

CEP ONLINE: A WEB-ORIENTED EXPERT SYSTEM FOR STATISTICAL PROCESS CONTROL

Francisco Louzada¹, Paulo Ferreira^{2*}, Anderson Ara² and Caroline Godoy³

Received July 26, 2017 / Accepted February 13, 2019

ABSTRACT. In this paper, a new software for Statistical Process Control (SPC) is proposed. The system, the so-called CEP Online, was developed based on statistical computing resources of well-known free softwares, such as HTML, PHP, R and MySQL under an online server with operating system Linux Ubuntu. The main uni and multivariate SPC tools are available for monitoring and evaluation of manufacturing and non-manufacturing production processes over time. Some advantages of the new software are: (i) low operational cost, since it is cloud-based, only needing a computer connected to the Internet; (ii) easy to use with great interaction with the user; (iii) it does not require investment in any specific hardware or software; (iv) real time reports generation on process condition monitoring and process capability. Thus, the CEP Online offers for SPC practitioners fast, efficient and accurate SPC procedures. Therefore, CEP Online becomes an important resource for those who have no access to non-free softwares, such as SAS, SPSS, Minitab and STATISTICA. To the best of our knowledge, the CEP Online is unique with respect to its characteristics.

Keywords: Control Charts, Capability Indices, Instantaneous Reports.

1 INTRODUCTION

The increasing enterprise competitiveness and increasing consumer requirements, along with the globalization and world computerization, have caused significant changes in production of manufacturing and non-manufacturing environments worldwide. Many enterprises, particularly industries, have been faced with the need for improving their products. As a consequence, the quality control of their products has become extremely important to have customer satisfaction and to generate profit.

*Corresponding author – <https://orcid.org/0000-0001-6312-6098>

¹Departamento de Matemática Aplicada e Estatística, Instituto de Ciências Matemáticas e de Computação, Universidade de São Paulo, São Carlos, SP, Brasil – E-mail: louzada@icmc.usp.br

²Departamento de Estatística, Instituto de Matemática e Estatística, Universidade Federal da Bahia, Salvador, BA, Brasil – E-mail: paulohenri@ufba.br; anderson.ara@ufba.br

³ Centro de Ciências Matemáticas Aplicadas à Indústria, Instituto de Ciências Matemáticas e de Computação, Universidade de São Paulo, São Carlos, SP, Brasil – E-mail: carolinegodoy.est@gmail.com

The standard methodology for quality improvement is the Statistical Process Control (SPC), which, according to [17], consists of a powerful set of tools used in achieving process stability and improving capability through the reduction of variability. SPC can be applied in any process involving a repetitive sequence of steps, i.e. it can be applied in both manufacturing and non-manufacturing processes.

Nowadays, there are several statistical softwares that can be used to generate SPC analysis (e.g., R, SAS, SPSS, Minitab, STATISTICA). However, with the exception of R, most of these softwares can only be used by purchasing their license, which has relatively high cost for certain sizes and types of enterprises. In this paper, we introduce a new software for SPC, the so-called CEP Online (CEP stands for “Controle Estatístico de Processo”, which is the Portuguese translation of Statistical Process Control). As similarly described in [13], its development is based on statistical computing resources of well-known free softwares, such as HTML, PHP, R and SQL under an online server with operating system Linux Ubuntu. It is aimed at enabling access to the main uni and multivariate SPC tools by mainly small and medium-sized enterprises.

To the best of our knowledge, there are no online softwares available in the literature that work and perform similarly to ours¹. This leads us to an innovative structure in applied statistics of SPC that will possibly be common in the near future.

It is worth noting that the R software itself (which is robust, rigorous, efficient and free, as pointed out by [5]) has statistical packages that allow one to apply many of the uni and multivariate SPC methods described in this paper (see Section 2). Among these packages are *SixSigma* [5], *qcc* [25], *qualityTools* [20], *lattice* [24], *spc* [11], *spcadjust* [8], *IQCC* [2], *MSQC* [22] and *MPCI* [23], all available on CRAN (<https://cran.r-project.org/>). However, we reinforce that none of them has all the statistical tools (descriptive statistics and graphics, goodness-of-fit tests, control charts and capability indices, for both univariate and multivariate designs) usually required to perform a complete SPC analysis. Thus, it is common for a SPC analyst to use different functions from different R packages to perform the desired tasks to produce desired output, which often requires significant time and effort of the analyst. Moreover, one needs programming skills to be able to use appropriately the R software. Compared to R, CEP Online aims to offer a reasonable set of SPC tools, which can be used in a faster, easier and friendlier way.

