quality of the manpower that shoulder the task of maintaining quality of its production processes. Exposure must be given to the management of industries and their workforce to the tools that are needed to do so. They should be knowledgeable in the theoretical and practical knowledge of the task they are shouldering. At the management level at least visionary personnel are needed who have the ability to forecast the future needs of the communities and design plans and programs to meet the forthcoming demands of the consumers. In this respect to have the deep knowledge, skill and ability to measure the quality of production process is indeed essential.

In this book the authors highlight the essential ingredients in multivariate statistical process control. With ample background to the development of measuring the quality of production process since early twenties of the last century to the present the authors have proposed a more refined method to do so. Led by Maman Djauhari the authors have successfully present their thoughts in clarity which I feel is easily understood by practitioners in this field. It offers an understanding on the current practice in multivariate statistical process control and its limitations, proposing the

use of reliable *GV*-chart and *VV*-chart for small sub-group size which previously had been treated heuristically.

To succeed in today's global economy players in industry especially the production sector need to keep abreast with the necessary knowledge development practically and theoretically. This is a book that entities in industry, executives, and students of higher learning aspiring to excel in this area should have in their possession and should be on the shelves of any reputable institutions of higher learning.

Maman Djauhari is a humble man of knowledge very passionate about rudiments of statistics who is prolific in his writings and frequently become main speakers at conferences in his field of expertise. His presentations often are very well received by the audience and spark wide interest in the topic of his discussion. He is very skilful and with the depth of his knowledge he is able to impress well those who read his writings and attend his talks on the importance of rudiments of statistics employable in finding solutions to problems besetting the society and environment. This book is the latest that he leads in authoring and it is indeed another that experts in this field should have in their possession.

Kamel Ariffin Mohd Atan
Emeritus Professor at Universiti Putra Malaysia,
Fellow of Academy of Sciences Malaysia,
President of Malaysian Academy of Mathematical Scientists

FOREWORD FROM THE DIRECTOR OF INSTITUTE FOR MATHEMATICAL RESEARCH (INSPEM)

Maintaining high level product and service quality are the main endeavours of many organizations. Thus there is a great need for these organizations to control and improve quality. Shewhart introduced the univariate statistical process control which involved only one characteristic to measure the quality of a production process. As time goes by there is a need in the industry to monitor two or more related characteristics simultaneously. Process monitoring problems in this nature is known as multivariate statistical process control. The multivariate process control techniques were established by Harold Hotelling in 1947. The most useful tool of multivariate statistical process control is the quality control chart.

In the authors' views, both the multivariate process variability monitoring and the multivariate process target monitoring are important in the production process. However, not much has been emphasized on the multivariate process variability monitoring. Thus, the special feature of this book is on the multivariate process variability monitoring. Currently the most adopted tool utilized to monitor process variability is the generalized variance-based control chart (*GV*-chart) which unfortunately has serious limitation. Complementary to this chart is the vector variance chart (*VV*-chart).

The reliability of both charts can be determined for large sub-group. For small sub-group size the only related chart available in the literature is the one constructed based on heuristic approach which means that the probability of false alarm (PFA) cannot be determined. This book will thus address this issue and provide a reliable *GV*-chart and a reliable *VV*-chart when the sub-group size is small. For practicality, this book also provides the readers with statistical tables of reliability constant.

Much credit should be given to the authors for their great effort and contribution in keeping the goal to bring a process into a state of statistical control. I believe many will gain benefit from this book and enjoy reading it. The authors' contribution in this field is indeed laudable.

Noor Akma Ibrahim
Professor of Statistics,
Director,
Institute for Mathematical Research (INSPEM)
Universiti Putra Malaysia

PREFACE FROM THE AUTHORS

A one page proposal submitted by Walter Andrew Shewhart to Western Electric Company on May 16, 1924 is the pioneering work that changed the way industries treat themselves. That paper is the origin of what we call now control chart diagram. It changed the paradigm on the way industry deals with production process. It is a tool for distinguishing chance-cause variation from assignable-cause variation. The goal is to bring a process into a state of statistical control. That's it! It is the state where there is only chance-cause variation occurs. Keeping the process is this state is necessary for quality professionals to predict what will happen in the future.

In Shewhart's proposal, only one characteristic was involved to measure the quality of the production process. Today, it is termed as univariate statistical process control (USPC). In this circumstance, Shewhart came up with this famous jargon: "quality is reciprocal to variability." It means that, to improve the process quality is to reduce the process variation. In William Edwards Deming's words, as cited in Neave (1990), "If I had to reduce my message for management to just a few words, I would say it all had to do with reducing variation."

Around a decade later, Harold Hotelling introduced the "T-squared distribution" which is a multivariate generalization of student-t distribution. The marriage of Walter Shewhart's work with this Hotelling's innovation has made complex process control operation into practice. By complex process is meant the process where its quality is measured by two or more correlated characteristics. Statistically speaking, this is the so-called multivariate statistical process control (MSPC).

This book is a special literature focusing on MSPC and more specifically on multivariate process variability monitoring. Why variability? There are two main reasons. *First*, variability is difficult to understand and thus to measure and manage. However, customer only feels the variability of the process and/or products and not the other parameters. See General Electric (1998) for an example of this phenomenon. In this regards, Snee (2006) has remarked variability as a venerable subject in statistical literature. That the variability is difficult to manage but important is described by Lara Boyd, a brain researcher at University of British Columbia, Vancouver, Canada, in her TEDx Talks, published on15 December 2015. Her talk is

available in https://www.youtube.com/watch?v=LNHBMFCzznE (accessed on 15 March 2018). Boyd said: "As a researcher, variability used to drive me crazy. It makes it very difficult to use statistics to test your data and ideas. And because of this, medical intervention studies are specifically designed to minimize variability." Furthermore, she added "In my research it is becoming really clear that the most important, the most informative data we collect is showing this variability." *Second*, as Hoerl and Snee (2012) have remarked, understanding variation should be a core competency of statisticians and all other quality professionals. These professionals should think deeply about variation and its effects.

In what follows, we begin with a brief discussion on the current practice of multivariate process target monitoring which is as important as multivariate process target monitoring but has received less attention in the literature. Then, we show the need of reliable control chart.

1. Current practice

Quality professionals agree that process variability and its counterpart, process target, have the same importance in production process (e.g. Montgomery, 2005, 2009). They have to be monitored continuously. However, the former has received less attention compared to the latter. This is perhaps due to the fact that any scalar measure of multivariate variability cannot explain the complex structure of process variability. As remarked by Reis and Rato (2013), the most adopted tool to monitor process variability is the generalized variance-based control chart (*GV*-chart in brief). However, as will be discussed later, this chart has a serious limitation. This leads Djauhari et al. (2008) to introduce vector variance chart (*VV*-chart) which is complementary with generalized variance chart. See Ryan (2011) for a discussion on *VV*-chart.

In the current practice, the use of *GV*-chart and of *VV*-chart for large sub-group size is distinguished from that for small sub-group size. For large sub-group size, the control limits of these charts are determined by using normal approximation. Thus, the control limits which refer to probability of false alarm (PFA) are approximated. Meanwhile, for small sub-group size, they are defined by using heuristic approach. As a consequence, for small sub-group size, the PFA cannot be specified. Consequently, the reliability of these charts cannot be determined.

Grammatically, the word "reliable" is an adjective. It means "consistently good in performance or quality." In other words, it means "able to be trusted." In relation with SPC, since control chart is a tool, it must be reliable. It is a tool which provides the main source of information about the history of process performance. In its turn, it leads to a decision making on the status of process quality. It then must be trusted. In practice, its reliability is determined by (i) the appropriateness of statistical test used in the construction of control chart, and (ii) the control limits which reflect the desired PFA.

How to determine the reliability of control chart if the sub-group size is small? This is the topic of the book.

2. Reliable control chart

The book consists of two proposals to handle the case where the sub-group size is small. The first proposal is to use a reliable *GV*-chart that will be described later. The second one is to use a reliable *VV*-chart. In order to construct these two reliable control charts, the notion of reliability constant will be introduced. In both charts, this constant reflects the probability of false alarm (PFA). It depends on the sub-group size n and the number of quality characteristics p.

In this book, the reliability constant is determined by using simulation experiment for each pair of n and p. Furthermore, to make it practical, this book is completed with the statistical tables of reliability constant.

Since both charts are complementary to each other, a simultaneous use of them is recommended. Have a wonderful journey!

Maman Abdurachman Djauhari
Zunna'aim Zolkeply
Dyah Erny Herwindiati

ACKNOWLEDGEMENT

The first author is very grateful to (Late) Professor Dr. Achmad Arifin, Department of Mathematics, Institut Teknologi Bandung, Indonesia, for the discussion on research methodology 36 years before Malcolm Gladwell discussed the "10K Hours Formula" in his book "*Outliers*" published in 2008. Special thanks go to Emeritus Professor Dato' Dr. Hj. Kamel Ariffin Mohd Atan, the Director of Institute for Mathematical Research (INSPEM), Universiti Putra Malaysia, in the period of April 1, 2002 until February 15, 2016 and Professor Dr. Noor Akma Ibrahim, the current Director, for giving the opportunity to him to be part of INSPEM where he had initiated this research. We are in debt to Professor Dr. Yves Escoufier, Université Montpellier 2, France, for an in-depth discussion related to Chapter 5.

The second author is in debt to the Faculty of Science, Technology and Human Development, Universiti Tun Hussein Onn, Malaysia, for supporting him with all the facilities to conduct his PhD program. Special thanks from the third author go to the Faculty of Information Technology, Universitas Tarumanagara, Indonesia, for the opportunity to contribute in this research.

All authors also thank:

1. Centre for Research in Statistics and Data Analysis, Tjahaja Bina Statistika Indonesia (TBSI), Ltd. P. (*Commanditaire Vennotschap*) for the facility to do the important parts of this research.
2. Rian F. Umbara, staff of TBSI for reading and preparing the final manuscript.
3. All staff of TBSI for providing excellent work environment.

Their full supports have made this book's writing a wonderful and fruitful adventure. Last but not least, the authors would like to thank the publisher for handling the mandscript professionally.

Maman Abdurachman Djauhari
Zunna'aim Zolkeply
Dyah Erny Herwindiati

DEDICATION

This book is dedicated to those who strive for the best quality of life.

Maman Abdurachman Djauhari
Zunna'aim Zolkeply
Dyah Erny Herwindiati

TABLE OF CONTENTS

CHAPTER 1
INTRODUCTION

In statistical process control (SPC), control chart is the main source of information about the history of process performance. This information leads to a decision making on the process quality. Therefore, control charts must be reliable. In practice, their reliability is determined by,

(i) The appropriateness of statistic used to construct the control chart,

(ii) The control limits which reflect the desired probability of false alarm (PFA).

It is worth noting that the word "reliable" is an adjective. It means "consistently good in quality or performance." Therefore, a reliable control chart is the one which is "able to be trusted."

This book is focused on multivariate statistical process control (MSPC), a branch of SPC dealing with controling/monitoring simultaneously several correlated variables (also called quality characteristics). And more specifically, it is focused on multivaiate process variability (MPV) monitoring. As remarked in the literature such as Montgomery (2001, 2005, 2009), all quality professionals know that MPV is as important as process target to be monitored. The techniques available to conduct this monitoring operation will help us to detect the special causes of MPV. When a process is statistically under control, then only common causes affect MPV.

In the last few years MPV monitoring operation has been thoroughly studied. And nowadays, many techniques and tools for this operation are available in the literature. However, according to Ries and Rato (2013), GV-based control charts are still the most adopted control charts to do this job. For example, recently Carlos Garcia-Diaz (2007) has developed a control chart based on the so-called effective variance (EV). A special feature of this chart is that it can be used to compare the variability of two processes with different number of quality characteristics. And we know that EV is a variant of generalized variance (GV). Due to its importance, we start our discussion on the *GV*-chart.

GV-chart is easy to construct and to interprete. The geometrical interpretation and distributional behavior of GV are interesting subject of discussion in the literature since the last five decades. See, for exampe, Anderson (1966, 1984) for an early development of the geometrical interpretation, Mason et al. (2009) the geometrical interpretation in MSPC, and Djauhari (2009) for a distributional behavior. This is perhaps the reason why, despite of serious limitations of GV as a measure of MPV, it becomes the most popular and widely used in practice.

Alt and Smith (1988), see also Montgomery (2001, 2005, 2009) among others, have specified the limitations of GV. This motivates Djauhari (2007) to questioning the appropriateness of GV as a measure of MPV. Then, in that paper he introduces the so-called vector variance (VV) as another measure and shows that VV and GV are complementary. The advantage of *VV*-chart in MPV monitoring compared to *GV*-chart, is demonstrated in Djauhari et al. (2008) and a recommendation to use them simultaneously is given in Djauhari and Mohamad (2010). This leads us to put *VV*-chart as the second topic of discussion in this book.

The problem reveals when we deal with small sample size. A preliminary study in Djauhari et al. (2016), involving samples of small, moderate and large (but not sufficiently large) size, shows that the control limits of both *GV*-chart and *VV*-chart have nothing to do with the probability of false alarm (PFA); they do not reflect the PFA. This means that the presence of an out-of-control (OOC) signal cannot probabilistically be trusted. Its reliability is unknown. Through simulation experiments, in this book we show that the results of this preliminary study are justified. Therefore, we expand this study to a more general sample size with general number of quality characteristics. Furthermore, to overcome the non-reliability problem, a reliability constant is introduced and simulation experiments were conducted to find its value. This will leads us to the construction of a reliable *GV*-chart and a reliable *VV*-chart. Finally, to facilitate the practitioners with these reliable charts, tables of reliability constant are provided and industrial examples are presented.

The remaining chapters are organized as follows. Chapter 2 is devoted to recall the most popular and important variability measures in univariate case and multivariate case as well. It is followed by Chapter 3 which focuses on classical *GV*-chart. "Classical" means that the control limits for small sample size are determined heuristically while normal approximation is used for large sample size. Later on, a reliable *GV*-chart for small sample size is discussed in Chapter 4. In this chapter, tables of reliability constant for *GV*-chart are provided for some selected PFAs. Then, an industrial example is presented to illustrate the need for reliable control chart. Our discussion is continued with *VV*-chart in Chapter 5. The book ends in Chapter 6 with a reliable *VV*-chart for small sample size, tables of reliability constant for *VV*-chart for some selected PFAs, and an industrial example to illustrate the advantage of this chart.

CHAPTER 2
MULTIVARIATE PROCESS VARIABILITY MEASURES

Process is anywhere and everywhere. Statistically, its behavior is represented by the so-called probabilistic distribution that governs the process. In this language, process target and process variability are represented by the so-called measure of central tendency and measure of dispersion, respectively. The central tendency refers to a location at which the distribution is concentrated. Meanwhile, the dispersion refers to the way all possible values of the variable(s) under study are scattered. Since the concern of this book is on process variability, in this chapter we recall some important measures of variability in multivariate process. Just for the sake of continuity, some measures of univariate process variability will also be recalled but not discussed. By univariate process we mean the process where only one quality characteristic is to be monitored. If there are more than one quality characteristic, and their correlations are taken into account, the process is called a multivariate process.

In univariate process, range, variance and standard deviation are the most widely used variability measures in SPC. There are also other measures such as Interquartile range, Luceno's measure, and quantile-based measure. Interquartile range is the range between the third quartile Q_3 and the first Q_1. What are the last two measures? Luceno's measure is defined as

$$L = \frac{1}{n}\sqrt{\frac{\pi}{2}}\sum_{i=1}^{n}\left|X_i - \bar{X}\right| \text{ where } \bar{X} = \frac{1}{n}\sum_{i=1}^{n}X_i,$$

where $X_1, X_2, \ldots, X_n$ is a sample of size n. Meanwhile, quantile-based measure is

$$\frac{x_{0.99865} - x_{0.00135}}{6}$$

where x_α is the $(1-\alpha)$-th quantile. These last two measures are derived from the capability indices C_{pc} and $C_p(q)$, respectively. A good discussion of these measures can be found in Montgomery (2001, p. 364-365).

In multivariate process, the variability of the process is difficult to measure. Numerically, it is represented in the form of covariance matrix. The difficulty appears because there is no single scalar measure that can be used to represent the complex covariance structure contained in this matrix. However, the following two measures are the important ones that usually used in multivariate statistical process control (MSPC); generalized variance (GV) and vector variance (VV). Other measures are also available in the literature of multivariate analysis but not very much appreciated in MSPC. We can say, for example, total variance (TV), square root of generalized variance (SRGV), relative generalized variance (RGV) and minimum volume ellipsoid (MVE). Some authors such as Alt and Smith (1988) and Montgomery (2001, 2005) have discussed the application of SRGV in MSPC. See also Djauhari (2005) for an improved version of SRGV. However, in practice, SRGV is dominated by GV. Meanwhile, an application of RGV can be found in Tang and Barnett (1996). Regarding MVE, it is usually empoyed in the literature of robust statistics. See, for example, Rousseeuw (1985), and Grambow and Stromberg (1998) for the details.

Another example is effective variance (EV). It is the geometric mean of all eigenvalues of Σ and defined as $\sqrt[p]{|\Sigma|}$. See Serfling (1980) and Pena and Rodriguez (2005) for the definition and Carlos Garcia-Diaz (2007) for its application in MSPC. This measure is usually used to compare two processes, in terms of their variability, when these processes involve different number of quality characteristics. It is worth noting that EV, is a variant of GV. Therefore, it inherits the property of GV.

Due to the limitation of EV, caused by the limitation of GV, in this book we introduce another measure based on VV called "effective vector variance (EVV)." We define $EVV = \frac{1}{p}Tr\left(\Sigma^2\right)$. It is the arithmatics mean of the squared eigenvalues of Σ. Here, p is the number of quality characteristics involved in the process monitoring. Like EV, EVV can also be used to compare two processes involving different number of quality characteristics.

In the rest of the book the discussion will be focused on GV and VV. We start with GV, the most popular and widely used measure in MSPC. Its advantages as well as its weaknesses will be discussed. It is to handle its weaknesses that VV is introduced as an alternative variability measure. We show that a simultaneous use of both GV and VV will be advantageous.

2.1. Generalized variance

There is no doubt that GV has a special role in monitoring covariance structure stability in a wide range of scientific investigations; from soft sciences to hard sciences, and from service industry to manufacturing industry. We can mention, for example, Beamon and Ware (1998) who employ GV in supply chain management, Edelman and Rao (2005) in theoretical physics, and Hubert et al. (2008) in astronomical data. We can also mention Wood (1994), Roes and Dorr (1997), and Sulek (2004) for its application in service industry, Sellick (1993), Shahian et al. (1996), Hanslik et al. (2001), and Woodall (2006) in health care industry, Ye et al. (2001) and Kruegel et al. (2005) in information industry, Florac et al. (2000) and Jakolte and Saxena (2002) in software industry, Ragea (2003) and Da Costa et al. (2005) in financial industry, and Woodall and Montgomery (1999), Djauhari (2005), Sullivan et al. (2007), and Mason et al. (2009) in manufacturing industry. Many more applications can easily be found in the literature showing the importance role of GV in real application.

