Bulletin of the *Transilvania* University of Braşov • Vol. 17(66) No. 1 – 2024
Series I: Engineering Sciences
https://doi.org/10.31926/but.ens.2024.17.66.1.2

# R COMPUTER PROGRAM FOR THE $s^2$ CONTROL CHART

**Nicolae EFTIMIE**[1]

***Abstract:*** *To monitor the process variability, the s chart is frequently used. However, the center line of this chart is not an unbiased estimator of the process standard deviation. To avoid this problem, the $s^2$ chart may be used instead of the s chart because its center line is an unbiased estimator for the process variance. The paper presents the conceiving of a computer program for the $s^2$ chart. The program is developed by means of the R language. After presenting the way in which the program was developed, the paper presents the results obtained after running it.*

***Key words:*** *$s^2$ chart, statistical process control, R language.*

## 1. Introduction

The $s^2$ chart for variances may be used for controlling the process variability. Even though the mean of the sample variances is an unbiased estimator of the process variance, this chart is little used in practice and relatively few papers have been published on the subject.

In one of these works, Saeed, Kamal, et al. (2024) used the bootstrap method for creating a control chart that can be used for monitoring the process variability [8]. Another paper, elaborated by Zaagan, Khan, et al. (2024), presents a new parameter free adaptive exponentially weighted moving average control chart for monitoring process dispersion [11]. Regarding the research on examination the *ARL* properties of the $s^2$ Control Chart, it can be mentioned the work of Zhou, Ng, et al. (2022) for performing a study on asymmetric control limits for a weighted-variance mean control chart for a Weibull distributed process [13] and the researches of Zhang, Bebbington, et al. (2005) [12]. Landim, Jardim, et al. (2021) proposed a modified $s^2$ control chart, which allows that as long as the process remains capable, the dispersion of the process can exceed the in-control value, until it reaches a specified maximum value [5]. Ajadi and Riaz (2017) proposed four control charts that may be used for monitoring the process mean and the process dispersion simultaneously [1]. Ho and da Costa Quinino (2016) combined attribute and variable data to monitor the process variability [3]. Among the papers with the objective to study the phase-I design structure of a bayesian variance chart it is

[1] Manufacturing Engineering Department, Faculty of Technological Engineering and Industrial Management, "Transilvania" University – Brasov, Romania, eftimien@unitbv.ro

mentioned the research of Saghir (2016) [9]. Lazariv, Schmid, et al. (2013) realized a study on control charts used for monitoring the variance of a timeseries [6]. Castagliola, Celano, et al. (2009) introduced and investigated the performance of a *CUSUM*-$s^2$ control chart that may be used for monitoring the variance of samples originated from a normally distributed population [2]. There can be also mentioned the studies of Kao and Ho (2006) regarding the monitoring of the sample variances by mean of an optimal normalizing transformation [4].

Unlike the papers mentioned above, the present study proposes the conceiving and development of a computer program for the $s^2$ control chart.

The work is organized in a manner that offers to researchers a literature review regarding the $s^2$ chart, the basic theoretical aspects of this chart, the statements necessary for the computer program development and information about how to run and to use the program results.

## 2. Theoretical aspects regarding s$^2$ chart

This chapter is going over some basics aspects necessary for applying the $s^2$ chart.

The center line of the $s^2$ chart is given by the following formula:

$$CLs^2 = \overline{s^2} = \frac{1}{m}\sum_{i=1}^{m} s_i^2 \tag{1}$$

where $m$ is the number of samples used to prepare the control chart.

The upper control limit and the lower control limit for the $s^2$ chart are calculated using the formulas (2) and (3) respectively [7]:

$$UCLs^2 = \frac{\overline{s^2}\chi^2_{1-\frac{\alpha}{2},n-1}}{n-1} \tag{2}$$

$$LCLs^2 = \frac{\overline{s^2}\chi^2_{\frac{\alpha}{2},n-1}}{n-1} \tag{3}$$

where $n$ is the size of samples, $\chi^2_{1-\frac{\alpha}{2},n-1}$ and $\chi^2_{\frac{\alpha}{2},n-1}$ are the (1 – α/2)-th and the (α/2)-th quantiles of the $\chi^2_{n-1}$ distribution and $\alpha$ is the probability of the error of type I.