In order to exemplify some of the methodologies used, we apply (univariate) SPC tools to evaluate the performance of a production process of chocolate bars by certain toy food company (the quality characteristic to be monitored is the chocolate bar’s weight), and present the different reports generated by the proposed CEP Online system, available on <http://www.mwstat.com/novocep>.

In general, the web-based approach proposed here has the following characteristics: focused on SPC implementation, built on cloud and evaluates the process in- or out-of-control condition

¹ However, it deserves mentioning here the work by [7], which describes a real time on-line control system able to find the optimal dosing of input aggregates in a Hot Mix Asphalt process.

by constructing control charts, as well as in- or out-of-specification condition by calculating capability indices. It is built from different and free languages and softwares, having a structure that instantly connects with the R software to generate all necessary calculations.

It is important to note that updated versions of the CEP Online system can be produced (e.g., by other researchers/practitioners) to incorporate new statistical tools into the available modules (univariate and multivariate SPC modules), as well as into new modules (e.g., process design and improvement with designed experiments, and acceptance-sampling techniques). Some of our current developments (which are the subjects of our actual researches) include: (i) new variables and attributes control charts that are better alternatives to the traditional univariate Shewhart ones (see, e.g., [21]); (ii) robust non-parametric versions of the multivariate T^2 control chart (see, e.g., [6, 18]); (iii) and the application of copulas to multivariate control charts and capability indices (see, e.g., [31, 4]). Since these SPC tools are not available in any of the existing statistical softwares that perform SPC analysis, their insertion into CEP Online also brings modernity, originality and exclusivity to our proposed system.

The remainder of the paper is organized as follows. In Section 2, we present the SPC procedures available at CEP Online. In Section 3, the system architecture is described and explained. Section 4 shows the implementation and evaluation of the CEP Online system, including the generation of control charts and the calculation of capability indices for univariate processes only (chocolate bars example). Finally, Section 5 concludes the paper.

2 SYSTEM ELEMENTS AND METHODOLOGY

In this section, we present the SPC tools used to analyze the process performance over time. More specifically, a set of uni and multivariate SPC techniques were applied in order to generate control charts and compute capability indices.

2.1 Univariate statistical analysis

In this subsection, we briefly describe some of the most widely used SPC tools for analyzing single-variable processes, i.e. processes with a single measurable quality characteristic, such as length, diameter, volume (unrealistic scenario in many real-world SPC applications).

2.1.1 Shewhart control charts

The Shewhart control charts are undoubtedly the most widely known and used statistical tools for process monitoring or supervision. They were developed in the 1920s by Walter A. Shewhart, a physicist, engineer and statistician of the Bell Telephone Laboratories.

As stated in [17], Shewhart proposed a general theory/model for generating control charts, which can be simply explained as follows. Let ω be a sample statistic that measures some quality characteristic of interest, and suppose that μ_ω and σ_ω are, respectively, the mean and standard

deviation of ω . Then, the control chart's center line (*CL*), lower (*LCL*) and upper (*UCL*) control limits, are given by

$$\begin{aligned} LCL &= \mu_\omega - L\sigma_\omega, \\ CL &= \mu_\omega, \\ UCL &= \mu_\omega + L\sigma_\omega, \end{aligned} \tag{1}$$

where *L* is the “distance” between the *CL* and the control limits, given in terms of standard deviation units. In the case of the well-known Six Sigma policy (adopted by us in our study), we have that *L* = 3 (this means that, if we assume a normal probability distribution as a model for a quality characteristic, then it turns out that the probability of producing an item within the control limits is 0.9973, which corresponds to 2,700 parts per million defective).

Still according to [17], the Shewhart control charts may be classified into two general types: **variables control charts**, if the quality characteristic can be measured and expressed as a number on some continuous scale of measurement; or **attributes control charts**, if the quality characteristic is not measured on a continuous scale or even a quantitative scale, i.e. we may judge each unit of product as either conforming or nonconforming on the basis of whether or not it possesses certain attributes, or we may count the number of nonconformities (defects) appearing on a unit of product.

Among the variables control charts, we highlight the \bar{x} , *S* and *R* charts, which are the most widely used charts for controlling central tendency based on the sample mean (\bar{x} chart), and process variability via the sample standard deviation (*S* chart) or sample range (*R* chart).