In the field of MSPC, all quality professionals know that process variability is very important to be monitored. It is as important as monitoring the process target. They understand that the former is more difficult than the latter. This is most probably caused by the fact that a scalar measure, and even several scalar measures used simultaneously, cannot be used to explain the whole covariance structure summarized in Σ. Take GV as an example. It is the determinant Σ. Therefore, according to the property of determinant operator, two different processes could have the same GV. In other words, according to GV, two different processes could have the same variability. Thus, when *GV*-chart is used to monitor the process variability, the chart is only a necessary chart. In other words if the process is in-control, then *GV*-chart is in-control but not vice versa. Furthermore, its sampling distribution tends very slowly to normality. Thus, it is not apt for practical purpose because in real application the sub-group size (i.e. the sample size) n is very limited.

One of the most important advantages of GV lies in its sampling distribution. As we will see in Sub-section 2.1.4, for $p > 2$, its distribution can be approximated by a very familiar distribution, say, normal distribution when n is sufficiently large. See also Muirhead (1982) and Anderson (1984) for the proof. However, when n is small, normal approximation is no longer appropriate. In this case, as can be seen in the literature such as Montgomery (2005, 2009), the current practice of process variability monitoring operation is by using *GV*-chart based on heuristic approach. The control limits are derived heuristically from the property that "*most of the probability distribution of* $|S|$ *is contained in the interval* $E(|S|) \pm 3\sqrt{Var(|S|)}$."

In the next chapter, we show that this property will lead to a non-reliable control chart. This is a serious weakness of *GV*-chart for small n. For further discussions on the advantages and weaknesses of GV, the readers are suggested to consult Alt and Smith (1988), Montgomery (2005) and Djauhari (2005).

Despite its weaknesses, due to the above desirable property, in the next paragraph we study the sampling distribution of GV. Before doing so, we first recall the classical asymptotic distribution. Later on, the parameter estimates issued from several independent random samples will also be derived. These estimates will provide the basic tools in process variability monitoring.

In the next sub-section, we start our presentation with geometrical interpretation of GV followed consecutively by its distributional behavior, delta method to derive the distribution of a function of random variables, and finally the asymptotic sampling distribution of GV.

2.1.1. Geometrical interpretation

In this book, unless otherwise, a random sample $X_1, X_2, \ldots, X_n$ of size n will always be considered drawn from $N_p(\mu,\Sigma)$; a p-variate normal distribution with mean vector μ and covariance matrix Σ positive definite. We assume $n > p$. The mean vector and covariance matrix of this sample are,

$$\bar{X} = \begin{pmatrix} \bar{x}_1 \\ \bar{x}_2 \\ \cdot \\ \cdot \\ \cdot \\ \bar{x}_p \end{pmatrix} \text{ and } S = \begin{pmatrix} s_{11} & s_{12} & \cdots & s_{1p} \\ s_{21} & s_{22} & \cdots & s_{2p} \\ \cdot & \cdot & \cdots & \cdot \\ \cdot & \cdot & \cdots & \cdot \\ \cdot & \cdot & \cdots & \cdot \\ s_{p1} & s_{p2} & \cdots & s_{pp} \end{pmatrix}.$$

where $\bar{x}_i$ is the sample mean of the i-th variable and s_{ij} is the sample covariance of

the i-th and j-th variables. Thus, s_{ii} is the variance of the i-th variable. These are computed as follows,

$$\bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i \text{ and } S = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \bar{X})(X_i - \bar{X})^t.$$

The matrix S is symmetric, i.e. $s_{ij} = s_{ji}$. Since $n > p$, S is also positive definite. It is a numerical representation of process variability in sample version. To understand its geometrical interpretation, we need to compute the so-called correlation matrix denoted by R,

$$R = \begin{pmatrix} r_{11} & r_{12} & \cdots & r_{1p} \\ r_{21} & r_{22} & \cdots & r_{2p} \\ \cdot & \cdot & \cdots & \cdot \\ \cdot & \cdot & \cdots & \cdot \\ \cdot & \cdot & \cdots & \cdot \\ r_{p1} & r_{p2} & \cdots & r_{pp} \end{pmatrix}.$$

This matrix is obtained from S through the transformation $r_{ij} = \dfrac{s_{ij}}{\sqrt{s_{ii}s_{jj}}}$.

Algebraically, if θ_{ij} is the angle between the i-th and j-th centered observed variables, $r_{ij} = \cos(\theta_{ij})$. Therefore,

1. $r_{ii} = 1$ for all $i = 1, 2, \ldots, p$
2. $-1 \le r_{ij} \le 1$

The two matrices, R and S, provide information about the geometric configuration of n data points in a real vector space $E = \mathbb{R}^p$ of p dimension and of p centered variable points in another real vector space $F = \mathbb{R}^n$ of n dimension. In F, the length of the i-th variable is $\sqrt{s_{ii}}$ (standard deviation) and the angle between the i-th and j-th centered variables is $\theta_{ij} = \arccos(r_{ij})$. The correlation matrix R alone can then be considered as a numerical representation of p centered variable points on the surface of a p-dimensional ball of radius 1 in F. Each point represents a standardized variable. Thus, by looking at R and S, we see the cloud of n centered data points in E and the cloud of p standardized variable points in F.

The form of data cloud in E reflects the variability of the process whether it is dispersed or concentrated. Anderson (1984) suggests to measure the dispersion level of the cloud by using the volume of *parallelotope* spanned by the p variables. The larger the volume, the higher the level of process variability. He shows that, see Anderson (1984, Theorem 7.5.1, p. 260), the sample GV, denoted by $|S|$, is the squared volume of parallelotope divided by $(n-1)^p$. Accordingly, GV satisfies these properties,

1. It is non-negative. It is 0 if S is a zero matrix or if there is a variable which can be written as a linear combination of the other(s),
2. If at least one variable is near the hyperplane spanned by other variables, then GV is small. In other words, GV is small if there is a variable which is almost a linear combination of the other(s).
3. Suppose all diagonal elements of S are fixed. Then, GV will be maximum if all variables are independent,
4. Suppose all off diagonal elements of S are fixed. If at least one diagonal element becomes smaller (or larger), then GV gets smaller (or larger),

2.1.2. Distributional behavior

Anderson (1984) provides a proof that sample GV follows this exact distribution,

$$|S| \sim \frac{|\Sigma|}{(n-1)^p} \prod_{k=1}^{p} \chi^2_{n-k} \tag{2.1}$$

where $\chi^2_{n-1}, \chi^2_{n-2}, \ldots, \chi^2_{n-p}$ are independent each of which is distributed as chi-squared distribution with degrees of freedom $(n-1), (n-2), \ldots, (n-p)$, respectively. In this book, see Sub-section 2.3.5, we use Cholesky decomposition to derive this distribution. We will see that this approach is much more attractive than the classical approach.

It is very unfortunate that for $p > 2$ the distribution (2.1) cannot be written in a closed form. As a Conseqcence, it is only apt in practice for $p = 1$ and 2. In univariate case, i.e. $p = 1$, it is well known that the sample variance folows $\frac{\sigma^2}{n-1}\chi^2_{n-1}$ where σ^2 is the population variance. Meanwhile, in bivariate case, i.e. $p = 2$,

$$|S| \sim \frac{|\Sigma|}{(n-1)^2} \chi^2_{n-1} \cdot \chi^2_{n-2}$$

can be simplified as $|S| \sim \frac{|\Sigma|}{(2n-2)^2}\left(\chi^2_{2n-4}\right)^2$. See again Anderson (1984) for the proof. Thus, in these two cases, (2.1) can be expressed in a closed form and its quantile can well be determined. But, when $p > 2$, this is not the case.

Nevertheless, (2.1) provides the moments of $|S|$. In fact, the the k-th moment, see Anderson (1984), is

$$E\left(|S|^k\right) = E\left[\left(\frac{|\Sigma|}{(n-1)^p} \prod_{k=1}^{p} \chi^2_{n-k}\right)^k\right]$$

$$= \left(\frac{2}{n-1}\right)^{pk} |\Sigma|^k \prod_{j=1}^{p} \frac{\Gamma\left(k + \frac{n-2j+1}{2}\right)}{\Gamma\left(\frac{n-2j+1}{2}\right)}.$$

Here, the gamma function $\Gamma(x)$ is,

$$\Gamma(x) = \int_0^\infty t^{x-1} e^{-t} dt.$$

Accordingly, the first moment is,

$$E(|S|) = E\left(\frac{|\Sigma|}{(n-1)^p} \prod_{i=1}^{p} \chi^2_{n-k}\right)$$

$$= \frac{|\Sigma|}{(n-1)^p} \prod_{k=1}^{p} (n-k)$$

and the second is,

$$E\left(|S|^2\right)=\left(\frac{2}{(n-1)}\right)^{2p}|\Sigma|^2\prod_{k=1}^{p}\frac{\Gamma\left(\frac{1}{2}(n-k)+2\right)}{\Gamma\left(\frac{1}{2}(n-k)\right)},$$

or simply,

$$E\left(|S|^2\right)=\frac{|\Sigma|^2}{(n-1)^{2p}}\prod_{k=1}^{p}((n-k)+2)(n-k)).$$

These two moments give us the variance of sample GV,

$$Var(|S|)=E\left(|S|^2\right)-\{E(|S|)\}^2$$

$$=\frac{|\Sigma|^2}{(n-1)^{2p}}\prod_{k=1}^{p}(n-k)\left(\prod_{j=1}^{p}(n-j+2)-\prod_{i=1}^{p}(n-i)\right).$$

Knowing the mean $E(|S|)$ and variance $Var(|S|)$ we can use these parameters to derive unbiased estimates of $|\Sigma|$ and $|\Sigma|^2$. These estimates will be required in process variability monitoring operation. For this purpose, let us denote,

$$b_1=\frac{1}{(n-1)^p}\prod_{k=1}^{p}(n-k)\text{ and }b_2=b_1\left\{\frac{1}{(n-1)^p}\prod_{k=1}^{p}(n-k+2)-b_1\right\}.$$

We learn from the literature that, based on the above sample, $\frac{|S|}{b_1}$ and $\frac{|S|^2}{(b_1^2+b_2)}$ are unbiased estimates of of $|\Sigma|$ and $|\Sigma|^2$, respectively. However, in MPV monitoring, we deal with several m samples; $m>1$. So, how can we estimate $|\Sigma|$ and $|\Sigma|^2$ based on m samples? This problem will be discussed in Chapter 3. And to solve this problem, in what follows we discuss sampling distributional behavior of GV and VV.

2.1.3. Delta method

For practical purposes, the use of distribution (2.1) is not working well. It is not capable to provide the exact quantiles. Therefore, we will be happy to have the asymptotic ones. To study the asymptotic form of (2.1), we start by using the classical approach based on delta method. Later on, we use non-classical method.

To derive the asymptotic distribution of the function $f(S)=|S|$ using delta method, the standard procedure is as follows:

1. Find the asymptotic distribution of S. For this purpose, since S is a function of random sample, we start with a discussion on the distribution of a function of random variables, and then a function of random vectors.
2. Since $f(S)$ is a real valued function, continuous, and the first and second derivatives exist, then we can use Theorem 4.2.5 in Anderson (1966) to solve the problem.

2.1.3.1. Distribution of a function of random variables

In this paragraph, our discussion is focused on univariate case. Suppose the sequence of random variables $X_1, X_2, \dots, X_n$ converges in probability to a constant c and converges in distribution to a probability distribution with distribution function $F(x)$. Let $F_n(x)$ be the distribution function of X_n and $Y_n=u(X_n)$ be a real valued function of random variable X_n. Assume the first derivative u' exists and $u'(x)\neq 0$ for all x in the neighborhood of c. Under this assumption, Y_n can be written in Taylor series expansion,

$$Y_n=u(c)+u'(c)(X_n-c)+\frac{u''(\xi)}{2}(X_n-c)^2$$

for a given ξ in the neighborhood of c.

Since the third term on the right hand side converges to 0 faster than the second, then Y_n converges in distribution to a distribution with distribution function $F\left(c+\frac{y-u(c)}{u'(c)}\right)$. This property leads us to the following results.

1. Let μ_X and σ_X^2 be the mean and variance of $F(x)$. Then, the mean μ_Y and variance σ_Y^2 of the limiting distribution of Y_n are $\mu_Y=u(c)+u'(c)(\mu_X-c)$ and $\sigma_Y^2=\{u'(c)\sigma_X\}^2$. Furthermore, since $X_n\xrightarrow{p}c$ which implies $c=\mu_X$, we have $\mu_Y=u(\mu_X)$ and $\sigma_Y^2=\{u'(\mu_X)\sigma_X\}^2$.
2. In special case, if $X_n\xrightarrow{d}N\left(\mu_X,\sigma_X^2\right)$, then $Y_n\xrightarrow{d}N\left(\mu_Y,\sigma_Y^2\right)$.

The second result is important in the study of asymptotic distribution of a function of sample mean. Suppose the population mean and variance are μ and σ^2. We denote $\bar{X}_n$ the sample mean. Accordingly, if $X_n \xrightarrow{d} N(\mu,\sigma^2)$, then $\bar{X}_n \xrightarrow{d} N\left(\mu,\frac{\sigma^2}{n}\right)$. In general, if $Y_n = u(\bar{X}_n)$ is a function of $\bar{X}_n$ satisfying the condition mentioned above, we have $Y_n \xrightarrow{d} N(\mu_Y,\sigma_Y^2)$ with $\mu_Y = u(\mu)$ and $\sigma_Y^2 = \frac{\{u'(\mu)\sigma\}^2}{n}$.

In the next paragraph, the above results will be generalized to a function of a sequence of random vectors.

2.1.3.2. Distribution of a function of random vectors

Now, we consider the multivariate case. Suppose a sequence $X_1, X_2, \ldots, X_n$ of random vectors converges in probability to a scalar vector c and converges in distribution to a p-variate normal distribution $N_p(c,\Sigma)$. We are interested to find the asymptotic distribution of a random variable $Y_n = u(X_n)$; a function of X_n where u' exists and $u'(x) \neq 0$ for all x in the neighborhood of c. For this purpose, we consider the Taylor series of Y_n,

$$Y_n = u(c) + \left(\frac{\partial u(c)}{\partial X_n}\right)^t (X_n - c) + R_\xi.$$

The remaining term on the right hand side is $R_\xi = \frac{1}{2}(X_n - c)^t A_\xi (X_n - c)$ where A_ξ is a symmetric matrix of size $(p \times p)$ with general element $a(i,j) = \frac{\partial^2 u(\xi)}{\partial x_i \partial x_j}$; $i, j = 1, 2, \ldots, p$, and x_i is the i-th element of X_n.

Since the quadratic form R_ξ converges faster to 0 than the linear form $\left(\frac{\partial u(c)}{\partial X_n}\right)^t (X_n - c)$, and $X_n \xrightarrow{d} N_p(c,\Sigma)$, then

$$Y_n \xrightarrow{d} N(\mu_Y,\sigma_Y^2) \tag{2.2}$$

with $\mu_Y = u(c)$ and $\sigma_Y^2 = \left(\left.\frac{\partial u(X_n)}{\partial X_n}\right|_{X_n=c}\right)^t \Sigma \left(\left.\frac{\partial u(X_n)}{\partial X_n}\right|_{X_n=c}\right)$.

It is worth noting that the result (2.2) can also be obtained by using different approach. See, for example, Wilks (1962), Anderson (1966), and Larzac and Cleroux (1992) for further discussions on that approach. The distribution (2.2) allows us to study the asymptotic distribution of GV in the next sub-section.

2.1.4. Asymptotic sampling distribution

By using the results in (2.1), the asymptotic distribution of S will be studied. Suppose $X_1, X_2, \ldots, X_n$ is a random sample from a p-variate normal distribution, $N_p(0,\Sigma)$. Then, see Anderson (1966), S can be expressed as the sum of $(n-1)$ random matrices. That is,

$$S = \frac{1}{n-1}\sum_{i=1}^{n-1} Z_i Z_i^t \text{ where } Z_1, Z_2, \ldots, Z_{n-1} \text{ are i.i.d. } N_p(0,\Sigma).$$

Accordingly, based on central limit theorem, S converges in distribution to a p^2-variate normal distribution. Furthermore, since S converges in probablity to Σ, the result (2.2) leads us to this asymptotic distribution of S,

$$\sqrt{n-1}(S-\Sigma) \xrightarrow{d} N_{p^2}(0,\Gamma). \tag{2.3}$$

Here Γ is the covariance matrix of S. It is of size $(p^2 \times p^2)$ and, see Larzac and Cleroux (1992), can be written simply $\Gamma = (I_{p^2} + K)(\Sigma \otimes \Sigma)$ with $K =$

$\sum_{i=1}^{p}\sum_{j=1}^{p} N_{ij} \otimes N_{ij}^{t}$, and N_{ij} is a matrix of size $(p \times p)$ where all of its elements are 0 except its (i,j)-th element equals 1. The matrix K is the so-called commutation matrix.