The error of type I is committed when an in-control process is considered to be an output of control process.

When the process is controlled using variable control charts, it is monitored from two points of view: from the point of view of variability and from the point of view of its location. To monitor the location of the process, an $\bar{x}$ chart, may be used. By $\bar{x}$ it is denoted the sample mean.

The center line of the $\bar{x}$ chart is given by the following formula:

$$CL\bar{x} = \bar{\bar{x}} = \frac{1}{m}\sum_{i=1}^{m} \bar{x}_i \tag{4}$$

The upper control limit and the lower control limit for the $\bar{x}$ chart are calculated using the formulas (5) and (6) respectively [7]:

$$UCL\bar{x} = \bar{\bar{x}} + Z_{1-\alpha/2}\frac{\sqrt{\overline{s^2}}}{\sqrt{n}} \tag{5}$$

$$LCL\bar{x} = \bar{\bar{x}} - Z_{1-\alpha/2}\frac{\sqrt{\overline{s^2}}}{\sqrt{n}} \tag{6}$$

where $\sqrt{\overline{s^2}}$ is the unbiased estimator of the process standard deviation provided by the $s^2$ chart and $Z_{1-\alpha/2}$ is the $(1 - \alpha/2)$-th quantile of the normal distribution with parameters $\mu = 0$ and $\sigma = 1$.

## 3. Objectives

Due to difficult calculations and graphical representations, the practical application of the $s^2$ control chart requires a suitable computer program. It should be noted that there is not a software application presented in the specialized literature that can be used for this purpose.

The aim of the present paper is to develop a computer program for the $s^2$ chart, to provide an alternative to monitoring process variability using the *s* chart. The process location will be monitored by means of the $\bar{x}$ chart. This program can be used both in research activity and for process control in industrial practice.

The programming language used for the development of the program is *R* [14]. This powerful programming language (used in the field of statistics) is currently being used by more and more researchers and scientists.

## 4. Methods

The computer program was developed in *R* language, version 4.3.3 [14].

The first stage in developing the program is to load the necessary libraries (Figure 1).

```r
library(openxlsx)
library(ggplot2)
library(dplyr)
library(nortest)
```

Fig. 1. *Loading the libraries*

The library *openxlsx* contains functions that allow the program to work with *Excel .xlsx* files [15]. The functions from the next library loaded, *ggplot2*, are used for creating data visualizations by means of the *Grammar of Graphics* [10]. The library *dplyr* is *a grammar of data manipulation* [16]. The last library loaded by the program, *nortest*, contains functions that are used for data normality testing [17].

The control chart can be plotted for any value of the *type I error, alpha*.

The following commands allow the user of the program to enter the value of *alpha* from the keyboard (Figure 2). The *scan* command allows the read of one single value of numeric type.

```
cat("Please enter the value of alpha:\n")
alpha <- scan(what = numeric(), nmax = 1)
```

Fig. 2. *Reading the value of alpha*

The measurements of the samples necessary for the graphical representation of the control chart are recorded in an Excel file.

For a better understanding of the program proposed by this paper it is used an Excel file that contains data obtained by means of the Excel's random numbers generator. These values are normally distributed, and they are simulating a process with the mean $\mu$ = 250.026 and the standard deviation $\sigma$ = 0.008667. These data are recorded in the file *xbars2.xlsx*, which contains data for $m$ = 25 samples, each of size $n$ = 10. The headings of columns are *SampleNumber* and *x*.

Figure 3 presents the sequence of required commands that prepare the data for the control chart.

At the beginning, a *data frame* named *df_xbar_s2* is created from the data recorded in the Excel file *xbars2.xlsx*. Then, by means of the "%>% pipe operator" two operations are combined. First, the data are grouped by the variable *SampleNumber* and then, the grouped data are summarized to obtain for each sample the following values: the size *n*, the average $\bar{x}$ and the variance $s^2$.

```
df_xbar_s2<-read.xlsx("xbars2.xlsx")
chart_data <- df_xbar_s2 %>% group_by(SampleNumber)
chart_data <- chart_data %>%
    summarise(
      n=n(),
      xbar=mean(x),
      s2=var(x)
    )
n <- chart_data$n[1]
```

Fig. 3. *Preparing the data for the control chart*

In order to plot the $s^2$ chart, a new function *Plot_s2_chart()* was created (Figure 4).