By considering the Shewhart's general model given by (1), we obtain the following control charts for variables (for details on the development of these graphics, see [17]):

- \bar{x} and *S* charts:

$$\mu_0 \pm 3 \frac{\sigma_0}{\sqrt{n}} \quad \text{and} \quad c_4 \sigma_0 \pm 3 \sigma_0 \sqrt{1 - c_4^2},$$

respectively, where *n* is the sample size (which is the same for the *m* collected samples/data groups), $c_4 = \frac{\Gamma(n/2)}{\Gamma((n-1)/2)} \sqrt{\frac{2}{n-1}}$, with $\Gamma(\cdot)$ the gamma function, and μ_0 and σ_0 are the (known/specified) process mean and standard deviation, respectively. Otherwise (i.e., if there is no parameter specification), the \bar{x} and *S* charts' control limits can be rewritten as

$$\bar{\bar{x}} \pm 3 \frac{\bar{S}}{c_4 \sqrt{n}} \quad \text{and} \quad \bar{S} \pm 3 \frac{\bar{S}}{c_4} \sqrt{1 - c_4^2},$$

respectively, where $\bar{\bar{x}} = \frac{\sum_{i=1}^m \bar{x}_i}{m} = \frac{\sum_{i=1}^m \sum_{j=1}^n x_{ij}}{nm}$ (overall mean), with x_{ij} being the quality characteristic value of the *j*th item of the *i*th sample, and $\bar{S} = \sqrt{\frac{\sum_{i=1}^m S_i^2}{m}}$, with $S_i^2 = \frac{\sum_{j=1}^n (x_{ij} - \bar{x}_i)^2}{n}$.

As suggested by their names, the statistics plotted on the \bar{x} and *S* charts are the sample mean, \bar{x}_i , and the sample standard deviation, $S_i = \sqrt{S_i^2}$, respectively.

These charts are also relatively easy to apply in the cases where the sample sizes are unequal/variable. For details, see [17].

• **\bar{x} and R charts:**

$$\mu \pm 3 \frac{\sigma_0}{\sqrt{n}} \quad \text{and} \quad d_2 \sigma_0 \pm 3 d_3 \sigma_0,$$

respectively, where $E[R_i] = d_2 \sigma_0$ and $\sqrt{Var[R_i]} = d_3 \sigma_0$ (see, e.g., Table 2.3 of [14], for estimates of d_2 and d_3 for some values of n), with $R_i = \max_j \{x_{ij}\} - \min_j \{x_{ij}\}$ being the range of sample i , for $i = 1, 2, \dots, m$, and \min and \max the minimum and maximum functions, respectively. If the process mean and standard deviation are not known/specified, then the control limits of the \bar{x} and R charts are revised as follows:

$$\bar{\bar{x}} \pm 3 \frac{\bar{R}}{d_2 \sqrt{n}} \quad \text{and} \quad \bar{R} \pm 3 \frac{d_3 \bar{R}}{d_2},$$

where $\bar{R} = \frac{1}{m} \sum_{i=1}^m R_i$.

Again, as suggested by their names, the statistics plotted on the \bar{x} and R charts are the sample mean, \bar{x}_i , and the sample range, R_i , respectively.

The R chart can also be used for samples of size $n = 1$ (individual units). In this case, we simply replace the sample range R_i by the moving range of two successive observations, that is $MR_i = |x_i - x_{i-1}|$, for $i = 2, \dots, m$.

Finally, [17] recommends the use of the sample range, instead of the sample standard deviation, as a measure of the process dispersion when the collected samples are of equal (i.e. $n_1 = \dots = n_m = n$) and small (i.e. $n < 10$) sizes.

Among the attributes control charts, we highlight those for fraction nonconforming (p and np charts) and for nonconformities or defects (c and u charts), which are also based on the same general statistical principles of the Shewhart control charts. Then, from (1), the center line and control limits of the main control charts for attributes would be as follows (again, for details on the development of these graphics, see [17]):

• **p chart:**

$$p_0 \pm 3 \sqrt{\frac{p_0(1-p_0)}{n}} \quad (\text{with parameter specification}) \tag{2}$$

or

$$\bar{p} \pm 3 \sqrt{\frac{\bar{p}(1-\bar{p})}{n}} \quad (\text{without parameter specification}), \tag{3}$$

where p_0 is the known/specified fraction nonconforming in the production process, and $\bar{p} = \frac{1}{m} \sum_{i=1}^m \hat{p}_i = \frac{1}{mm} \sum_{i=1}^m D_i$, with D_i being the nonconforming units in sample i . Note that \hat{p}_i is the statistic plotted on the p chart.

In order to deal with variable sample size, we replace n by n_i in (2) and (3), as well as in the calculus of \hat{p}_i (thus, $\hat{p}_i = \frac{D_i}{n_i}$). Moreover, [17] suggests the use of a standardized p control chart, where $CL = 0$, $LCL = -3$, $UCL = +3$ and the variable plotted is

$$z_i = \frac{\hat{p}_i - p_0}{\sqrt{\frac{p_0(1-p_0)}{n_i}}} \quad (\text{with parameter specification})$$

or

$$z_i = \frac{\hat{p}_i - \bar{p}}{\sqrt{\frac{\bar{p}(1-\bar{p})}{n_i}}} \quad (\text{without parameter specification}).$$

According to the author, the major advantage of the latter approach is that tests for runs and pattern-recognition methods can safely be applied to the standardized chart.