The matrix K can be written in a simple manner. In each of its row and each of its column, there is only one element equals 1 with certain regularity and the other elements are 0. Precisely, if $k(i,j)$ is the general element of K, all elements which equals 1 are,

$$
\begin{aligned}
&k(1,1),\ k(2,p+1),\ \ldots,\ k(p,(p-1)p+1),\\
&k(p+1,2),\ k(p+2,p+2),\ \ldots,\ k(2p,(p-1)p+2),\\
&\ldots\ldots\\
&k((p-1)p+1,p),\ k((p-1)p+2,2p),\ \ldots,\ k(p^2,p^2).
\end{aligned}
$$

For example, for $p = 2$, those elements are $k(1,1) = k(2,3) = k(3,2) = k(4,4) = 1$. Thus,

$$
K = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}.
$$

As another example, if $p = 3$, there are nine elements equal 1, i.e. $k(1,1) = k(2,4) = k(3,7) = k(4,2) = k(5,5) = k(6,8) = k(7,3) = k(8,6) = k(9,9) = 1$. In this case,

$$
K = \begin{pmatrix}
1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 \\
0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \\
0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1
\end{pmatrix}.
$$

To simplify the problem to find the distribution of $|S|$, let us consider $vec(S)$ the vector representation of S. It is a vector in real vector space of p^2 dimension obtained from S by stacking each column underneath the others. Therefore the k-th component of $vec(S)$ is,

$$
vec(S)_k = \begin{cases}
s_{1k} & ;1 \le k \le p \\
s_{2(k-p)} & ;p+1 \le k \le 2p \\
s_{3(k-2p)} & ;2p+1 \le k \le 3p \\
 & \ldots\ldots\ldots \\
s_{p(k-(p-1)p)} & ;(p-1)p+1 \le k \le p^2
\end{cases}
$$

Instead of working with S, we can simply work with $vec(S)$. With this notation, (2.3) can be written as,

$$
\sqrt{n-1}\left(vec(S) - vec(\Sigma)\right) \xrightarrow{d} N_{p^2}(0,\Gamma). \tag{2.4}
$$

By using the result in 2.1.3.2, the asymptotic distribution of sample GV can immediately be derived from (2.4). Let $u(S) = |S|$ and suppose at least one of its cofactors is not 0. According to (2.2), $|S| \xrightarrow{d} N\left(\mu,\sigma^2\right)$ with

$$
\mu = |\Sigma| \text{ and } \sigma^2 = \frac{1}{n-1}\left(\left.\frac{\partial u(vec(S))}{\partial S}\right|_{S=\Sigma}\right)^{t} \Gamma \left(\left.\frac{\partial u(vec(S))}{\partial S}\right|_{S=\Sigma}\right).
$$

The partial derivative on the right hand side, see Muirhead (1982) and Kollo and von Rossen (2005), is equal to $\left.\frac{\partial u(vec(S))}{\partial s_{ij}}\right|_{S=\Sigma} = (-1)^{i+j}|\Sigma_{ij}|$ where $|\Sigma_{ij}|$ is the (i,j)-th cofactor of Σ. It is worth noting that Anderson (1966) and also Muirhead (1982) come up with this result,

$$
\sqrt{n}\left(|S| - |\Sigma|\right) \xrightarrow{d} N\left(0, 2p|\Sigma|^2\right). \tag{2.5}
$$

In what follows we show that the convergence is very slow and that the parameters $|\Sigma|$ and $\frac{2p}{n}|\Sigma|^2$ in (2.5) are not the true mean and variance of $|S|$. The true parameters are given by the exact distribution of $|S|$ in (2.1) while these parameters are the limit of the true ones when the sample size n tends to infinity.

Now, we show the first assertion that the convergence is very slow. As we have mentioned in Sub-section 2.1.2, from (2.1) we have derived that the true mean of $|S|$ is $b_1|\Sigma|$ while the true variance is $b_2|\Sigma|^2$. Interestingly, $\lim_{n\to\infty} b_1 = 1$ and $\lim_{n\to\infty}\frac{\frac{2p}{n}}{b_2} = 1$. A numerical experiment for $p = 2$ shows that n more than 10.000 is required in order for the true mean $b_1|\Sigma|$ differs from $|\Sigma|$ by no more than 10^{-5}. Larger value of n is required for larger p. Thus, the convergence is very slow.

To show the second assertion, we use the result (2.5) to derive the asymptotic distribution of sample GV. Since,

1. $\lim_{n\to\infty}\frac{b_1|\Sigma|-|\Sigma|}{\sqrt{\frac{2p}{n}|\Sigma|^2}} = 0$ (the numerator is the true mean of $|S|$ subtracted by the mean in (2.5) while the denominator is the standard deviation in (2.5)), and
2. $\lim_{n\to\infty}\frac{\sqrt{b_2|\Sigma|^2}}{\sqrt{\frac{2p}{n}|\Sigma|^2}} = 1$ (the numerator is the true standard deviation of $|S|$ and the denominator is as in point 1),

by using Lemma A in Serfling (1980, p. 20), these two limits gives us this limiting distribution,

$$\frac{1}{\sqrt{b_2}}(|S|-b_1|\Sigma|) \xrightarrow{d} N\left(0,|\Sigma|^2\right). \qquad (2.6)$$

This distribution proves the second assertion that the parameters in (2.5) are the limit of the true ones. In Chapters 3 and 4, the result in (2.6) will be used as the theoretical basis in the construction of *GV*-chart.

2.2. Vector variance

Due to its commendable properties such as mentioned in Reis and Rato (2013), GV is still the most adopted measure in applications. However, to handle its limitations, Djauhari (2007) introduces an alternative measure called vector variance (VV). This measure and GV are complementary. Accordingly, if they are used simultaneously, we will have a better understanding about the process variability. See Djauhari et al. (2008) for an application of VV in MSPC and also Djauhari and Mohamad (2010) for a simultaneous application of GV and VV. For this reason, in this section we investigate the sampling distribution of VV.

2.2.1. Basic notion

Let $X^{(1)}$ and $X^{(2)}$ be two random vectors of dimension p and q, respectively. We write $\mu^{(i)} = E\left(X^{(i)}\right)$; $i=1,2$, and $\Sigma_{ij} = E\left(\left(X^{(i)}-\mu^{(i)}\right)\left(X^{(j)}-\mu^{(j)}\right)^t\right)$; $i, j = 1, 2$. With this notations, the covariance matrix Σ of their superposition is partitioned as,

$$\Sigma = \begin{pmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{pmatrix}.$$

In bivariate case, i.e. $p = q = 1$, Σ_{ij} is the covariance of two variables. It measures the linear relationship of $X^{(1)}$ and $X^{(2)}$. In other cases, to study that linear relationship, Escoufier (1973) introduces the so-called vector covariance,

$$CV\left(X^{(1)}, X^{(2)}\right) = Tr(\Sigma_{12}\Sigma_{21}).$$

The symbol "*Tr*" represents the *trace* operator; *Tr*(A) is the sum of all diagonal elements of a matrix A. Accordingly,

$$VV\left(X^{(1)}\right) = Tr\left(\Sigma_{11}^2\right) \text{ and } VV\left(X^{(2)}\right) = Tr\left(\Sigma_{22}^2\right)$$

are the vector variance (VV) of $X^{(1)}$ and of $X^{(2)}$, and

$$RV\left(X^{(1)}, X^{(2)}\right)\frac{Tr(\Sigma_{12}\Sigma_{21})}{\sqrt{Tr\left(\Sigma_{11}^2\right)Tr\left(\Sigma_{22}^2\right)}}$$

is their vector correlation (RV).

In mathematical language, RV is the cosine of the angle between two spaces each of which is spanned by random vectors $X^{(1)}$ and $X^{(2)}$.

2.2.2. Geometric interpretation

Let X be a random vector having positive definite covariance matrix Σ. As remarked in Djauhari (2007), $Tr\left(\Sigma^2\right)$, the VV of X, satisfies these properties.

1. VV ≥ 0. It is equal to 0 if and only if $\Sigma = 0$,
2. VV is small if and only if all elements of Σ are small,
3. VV is large if at least one element of Σ is large in magnitude,
4. If all diagonal elements are fixed, then VV is maximum if all variables are perfectly correlated to each other.

To understand its geometrical interpretation, consider the cloud of data points in E. Suppose all variables are centered and let W denote the matrix of size $(n \times n)$ defined by,

(i) w_{ij} is the dot product of the i-th and j-th data points in E divided by $(n-1)$.

(ii) w_{ii} is then the squared Euclidean length of the i-th data point divided by $(n-1)$.

Thus, W is a symmetric matrix. It summarizes the relative position of each data points in E around the sample mean vector $\overline{X}$ as the center of gravity of the cloud. The length of the i-th point from the center is given by $\sqrt{w_{ii}}$, and the angle between two points is $\arccos\left(\frac{w_{ij}}{\sqrt{w_{ii}w_{jj}}}\right)$.

Under this consideration, Djauhari (2016) show that $Tr\left(S^2\right) = Tr\left(W^2\right)$. This means that the sample VV is a numerical indicator of how the cloud of n data points is dispersed around the center. Its value becomes larger when and only when the cloud becomes more dispersed around the center. The smaller its value the more concentrated the cloud and the larger its value the more dispersed the cloud in a subspace of E of dimension k; $k \leq p$. It equals 0 if and only if the n data points are coincide; they are equal to each other. In the sense of the above properties, VV can be

used as a measure of multivariate dispersion.

It is worth noting that, unlike GV, VV can still be used to measure the variability when the covariance matrix is singular. However, like GV, two different processes might have the same variability in terms of VV. The above analysis shows that GV and VV are complementary. Therefore, their simultaneous use is recommended.

2.2.3. Asymptotic distribution

In bivariate and multivariate cases, only an asymptotic distribution is available for VV. This will be highlighted in this sub-section. Those who are interested in the details are suggested to consult Djauhari (2007). First, we consider the function u discussed in the previous section and define $u(S) = Tr\left(S^2\right)$. Thus,

$$u(S) = \sum_{i=1}^{p}\sum_{j=1}^{p} s_{ij}^2 .$$

To study the distribution of $u(S)$, here we use again the results (2.2). First, since $S \xrightarrow{p} \Sigma$, then $Tr\left(S^2\right) \xrightarrow{p} Tr\left(\Sigma^2\right)$. Second, the partial derivative of the quadratic form $Tr\left(S^2\right)$ is simply $\left.\frac{\partial u(vec(S))}{\partial s_{ij}}\right|_{S=\Sigma} = 2vec(\Sigma)$. These two results lead us to the following limiting distribution similar to (2.4),

$$\sqrt{n-1}\left(Tr\left(S^2\right) - Tr\left(\Sigma^2\right)\right) \xrightarrow{d} N\left(0, \sigma^2\right) \tag{2.7}$$

where $\sigma^2 = \left(\left.\frac{\partial u(vec(S))}{\partial S}\right|_{S=\Sigma}\right)^t \Gamma \left(\left.\frac{\partial u(vec(S))}{\partial S}\right|_{S=\Sigma}\right) = 4vec(\Sigma)^t \Gamma vec(\Sigma)$.

Since $vec(\Sigma)^t \Gamma vec(\Sigma) = 2Tr\left(\Sigma^4\right)$, see Djauhari (2016) for the proof, finally we have,

$$\sqrt{\frac{n-1}{8}}\left(Tr\left(S^2\right) - Tr\left(\Sigma^2\right)\right) \xrightarrow{d} N\left(0, Tr\left(\Sigma^4\right)\right). \tag{2.8}$$

We see that $Tr\left(S^2\right)$ is an asymptotically unbiased estimate of $Tr\left(\Sigma^2\right)$. Another asymptotic unbiased estimator with smaller variance, see Djauhari (2016), is given in (2.9).

$$\frac{n-1}{\sqrt{8n}}\left\{Tr\left(S^2\right)-\frac{n+1}{n-1}Tr\left(\Sigma^2\right)\right\} \xrightarrow{d} N\left(0,Tr\left(\Sigma^4\right)\right). \quad (2.9)$$

Accordingly, we will always refer to (2.9) when we construct *VV*-chart in Chapter 5.

2.3. Message to the practitioners

Two measures of multivariate variability, GV and VV, are presented and discussed. The exact sampling distribution of GV is only applicable in practice for $p < 3$. For $p \geq 3$, we are happy with its asymptotic distribution. It is so with VV when $p \geq 2$. Therefore,

1. *GV*-chart can be used for small n if $p = 2$. If $p \geq 3$, this chart is only apt for large n.
2. *VV*-chart is to monitor the process when n is large whatever $p > 1$.

For practical application, since GV and VV are complementary to each other, the simultaneous use of both *GV*-chart and *VV*-chart in MSPC is strongly recommended.

CHAPTER 3
GENERALIZED VARIANCE CHART

In the rest of the book, process variability monitoring will always be based on m independent samples of the same size n, each of which is drawn from a p-variate normal distribution with positive definite covariance matrix Σ. And we focus in this chapter on monitoring operation using *GV*-chart discussd in Chapter 2. For this purpose, we start our discussion with the classical technique of the *GV*-chart construction for both large and small sample size n. Then, we show that in those two cases the same control chart is employed. This is presented in Section 3.1.

After showing that the control limits of *GV*-chart are not unbiased, in Section 3.2 a new version of *GV*-chart, called improved *GV*-chart or simply denoted by *IGV*-chart, is presented. It is followed by a sensitivity analysis in terms of average run length (ARL). Later on, we show that *IGV*-chart has better ARL than *GV*-chart.

To illustrate the advantages of *IGV*-chart, an industrial example will be studied, and to make the readers comfortable with the practical implementation of this chart, in the last section we address a message for practitioners.

3.1. Classical technique

Let S_i be the covariance matrix of the i-th sample; $i = 1, 2, \ldots, m$, and $\bar{S}$ be their average. *GV*-chart is a control chart constructed by plotting $|S_i|$ and the control limits in the same chart. Theoretically, for the i-th sample, the control limits are derived from the distribution of $|S_i|$. In what folows, these control limits are studied.

We know from (2.6) that,

$$\frac{1}{\sqrt{b_2}}\left(|S_i|-b_1|\Sigma|\right) \xrightarrow{d} N\left(0,|\Sigma|^2\right). \quad (3.1)$$

Therefore, for n sufficiently large, (3.1) allows us to approximate the distribution of $|S_i|$ by normal distribution with mean $b_1|\Sigma|$ and variance $b_2|\Sigma|^2$. In practice, this approximation is far from the limit because the convergence is very slow. This indicates that the practicality of the *GV*-chart for small n will differ considerably from the one where n is large.

3.1.1. Case of large sample

If Σ is unknown, the parameters mean and variance in (3.1) must be estimated. Based on the i-th sample alone, in Chapter 2 we show that $\frac{|S_i|}{b_1}$ and $\frac{|S_i|^2}{(b_1^2+b_2)}$ are unbiased estimates of $|\Sigma|$ and $|\Sigma|^2$, respectively. What wil happen with these estimates if all m samples are involved? Since we deal with m independent samples, in classical technique such as presented in the literature, $|\Sigma|$ is estimated by $\frac{|\bar{S}|}{b_1}$ and $|\Sigma|^2$ by $\frac{|\bar{S}|^2}{(b_1^2+b_2)}$. Then, these estimates are used directly to define the control limits.

Specifically, for probability of false alarm (PFA) α = 0.0027, this technique defines the upper control limit (UCL), central line (CL), and lower control limit (LCL) as follows,

$$\begin{aligned} \text{UCL} &= \frac{|\bar{S}|}{b_1}(b_1+3\sqrt{b_2}) \\ \text{CL} &= |\bar{S}| \\ \text{LCL} &= \max\left\{0, \frac{|\bar{S}|}{b_1}(b_1-3\sqrt{b_2})\right\}. \end{aligned} \tag{3.2}$$

The *GV*-chart is then constructed by plotting, in the same diagram, the value of GV for each sample and the control limits (3.2).

In the next two sub-sections, we discuss the case of small sample size and then present an example. Later on, in Section 3.2, we show that these control limits are not unbiased.

3.1.2. Case of small sample

In this case, approximation to normality in 3.1.1 is far from the limit because the convergence is very slow. A simulation study in Chapter 4, as displayed in Table 4.6, supports this claim. Consequently, the control limits in (3.2) are not suitable. To overcome this problem, we learn from the literature such as Montgomery (2001, 2005, 2009) that the control limits when n is small are defined heuristically using the property that "most of the probability distribution of $|S|$ is contained in the interval $E(|S|) \pm 3\sqrt{Var(|S|)}$." Therefore, based on this property the control limits are defined as,

$$\begin{aligned} \text{UCL} &= E(|S|)+3\sqrt{Var(|S|)} \\ \text{CL} &= E(|S|) \\ \text{LCL} &= \max\left\{0, E(|S|)-3\sqrt{Var(|S|)}\right\}. \end{aligned}$$

But, we know that $E(|S|)$ is estimated by $|\bar{S}|$ and $Var(|S|)$ by $\frac{b_2}{b_1^2}|\bar{S}|^2$. Therefore, these control limits are nothing more than (3.2) which are defined for large n. This is surprising! Furthermore, a serious problem occurs when we relate these control limits with the PFA. They do not reflect PFA since the word "most" in the above property signifies that they have nothing to do with PFA. To handle this problem, in Chapter 4 an adjusted PFA and a reliable *GV*-chart will be introduced. This will hopefully be a contribution of this book.

3.1.3. Example

The Center for Indonesian Army Industry (PINDAD) is to study the stability of flange production process variability. For this purpose, a preliminary analysis was conducted involving m = 20, n = 5, and p = 3. The three quality characteristics are X_1 (the diameter of the nozzle), X_2 (the thickness of the wall), and X_3 (the thickness of the base). Observation on these quality characteristics gives us 20 covariance matrices in Table 3.1 (written as lower triangular matrix).

Table 3.1. Sample covariance matrix

No.	S_i			No.	S_j		
	0,0680				0,0150		
1	-0.0260	0.0145		11	0.0156	0.0870	
	-0.0356	-0.0076	0.1445		-0.0605	-0.0944	0.2669
	0.0020				0.0150		
2	0.0014	0.0048		12	0.0069	0.0713	
	-0.0069	0.0040	0.1195		-0.0500	-0.0566	0.1989
	1.3568				0.1058		
3	-0.1143	0.0130		13	-0.0540	0.0805	
	0.1970	-0.0158	0.0403		-0.0815	0.0363	0.1048
	0.2655				0.0805		
4	0.0886	0.0483		14	0.0479	0.0389	
	0.2814	0.0832	0.3197		-0.0708	-0.0798	0.2331
	0.0508				0.0300		
5	0.0024	0.0764		15	-0.0069	0.0179	
	0.0471	0.0146	0.0745		0.0783	-0.0146	0.2107
	0.5788				3.5868		
6	0.0369	0.0183		16	0.4610	0.1542	
	-0.6935	-0.0787	1.0467		-0.3907	-0.0686	0.0679
	0.1030				0.0343		
7	0.0003	0.0233		17	0.0136	0.0274	
	0.0020	0.0428	0.1208		-0.0469	-0.0528	0.1189
	0.1630				0.0163		
8	0.0286	0.0433		18	-0.0119	0.0323	
	-0.2100	-0.0171	0.3442		-0.0115	0.0270	0.0587
	0.1920				0.0293		
9	0.0108	0.0292		19	0.0163	0.0129	
	-0.1290	-0.0100	0.2252		-0.0010	0.0065	0.0319
	0,2725				0.3888		
10	0.0994	0.1050		20	0.0206	0.0089	
	-0.1340	-0.0738	0.1169		-0.1938	-0.0170	0.1034

The average of these covariance matrices is,

$$\bar{S} = \begin{pmatrix} 0.3677 & & \\ 0.0319 & 0.0454 & \\ -0.0755 & -0.0186 & 0.1974 \end{pmatrix},$$

and the determinant of $\bar{S}$ is $|\bar{S}| = 0.0028$. Furthermore, the generalized variance $|S_i|$ of each of the 20 samples is in Table 3.2.

Table 3.2. Sample generalized variance

No.	$\|S_i\|$	No.	$\|S_i\|$
1	8.57E-06	11	9.51E-06
2	5.87E-07	12	1.59E-05
3	5.28E-05	13	2.32E-04
4	7.59E-05	14	2.87E-05
5	1.11E-04	15	2.90E-06
6	1.30E-03	16	7.43E-03
7	1.01E-04	17	1.21E-06
8	3.96E-04	18	1.38E-05
9	7.59E-04	19	2.15E-06
10	7.87E-04	20	3.14E-06

These results and Equation (3.1) give us $b_1 = 0.3750$, $b_2 = 0.5625$, and from (3.2) we get LCL = 0, CL = 0.0028 and UCL = 0.0196. Therefore, we obtain the *GV*-chart as presented in Figure 3.1. The horizontal axis is the sample number and the vertical axis refers to the value of sample GV.