This function allows the receiving as an argument an object of type *tibble* that contains the summarized data necessary for the chart plotting. The means of the sample averages and of the sample variances are calculated using the formulas (4) and (1). Next, the number of samples is determined. The control limits for the $s^2$ chart are calculated by means of formulas (2) and (3). In order for the user of the program to notice on the chart the points that signal special causes of variation, the vector *PointsColor* is created. At the beginning, the color of all points on the chart is set to green, and then, the color of the points that are beyond the control limits is set to red.

The argument of the function *ggplot()* specifies that for the graphical representation, the data from the object of type *tibble* named *chart_elements* will be used. The plotting of the chart is realised by means of a series of functions *geom_line(), geom_point(), geom_hline()* and *geom_segment().* At the end of the function, the scales and the labels for the chart axes are established.

At this point, the plot of the s2 chart becomes simpler by using the following command: Plot_s2_chart(chart_elements = chart_data)

```
Plot_s2_chart <- function(chart_elements){
  xdoublebar = mean(chart_elements$xbar)
  s2bar=mean(chart_elements$s2)
  m<-nrow(chart_elements)
  UCLs2 = s2bar/(n-1)*qchisq(p=1-alpha/2, df=n-1,
      lower.tail = TRUE)
  LCLs2 = s2bar/(n-1)*qchisq(p=alpha/2, df=n-1,
      lower.tail = TRUE)
  PointsColor<-rep("darkgreen", m)
  PointsColor[chart_elements$s2>UCLs2]="red"
  PointsColor[chart_elements$s2<LCLs2]="red"
  ggplot(data = chart_elements)+
      geom_line(mapping = aes(x=SampleNumber, y=s2), linewidth = 1) +
      geom_point(mapping = aes(x=SampleNumber, y=s2),
               color=PointsColor, size = 2.2)+
      geom_hline(yintercept = s2bar, linewidth=1.2, colour="blue")  +
      geom_hline(yintercept = UCLs2, linewidth=1.2, colour="blue") +
      geom_hline(yintercept = LCLs2, linewidth=1.2, colour="blue") +
      geom_segment(mapping=aes(x=SampleNumber, y=rep(x=LCLs2,times = m),
                     xend=SampleNumber,
                     yend = rep(x=UCLs2,times=m)), linewidth=0.2)+
      scale_x_continuous(breaks = chart_elements$SampleNumber)+
      scale_y_continuous(breaks = c(LCLs2, s2bar, UCLs2))+
      labs(x = "Sample number", y = "Sample variance")
}
```

Fig. 4. *The code of the function Plot_s2_chart()*

After the graphical representation, the control chart must be analyzed. If the process was influenced by special causes of variation, they must be detected and, if they have a negative effect, they must be eliminated from the process. In addition, the points corresponding to them must be removed from the chart. To perform this operation, the *tibble* named *chart_data* is filtered. In the presented example, the sample number 23 was eliminated from the *tibble*. Then, the function *Plot_s2_chart()* is run again, having as an argument the filtered *tibble* (Figure 5).

```
chart_data<- chart_data %>%
  filter(SampleNumber != 23)
Plot_s2_chart(chart_elements = chart_data)
```

Fig. 5. *Plotting the revised $s^2$ control chart – calling the function Plot_s2_chart*

For the graphical representation of the $\bar{x}$ chart, the function *Plot_xbar_chart()* was created in an analogous way as the function *Plot_s2_chart().*

The formulas used to calculate the control limits of the $\bar{x}$ chart are (5) and (6).
Before the capability analysis, the test of normality for data is required.
Figure 6 shows the commands used for performing this task.

```
df_tn <- df_xbar_s2[df_xbar_s2$SampleNumber != 23, ]
ad.test(df_tn$x)
```