- **np chart:**

$$np_0 \pm 3\sqrt{np_0(1-p_0)} \quad (\text{with parameter specification})$$

or

$$n\bar{p} \pm 3\sqrt{n\bar{p}(1-\bar{p})} \quad (\text{without parameter specification}).$$

The statistic plotted on this chart is $n\hat{p}_i = D_i$, i.e. the nonconforming units in sample i , for $i = 1, 2, \dots, m$. This is the reason why the np chart is often called a **number nonconforming control chart**.

As pointed out by [17], the np chart has the advantage of being easier to interpret, especially by non-statisticians, than the p chart. However, the p chart is recommended over the np chart for variable sample size, since the former is easier to interpret for this scenario (the center line of the p chart will not vary across samples).

- **c chart:**

$$c_0 \pm 3\sqrt{c_0} \quad (\text{with parameter specification})$$

or

$$\bar{c} \pm 3\sqrt{\bar{c}} \quad (\text{without parameter specification}),$$

where c_0 is the expected number of nonconformities/defects in an inspection unit of product (in general, the inspection unit will be a single unit of product), and $\bar{c} = \frac{1}{m} \sum_{i=1}^m \hat{c}_i$, with \hat{c}_i being the number of nonconformities in sample i (this is also the statistic plotted on the c chart).

The c chart, as well as the u chart (described in the sequence), are particularly useful over the p and np charts in the case where an unit may contain several nonconformities and not be classified as nonconforming. For instance, manufacturing personal computers could have one or more very minor flaws (e.g., in the cabinet finish), but since these flaws do not seriously affect the unit's functional operation, it could be classified as conforming [17].

- **u chart:**

$$u_0 \pm 3\sqrt{\frac{u_0}{n}} \quad (\text{with parameter specification}) \quad (4)$$

or

$$\bar{u} \pm 3\sqrt{\frac{\bar{u}}{n}} \quad (\text{without parameter specification}), \quad (5)$$

where u_0 represents the observed average number of nonconformities per unit in a preliminary set of data, and $\bar{u} = \frac{1}{m} \sum_{i=1}^m \hat{u}_i$, with \hat{u}_i being the average number of nonconformities per inspection unit in sample i (\hat{u}_i is also the variable plotted on the u chart).

We can handle the issue of variable sample size by replacing n by n_i in (4) and (5), or, as suggested by [17] (which is also the preferred option), by using a standardized u control chart, where $CL = 0$, $LCL = -3$, $UCL = +3$ and the statistic plotted is

$$z_i = \frac{u_i - u_0}{\sqrt{\frac{u_0}{n_i}}} \quad (\text{with parameter specification})$$

or

$$z_i = \frac{u_i - \bar{u}}{\sqrt{\frac{\bar{u}}{n_i}}} \quad (\text{without parameter specification}).$$

The u chart is recommended over the c chart for variable sample size, since the latter would be very difficult to interpret for this scenario (both the center line and the control limits of the c chart will vary with the sample size).

2.1.2 CuSum and EWMA charts

The Shewhart control charts, described in Subsubsection 2.1.1, are useful to detect large shifts in the monitored process parameters. However, when small shifts are of great interest, we can use two effective alternative procedures: the Cumulative Sum (CuSum) and the Exponentially Weighted Moving Average (EWMA) control charts. Although they are not really new (both date from the 1950s), they are usually considered to be more advanced techniques than the Shewhart control charts. Next, we describe the CuSum and EWMA control charts for monitoring the process mean and when specifications for the process parameters are available/provided.

- **CuSum chart:** consists of plotting the two statistics:

$$C_i^+ = \max\{0, \bar{x}_i - (\mu_0 + K) + C_{i-1}^+\} \quad \text{and} \\ C_i^- = \max\{0, (\mu_0 + K) - \bar{x}_i + C_{i-1}^-\},$$

for $i = 1, 2, \dots, m$, where $K = 0.5 \frac{\sigma_0}{\sqrt{n}}$ and the starting values are $C_0^+ = C_0^- = H/2$ (50% fast initial response or headstart; see [17]), where $H = 5 \frac{\sigma_0}{\sqrt{n}}$ is the decision interval (control

limit). These values of K and H are the recommended ones to detect a shift in the process mean of 1σ .