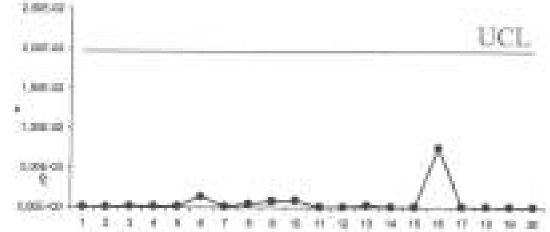


Figure 3.1. *GV*-chart

This chart indicates that there is no out-of-control (OOC) signal occurs during the production process.

Before we leave this section, it is worth noting that,

1. The form of *GV*-chart is the same whether we consider $n = 5$ as large or small. However, if we consider $n = 5$ as large, PFA = 0.0027 is far from reality. On the other hand, if $n = 5$ is considered as small, than we do not know the PFA.
2. The computation process and the drawing are realized using Microsoft Excel. This is to illustrate the simplicity of *GV*-chart construction.

3.2. Recent development

In the previous section, the control limits of *GV*-chart in (3.2) are determined by using these statistics $\frac{|\bar{S}|}{b_1}$ and $\frac{|\bar{S}|^2}{(b_1^2+b_2)}$ as the estimate of $|\Sigma|$ and of $|\Sigma|^2$. In the next sub-section we show that these estimates are not unbiased. Later on, we present the unbiased ones.

3.2.1. IGV-chart construction

In one sample case, see Montgomery (2001, 2005, 2009), we have $S \sim W_p(\Sigma,(n-1))$ which leads to the distribution of $|\bar{S}|$ in (2.1). Since the m samples are independent, then $\bar{S} \sim W_p(\Sigma, m(n-1))$ and analogous to (2.1), we have the following distribution of $|\bar{S}|$,

$$|\bar{S}| \sim \frac{|\Sigma|}{(m(n-1))^p}\prod_{k=1}^{p}\chi^2_{m(n-1)-k+1} \tag{3.3}$$

where $\chi^2_{m(n-1)}, \chi^2_{m(n-1)-1}, \ldots,$ and $\chi^2_{m(n-1)-p+1}$ are independent. This distribution leads us to the following unbiased estimates of $|\Sigma|$ and $|\Sigma|^2$,

$$|\bar{S}|\frac{b_1}{b_3} \text{ and } \frac{|\bar{S}|^2}{(b_3^2+b_4)} \tag{3.4}$$

where

(i) $b_3 = \frac{1}{(m(n-1))^p}\prod_{k=1}^{p} m(n-1)-k+1$

(ii) $b_4 = b_3\left\{\frac{1}{(m(n-1))^p}\prod_{j=1}^{p}[m(n-1)-j+3]-b_3\right\}$.

Accordingly, $|\bar{S}|$ converges in distribution as follows.

$$\frac{1}{\sqrt{b_4}}(|\bar{S}|-b_3|\Sigma|) \xrightarrow{d} N\left(0,|\Sigma|^2\right). \tag{3.5}$$

This result gives us, see Djauhari (2005), the following unbiased control limits of *GV*-chart for PFA $\alpha = 0.0027$.

$$\begin{aligned} \text{UCL} &= |\bar{S}|\left(\frac{b_1}{b_3}+3\sqrt{\frac{b_2}{b_3^2+b_4}}\right)\\ \text{CL} &= |\bar{S}|\frac{b_1}{b_3}\\ \text{LCL} &= \max\left\{0,|\bar{S}|\left(\frac{b_1}{b_3}-3\sqrt{\frac{b_2}{b_3^2+b_4}}\right)\right\}\end{aligned} \tag{3.6}$$

We see that when $m = 1$, then $b_3 = b_1$ and $b_4 = b_2$. Thus, (3.6) becomes (3.2). In the rest of the book, we call *IGV*-chart the *GV*-chart with these control limits.

3.2.2. Sensitivity analysis

Both *GV*-chart (3.2) and *IGV*-chart (3.6) will be compared in terms of their ARL. For this purpose, let σ be the in-control value of *GV*-chart. We denote $\Delta = k\sigma$ the shift in GV from its in-control value where k is a positive constant, and $\hat{\Delta}$ its estimator. For *IGV*-chart, $\hat{\Delta} = k\hat{\sigma}_I$ where $\hat{\sigma}_I$ is the estimator of σ used in Equation (3.6). Meanwhile, for *GV*-chart in (3.2), $\hat{\Delta} = k^*\hat{\sigma}_K$, where $\hat{\sigma}_K$ is the estimator of σ

used in (3.2) and k^* is another positive constant. Thus, the two constants satisfy $k^* = k\frac{\hat{\sigma}_I}{\hat{\sigma}_s}$. But, from (3.2) and (3.6), we have

$$\hat{\sigma}_s = |\bar{S}|\frac{\sqrt{b_2}}{b_1} \text{ and } \hat{\sigma}_I = |\bar{S}|\sqrt{\frac{b_2}{b_3^2 + b_4}}.$$

Therefore, k and k^* have the following relationship in terms of b_1, b_3, and b_4,

$$k^* = k\sqrt{\frac{b_1^2}{b_3^2 + b_4}} \text{ or } k = k^*\sqrt{\frac{b_3^2 + b_4}{b_1^2}}.$$

Based on these relationships, we compute ARL_{GV} and ARL_{IGV}; the ARL of *GV*-chart and of *IGV*-chart. Let $\Phi(x)$ be the cumulative distribution function of the standard normal distribution and PFA $\alpha = 0.0027$. Then,

$$ARL_{IGV} = \frac{1}{1-\beta_I} \text{ and } ARL_{GV} = \frac{1}{1-\beta_S},$$

where

$$\beta_I = \Phi(3-k) - \Phi(-3-k) \text{ and } \beta_S = \Phi\left(3-k^*\right) - \Phi\left(-3-k^*\right).$$

By using these equations, a numerical solution for ARL_{GV} and ARL_{IGV} is given in Table 3.3 for some selected values on n and p.

Table 3.3. ARL of *GV*-chart given ARL of *IGV*-chart

	$n=5$		$n=7$		$n=10$	
		$p=2$	$p=3$	$p=2$	$p=3$	$p=2$	$p=3$
k	ARL_{IGV}	ARL_{GV}	ARL_{GV}	ARL_{GV}	ARL_{GV}	ARL_{GV}	ARL_{GV}
0.00	370.38	370.38	370.38	370.38	370.38	370.38	370.38
0.25	281.14	316.07	354.94	304.91	337.96	297.16	322.26
0.50	155.22	215.06	314.87	193.60	265.54	180.10	228.09
0.75	81.22	134.34	263.56	113.63	191.92	101.58	147.94
1.00	43.89	83.13	212.62	66.95	134.33	58.06	94.39
1.25	24.96	52.39	168.15	40.61	93.62	34.40	60.95
1.50	14.97	33.92	131.88	25.51	65.79	21.24	40.23
1.75	9.47	22.60	103.31	16.63	46.87	13.68	27.22
2.00	6.30	15.51	81.15	11.24	33.91	9.18	18.89
2.25	4.41	10.96	64.08	7.87	24.94	6.41	13.45
2.50	3.24	7.97	50.92	5.71	18.65	4.66	9.83
2.75	2.49	5.96	40.75	4.28	14.18	3.51	7.35
3.00	2.00	4.58	32.85	3.32	10.96	2.75	5.64

This results show that, in OOC situations, ARL_{IGV} is always smaller than ARL_{GV}. The same conclusion is apparent in Table 3.4 when ARL_{GV} is fixed.

Table 3.4. ARL of *IGV*-chart given ARL of *GV*-chart

	$n=5$		$n=7$		$n=10$	
		$p=2$	$p=3$	$p=2$	$p=3$	$p=2$	$p=3$
k^*	ARL_{GV}	ARL_{IGV}	ARL_{IGV}	ARL_{IGV}	ARL_{IGV}	ARL_{IGV}	ARL_{IGV}
0.00	370.38	370.38	370.38	370.38	370.38	370.38	370.38
0.25	281.14	232.54	100.69	251.75	176.51	262.75	219.68
0.50	155.22	98.50	20.94	118.13	55.85	130.89	86.99
0.75	81.22	42.64	6.31	54.82	20.22	63.39	36.04
1.00	43.89	20.22	2.71	27.22	8.69	32.42	16.63
1.25	24.96	10.57	1.59	14.61	4.41	17.74	8.58
1.50	14.97	6.07	1.19	8.48	2.62	10.39	4.93
1.75	9.47	3.82	1.05	5.30	1.79	6.50	3.13
2.00	6.30	2.62	1.01	3.56	1.38	4.34	2.19
2.25	4.41	1.94	1.00	2.56	1.17	3.08	1.67
2.50	3.24	1.55	1.00	1.96	1.07	2.32	1.37
2.75	2.49	1.31	1.00	1.59	1.03	1.84	1.20
3.00	2.00	1.17	1.00	1.36	1.01	1.53	1.10

3.2.3. Case of small sample

The practical implementation of *IGV*-chart needs sufficiently large n because the control limits (3.6) are derived from asymptotic distribution of $|S|$. Unfortunately, the convergence to normality is very slow. Thus, *IGV*-chart in (3.6) is not apt for n small. To construct *IGV*-chart when n is small, literature suggests to use again the property that "most of the distribution of $|S|$ is contained in the interval $E(|S|)$ $\pm 3\sqrt{Var(|S|)}$." Accordingly, the control limits are,

$$\begin{aligned} \text{UCL} &= E(|S|) + 3\sqrt{Var(|S|)} \\ \text{CL} &= E(|S|) \\ \text{LCL} &= \max\left\{0, E(|S|) - 3\sqrt{Var(|S|)}\right\} \end{aligned} \tag{3.7}$$

But, we have remarked in 3.2.1 that, based on m independent samples, $|\bar{S}|\frac{b_1}{b_3}$ and $\frac{b_2}{b_3^2 + b_4}|\bar{S}|^2$ are unbiased estimates of $E(|S|)$ and $Var(|S|)$. Therefore, these control limits are equal to those when n is large in (3.6). And as previously, they have nothing to do with PFA; they do not reflect the PFA.

3.2.4. Example

Consider again the example in 3.1.3. To construct *IGV*-chart, we compute b_3 = 0.9628 and b_4 = 0.0722. Then, by using (3.6) we find LCL = 0, CL = 0.0011 and UCL = 0.0074. Accordingly, by plotting these control limits and the value of $|S_i|$ in Table 3.2 in the same chart, the history of process variability based on *IGV*-chart is presented in Figure 3.2.

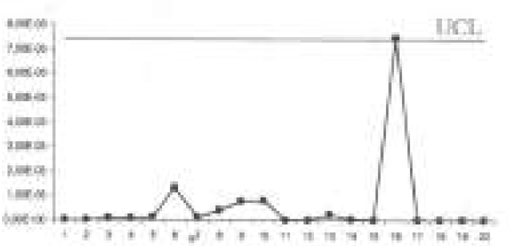


Figure 3.2. *IGV*-chart

In this figure, an OOC signal occurs at sample 16. This signal is not detected by *GV*-chart. This is an advantage of unbiased control limits in *IGV*-chart. Again, if we consider $n = 5$ as small, the PFA of this signal is remain unknown.

At the monitoring period where OOC signal occurs, root causes analysis is required. However, this problem is beyond the scope of this book. Thus, we leave it for future book.

3.3. Message to the practitioners

Whether n is large or small, *IGV*-chart is constructed as follows,

Step 1. Calculate covariance matrix S_i and its determinant $|S_i|$; $i = 1, 2, \ldots, m$.

Step 2. Find $\bar{S}$ the average of all covariance matrices, and its determinant $|\bar{S}|$.

Step 3. Calculate b_1, b_2, b_3, and b_4. Then, calculate the control limits in (3.6).

Step 4. Construct the *IGV*-chart by plotting $|S_i|$; $i = 1, 2, \ldots, m$, and the control limits obtained in *Step 3* in the same chart.

If n is large, the control limits refer approximately to a desired PFA. Meanwhile, if n is small, they have nothing to do with PFA. Thus, when n is small, the reliability of *IGV*-chart is unknown. This is the problem that will be handled in the next chapter.

CHAPTER 4
RELIABLE GENERALIZED VARIANCE CHART

The *IGV*-chart discussed in Chapter 3 has unbiased control limits whatever the sample size n as long as $n > p$. This is an advantage of this chart. However, for small n, the PFA is unknown. On the other hand, for a given PFA, its practical implementation needs sufficiently large n. The control limits (3.6) are derived from normal approximation to the distribution of $|S|$. Unfortunately, since the convergence to normality is very slow, the constant multiplier "3" on the right hand side of (3.6) will make sense only when n is sufficiently large. It refers to PFA $\alpha = 0.0027$ (i.e. 2700 defects per million opportunity or, in brief, 2700 DPMO). In this book, the term "large" means that it is not suitable for industrial environment which requires small n. Remember that industry needs fast and cheap data collection process.

What is happening if n is small? By construction of *IGV*-chart, the use of normal approximation will be misleading for small n. Consequently, the constant "3" has nothing to do with PFA $\alpha = 0.0027$. This PFA value is thus no longer appropriate but still unknown. Thus, the reliability of *IGV*-chart is also unknown.

So, what is the appropriate constant which reflects a desired PFA? Let us denote it by K. With this notation, instead of using the control limits in (3.6), see also Sub-section 3.2.3, we use

$$\begin{aligned} \text{UCL} &= |\bar{S}|\left(\frac{b_1}{b_3} + \text{K}\sqrt{\frac{b_2}{b_3^2 + b_4}}\right) \\ \text{CL} &= |\bar{S}|\frac{b_1}{b_3} \qquad\qquad (4.1) \\ \text{LCL} &= \max\left\{0, |\bar{S}|\left(\frac{b_1}{b_3} - \text{K}\sqrt{\frac{b_2}{b_3^2 + b_4}}\right)\right\} \end{aligned}$$

We recall that the distribution of $|S_i|$ in (2.1) cannot be expressed in a closed form. Its quantiles remain unknown. Hence, given a desired PFA, K also remains unknown. For this reason, in the next section we conduct some simulation

experiments to find the experimental quantiles and the experimental value of K for some selected n, p and PFA = α. Once K has been simulated for a desired PFA, *IGV*-chart can be constructed by using the control limits in (4.1). With this new definition of control limits, the reliability of *IGV*-chart can be determined in terms of PFA. In the rest of the book, we call K the reliability constant.

The first problem that we encounter is as follows. If for small n we fix K = 3, then what is the PFA? The second problem is, if for small n we set a desired value of PFA = α, then what is the reliability constant K? These problems are the subject of discussion in the rest of this chapter; some important properties of the new control limits (4.1) will be derived and the computational technique of PFA and K will be delivered.

The simulation experiment and its results are presented in Section 4.2. And an example of application is given in Section 4.3. To close this chapter, a message to the practitioners in the last section is addressed. This is to make clear the procedure of control chart construction for limited sample size.

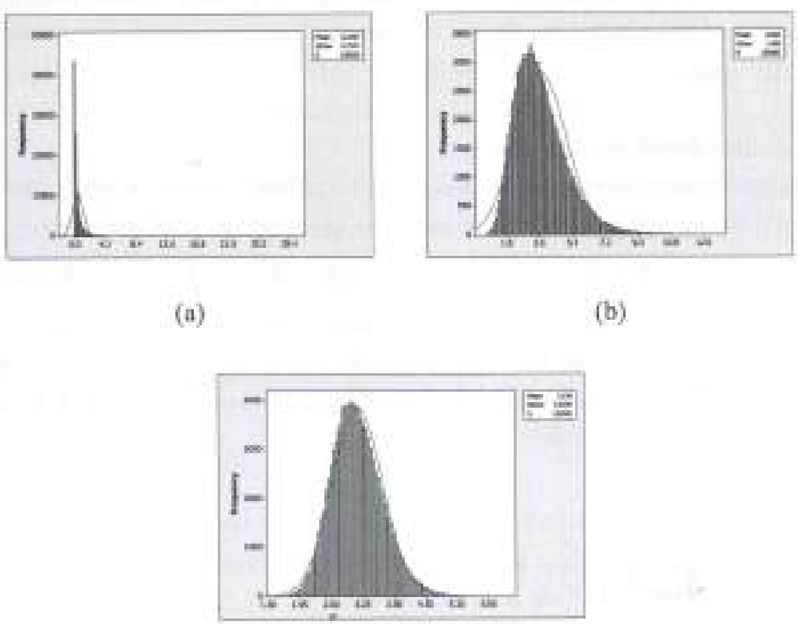
4.1. Simulation experiments

In the literature such as Montgomery (2001, 2009), when n is small, the control limits of *IGV*-chart are defined by using the property "most of the distribution $|S|$ lies between $E(|S|) - 3\sqrt{Var(|S|)}$ and $E(|S|) + 3\sqrt{Var(|S|)}$." The left end point is the LCL while the right is UCL. The middle point is CL. Through simulation experiment, in this section we study the distribution of GV and then illustrate that this definition of control limits might be misleading in the sense that it has nothing to do with PFA. Thus, its reliability is questionable.

As suggested in Djauhari et al. (2017), for each pair of (n, p) we simulate n random data from p-variate normal distribution with zero mean vector and identity covariance matrix. Then, we compute the value of GV. This experiment is repeated 100,000 times to collect 100,000 simulated data of GV. In order to get a better understanding about the distribution of GV, this experiment is replicated 100 times and then we compute their average. At the end, we have 100,000 averages of simulated data of GV which will be used to study that distribution. Just for illustration, let us choose $p = 3$ and $n = 5$, 20 and 100 (representing small, moderate and large sample size) and observe the histogram and numerical summary.

Below, we see the results. In Figure 4.1, the histogram of 100,000 averages of simulated data of GV is presented. Figure 4.1(a) is for $n = 5$ while Figures 4.1(b) and 4.1(c) are for $n = 20$ and 100. Their numerical summaries consisting of the smallest data (Min), first quartile (Q_1), Median, third quartile (Q_3) and largest data (Max) are given in Table 4.1.

Figure 4.1 illustrates that, for $n = 5$, the empirical distribution of GV is far from being normal. It is strongly skewed to the right. According to Anderson-Darling's test of normality, see Razali and Wah (2011), we get $AD = 14561.56$ and p-value < 0.005. This means that normality is rejected. This is also so for $n = 20$. For this moderate value of n too, the normality assumption is rejected ($AD = 2508.491$ and p-value < 0.005).