Fig. 6. *Testing the normality of data*

First, the object of type *data frame* named *df_tn* is created by removing the record of point 23 from the data frame *df_xbar_s2*. The function *ad.test()* performs the Anderson-Darling test of normality on variable *x* of this *data frame*. The next step to be achieved is the process capability analysis. Figure 7 presents the commands used for this purpose. First there are calculated the process mean and, after that, the process standard deviation, which is the square root of the process variance. In the next step, the values for the capability indexes and the proportions of nonconforming items are calculated.

```
miu <- mean(chart_data$xbar)
sigma <- sqrt(mean(chart_data$s2))
USL <- 250.052
LSL <- 250
Cp <- (USL-LSL)/(6*sigma)
CPU <- (USL - miu)/(3*sigma)
CPL <- (miu - LSL)/(3*sigma)
Cpk <- min(CPU, CPL)
pL <- pnorm(q = LSL, mean = miu, sd = sigma, lower.tail = TRUE)
pU <- pnorm(q = USL, mean = miu, sd = sigma, lower.tail = FALSE)
ptot <- pL + pU
cat("The capability indices are:","\n")
cat("Cp=", Cp,"\n")
cat("CPL=", CPL,"\n")
cat("CPU=", CPU,"\n")
cat("Cpk=", Cpk,"\n")
cat("The proportions of nonconforming items are:","\n")
cat("pL=", pL,"\n")
cat("pU=", pU,"\n")
cat("ptot=", ptot,"\n")
```

Fig. 7. *The process capability analysis*

## 5. Results and Discussion

The program was run step by step, and the results are presented below. First, the commands in Figures 1 and 2 are run. The necessary libraries were loaded, and the *type I error* value was entered by the user of the program from the keyboard. Then the commands in Figures 3 and 4 are executed. In the next stage, the run of the conceived function *Plot_s2_chart()* will plot the $s^2$ chart (Figure 8). It can be observed that the point 23 is above the upper control limit. This means that the process was influenced by a special cause of variation with a negative effect. After finding and eliminating the special cause of variation, the point 23 will be removed from the chart, and then, the revised chart will be plotted (Figure 9). After the process variability have been brought under control, the $\bar{x}$ chart may be plotted For the presented example, the $\bar{x}$ chart does not point out any special cause of variation, so the process is considered under control. After the run of Anderson-Darling normality test, the *P* value resulted equal to 0.2361, so the data distribution is normal. Finally, the process capability analysis was performed. The process capability index resulted equal to 1.3194, a value that is greater than the minimum theoretical accepted value 1, so the process is capable.

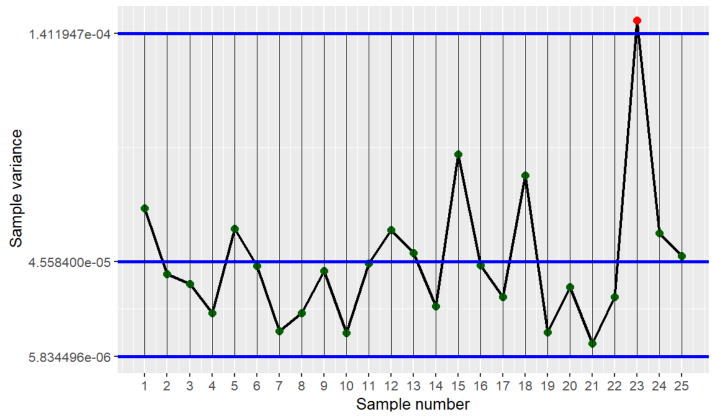


Fig. 8. *The $s^2$ chart*

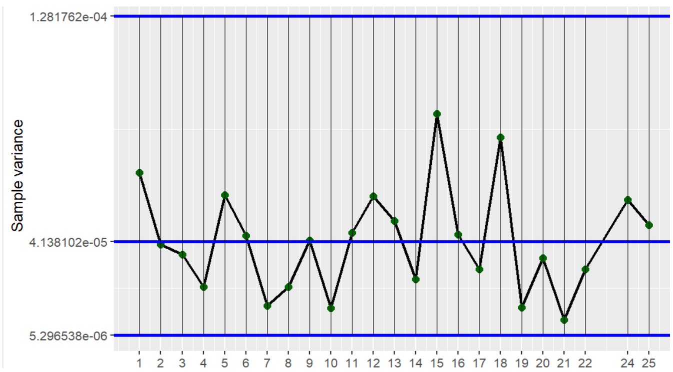


Fig. 9. *The revised $s^2$ chart*

## 6. Conclusions

One of the advantages of using the $s^2$ chart consists in the fact that its center line is an unbiased estimator of the process variance and is consequently a preferred alternative to the *s* chart. The use of modern current computation techniques and the development of programs to help the implementation of $s^2$ chart represents an important objective.