The CuSum control chart, as described above, is also known as **tabular or algorithmic CuSum**. It may also be constructed for individual observations (for details, see [17]).

In order to improve the ability of the CuSum chart to detect large process shifts, [17] suggests the use of a combined CuSum-Shewhart procedure for online control (the Shewhart limit is at $H = 3.5 \frac{\sigma_0}{\sqrt{n}}$).

- **EWMA chart:** the exponentially weighted moving average is defined as

$$W_i = \lambda \bar{x}_i + (1 - \lambda) W_{i-1},$$

for $i = 1, 2, \dots, m$, where $0 < \lambda \leq 1$ is a constant (smoothing factor) and the starting value is the process target, i.e. $W_0 = \mu_0$. Exact and asymptotic (steady-state values) control limits are given by

$$\mu_0 \pm \gamma \sqrt{\frac{\lambda [1 - (1 - \lambda)^{2i}] \sigma_0^2}{(2 - \lambda)n}} \quad \text{and} \quad \mu_0 \pm \gamma \sqrt{\frac{\lambda \sigma_0^2}{(2 - \lambda)n}},$$

respectively, where γ is a positive constant.

[17] strongly recommends the use of the exact control limits for small values of i , e.g. $i \leq 10$. The author also suggests the use of $0.05 \leq \lambda \leq 0.25$, which works well in practice, with $\lambda = 0.05, 0.10$ and 0.20 being popular choices, as well as $\gamma = 3$ (the usual 3σ limits).

The EWMA chart is also very insensitive (robust) to the normality assumption, which makes it an ideal control chart to use with individual observations [17].

2.1.3 Process capability analysis

It is widely known that being in-control is not enough for a production process, since even an in-control process may produce bad (useless) items. A process must also be capable to achieve customer and/or product requirements or specifications.

One simple and quantitative way to express process capability is through the **capability indices**. Among them, we highlight the capability indices for on-center (C_p and P) and off-center (C_{pk} and C_{pm}) processes. Such indices are defined as follows.

- C_p :

$$C_p = \frac{USL - LSL}{6\sigma_0} \quad (\text{with parameter specification})$$

or

$$\hat{C}_p = \frac{USL - LSL}{6S} \quad (\text{without parameter specification}),$$

where USL and LSL are, respectively, the upper and lower specification limits for the measurable quality characteristic, and S is the sample standard deviation.

According to [17], this kind of index (known as **process capability ratios**, PCRs) is extensively used in industry. It measures the spread of the specifications relative to the 6σ spread in the process, or, as we usually say, the **potential capability** in the process.

The C_p index, as well as the P index (described in the sequence), implicitly assume that the process is centered between the upper and lower specification limits, i.e. that the process mean (μ_0 or \bar{x}) coincides with the center (target value, $T = \frac{1}{2}(USL + LSL)$) of the specification interval ($[LSL, USL]$) (on-center process).

Besides computing and interpreting the point estimate of C_p , we can also report a confidence interval for it. If the quality characteristic is normally distributed, then a $100(\alpha)\%$ confidence interval on C_p is obtained from

$$\left[\hat{C}_p \sqrt{\frac{\chi_{n-1; (1-\alpha)/2}^2}{n-1}}; \hat{C}_p \sqrt{\frac{\chi_{n-1; (1+\alpha)/2}^2}{n-1}} \right], \quad (6)$$

where $\chi_{n-1; (1-\alpha)/2}^2$ and $\chi_{n-1; (1+\alpha)/2}^2$ are, respectively, the $(1-\alpha)/2$ and $(1+\alpha)/2$ percentile points of the chi-square distribution with $n-1$ degrees of freedom.

• P :

$$P = \left(\frac{1}{C_p} \right) \times 100 \quad (\text{with parameter specification})$$

or

$$\hat{P} = \left(\frac{1}{\hat{C}_p} \right) \times 100 \quad (\text{without parameter specification}).$$

The P index has a useful practical interpretation - it is the percentage of the specification band used up by the process [17].

A $100(\alpha)\%$ confidence interval on P is obtained from (6) as follows:

$$\left[\left(\hat{C}_p \sqrt{\frac{\chi_{n-1; (1+\alpha)/2}^2}{n-1}} \right)^{-1} \times 100; \left(\hat{C}_p \sqrt{\frac{\chi_{n-1; (1-\alpha)/2}^2}{n-1}} \right)^{-1} \times 100 \right].$$

• C_{pk} :

$$C_{pk} = \min \left\{ \frac{USL - \mu_0}{3\sigma_0}; \frac{\mu_0 - LSL}{3\sigma_0} \right\} \quad (\text{with parameter specification})$$

or

$$\hat{C}_{pk} = \min \left\{ \frac{USL - \bar{x}}{3S}; \frac{\bar{x} - LSL}{3S} \right\} \quad (\text{without parameter specification}).$$

An advantage of C_{pk} , as well as of C_{pm} (described in the sequence), over the C_p and P indices (described before), is that the former ones take process centering into account for the process capability estimation [17]. Thus, C_{pk} can also be used in the situation where the process is not operating at the midpoint of the specification interval (off-center process). In this case, $C_{pk} < C_p$. The C_{pk} index is usually said to measure the **actual capability** in the process.