(a) (b) (c)

Figure 4.1. Histogram of GV for $p = 3$ and $n = 5$, 20 and 100

As can be seen in Figure 4.1(c), even for $n = 100$, the distribution is still far from being normal ($AD = 498.787$ and p-value < 0.005). We will see in Table 4.7 (Appendix B) that for $p = 3$ and $n = 5$, with PFA = 0.0027, we have K = 3.9663. This value is larger than 3 as required in (3.6). There is no doubt that the larger the sample size, the closer the distribution to normality. But, large sample size is not interesting in industrial application.

The above situation is illustrated in Table 4.1 where the distribution of $|S|$ in (2.1) is represented in the form of numerical summary. This table is in accordance with the results in Figure 4.1.

Table 4.1. Numerical summary of VV for $p = 3$ and $n = 5$, 20 and 100

Statistic	$n = 5$	$n = 20$	$n = 100$
Max	3.07E+01	6.27E+00	2.71E+00
Q_3	3.89E-01	1.08E+00	1.12E+00
Median	1.32E-01	7.30E-01	9.45E-01
Q_1	3.82E-02	4.84E-01	7.98E-01
Min	1.72E-07	2.65E-02	2.56E-01

The simulation results in this section suggest that when $p = 3$, even for sample size n as large as 100, the *IGV*-chart with control limits (3.6) is not apt for monitoring the process variability. It is for this reason that we suggest (4.1).

The problem in using (4.1) is to find (i) PFA for K = 3, and (ii) K if PFA = α. To find PFA when K = 3, we proceed as follows.

Step 1. Find the average and standard deviation of 100,000 averages of simulated data of GV.

Step 2. Calculate UCL = average + 3*(standard deviation)

Step 3. Find α where the $(1-\alpha)$-th quantile of the simulated data equals UCL.

Step 4. Then, PFA = 2α.

The results for $p = 3$ and $n = 5$, 20 and 100 are given in Table 4.2.

Table 4.2. PFA of *GV*-chart for $p = 3$

n	PFA
5	0.0374
20	0.0311
100	0.0146

The PFA in the second column is far greater than 0.0027 even for $n = 100$. This strongly suggests not to use the control limits (3.6) with PFA = 0.0027. According to this table, if $p = 3$ and $n = 5$, then (3.6) does not reflect PFA = 0.0027. We find that the corresponding PFA is $\alpha = 0.0374$ (37400 DPMO).

Now, we can proceed the following steps to find K for PFA = α (= 0.0027).

Step 1. Find the average and standard deviation of 100,000 averages of simulated data of GV.

Step 2. Sort the 100,000 simulated data in ascending order

Step 3. Find $x_{0.99865}$ the 0.99865-th quantile of the 100,000 simulated data.

Step 4. Then, K = ($x_{0.99865}$ − average)/(standard deviation).

The results for $p = 3$ and $n = 5$, 20 and 100 are given in Table 4.3.

Table 4.3. The multiplier K in (4.1)

n	K
5	9.2589
20	5.4568
100	3.9663

This table shows again that K = 3 in (3.6) is not appropriate. If PFA = 0.0027 and $n = 5$, then K = 9.2589. Thus, if PFA = 0.0027 is preferable, for $p = 3$ and $n = 5$, the control limits are,

$$\begin{aligned} \text{UCL} &= |\bar{S}|\left(\frac{b_1}{b_3} + 9.2589\sqrt{\frac{b_2}{b_3^2 + b_4}}\right) \\ \text{CL} &= |\bar{S}|\frac{b_1}{b_3} \\ \text{LCL} &= \max\left\{0, |\bar{S}|\left(\frac{b_1}{b_3} - 9.2589\sqrt{\frac{b_2}{b_3^2 + b_4}}\right)\right\} \end{aligned} \tag{4.2}$$

The above illustrations lead us to continue the simulation experiment to find K for a given PFA, and to find PFA for a given K for general values of n and p. This is discussed in the next section.

4.2. A more general result

The practical implementation of *IGV*-chart for small n requires the appropriate value of K for a given value of p and a desired PFA. First of all, we look at the graphic of numerical summary of GV as a function of n for $p = 3$ and 5 resulted from simulation experiment described in the previous section. This graphic is presented in Figure 4.2.

(a) (b)

Figure 4.2. Numerical summary of GV as function of n for (a) $p = 3$ and (b) $p = 5$

In this figure, the horizontal axis represents the sample size n while the vertical axis is consecutively (from the top to the bottom); Max, Q_3, Median, Q_1 and Min. We see that even for n as large as 100, the numerical summary indicates that the distribution of GV is skewed to the right and cannot be considered as normal. This is true for all values of p in the experiment.

To facilitate the readers, in this section we provide statistical tables for K when n, p and PFA are known. Table 4.6 in Appendix A provides the value of PFA if we use the control limits in (3.6). Furthermore, Tables 4.7 – 4.12 in Appendix B give the values of K in (4.1) for some selected pairs of (n, p) and different PFA. All these tables were obtained from the same simulation experiment discussed in the previous section.

4.3. Example

Let us restudy again the example in 3.1.3 and in 3.2.4. Since $p = 3$ and $n = 5$, with PFA = 0.0027 we find from Table 4.7 in Appendix B the value of K in (4.1) is equal to 9.2589. Accordingly, the control limits for *IGV*-chart are LCL = 0, CL = 0.0011 and UCL = 0.0228. With this result, the *IGV*-chart is presented in Figure 4.3.

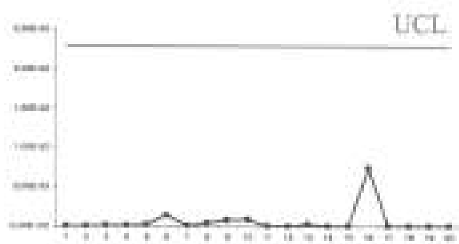


Figure 4.3. Reliable *IGV*-chart

No OOC signal is detected by this chart. This is most probably due to the use of too small sample size or too small PFA or both. For PFA = 0.0027, the larger the sample size, the closer the UCL to that in (3.6). Thus, to optimize the performance of reliable *IGV*-chart for a desired PFA, the use of large enough (but still manageable) sample size is recommended. If the sqmpe size is too small, the chart is very risky of not being able to detect the shift in covariance structure.

4.4. Message to the practitionners

GV-chart is the most adopted chart in monitoring operation of process variability. However, instead of using *GV*-chart, we suggest to use *IGV*-chart since its control limits are unbiased. But, it is only apt when the sample size n is sufficiently large which is not suitable in manufacturing industry.

When n is small, the control limits in (3.6) is misleading. It must be changed with (4.1). For this purpose, the reliability constant K is provided in Tables 4.7 – 4.12 as a function of n, p and PFA.

APPENDIX A: The value of PFA related to the control limits (3.6)

Table 4.6. PFA if K = 3

	Number of quality characteristics (p)							
n	3	4	5	6	7	8	9	10
4	0.0315							
5	0.0374	0.0257						
6	0.0393	0.0325	0.0217					
7	0.0399	0.0363	0.0288	0.0192				
8	0.0398	0.0380	0.0328	0.0260	0.0171			
9	0.0393	0.0389	0.0354	0.0300	0.0233	0.0154		
10	0.0385	0.0392	0.0368	0.0330	0.0278	0.0212	0.0137	
11	0.0377	0.0392	0.0380	0.0350	0.0309	0.0256	0.0193	0.0123
12	0.0369	0.0390	0.0383	0.0363	0.0329	0.0287	0.0237	0.0176
13	0.0362	0.0385	0.0387	0.0373	0.0347	0.0312	0.0265	0.0220
14	0.0354	0.0381	0.0388	0.0380	0.0357	0.0327	0.0294	0.0253
15	0.0345	0.0376	0.0387	0.0382	0.0367	0.0343	0.0312	0.0279
20	0.0311	0.0351	0.0372	0.0383	0.0383	0.0377	0.0366	0.0352
30	0.0261	0.0304	0.0334	0.0354	0.0366	0.0375	0.0379	0.0380
40	0.0230	0.0269	0.0301	0.0324	0.0341	0.0356	0.0365	0.0373
50	0.0207	0.0244	0.0275	0.0298	0.0318	0.0334	0.0344	0.0356
60	0.0188	0.0224	0.0251	0.0277	0.0297	0.0313	0.0326	0.0340
70	0.0174	0.0208	0.0235	0.0259	0.0278	0.0295	0.0310	0.0322
80	0.0164	0.0194	0.0221	0.0243	0.0262	0.0279	0.0294	0.0308
90	0.0154	0.0183	0.0207	0.0230	0.0250	0.0267	0.0281	0.0294
100	0.0146	0.0175	0.0198	0.0219	0.0237	0.0254	0.0268	0.0281

APPENDIX B: The constant K in (4.1) related to selected PFAs

Table 4.7. K if PFA = 0.0027

	Number of quality characteristics (p)							
n	3	4	5	6	7	8	9	10
4	10.3283							
5	9.2589	10.5126						
6	8.5119	9.9147	10.4343					
7	7.9459	9.2809	10.1479	10.2630				
8	7.4908	8.7717	9.7243	10.3160	9.9143			
9	7.1348	8.3042	9.2773	9.9143	10.2481	9.5782		
10	6.8567	7.9557	8.8941	9.6572	10.1798	10.0235	9.0616	
11	6.5967	7.6055	8.5325	9.3216	9.8647	10.1392	9.9722	8.6011
12	6.3966	7.3482	8.2278	8.9707	9.6038	10.0213	10.1061	9.5263
13	6.2231	7.1428	7.9705	8.6788	9.2617	9.7639	10.0970	10.0040
14	6.0879	6.9268	7.6769	8.3836	9.0407	9.5815	9.9631	10.0556
15	5.9391	6.7483	7.4894	8.1931	8.7666	9.3013	9.6917	9.9829
20	5.4568	6.0934	6.6973	7.2742	7.7976	8.2956	8.7108	9.0742
30	4.9219	5.4069	5.8418	6.2717	6.6567	7.0336	7.4028	7.7249
40	4.6172	5.0222	5.3700	5.7017	6.0193	6.3339	6.6345	6.9339
50	4.4267	4.7572	5.0666	5.3577	5.6370	5.8852	6.1434	6.3915
60	4.2830	4.5798	4.8496	5.0953	5.3335	5.5684	5.7912	5.9772
70	4.1760	4.4554	4.6822	4.9069	5.1488	5.3284	5.5297	5.7183
80	4.0813	4.3348	4.5586	4.7648	4.9674	5.1379	5.3295	5.5050
90	4.0115	4.2430	4.4581	4.6474	4.8263	5.0097	5.1644	5.3271
100	3.9663	4.1802	4.3756	4.5528	4.7116	4.8789	5.0302	5.1676

Table 4.8. K for PFA = 0.005

n	Number of quality characteristics (p)							
	3	4	5	6	7	8	9	10
4	7.9727							
5	7.4666	7.8516						
6	6.9881	7.7474	7.5211					
7	6.6217	7.4477	7.7386	7.2494				
8	6.2987	7.1195	7.6054	7.7050	6.9182			
9	6.0590	6.8324	7.4090	7.6281	7.5164	6.5461		
10	5.8538	6.6050	7.1756	7.5566	7.7017	7.2797	6.1465	
11	5.6779	6.3774	6.9808	7.4094	7.6158	7.5478	7.1133	5.7603
12	5.5305	6.1938	6.7745	7.2231	7.5215	7.6328	7.4377	6.7564
13	5.3921	6.0540	6.6050	7.0445	7.3617	7.5927	7.5214	7.3554
14	5.2943	5.9021	6.4164	6.8684	7.2296	7.5057	7.6219	7.5128
15	5.1788	5.7741	6.2870	6.7309	7.0895	7.3730	7.5337	7.5569
20	4.8154	5.2980	5.7343	6.1126	6.4872	6.7991	7.0614	7.2537
30	4.3941	4.7675	5.0999	5.4081	5.6841	5.9515	6.2101	6.4244
40	4.1494	4.4667	4.7385	4.9960	5.2252	5.4590	5.6694	5.8827
50	3.9958	4.2569	4.5012	4.7268	4.9276	5.1251	5.3186	5.4865
60	3.8785	4.1183	4.3373	4.5259	4.7124	4.8872	5.0623	5.2014
70	3.7903	4.0095	4.1996	4.3798	4.5629	4.7033	4.8628	4.9997
80	3.7233	3.9226	4.0996	4.2673	4.4210	4.5637	4.7013	4.8300
90	3.6609	3.8518	4.0195	4.1686	4.3095	4.4533	4.5743	4.7014
100	3.6212	3.7949	3.9510	4.0927	4.2229	4.3507	4.4772	4.5888

Table 4.9. K for PFA = 0.01

n	Number of quality characteristics (p)							
	3	4	5	6	7	8	9	10
4	5.7636							
5	5.6782	5.4186						
6	5.4694	5.6735	5.0288					
7	5.2648	5.6410	5.5268	4.7227				
8	5.0984	5.4982	5.6058	5.3693	4.4227			
9	4.9401	5.3703	5.5851	5.5048	5.1306	4.1270		
10	4.8125	5.2469	5.5016	5.5894	5.4451	4.8852	3.7903	
11	4.7010	5.1250	5.4152	5.5657	5.5401	5.2762	4.6763	3.5271
12	4.6068	5.0066	5.3077	5.5088	5.5688	5.4480	5.1200	4.4201
13	4.5176	4.9195	5.2108	5.4336	5.5362	5.5354	5.3045	4.9716
14	4.4434	4.8299	5.1314	5.3600	5.4996	5.5319	5.4737	5.2154
15	4.3672	4.7472	5.0547	5.2819	5.4433	5.5379	5.4866	5.3922
20	4.1100	4.4325	4.7226	4.9482	5.1586	5.3087	5.4293	5.5056
30	3.8027	4.0711	4.2984	4.5057	4.6755	4.8510	4.9825	5.1158
40	3.6262	3.8541	4.0493	4.2279	4.3773	4.5308	4.6697	4.7997
50	3.5110	3.7069	3.8852	4.0385	4.1778	4.3111	4.4352	4.5510
60	3.4264	3.6035	3.7581	3.9030	4.0321	4.1519	4.2648	4.3668
70	3.3537	3.5208	3.6631	3.7944	3.9145	4.0204	4.1301	4.2239
80	3.3026	3.4541	3.5864	3.7146	3.8187	3.9245	4.0224	4.1127
90	3.2573	3.4018	3.5256	3.6404	3.7431	3.8436	3.9292	4.0258
100	3.2239	3.3547	3.4765	3.5774	3.6760	3.7663	3.8555	3.9366

Table 4.10. K for PFA = 0.025

	Number of quality characteristics (*p*)							
n	3	4	5	6	7	8	9	10
4	3.4778							
5	3.7246	3.0606						
6	3.7375	3.5026	2.6988					
7	3.7071	3.6750	3.2828	2.4485				
8	3.6582	3.6988	3.5133	3.0758	2.2189			
9	3.6048	3.7029	3.6263	3.3498	2.8530	2.0182		
10	3.5467	3.6790	3.6606	3.5166	3.2153	2.6592	1.8162	
11	3.5075	3.6491	3.6836	3.6057	3.4073	3.0453	2.4810	1.6522
12	3.4583	3.6149	3.6705	3.6435	3.5091	3.2709	2.8956	2.3063
13	3.4208	3.5754	3.6541	3.6582	3.5830	3.4200	3.1139	2.7467
14	3.3852	3.5433	3.6392	3.6644	3.6173	3.4947	3.3164	3.0209
15	3.3447	3.5101	3.6170	3.6562	3.6409	3.5647	3.4131	3.2157
20	3.2100	3.3752	3.4909	3.5784	3.6260	3.6429	3.6316	3.6005
30	3.0377	3.1851	3.3015	3.3937	3.4678	3.5310	3.5747	3.6047
40	2.9348	3.0639	3.1734	3.2619	3.3359	3.4073	3.4613	3.5138
50	2.8659	2.9802	3.0823	3.1656	3.2395	3.3050	3.3558	3.4089
60	2.8093	2.9189	3.0053	3.0888	3.1601	3.2196	3.2762	3.3280
70	2.7657	2.8699	2.9520	3.0300	3.0928	3.1535	3.2091	3.2540
80	2.7331	2.8255	2.9073	2.9795	3.0407	3.0994	3.1494	3.2003
90	2.7060	2.7942	2.8671	2.9378	2.9999	3.0546	3.1042	3.1528
100	2.6816	2.7667	2.8363	2.9002	2.9593	3.0136	3.0613	3.1061

Table 4.11. K for PFA = 0.05

	Number of quality characteristics (*p*)							
n	3	4	5	6	7	8	9	10
4	2.1594							
5	2.5230	1.8002						
6	2.6356	2.2639	1.5227					
7	2.6835	2.4778	2.0393	1.3358				
8	2.6948	2.5760	2.2941	1.8546	1.1757			
9	2.6889	2.6306	2.4383	2.1274	1.6772	1.0474		
10	2.6822	2.6554	2.5237	2.3145	1.9939	1.5252	0.9252	
11	2.6689	2.6708	2.5873	2.4230	2.1869	1.8491	1.3974	0.8218
12	2.6542	2.6758	2.6157	2.5013	2.3143	2.0571	1.7278	1.2751
13	2.6425	2.6713	2.6414	2.5533	2.4132	2.2134	1.9297	1.6125
14	2.6292	2.6670	2.6555	2.5946	2.4727	2.3116	2.1112	1.8392
15	2.6141	2.6607	2.6613	2.6149	2.5299	2.3999	2.2209	2.0174
20	2.5512	2.6200	2.6509	2.6547	2.6358	2.5929	2.5263	2.4493
30	2.4652	2.5378	2.5878	2.6177	2.6392	2.6456	2.6382	2.6246
40	2.4069	2.4760	2.5291	2.5681	2.5988	2.6219	2.6352	2.6422
50	2.3643	2.4317	2.4816	2.5240	2.5578	2.5850	2.6028	2.6190
60	2.3336	2.3966	2.4431	2.4870	2.5208	2.5510	2.5714	2.5921
70	2.3058	2.3675	2.4144	2.4563	2.4881	2.5174	2.5433	2.5640
80	2.2875	2.3438	2.3885	2.4263	2.4619	2.4921	2.5158	2.5381
90	2.2709	2.3222	2.3657	2.4059	2.4394	2.4670	2.4935	2.5174
100	2.2556	2.3057	2.3495	2.3858	2.4174	2.4454	2.4705	2.4933

Table 4.12. K for PFA = 0.1

	Number of quality characteristics (p)							
n	3	4	5	6	7	8	9	10
4	1.1740							
5	1.5345	0.9114						
6	1.6938	1.3014	0.7274					
7	1.7821	1.5121	1.1175	0.6111				
8	1.8338	1.6357	1.3440	0.9787	0.5176			
9	1.8633	1.7180	1.4915	1.2028	0.8586	0.4471		
10	1.8842	1.7675	1.5923	1.3684	1.0965	0.7575	0.3829	
11	1.8958	1.8088	1.6670	1.4812	1.2611	0.9912	0.6779	0.3304
12	1.9033	1.8349	1.7190	1.5698	1.3796	1.1601	0.9050	0.6045
13	1.9102	1.8525	1.7602	1.6305	1.4769	1.2895	1.0651	0.8271
14	1.9120	1.8703	1.7914	1.6882	1.5449	1.3824	1.2073	0.9936
15	1.9144	1.8801	1.8150	1.7244	1.6075	1.4706	1.3052	1.1308
20	1.9115	1.9069	1.8784	1.8327	1.7742	1.6990	1.6119	1.5236
30	1.8898	1.9046	1.9037	1.8919	1.8756	1.8472	1.8101	1.7705
40	1.8697	1.8898	1.9008	1.9023	1.8987	1.8892	1.8727	1.8528
50	1.8529	1.8773	1.8899	1.8980	1.9007	1.8994	1.8941	1.8831
60	1.8417	1.8645	1.8804	1.8908	1.8968	1.9007	1.8988	1.8961
70	1.8295	1.8534	1.8709	1.8828	1.8910	1.8957	1.8983	1.8986
80	1.8200	1.8455	1.8612	1.8752	1.8850	1.8911	1.8948	1.8971
90	1.8128	1.8353	1.8533	1.8666	1.8779	1.8849	1.8905	1.8944
100	1.8050	1.8278	1.8450	1.8604	1.8703	1.8789	1.8850	1.8902

CHAPTER 5
VECTOR VARIANCE CHART

As remarked in Cleroux and Ducharme (1989), VV is simply the variance of a random vector. As a measure of multivariate variability, its introduction can be seen in Djauhari (2007) and its application in MSPC is discussed in Djauhari et al. (2008). Since it is a complementary measure of GV and it features good properties, when GV fails to detect that two covariance matrices are different to each other, VV might be able to do the job. The converse is also true. Thus, a simultaneous use of both measures is recommended to get a better understanding of process variability.