Therefore, the present work proposes an original computer program that can be applied to control the process variability by means of the $s^2$ chart.

The program was developed using *R,* a powerful language for statistics, which is used at present by more and more scientists.

The paper presents in detail all the stages of the program development.

The program may be used for any value of the type I error value, and besides plotting of the control chart it also performs the data test of normality and the process capability analysis.

The paper also presents the application of the program on a concrete example.

**References**


1. Ajadi, J.O., Riaz, M.: *New memory-type control charts for monitoring process mean and dispersion*. In: Scientia Iranica E 24(6) (2017), p. 3423-3438, DOI: 10.24200/sci.2017.4483.
2. Castagliola, P., Celano, G., *et al.*: *A New CUSUM-$S^2$ Control Chart for Monitoring the Process Variance*. In: Journal of Quality in Maintenance Engineering 15(4) (2009), p. 344-357, DOI 10.1108/13552510910997724.
3. Ho, L.L., da Costa Quinino, R.: *Combining Attribute and Variable Data to Monitor Process Variability: MIX $S^2$ Control Chart*. In: Journal of Advanced Manufacturing Technology 87 (2016), p. 3389–3396, DOI 10.1007/s00170-016-8702-5.
4. Kao, S.-C., Ho, C.: *Process Monitoring of the Sample Variances Through an Optimal Normalizing Transformation*. In: Journal of Advanced Manufacturing Technology (2006) Vol. 30, p. 459–469, DOI 10.1007/s00170-005-0092-z.
5. Landim, T.R., Jardim, F.S., *et al.*: *Modified Control Chart for Monitoring the Variance*. In: Brazilian Journal of Operations & Production Management 18(03), (2021), e20211129, DOI: https://doi.org/10.14488/BJOPM.2021.015.
6. Lazariv, T., Schmid, W., et al.: *On Control Charts for Monitoring the Variance of a Timeseries*. In: Journal of Statistical Planning and Inference 143 (2013), p. 1512–1526. DOI: http://dx.doi.org/10.1016/j.jspi.2013.04.001.
7. Qiu, P.: *Introduction to Statistical Process Control*. New York. Chapman and Hall/CRC, 2014.
8. Saeed, N., Kamal, S., et al.: *Percentile bootstrap control chart for monitoring process variability under non-normal processes*. In: Scientia Iranica (2024), Vol. 31(15), p. 1282-1292, DOI:https://doi.org/ 10.24200/sci.2021.58118.5573.
9. Saghir, A.: *Phase-I Design Structure of Bayesian Variance Chart*. In: Cogent Mathematics, 3:1172403, (2016), DOI: http://dx.doi.org/10.1080/ 23311835.2016.1172403.
10. Wickham, H.: *ggplot2: Elegant Graphics for Data Analysis*, *Second edition*. New York. Springer-Verlag, 2016.
11. Zaagan, A. A., Khan, I., et al.: *Parameter free AEWMA control chart for dispersion in semiconductor manufacturing*. In: Scientific Reports (2024), Vol. 14(1), 10512.ISO 690, DOI: https://doi.org/10.1038/s41598-024-61408-5.
12. Zhang, L., Bebbington, M.S., *et al*. *On Statistical Design of the $S^2$ Control Chart*. In: Communications in Statistics - Theory and Methods (2005), Vol. 34:1, p. 229-244, DOI: 10.1081/STA-200045817.
13. Zhou, J.J., Ng, K.H., *et al.*: *Asymmetric Control Limits for Weighted-Variance Mean Control Chart with Different Scale Estimators under Weibull Distributed Process*. In: Mathematics (2022), Vol. 10, 4380 DOI: https://doi.org/10.3390/math10224380.
14. https://www.R-project.org/. Accessed: 22.05.2024.
15. https://CRAN.R-project.org/package=openxlsx. Accessed: 20.08.2024.
16. https://CRAN.R-project.org/package=dplyr. Accessed: 20.08.2024.
17. https://CRAN.R-project.org/package=nortest. Accessed: 20.08.2024.