Assuming that the quality characteristic follows a normal distribution, we obtain an approximate $100(\alpha)\%$ confidence interval on C_{pk} as follows:

$$\left[\hat{C}_{pk} \left(1 - z_{(1+\alpha)/2} \sqrt{\frac{1}{9n\hat{C}_{pk}^2} + \frac{1}{2(n-1)}} \right); \hat{C}_{pk} \left(1 + z_{(1+\alpha)/2} \sqrt{\frac{1}{9n\hat{C}_{pk}^2} + \frac{1}{2(n-1)}} \right) \right], \quad (7)$$

where $z_{(1+\alpha)/2}$ represents the $(1 + \alpha) / 2$ percentile point of the standard normal distribution.

- C_{pm} :

$$C_{pm} = \frac{C_p}{\sqrt{1 + V^2}} \quad (\text{with parameter specification})$$

or

$$\hat{C}_{pm} = \frac{\hat{C}_p}{\sqrt{1 + \hat{V}^2}} \quad (\text{without parameter specification}),$$

where $V = \frac{\mu_0 - T}{\sigma_0}$ and $\hat{V} = \frac{\bar{x} - T}{s}$.

According to [17], the C_{pk} index does not tell us about the location of the mean in the specification interval, whereas the C_{pm} index is a better indicator of centering.

Moreover, [3] shows that

$$C_{pm} \geq k \implies |\mu_0 - T| < \frac{1}{6k} (USL - LSL).$$

Thus, from a given value of C_{pm} , we can place a constraint on the difference between μ_0 and T . For instance, $C_{pm} \geq 2$ implies that $|\mu_0 - T| < \frac{1}{12} (USL - LSL)$.

Note that C_p , C_{pk} and C_{pm} values greater than 1, as well as P values lower than 100, are desired.

For an extensive summary of confidence intervals for various capability indices of kind PCRs (including C_p , C_{pk} and C_{pm}), see [12].

Finally, it is also important to note that a process capability analysis can also be performed via graphical techniques, such as histograms and control charts.

2.2 Multivariate statistical analysis

In practice, most process monitoring and control scenarios involve various related variables. Applying univariate control charts to each individual variable is not a good solution, since it is inefficient and can lead to erroneous conclusions [17]. Thus, multivariate methods that consider the variables jointly are needed.

In multivariate statistical quality control, we generally use the multivariate normal distribution to describe the joint behavior of continuous quality characteristics (multivariate normal process), i.e. $X = (X_1, \dots, X_p)' \sim N_p(\mu, \Sigma)$, where X_1, \dots, X_p denote the p variables (quality characteristics of interest), $\mu = (\mu_1, \dots, \mu_p)'$ is the vector of the means of the X 's, and Σ is the $p \times p$ covariance matrix (whose main diagonal elements are the variances of the X 's and the off-diagonal elements are the covariances).

Some of the most widely used control charts for monitoring the variability (two approaches) and mean (Hotelling T^2 chart) in the multivariate case, as well as capability indices (MC_p , MC_{pk} , MC_{pm} , MP_{PC} , Mp_1 and Mp_2), are briefly described in the following subsections.

2.2.1 Variability control charts

In this subsection, we present control charts for monitoring the multivariate normal process variability, that are based on two different approaches discussed in [17].

- **First approach:** the statistic plotted on the control chart for the i^{th} sample, $i = 1, \dots, m$, is

$$W_i = -pn + pn \ln(n) - n \ln \left(\frac{|A_i|}{|\Sigma_0|} \right) + tr(\Sigma_0^{-1} A_i), \quad (8)$$

where Σ_0 is the (known/specified) $p \times p$ covariance matrix, $A_i = (n-1)S_i$, with S_i being the sample covariance matrix for sample i , and tr is the trace operator. The control chart only has an upper control limit given by $UCL = \chi_{p(p+1)/2; 0.9973}^2$.

This first approach is a direct extension of the univariate S^2 control chart. For details on the S^2 control chart, see [17].

In practice, Σ (the true covariance matrix) usually will be estimated from preliminary samples. In this case, we simply replace Σ_0 by S (the sample covariance matrix) in (8).