In this chapter we recall VV briefly. Then, we discuss the construction of *VV*-chart and its use in process variability monitoring. It visualizes the history of process variability and is constructed by plotting the value of VV for each sample and the control limits in the same diagram. Later on, we study its sampling distribution and show that this chart has a much better ARL than *IGV*-chart.

We begin our discussion by recalling in Section 5.1 the basic notion of VV followed by its geometric interpretation. Then, Section 5.2 focuses on its sampling distributional behavior. This will allow us in Section 5.3 to construct *VV*-chart for large as well as small sample size. Its sensitivity analysis is discussed in Section 5.4 and an industrial example is presented in Section 5.5. This chapter ends with a practical message to the practitioners in Section 5.6.

5.1. Basic notion

In bivariate case, we use to talk about the variance of each variable, covariance and correlation of the two variables. What can we say when we deal with multivariate case? Let us consider two random vectors X and Y of p and q dimension, respectively. Here, p and q might be different. If we denote Σ_{11} and Σ_{22} the covariance matrices of X and of Y, and $\Sigma_{21} = \Sigma_{12}^t$ the covariance matrix between Y and X, then their joint covariance matrix is,

$$\Sigma = \begin{pmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{pmatrix}.$$

Analogous to the bivariate case, the coefficient

$$\rho V = \frac{Tr(\Sigma_{12}\Sigma_{21})}{\sqrt{Tr(\Sigma_{11}^2)Tr(\Sigma_{22}^2)}}.$$

measures the linear relationship of the two random vectors. This is the so-called "vector correlation." Accordingly, we call $Tr(\Sigma_{12}\Sigma_{21})$ vector covariance of X and Y and $Tr(\Sigma_{11}^2)$ and $Tr(\Sigma_{22}^2)$ vector variance of X and of Y, respectively.

In the rest of this chapter we focus our discussion on VV of a random vector, say X, and its role as a multivariate variability measure. Let X be of dimension p and have covariance matrix Σ positive definite. The VV of X, $Tr(\Sigma^2)$, is the sum of all diagonal elements of Σ^2. It is then simply equal to the sum of squares of all elements of Σ or, equivalently, the sum of squares of all eigenvalues of Σ. This is different from GV which is the product of all eigenvalues of Σ.

From this definition, we clearly see that,

1. In univariate case, VV is the square of the variance while GV is the variance itself.
2. In univariate case, VV = 0 if and only if GV = 0 if and only if the distribution of X is degenerate at its mean.
3. In multivariate case, VV = 0 if and only if the distribution is degenerate at the mean vector. On the other hand, GV = 0 if the distribution is degenerate at the mean vector or if the data dimension is actually less than p.
4. VV is small if each of the p variables has small variance and it is large if at least one variable has a large variance. However, GV is small is there is a variable which is nearly a linear combination of the other(s).
5. When VV is large, it does not mean that all variables have large variances. It does not also mean that there is no variable near the hyperplane formed by the other variables.

Thus, VV generalizes the variability measure from univariate into multivariate setting, and both VV and GV have different but complementary properties. Furthermore, algebraically GV is a multilinear form and VV is a quadratic form. This suggests that simultaneous use of both GV and VV will enrich our understanding of multivariate variability. Due to these properties, in what follows VV will be discussed more in-depth.

5.2. Recall on distributional behavior

As we have mentioned in Chapter 2, see (2.9), the asymptotic distribution of sample VV is,

$$\frac{n-1}{\sqrt{8n}}\left\{Tr\left(S^2\right) - \frac{n+1}{n-1}Tr\left(\Sigma^2\right)\right\} \xrightarrow{d} N\left(0, Tr\left(\Sigma^4\right)\right). \quad (5.1)$$

In that chapter, the distribution (5.1) is derived by using the Taylor series of $Tr\left(S^2\right)$ around Σ up to the second term. Since VV is a quadratic form, $Tr\left(S^2\right)$ can exactly be represented by a Taylor series up to the quadratic term. Therefore, before using the result in (5.1) in practice, we have to be sure that the quality of Taylor series approximation is acceptable. For this purpose, we conduct a simulation experiment to study the contribution of the constant term, linear term and quadratic term as a function of n and p.

A simulation experiment, see Djauhari (2007), gives the following results for n varies from 5 to 100 and p from 3 to 10.

1. It is clear that the contribution of the constant term is p. Meanwhile, the contribution of the linear term is generally of order 10^{-3} which is relatively small with respect to p.
2. The contribution of the quadratic term is very small, i.e. of order 10^{-5}.

These results indicate that the use of Taylor series of $Tr\left(S^2\right)$ around Σ up to the second term is justified even for n as small as 5.

5.3. *VV*-chart construction

In this section, we present an application of (5.1) in MSPC. First, we construct *VV*-chart. For this purpose, all what we need is to estimate these parameters,

$$\theta = \frac{n+1}{n-1}Tr\left(\Sigma^2\right) \text{ and } \eta^2 = \frac{8n}{(n-1)^2}Tr\left(\Sigma^4\right),$$

which will determine the control limits. These control limits and the values of all sample VVs define the *VV*-chart. Its performance or sensitivity to the shift in covariance structure will be analyzed in terms of its ARL in Section 5.4.

5.3.1. Parameter estimation

Consider again the m independent samples of size n drawn from a p-variate normal distribution with positive definite covariance matrix Σ; $n > p$. The sample covariance matrices S_i; i = 1, 2, ..., m, are independent and identically distributed and $(n-1)S_i$ is a Wishart distribution $W_p(\Sigma, n-1)$. This leads us to the following estimates of θ and η^2.

Suppose λ_1, λ_2, ..., λ_p are the eigenvalues of Σ and $\bar{\lambda}_k$ is the k-th eigenvalue of $\bar{S}$; the average of S_i. Since S_i; i = 1, 2, ..., m, are independent and $(n-1)S_i$ is a Wishart distribution $W_p(\Sigma, n-1)$, then $(mn-1)\bar{S}$ is also a Wishart matrix $W_p(\Sigma, mn-1)$. Therefore, see Mardia et al. (2000), $\bar{\lambda}_k$ is asymptotically $N\left(\lambda_k, \frac{2\lambda_k^2}{m(n-1)}\right)$. This implies that $Tr\left(\bar{S}\right)$ is an asymptotically unbiased estimate of $Tr(\Sigma)$. As a consequence, we have asymptotically,

$$E\left[Tr\left(\bar{S}^2\right)\right] = E\left[\sum_{k=1}^{p}\bar{\lambda}_k^2\right] = \left\{1+\frac{2}{m(n-1)}\right\}\sum_{k=1}^{p}\lambda_k^2 = \left\{1+\frac{2}{m(n-1)}\right\}Tr\left(\Sigma^2\right).$$

This means that $\left\{1+\frac{2}{m(n-1)}\right\}^{-1}Tr\left(\bar{S}^2\right)$, i.e. $\left\{1-\frac{2}{m(n-1)+2}\right\}Tr\left(\bar{S}^2\right)$, is an unbiased estimate of $Tr\left(\Sigma^2\right)$. Finally, if we write

$$\bar{\theta} = \frac{n+1}{n-1}\left\{1-\frac{2}{m(n-1)+2}\right\}Tr\left(\bar{S}^2\right) \tag{5.2}$$

then $\bar{\theta}$ is an asymptotically unbiased estimate of θ.

Now, we look for an unbiased estimate of η^2. We recall that $\bar{\eta}^2 = \frac{8n}{(n-1)^2}Tr\left(\bar{S}^4\right)$ is an estimate of η^2 but not unbiased because,

$$E\left(\bar{\eta}^2\right) = E\left[\frac{8n}{(n-1)^2}Tr\left(\bar{S}^4\right)\right] = \frac{8n}{(n-1)^2}E\left[\sum_{k=1}^{p}\bar{\lambda}_k^4\right]$$
$$= \frac{8n}{(n-1)^2}\left\{1+\frac{12}{m(n-1)}+\frac{12}{\{m(n-1)\}^2}\right\}Tr\left(\Sigma^4\right) \neq \eta^2.$$

However, the last equation implies that,

$$\bar{\eta}^2 = \frac{8n}{(n-1)^2}\left\{1+\frac{12}{m(n-1)}+\frac{12}{\{m(n-1)\}^2}\right\}^{-1}Tr\left(\bar{S}^4\right) \tag{5.3}$$

is an asymptotically unbiased estimate of η^2.

5.3.2. VV-chart for large sample size

VV-chart is constructed by plotting the value of $Tr\left(S_i^2\right)$ for all i = 1, 2, ..., m, together with the control limits in the same diagram. According to the distribution (5.1) and the estimates (5.2) and (5.3), for large n and PFA = 0.0027, the control limits are,

$$\begin{aligned} \text{UCL} &= \bar{\theta} + 3\bar{\eta} \\ \text{CL} &= \bar{\theta} \\ \text{LCL} &= \max\left\{0, \bar{\theta} - 3\bar{\eta}\right\} \end{aligned} \tag{5.4}$$

Like *IGV*-chart, this chart can also be used to monitor process variability. However, as displayed in Table 4.7 and Table 6.6, the convergence in distribution of VV to normality (5.1) is also slow but faster than that of GV (3.5). As a consequence, when n is not sufficiently large, PFA = 0.0027 for *VV*-chart is no longer suitable.

5.3.3. VV-chart for small sample size

As can be seen in Chapter 6, like GV, the distribution of VV is skewed to the right but its skewness is weaker than that of GV. Therefore, to construct *VV*-chart for small n, we adopt again the heuristic approach suggested in Montgomery (2005, 2009) and employed in Chapter 3 to construct small sample *IGV*-chart. In a similar manner, we can say that most of the probability distribution of VV is contained in the interval $E\left(Tr\left(S^2\right)\right) \pm 3\sqrt{Var\left(Tr\left(S^2\right)\right)}$. Therefore, the control limits are,

$$\begin{aligned}
\text{UCL} &= E\left(Tr\left(S^2\right)\right) + 3\sqrt{Var\left(Tr\left(S^2\right)\right)} \\
\text{CL} &= E\left(Tr\left(S^2\right)\right) \\
\text{LCL} &= \max\left\{0, E\left(Tr\left(S^2\right)\right) - 3\sqrt{Var\left(Tr\left(S^2\right)\right)}\right\}
\end{aligned} \tag{5.5}$$

But, $E\left(Tr\left(S^2\right)\right) = \hat{\theta}$ and $Var\left(Tr\left(S^2\right)\right) = \hat{\eta}^2$. This means that (5.5) is equal to the control limits for large n defined in (5.4). However, it must be underlined that the control limits (5.5) do not reflect PFA. In Table 6.5, Chapter 6, the adjusted PFA as function of n and p will be displayed.

5.4. Sensitivity analysis

The sensitivity of *VV*-chart is studied and compared with that of *IGV*-chart. For this purpose, a simulation experiment was conducted to compare their performance in terms of ARL. First, we generate n random data from a p-variate normal distribution $N_p(\mu,\Sigma)$. In this experiment, $p = 3$, $n = 500$ and, without loss of generality, μ is the

zero vector, and Σ is the identity matrix.

Let PFA = α for both charts. The goal of this experiment is to find α for which the two charts have the same in-control ARL. For this purpose, the experiment is repeated several times until an OOC signal occurs. The number of replications of this repeated experiment is 30. At the end, we get 30 random data of ARL.

Let us denote ARL_0(IGV) and ARL_0(VV) the average of the 30 random data of ARL related to the *IGV*-chart and to the *VV*-chart, respectively. The experiment gives ARL_0(IGV) = ARL_0(VV) = 75 at $\alpha = 0.0148$. This value of α is then used to determine the UCL of the two charts. Based on these UCL, we calculate the out-of-control ARL. We denote ARL_1(IGV) and ARL_1(VV) the out-of-control ARL of the *IGV*-chart and of *VV*-chart.

In Djauhari et al. (2008), two kinds of OOC situation are studied. First, is the situation where the shift in variance occurs. Second, is where we have the shift in covariance. For the first issue, two cases will be considered:

1. The shift of the variance of the first variable.
2. The shift of the variance of all variables.

For the second issue, the covariance shift is defined by shifting the value of all covariance σ_{ij} by the same magnitude for all $i \neq j$.

5.4.1. Variance shift

Table 5.1 presents the results when the shift in the variance of the first variable occurs. Meanwhile, Table 5.2 is the results when the shift of the variance of all variables occurs. The first table indicates that for small shifts in the variance of the first variable, ARL_1(VV) is smaller than ARL_1(IGV). Interestingly, the second table gives a similar indication for a small shift of the variance of all variables. However, ARL_1(VV) and ARL_1(IGV) are almost the same for a large shift.

Table 5.1. $ARL_1(IGV)$ and $ARL_1(VV)$ for variance shift of the first variable

Variance shift	$ARL_1(IGV)$	$ARL_1(VV)$
1.025	60.4333	44.0333
1.050	30.3333	19.0333
1.075	16.7000	16.6333
1.100	15.4333	7.9000
1.125	7.7667	6.4000
1.150	6.6000	4.6667
1.175	5.2000	4.2000
1.200	4.2667	3.3333
1.225	3.3667	1.8333

Table 5.2. $ARL_1(IGV)$ and $ARL_1(VV)$ for variance shift of all variables

Variance shift	$ARL_1(IGV)$	$ARL_1(VV)$
1.025	22.3333	17.2333
1.050	5.3333	4.4333
1.075	2.6667	2.5000
1.100	1.7667	1.6667
1.125	1.3000	1.2667
1.150	1.1667	1.0333
1.175	1.0333	1.0333
1.200	1.0333	1.0000
1.225	1.0000	1.0000

We conclude that, in general, *VV*-chart is more sensitive than *IGV*-chart to the shift in small variance.

5.4.2. Covariance shift

In this experiment, all variances are kept equal to 1 while the covariance σ_{ij} is shifted by the same magnitude which varies from 0 to 0.9 in step 0.1 for all $i \neq j$. The results are presented in Table 5.3.

Table 5.3. $ARL_1(IGV)$ and $ARL_1(VV)$ when σ_{ij} is shifted, $i \neq j$

Covariance shift	$ARL_1(IGV)$	$ARL_1(VV)$
0.1	75.2000	59.3333
0.2	68.3667	25.8000
0.3	22.9667	11.0333
0.4	6.7000	5.2333
0.5	2.0667	2.6000
0.6	1.1333	1.3667
0.7	1.0000	1.0667
0.8	1.0000	1.0000
0.9	1.0000	1.0000

This table shows a similar phenomenon as showed in the previous two tables. Therefore, in general, the performance of *VV*-chart is better than that of *IGV*-chart. This is very promising. It is more sensitive than *IGV*-chart to detect not only small variance but also small correlation.

5.5. Example

In this section we discuss an implementation of *VV*-chart in process variability monitoring of drive rib production process at Indonesian Aerospace Ltd., located in Bandung, Indonesia. Drive rib is a component of aircraft wing. The authors are in debt to Ir. Sutarno, an aircraft engineer at Indonesian Aerospace Ltd., for providing the data and giving permission to publish the results.

In this monitoring operation, $m = 24$, $n = 4$ and $p = 3$. For each sample, after having computed the covariance matrix, we compute the corresponding VV. For all samples, the observed value of VV, $Tr\left(S_k^2\right)$ for each k = 1, 2, ..., 24, are given in Table 5.4.

Table 5.4. Sample vector variance

k	$Tr\left(S_k^2\right)$	k	$Tr\left(S_k^2\right)$
1	6.40*E*-06	13	1.77*E*-05
2	8.77*E*-07	14	2.05*E*-05
3	4.47*E*-05	15	3.55*E*-05
4	8.17*E*-05	16	2.45*E*-05
5	4.68*E*-06	17	2.82*E*-05
6	1.05*E*-06	18	3.66*E*-05
7	1.23*E*-06	19	7.09*E*-05
8	1.90*E*-06	20	7.48*E*-06
9	4.41*E*-06	21	4.35*E*-06
10	5.26*E*-07	22	1.19*E*-06
11	6.20*E*-06	23	1.96*E*-06
12	2.78*E*-06	24	6.68*E*-08

We also compute the average of sample covariance matrices, $\bar{S}$, and its square, $\bar{S}^2$. And we find $Tr\left(\bar{S}^2\right) = 9.85E\text{-}06$ and $Tr\left(\bar{S}^4\right) = 9.14E\text{-}11$.

To construct the *VV*-chart, we need to find the UCL. Since $n = 4$ is small, we need to calculate the parameter estimates $\bar{\theta}$ in (5.2) and $\bar{\eta}^2$ in (5.3), and the UCL in (5.5). Here are the results; we have $\bar{\theta}$ = 1.60E-05, $\bar{\eta}^2$ = 2.78E-10 and UCL = 6.60*E*-05. Accordingly, the corresponding *VV*-chart is displayed in Figure 5.2. In this figure, the vertical axis refers to the value of VV and the horizontal axis is the sample number.