- **Second approach:** consists of plotting the sample generalized variance, $|S_i|$, $i = 1, \dots, m$, which is a widely used measure of multivariate dispersion. The control limits of the chart are given by

$$|\Sigma_0| \left(b_1 \pm 3b_2^{0.5} \right) \quad (\text{with parameter specification})$$

or

$$\frac{|S|}{b_1} \left(b_1 \pm 3b_2^{0.5} \right) \quad (\text{without parameter specification}),$$

where $b_1 = \frac{1}{(n-1)^p} \prod_{k=1}^p (n-k)$ and $b_2 = \frac{1}{(n-1)^{2p}} \prod_{k=1}^p (n-k) [\prod_{l=1}^p (n-l+2) - \prod_{l=1}^p (n-l)]$.

[17] suggests the use of univariate control charts for variability (e.g., the S and R charts, described in Section 2.1.1) in conjunction with the control chart for $|S|$.

2.2.2 Hotelling T^2 chart

The Hotelling T^2 control chart is the multivariate extension of the univariate Shewhart \bar{x} chart (described in Section 2.1.1). It is also the most used multivariate procedure for monitoring the mean vector of the process. In this subsection, we present two versions of the Hotelling T^2 chart: one for grouped data ($n > 1$), and another for individual observations ($n = 1$).

- **Grouped data:** the test statistic plotted on the control chart for sample $i, i = 1, 2, \dots, m$, is

$$T_i^2 = n(\bar{x}_i - \mu_0)' \Sigma_0^{-1} (\bar{x}_i - \mu_0) \quad (\text{with parameter specification})$$

or

$$T_i^2 = n(\bar{x}_i - \bar{\bar{x}})' S^{-1} (\bar{x}_i - \bar{\bar{x}}) \quad (\text{without parameter specification}),$$

where μ_0 is the (known/specified) vector of in-control means for each quality characteristic, $\bar{x}_i = (\frac{1}{n} \sum_{j=1}^n x_{1ij}, \dots, \frac{1}{n} \sum_{j=1}^n x_{pij})'$, with $x_{ij} = (x_{1ij}, \dots, x_{pij})'$ being the vector of quality characteristics for the j^{th} observation of the i^{th} sample, $\bar{\bar{x}} = (\frac{1}{m} \sum_{i=1}^m \bar{x}_{1i}, \dots, \frac{1}{m} \sum_{i=1}^m \bar{x}_{pi})'$, and S is the $(p \times p)$ average of sample covariance matrices S_i , with elements $\bar{S}_{kg} = \frac{1}{m} \sum_{i=1}^m S_{kgi} = \frac{1}{m(n-1)} \sum_{i=1}^m \sum_{j=1}^n (x_{kij} - \bar{x}_{ki})(x_{gij} - \bar{x}_{gi})$, for $k, g = 1, 2, \dots, p$.

For the first case (with parameter specification), the upper limit on the control chart is $\chi_{p; 0.9973}^2$. While for the second case (without parameter specification), there are two possible upper limits to be used, depending on the phase of control chart usage [1]. As explained in [17], phase I analysis consists of using the Hotelling T^2 chart for establishing control, i.e. for testing if the process was in control when the m groups were drawn and the sample statistics $\bar{\bar{x}}$ and S calculated (retrospective analysis); while in phase II the chart is used to monitor future production through an in-control set of observations obtained in phase I. The phase I upper limit for the Hotelling T^2 control chart is given by

$$UCL = \frac{p(m-1)(n-1)}{mn-m-p+1} F_{p, mn-m-p+1; 0.9973}$$

In phase II, the upper limit is

$$UCL = \frac{p(m+1)(n-1)}{mn-m-p+1} F_{p, mn-m-p+1; 0.9973}$$

where $F_{p, mn-m-p+1; 0.9973}$ is the 99.73th percentile of the F distribution with p numerator degrees of freedom and $mn - m - p + 1$ denominator degrees of freedom.

- **Individual observations:** in this situation (very common in the chemical and process industries), the Hotelling T^2 statistic becomes

$$T_i^2 = (x_i - \mu_0)' \Sigma_0^{-1} (x_i - \mu_0) \quad (\text{with parameter specification}),$$

for $i = 1, 2, \dots, m$, where $x_i = (x_{1i}, \dots, x_{pi})'$ and the upper limit of the control chart is still given by $\chi_{p; 0.9973}^2$; or

$$T_i^2 = (x_i - \bar{x})' S^{-1} (x_i - \bar{x}) \quad (\text{without parameter specification}),$$

where $\bar{x} = (\frac{1}{m} \sum_{i=1}^m x_{1i}, \dots, \sum_{i=1}^m x_{pi})'$ and the upper control limit is given by

$$UCL = \frac{(m-1)^2}{m} \beta_{p/2, (m-p-1)/2; 0.9973} \quad (\text{Phase I})$$