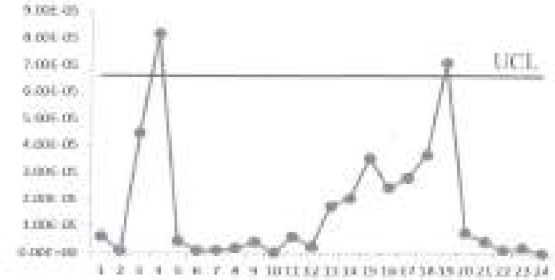


Figure 5.2. *VV*-chart

Two OOC signals occur in this chart, i.e. at the fourth and nineteenth samples. At these two monitoring periods, root causes analysis is required. However, since this problem is beyond the scope of this book, the solution is not discussed in this book.

The readers who are interested are invited to compare the history of process variability issued from *VV*-chart with that provided by *IGV*-chart (and even by *GV*-chart). Two different histories will occur showing the advantage that could be obtained when both charts are used simultaneously.

It is worth noting that all the computations in preparing and drawings the *VV*-chart are realized using Microsoft Excel 2000. Again, this illustrates the simplicity of *VV*-chart construction.

5.6. Message to the practitioners

This chapter presents the procedure to construct the *VV*-chart. It consists of;

Step 1. For each sample k, calculate S_k and $Tr\left(S_k^2\right)$; $k = 1, 2, \ldots, m$.

Step 2. Calculate $\bar{S}$ the average of all S_k, and then $\bar{S}^2$.

Step 3. Calculate $Tr\left(\bar{S}^2\right)$ and $Tr\left(\bar{S}^4\right)$; the sum of squares of all elements of $\bar{S}$ and that of $\bar{S}^2$, respectively.

Step 4. Calculate $\bar{\theta}$ in (5.2), $\bar{\eta}^2$ in (5.3), and the UCL in (5.4).

Step 5. Construct the *VV*-chart by plotting $Tr\left(S_k^2\right)$ issued from *Step 1* and the UCL obtained in *Step 4* in the same chart.

Remember that, when n is small, the *VV*-chart defined by the control limits in (5.5) has nothing to do with PFA. So, the reliability of *VV*-chart in Figure 5.2 is unknown. If PFA is important, we recommend to go directly to Chapter 6 where a reliable *VV*-chart is provided.

CHAPTER 6
RELIABLE VECTOR VARIANCE CHART

The *VV*-chart defined by the control limits in (5.4) is for large n with PFA = 0.0027. This PFA is reflected on the right hand side of (5.4) in the form of the mutiplier constant "3." In the current practice, these control limits are also used for small n as described in (5.5) but with unknown PFA. Thus, when n is small, the reliability of *VV*-chart is unknown. Why? Because, when PFA = 0.0027, the mutiplier constant "3" is no longer suitable.

Let us again denote K the reliability constant for a desired PFA. Then, instead of using the control limits in (5.5), we use these ones

$$\begin{aligned} \text{UCL} &= \bar{\theta} + \text{K}\bar{\eta} \\ \text{CL} &= \bar{\theta} \\ \text{LCL} &= \max\left\{0, \bar{\theta} - \text{K}\bar{\eta}\right\} \end{aligned} \tag{6.1}$$

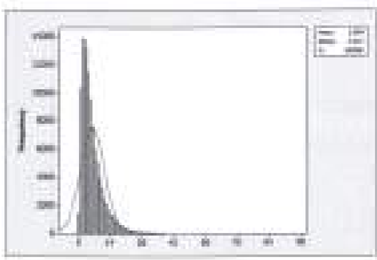
Well, the problem now is to find PFA (α) if we fix K = 3 and to find K if we fix a desired PFA. To solve this problem, in what follows the results of the same simulation study presented in Chapter 4 will be reported. We start our report by recalling in Section 6.1 the simulation experiment and its preliminary results. A more general results are presented in Section 6.2 and an example of application is given in Section 6.3. To close this chapter, a practical message is addressed in the last section to facilitate the practitioners.

6.1. Simulation experiments

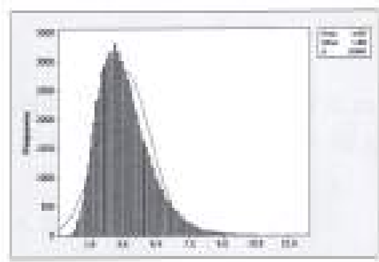
The simulation experiments in Chapter 4, Section 4.1, were actually conducted to generate not only random data of GV but also of VV. In this chapter we focus on random generation of VV. For each pair of (n, p) we simulate n random data from $N_p\left(0, I_p\right)$ and then compute the value of VV. This experiment is repeated 100,000 times to get 100,000 random data.

We replicate that experiment 100 times to get 100 datasets each of which consists of 100,000 random data. Finally, from all these datasets, we have 100,000 averages of simulated data which will be used to study the distribution of VV. For $p = 3$ and $n = 5$, 20 and 100, an illustration of VV distribution will be presented in the form of histogram and numerical summary.

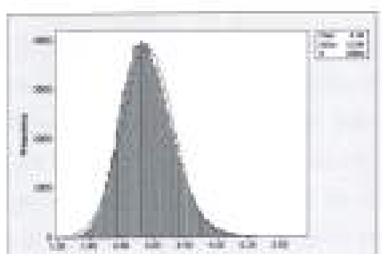
In Figure 6.1, the histogram of VV is presented. Figure 6.1(a) is for $n = 5$ while Figures 6.1(b) and 6.1(c) are for $n = 20$ and $n = 100$.



(a)



(b)



(c)

Figure 6.1. Histogram of VV for $p = 3$ and $n = 5$, 20 and 100

This figure illustrates a samilar situation as for GV in Figure 4.1. For $n = 5$, the empirical distribution of VV is far from being normal. It is strongly skewed to the right but weaker than GV. According to Anderson-Darling's test of normality, see Razali and Wah (2011), $AD = 4725.144$ and p-value < 0.005. Thus, VV is not normal. Nevertheless, its AD is far less than $AD = 14561.56$ for GV which means that VV tends to normality faster than GV. This situation is aso applied for moderate and large n. For moderate $n = 20$, $AD = 1012.027$ and p-value < 0.005 while $AD =$

2508.491 for GV. Meanwhile, for large $n = 100$, the histogram is seemingly close to normality. However, its $AD = 485.514$ and p-value < 0.005 indicate that it is still far from but closer to normality than GV which has $AD = 498.787$.

Regarding the numerical summary of VV, Table 6.1 supports the results given in Figure 6.1. If we compare this table with Table 4.1, we see that VV tends faster to normality than GV. However, it still needs sufficiently large n to use normal approximation.

Table 6.1. Numerical summary of VV for $p = 3$

Statistic	$n = 5$	$n = 20$	$n = 100$
Max	98.3179	13.4826	6.3190
Q_3	7.8010	4.4053	3.4498
Median	4.4828	3.4166	3.0860
Q_1	2.4621	2.6180	2.7601
Min	0.0167	0.4001	1.3673

To find PFA for K = 3 and to find K for a desired PFA, we proceed the same steps described in Section 4.1. Table 6.2 presents the PFA of *VV*-chart in (6.1) for K = 3, $p = 3$ and $n = 5$, 20 and 100.

Table 6.2. PFA of *VV*-chart

n	PFA
5	0.0371
20	0.0213
100	0.0096

Furthermore, when PFA is 0.0027, the constant K in (6.1) for $p = 3$ and $n = 5$, 20 and 100 is given in Table 6.3.

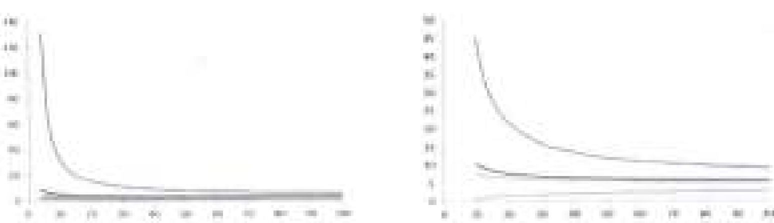
Table 6.3. The multiplier K of *VV*-chart

n	K
5	6.314590
20	4.486869
100	3.589542

These two tables strengthen the claim that the reliability constant K or the PFA needs to be adjusted. Therefore, in the next section, we report a more general result of simulation experiment.

6.2. A more general result

The practical implementation of *VV*-chart for small n requires the appropriate value of K for a given value of p and a desired PFA. First of all, we look at Figure 6.2 the numerical summary of VV as a function of n for $p = 3$ and 5.



(a) (b)

Figure 6.2. Numerical summary of VV as function of n for (a) $p = 3$ and (b) $p = 5$

The vertical axis represents the value of (from the top to the bottom) Max, Q_3, Median, Q_1 and Min. On the other hand, the horizontal axis refers to the sample size n. If this figure is compared with Figure 4.2, it is evident that VV needs less sample size than GV in order for a normal approximation can be justified. However, it is still not suitable for small n. Therefore, we continue our study to find the value of PFA if K = 3 in (6.1) and to find the value of K for a more general values of n, p, and PFA. The results are summarized in Table 6.4 in Appendix C and Tables 6.5 – 6.10 in

Appendix D. If K = 3 in (6.1), Table 6.5 provides the corresponding PFA as function of n and p. On the other hand, for a given PFA, the value of K as function of n and p are given in Tables 6.5 – 6.10.

6.3. Example

Consider again the example in Section 5.5 with $m = 24$, $n = 4$ and $p = 3$. The observed value of VV in each sample is given in Table 5.5. We recall that $Tr\left(\bar{S}^2\right)$ = 9.85*E*-06 and $Tr\left(\bar{S}^4\right)$ = 9.14*E*-11. If we choose PFA = 0.0027, Table 6.5 gives the value of K in (6.1). That is K = 6.7608. For $n = 4$ and $m = 24$, the calculation of $\bar{\theta}$ and $\bar{\eta}^2$ in (5.2) and (5.3) gives $\bar{\theta}$ = 1.60E-05 and $\bar{\eta}^2$ = 2.78E-10. Thus, UCL = 1.29E-04 and the corresponding *VV*-chart is in Figure 6.3.

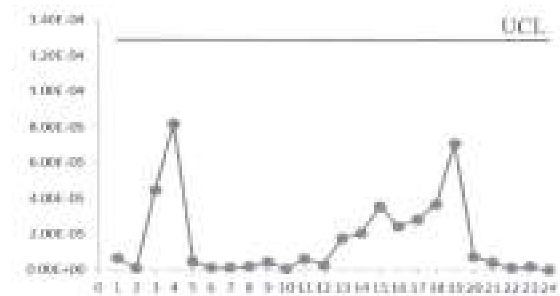


Figure 6.3. Reliable *VV*-chart

This is the reliable *VV*-chart for $m = 24$, $n = 4$, and $p = 3$. The message given by this chart is similar to the one given by *IGV*-chart in the example in Section 4.3. In this chart also, there is no OOC signal occurs. The message in Figure 6.3 is very different compared to the *VV*-chart in Figure 5.2. Again, this is most probably due to the use of too small sample size or too small PFA or both. For PFA = 0.0027, the larger the sample size, the closer the UCL to the one in (5.5). As mentioned in Section 4.3, to optimize the performance of reliable *VV*-chart for a desired PFA, it is recommended to use a large enough sample size but still manageable.

6.4. Message to the practitioners

The procedure in Section 5.6 can similarly be used to construct the reliable *VV*-chart. In this regards, *Step* 3 must be modified. Instead of using (5.4) to calculate the UCL, we use (6.1) with appropriate value of K and PFA given in Appendix D.

APPENDIX C: The value of PFA related to the control limits (6.1)

Table 6.4. PFA if K = 3

n	Number of quality characteristics (*p*)							
	3	4	5	6	7	8	9	10
4	0.0391							
5	0.0371	0.0348						
6	0.0352	0.0328	0.0305					
7	0.0335	0.0309	0.0287	0.0269				
8	0.0320	0.0292	0.0271	0.0253	0.0239			
9	0.0306	0.0278	0.0258	0.0240	0.0224	0.0213		
10	0.0294	0.0266	0.0245	0.0228	0.0215	0.0202	0.0192	
11	0.0282	0.0255	0.0234	0.0218	0.0204	0.0192	0.0182	0.0175
12	0.0272	0.0244	0.0225	0.0210	0.0196	0.0185	0.0175	0.0167
13	0.0262	0.0236	0.0216	0.0200	0.0188	0.0177	0.0168	0.0160
14	0.0253	0.0226	0.0208	0.0193	0.0181	0.0170	0.0162	0.0154
15	0.0245	0.0220	0.0201	0.0186	0.0174	0.0164	0.0156	0.0149
20	0.0213	0.0190	0.0173	0.0160	0.0150	0.0141	0.0135	0.0128
30	0.0172	0.0153	0.0139	0.0131	0.0121	0.0113	0.0109	0.0104
40	0.0149	0.0132	0.0120	0.0111	0.0104	0.0100	0.0095	0.0091
50	0.0133	0.0118	0.0108	0.0100	0.0094	0.0089	0.0085	0.0081
60	0.0121	0.0108	0.0098	0.0091	0.0086	0.0082	0.0078	0.0076
70	0.0112	0.0100	0.0091	0.0085	0.0081	0.0076	0.0073	0.0070
80	0.0106	0.0094	0.0086	0.0081	0.0076	0.0072	0.0069	0.0067
90	0.0100	0.0089	0.0082	0.0076	0.0072	0.0070	0.0066	0.0064
100	0.0096	0.0085	0.0078	0.0073	0.0070	0.0066	0.0064	0.0061

APPENDIX D: The constant K in (6.1) related to selected PFAs

Table 6.5. K if PFA = 0.0027

	Number of quality characteristics (p)							
n	3	4	5	6	7	8	9	10
4	6.7608							
5	6.3143	5.9339						
6	6.0178	5.6534	5.3890					
7	5.7563	5.4045	5.1732	5.0047				
8	5.5588	5.2313	5.0190	4.8399	4.6995			
9	5.3699	5.0942	4.8910	4.7218	4.6094	4.4883		
10	5.2127	4.9547	4.7702	4.6210	4.4998	4.4048	4.3145	
11	5.1175	4.8602	4.6720	4.5312	4.4226	4.3272	4.2474	4.1775
12	5.0022	4.7607	4.5869	4.4545	4.3444	4.2592	4.1894	4.1231
13	4.9145	4.6869	4.5062	4.3748	4.2872	4.2089	4.1356	4.0704
14	4.8258	4.6119	4.4387	4.3287	4.2302	4.1496	4.0819	4.0296
15	4.7593	4.5464	4.3943	4.2804	4.1825	4.0978	4.0464	3.9831
20	4.4868	4.2961	4.1606	4.0747	3.9944	3.9293	3.8816	3.8324
30	4.1648	4.0227	3.9162	3.8481	3.7812	3.7290	3.6917	3.6600
40	3.9822	3.8655	3.7652	3.7046	3.6497	3.6162	3.5861	3.5580
50	3.8700	3.7503	3.6743	3.6171	3.5726	3.5408	3.5160	3.4889
60	3.7750	3.6774	3.6096	3.5509	3.5182	3.4835	3.4570	3.4356
70	3.7191	3.6241	3.5517	3.5123	3.4791	3.4453	3.4167	3.3968
80	3.6625	3.5734	3.5135	3.4770	3.4420	3.4097	3.3872	3.3653
90	3.6113	3.5375	3.4859	3.4366	3.4085	3.3837	3.3617	3.3438
100	3.5895	3.5111	3.4580	3.4149	3.3839	3.3643	3.3409	3.3223

Table 6.6. K for VV-chart with PFA = 0.005

	Number of quality characteristics (p)							
n	3	4	5	6	7	8	9	10
4	5.8112							
5	5.4730	5.1825						
6	5.2378	4.9723	4.7560					
7	5.0448	4.7785	4.5972	4.4508				
8	4.8924	4.6385	4.4673	4.3303	4.2175			
9	4.7546	4.5278	4.3775	4.2295	4.1360	4.0495		
10	4.6304	4.4230	4.2666	4.1524	4.0525	3.9745	3.9019	
11	4.5488	4.3438	4.1925	4.0820	3.9903	3.9093	3.8465	3.7947
12	4.4611	4.2620	4.1279	4.0175	3.9319	3.8618	3.8038	3.7487
13	4.3882	4.1997	4.0659	3.9598	3.8801	3.8153	3.7547	3.7029
14	4.3184	4.1403	4.0064	3.9174	3.8334	3.7703	3.7173	3.6701
15	4.2601	4.0915	3.9727	3.8746	3.7944	3.7295	3.6852	3.6333
20	4.0469	3.8895	3.7740	3.7044	3.6441	3.5927	3.5486	3.5076
30	3.7838	3.6672	3.5782	3.5215	3.4657	3.4236	3.3914	3.3666
40	3.6352	3.5377	3.4584	3.4046	3.3599	3.3319	3.3041	3.2793
50	3.5418	3.4436	3.3798	3.3324	3.2937	3.2661	3.2425	3.2169
60	3.4651	3.3832	3.3248	3.2774	3.2490	3.2187	3.1980	3.1806
70	3.4137	3.3374	3.2786	3.2438	3.2152	3.1859	3.1633	3.1441
80	3.3685	3.2962	3.2456	3.2102	3.1800	3.1553	3.1365	3.1195
90	3.3317	3.2609	3.2206	3.1803	3.1563	3.1370	3.1148	3.1016
100	3.3105	3.2410	3.1960	3.1596	3.1379	3.1136	3.0988	3.0815

Table 6.7. K for VV-chart with PFA = 0.01

	Number of quality characteristics (p)							
n	3	4	5	6	7	8	9	10
4	4.8013							
5	4.5827	4.3686						
6	4.4064	4.2211	4.0677					
7	4.2751	4.0890	3.9578	3.8501				
8	4.1627	3.9800	3.8557	3.7583	3.6778			
9	4.0662	3.9071	3.7834	3.6845	3.6094	3.5520		
10	3.9820	3.8274	3.7141	3.6238	3.5542	3.4924	3.4401	
11	3.9227	3.7681	3.6547	3.5753	3.5065	3.4437	3.3928	3.3548
12	3.8623	3.7069	3.6093	3.5263	3.4627	3.4072	3.3622	3.3209
13	3.8031	3.6635	3.5607	3.4817	3.4209	3.3719	3.3254	3.2875
14	3.7514	3.6155	3.5183	3.4475	3.3896	3.3385	3.2969	3.2606
15	3.7064	3.5808	3.4892	3.4126	3.3549	3.3107	3.2697	3.2359
20	3.5449	3.4314	3.3465	3.2867	3.2391	3.2005	3.1663	3.1380
30	3.3458	3.2572	3.1887	3.1467	3.1016	3.0675	3.0456	3.0223
40	3.2333	3.1544	3.0969	3.0532	3.0200	2.9978	2.9744	2.9565
50	3.1588	3.0880	3.0378	3.0003	2.9689	2.9462	2.9253	2.9053
60	3.1049	3.0380	2.9913	2.9588	2.9326	2.9103	2.8920	2.8748
70	3.0596	3.0008	2.9562	2.9275	2.9045	2.8806	2.8629	2.8499
80	3.0275	2.9694	2.9308	2.9028	2.8795	2.8583	2.8427	2.8306
90	3.0007	2.9447	2.9094	2.8788	2.8593	2.8462	2.8264	2.8169
100	2.9782	2.9273	2.8904	2.8631	2.8428	2.8248	2.8129	2.7981

Table 6.8. K for VV-chart with PFA = 0.025

	Number of quality characteristics (p)							
n	3	4	5	6	7	8	9	10
4	3.5646							
5	3.4594	3.3555						
6	3.3696	3.2719	3.1871					
7	3.3015	3.1994	3.1233	3.0632				
8	3.2409	3.1430	3.0708	3.0087	2.9639			
9	3.1897	3.0940	3.0264	2.9657	2.9197	2.8843		
10	3.1455	3.0545	2.9851	2.9313	2.8904	2.8503	2.8199	
11	3.1051	3.0173	2.9491	2.9002	2.8565	2.8213	2.7899	2.7678
12	3.0701	2.9828	2.9208	2.8729	2.8324	2.7979	2.7667	2.7448
13	3.0401	2.9548	2.8931	2.8438	2.8070	2.7744	2.7477	2.7236
14	3.0111	2.9248	2.8673	2.8220	2.7844	2.7516	2.7278	2.7067
15	2.9829	2.9055	2.8479	2.8023	2.7644	2.7349	2.7104	2.6902
20	2.8858	2.8119	2.7620	2.7197	2.6913	2.6675	2.6438	2.6245
30	2.7613	2.7018	2.6595	2.6294	2.6027	2.5800	2.5651	2.5480
40	2.6898	2.6357	2.5993	2.5694	2.5473	2.5325	2.5156	2.5082
50	2.6409	2.5939	2.5604	2.5328	2.5131	2.4984	2.4831	2.4724
60	2.6059	2.5613	2.5296	2.5060	2.4894	2.4733	2.4612	2.4508
70	2.5752	2.5358	2.5070	2.4854	2.4682	2.4545	2.4435	2.4336
80	2.5551	2.5133	2.4881	2.4681	2.4533	2.4398	2.4293	2.4186
90	2.5381	2.4978	2.4725	2.4548	2.4402	2.4281	2.4173	2.4111
100	2.5225	2.4836	2.4599	2.4407	2.4266	2.4155	2.4089	2.4003

Table 6.9. K for VV-chart with PFA = 0.05

	Number of quality characteristics (p)							
n	3	4	5	6	7	8	9	10
4	2.7022							
5	2.6624	2.6198						
6	2.6276	2.5839	2.5418					
7	2.5961	2.5474	2.5098	2.4756				
8	2.5703	2.5190	2.4824	2.4466	2.4208			
9	2.5423	2.4934	2.4573	2.4262	2.3978	2.3740		
10	2.5224	2.4712	2.4335	2.4030	2.3779	2.3565	2.3375	
11	2.4998	2.4517	2.4136	2.3840	2.3607	2.3400	2.3215	2.3060
12	2.4804	2.4315	2.3964	2.3680	2.3458	2.3249	2.3047	2.2911
13	2.4650	2.4164	2.3810	2.3529	2.3299	2.3093	2.2948	2.2791
14	2.4507	2.4017	2.3680	2.3392	2.3170	2.2982	2.2830	2.2678
15	2.4328	2.3876	2.3558	2.3268	2.3043	2.2873	2.2713	2.2580
20	2.3785	2.3337	2.3038	2.2805	2.2600	2.2443	2.2298	2.2179
30	2.3036	2.2665	2.2405	2.2197	2.2045	2.1911	2.1790	2.1683
40	2.2597	2.2263	2.2025	2.1833	2.1683	2.1582	2.1476	2.1407
50	2.2298	2.1982	2.1772	2.1586	2.1462	2.1351	2.1255	2.1200
60	2.2057	2.1784	2.1559	2.1418	2.1299	2.1200	2.1112	2.1053
70	2.1871	2.1599	2.1419	2.1267	2.1143	2.1059	2.0994	2.0930
80	2.1729	2.1467	2.1294	2.1151	2.1053	2.0974	2.0902	2.0841
90	2.1627	2.1360	2.1183	2.1069	2.0970	2.0893	2.0821	2.0778
100	2.1517	2.1268	2.1103	2.0989	2.0885	2.0792	2.0754	2.0699

Table 6.10. K for VV-chart with PFA = 0.1

	Number of quality characteristics (p)							
n	3	4	5	6	7	8	9	10
4	1.9020							
5	1.9159	1.9195						
6	1.9194	1.9154	1.9103					
7	1.9183	1.9102	1.9013	1.8918				
8	1.9155	1.9053	1.8945	1.8844	1.8751			
9	1.9105	1.8978	1.8856	1.8766	1.8651	1.8565		
10	1.9072	1.8923	1.8795	1.8698	1.8571	1.8495	1.8425	
11	1.8995	1.8840	1.8726	1.8615	1.8517	1.8431	1.8345	1.8273
12	1.8947	1.8802	1.8655	1.8522	1.8448	1.8361	1.8276	1.8205
13	1.8919	1.8729	1.8591	1.8475	1.8376	1.8301	1.8227	1.8153
14	1.8871	1.8687	1.8548	1.8417	1.8333	1.8233	1.8176	1.8105
15	1.8807	1.8634	1.8495	1.8376	1.8276	1.8202	1.8109	1.8057
20	1.8591	1.8406	1.8274	1.8169	1.8082	1.7995	1.7930	1.7856
30	1.8295	1.8106	1.7978	1.7876	1.7796	1.7723	1.7662	1.7612
40	1.8073	1.7918	1.7793	1.7689	1.7628	1.7566	1.7504	1.7468
50	1.7936	1.7778	1.7653	1.7582	1.7510	1.7453	1.7399	1.7361
60	1.7820	1.7674	1.7553	1.7484	1.7406	1.7371	1.7309	1.7285
70	1.7728	1.7585	1.7481	1.7404	1.7343	1.7288	1.7254	1.7219
80	1.7652	1.7519	1.7409	1.7341	1.7284	1.7244	1.7200	1.7167
90	1.7585	1.7458	1.7358	1.7301	1.7237	1.7200	1.7155	1.7133
100	1.7529	1.7402	1.7306	1.7254	1.7191	1.7151	1.7120	1.7089

REFERENCES

1. Alt, F.B. and Smith, N.D. (1988). Multivariate process control. In Krishnaiah, P.R., Rao, C.R. (Ed.) *Handbook of Statistics* (p. 333-351). Elsevier Science.
2. Anderson, T.W. (1966). *Introduction to multivariate statistical analysis*, 1st Edition. John Wiley & Sons, New York.
3. Anderson, T.W. (1984). *Introduction to multivariate statistical analysis*, 2nd Edition, Wiley-Interscience, New York.
4. Beamon, B.M., and Ware, T.M. (1998). A process quality model for the analysis, improvement and control of supply chain systems. *International Journal of Physical Distribution & Logistics Management*, 28, p. 704-715.
5. Carlos Garcia-Diaz, J. (2007). The 'effective variance' control chart for monitoring the dispersion process with missing data. *European Journal of Industrial Engineering*, 1(1), DOI: 10.1504/EJIE.2007.012653.
6. Cleroux, R., Ducharme, G.R. (1989). Vector correlation for elliptical distributions. *Communications in Statistics – Theory and Methods*, 18(4), p. 1441-1454.
7. Da Costa, Jr. N., Nunes, S., Ceretta, P., and Da Silva, S. (2005). Stock-market comovements revisited. *Economics Bulletin*, 7, p. 1-9.
8. Djauhari, M.A. (2005). Improved monitoring of multivariate process variability. *Journal Quality of Technology*, 37(1), p. 32-39.
9. Djauhari, M.A. (2007). A measure of multivariate data concentration. *Journal of Applied Probability & Statistics*, 2(2), p. 139-155.
10. Djauhari, M.A. (2009). Asymptotic distribution of sample covariance determinant. *MATEMATIKA*, 25(1), p. 79-85.
11. Djauhari, M.A. (2016). *Advanced monitoring techniques for complex process variability in manufacturing industry*. Universiti Putra Malaysia Press, Serdang, Malaysia.
12. Djauhari, M.A., Mashuri, M., and Herwindiati, D.E. (2008). Multivariate process variability monitoring. *Communications in Statistics – Theory and Methods*, 37, p. 1742-1754.
13. Djauhari, M.A., and Mohamad, I. (2010). How to control process variability more effectively: the case in b-complex vitamin production process. *South African Journal of Industrial Engineering*, 21(21), p. 207-215.

14. Djauhari, M.A., Salleh, R.M., Zolkeply, Z., and Lee, S.L. (2017). On the reliability of Shewhart-type control charts for multivariate process variability. DOI: 10.1063/1.4982866.
15. Edelman, A., and Rao, N.R. (2005). Random matrix theory. *Acta Numerica*, 14, p. 233-297.
16. Escoufier, Y. (1973). Le traitement des variables vectorielles. *Biometrics* 29, p. 751-760.
17. Florac, A.W., Carleton, A.D., and Barnard, J.R. (2000). Statistical process control: Analyzing a space shuttle onboard software process. *IEEE Software*, July/August, p. 97-106.
18. General Electric (1998). Annual Report, accessed on 15 March 2018 from: www.annualreports.com/HostedData/AnnualReportArchive/g/NYSE_GE_1998.pdf
19. Grambow, S., Stromberg, A.J. (1998). Combining the EID and FSA for computing the minimum volume ellipsoid. Department of Statistics, University of Kentucky.
20. Hanslik, T., Boelle, P., and Flahault, A. (2001). The control chart: An epidemiological tool for public health monitoring. *Public Health*, 115, p. 277-281.
21. Hoerl, R.W., and Snee, R.D. (2012). *Statistical Thinking*. John Wiey & Sons, Inc., Hoboken, New Jersey.
22. Hubert, M., Rousseeuw, P.J., and van Aelst, S. (2008). High-breakdown robust multivariate methods. *Statistical Science*, 23(1), p. 92-119.
23. Jakolte, P., and Saxena, A. (2002). Optimum control limits for employing statistical process control in software process. *IEEE Transactions on Software Engineering*, 28, p. 1126-1134.
24. Kollo, T., and von Rossen, D. (2005). *Advanced multivariate statistics with matrices*. Springer, The Netherlands.
25. Kruegel, C., Valuer, F., and Vigna, G. (2005). *Intrusion detection and correlation, challenges and solutions*. Springer Science + Business Media, Inc., Boston.
26. Larzac, A., and Cleroux, R. (1992). Tests d'homogénéité entre indices de redondance pour des lois elliptiques. *Revue de statistique appliquée*, 40(3), p. 19-33.

27. Mardia, K.V., Kent, J.T., Bibby, J.M. 2000). *Multivariate Analysis*, 7th Printing. Academic Press, London.

28. Mason, R.L., Chou,Y.M., and Young, J.C. (2009) Monitoring variation in a multivariate process when the dimension is large relative to the sample size. *Communications in Statistics-Theory and Methods*, 38, p. 939-951.

29. Montgomery, D.C. (2001). *Introduction to statistical quality control*, 4th Edition. Wiley, New York.

30. Montgomery, D.C. (2005). *Introduction to statistical quality control*. 5th Edition. John Wiley and Sons, Inc., New York.

31. Montgomery, D.C. (2009). *Introduction to statistical quality control*. 6th Edition. John Wiley and Sons, Inc., New York.

32. Muirhead, R.J. (1982). *Aspect of multivariate statistical theory*. John Wiley & Sons, Inc., New York.

33. Neave, H.R. (1990). *The Deming dimension*, SPC Press, Knoxville, Tennessee.

34. Pena, D., and Rodriguez, J. (2005). Descriptive measures of multivariate scatter and linear dependence. Accessed on 4 April 2017 from: http://halweb.uc3m.es/esp/personal/dpena/article/JVMA03.PDF

35. Ragea, V. (2003). Testing correlation stability during hectic financial markets. *Financial Market and Portfolio Management*, 17, p. 289-380.

36. Razali, N., and Wah, B. (2011). Power comparison of Shaphiro-Wilk, Kolmogorov-Smirnov, Lilliefors and Anderson-Darling tests. *Journal of Statistical Modelling and Analytics*. 2, p.21-33.

37. Reis, M.S., and Rato, T.J. (2013). Large-scale statistical process control: Should we use partial or marginal correlations? *American Institute of Chemical Engineers (AIChE)*. Paper presented at the Annual Meeting, November 3-8, 2013, Hotel San Francisco Union Square, San Francisco, https://aiche.confex.com/aiche/2013/webprogram/Paper317381.html

38. Roes, K.C.B., and Dorr, D. (1997). Implementing statistical process control in service processes. *International Journal of Quality Science*, 2, p. 149-66.

39. Rousseeuw, P.J. (1985). Multivariate estimation with high breakdown point. In: Grossman, B. W., Pflug, G., Vincze, I., Wertz, W. (Ed). *Mathematical Statistics and Applications* (p. 283-297). D. Reidel Publishing Company.

40. Ryan, T.P. (2011). *Statistical methods for quality improvement*. 3rd Edition. John Wiley & Sons, Inc., Hoboken, New Jersey.

41. Sellick, J.A. Jr. (1993). The use of statistical process control charts in hospital epidemiology. *Infection Control and Hospital Epidemiology*, 14, p. 649-56.

42. Serfling, R.J. (1980). *Approximation Theorems of Mathematical Statistics*. John Wiley, New York.

43. Shahian, D.M., Williamson, W.A., Svensson, L.G., Restuccia, J.D., and d'Agostino, R.S. (1996). Applications of statistical quality control to cardiac surgery. *Annals of Thoracic Surgery*, 62, p. 1351-1359.

44. Snee, R.D. (2006). Process variation – Enemy or opportunity. *Quality Progress*, December 2006 Edition, p. 73-75.

45. Sulek, J.M. (2004). Statistical quality control in services. *International Journal of Services Technology and Management*, 5, p. 522-531.

46. Sullivan, J.H., Stoumbos, Z.G., Mason, R.L., and Young, J.C. (2007). Step-down analysis for changes in the covariance matrix and other parameters. *Journal of Quality Technology*, 39, p. 66-84.

47. Tang, P.F. and Barnett, N.S. (1996). Dispersion control for multivariate processes. *Australian Journal of Statistics*,38(3), p. 235-251.

48. Wilks, S. (1962). *Mathematical statistics*. John Wiley & Sons Inc., New York.

49. Wood, M. (1994). Statistical methods for monitoring service process. *International Journal of Service Industry Management*, 5, p. 53-68.

50. Woodall, W.H. (2006). Use of control charts in health care monitoring and public health surveillance. *Journal of Quality Technology*, 38, p. 89-104.

51. Woodall, W.H., and Montgomery, D.C. (1999). Research issues and ideas in statistical process control. *Journal of Quality Technology*, 31(4), p. 376-386.

52. Ye., N., Chen, Q. (2001). An anomaly detection technique based on a chi-square statistic for detecting intrusion into information system. *Quality and Reliability Engineering International*, 17, p.105-112.

ABOUT THE AUTHORS

1. Maman Abdurachman Djauhari

Maman A. Djauhari was born in Dayeuh Garut, The Land of Pasundan, Indonesia, on December 8, 1948. He got his Doctorandus in Mathematics from Institut Teknologi Bandung (ITB), Indonesia, in 1974. His Diplome d'Etudes Approfondies and Doctorat du 3ème cycle, both in Pure and Applied Mathematics, were obtained from Université de Montpellier 2, France, in 1977 and 1979, respectively.

His career in academic management started in 1983 when he was appointed as Secretary of Department of Mathematics ITB. Since then, during 25 years without interruption, he involved in ITB management. His last position at ITB before his retirement in 2009 was Chairman of the Council of Professors.

Highest rank of Professorship in Indonesia (IV/e) was conferred by the President of the Republic of Indonesia in 2007 for his scientific and academic achievement. Recently, in 2014, the A-grade UDQ05 of his Professorship which is similar to VK05 was conferred to him by Universiti Putra Malaysia. He was an Academic Advisor, Faculty of Mathematics and Computer Science, Universiti Teknologi Mara, Shah Alam, Malaysia, 2012-2014. At the same period, he was also an Academic Advisor, Faculty of Science, Technology and Human Development, Universiti Tun Hussein Onn, Malaysia.

In professional life, he had received many awards/recognitions including;

1. His post as Consultant at The World Bank in 2009,
2. Gold Medal Award for outstanding contribution in statistics conferred by the Islamic Countries Society of Statistical Sciences (ISOSS) in 2005,
3. Gold Medal Award for Fidelity (*Satya Lencana Karya Satya XXX*) conferred by the President of the Republic of Indonesia in 2005,
4. Excellent Service Award from Faculty of Science, Universiti Teknologi Malaysia, 2013,
5. Excellent Consultation Award, Universiti Teknologi Malaysia, 2011,

6. Recognition from UNESCO-AEGIS (Australian Experts Group in Industrial Study) as activist in scientific knowledge production in Asia-Pacific Region, 2006,
7. Recognition from Associate Editor of Journal "*BioPharm; The Applied Technology of Biopharmaceutical Development*", Milwaukee, USA, for consultancy in statistical analysis, 2001, and
8. Best Lecturer Award, Department of Mathematics, ITB, Indonesia, 2000.

His research interest includes (in alphabetical order) complex network analysis, complex systems, econophysics, financial market analysis, mathematical statistics, multivariate analysis, statistical process control. He can be reached at maman@math.itb.ac.id or maman_djauhari@yahoo.com.

2. Zunna'aim Zolkepley

Zunna'aim Zolkeply was born on August 29, 1989 in Malaysia. He got his Master degree in Statistics from the Department of Mathematical Sciences, Faculty of Science, Universiti Teknologi Malaysia, in 2014. Currently, he is a PhD student at the Faculty of Science, Technology and Human Development, Universiti Tun Hussein Onn, Malaysia.

Currently, his research is focused on multivariate statistical process control for service industry and its application in many areas including education management. He can be contacted at zunnaaim2@yahoo.com.

3. Dyah Erny Herwindiati

Dyah Erny Herwindiati is a Professor at Universitas Tarumanagara, Jakarta, Indonesia. Currently, she is the Dean of Faculty of Information Technology. She got her Doctorate degree in Robust Statistics from Institut Teknologi Bandung (ITB), Indonesia, in 2006.

Her current research is focused on applied robust algorithm for land cover mapping of remote sensing data, data mining, and content based image retrieval. Her email address is dyahh@fti.untar.ac.id.