[29], where $\beta_{p/2, (m-p-1)/2; 0.9973}$ is the 99.73th percentile of the beta distribution with parameters $p/2$ and $(m-p-1)/2$; or

$$UCL = \frac{p(m+1)(m-1)}{m^2 - mp} F_{p, m-p; 0.9973} \quad (\text{Phase II}).$$

However, in the case of individual observations, we need to use more sophisticated estimators of the covariance matrix Σ . One of them is the usual estimator obtained by simply pooling all m observations, i.e.

$$S_1 = \frac{1}{m-1} \sum_{i=1}^m (x_i - \bar{x})(x_i - \bar{x})'.$$

A second estimator, originally suggested by [9], considers the difference between successive pairs of observations, i.e.

$$S_2 = \frac{1}{2} \frac{V'V}{(m-1)},$$

where

$$V = \begin{bmatrix} v_1' \\ \vdots \\ v_{m-1}' \end{bmatrix} = \begin{bmatrix} (x_2 - x_1)' \\ \vdots \\ (x_m - x_{m-1})' \end{bmatrix}.$$

For other estimators of Σ , see, e.g., [27].

Interesting alternatives to the T^2 and generalized variance ($|S|$) control charts for monitoring bivariate processes (namely, the ZMAX, VMAX and MCMAX charts), are presented in [15].

2.2.3 Multivariate process capability analysis

Multivariate process capability indices (MPCIs) can be constructed using several different approaches. For a discussion on these approaches, see, e.g., [26]. In this subsection, we focus on MPCIs for multivariate normal processes data using principal component analysis (PCA). For details on this multivariate statistical technique, see, e.g., Chapter 8 of [10]. Among the PCA-based MPCIs, we highlight MC_p , MC_{pk} and MC_{pm} [33], MP_{PC} [30], Mp_1 and Mp_2 [26]. Such MPCIs are briefly described below.

- MC_p , MC_{pk} , MC_{pm} [33]:

$$MC_p = \left(\prod_{i=1}^v C_{p; PC_i} \right)^{1/v}, \tag{9}$$

where $C_{p; PC_i} = \frac{USL_{PC_i} - LSL_{PC_i}}{6\sigma_{PC_i}}$ represents the univariate measure of potential process capability C_p for the i^{th} principal component (PC_i), v denotes the number of principal components (PCs) comprising around 90% of the process variability, and LSL_{PC_i} , USL_{PC_i} and σ_{PC_i} represent the lower specification limit, upper specification limit and standard deviation of PC_i , respectively, where

$$LSL_{PC_i} = u_i' LSL \quad \text{and} \quad USL_{PC_i} = u_i' USL,$$

with u_1, u_2, \dots, u_p being the eigenvectors of Σ , and LSL and USL are, respectively, the upper and lower specification limits of $X = (X_1, X_2, \dots, X_p)'$.

Sometimes, we need to estimate σ_{PC_i} by S_{PC_i} (that is, the sample standard deviation of PC_i), thus obtaining

$$M\hat{C}_p = \left(\prod_{i=1}^v \hat{C}_{p; PC_i} \right)^{1/v},$$

where $\hat{C}_{p; PC_i} = \frac{USL_{PC_i} - LSL_{PC_i}}{6S_{PC_i}}$.

A 100(α)% confidence interval on MC_p is obtained from [34]:

$$\left[\left(\prod_{i=1}^v \hat{C}_{p; PC_i} \sqrt{\frac{\chi_{n-1}^2; (1-\alpha)/2}{n-1}} \right)^{1/v}; \left(\prod_{i=1}^v \hat{C}_{p; PC_i} \sqrt{\frac{\chi_{n-1}^2; (1+\alpha)/2}{n-1}} \right)^{1/v} \right].$$

Similarly, [33] have defined MPCIs MC_{pk} and MC_{pm} by replacing $C_{p; PC_i}$ with $C_{pk; PC_i}$ and $C_{pm; PC_i}$, respectively, in (9).

The MPCIs proposed by [33] have the advantage of being simple and easy to calculate. However, [26] show, through an example, that such indices are incorrect because they

assume that specification limits of different PCs are independent of each other (which is not true).

- MP_{PC} [30]:

$$MP_{PC} = \frac{1 + \sqrt{2}}{6\sqrt{\lambda_1}},$$

where λ_1 denotes the eigenvalue associated with the first PC of Σ_W (that is, the covariance matrix of W), with W being the (transformed) vector whose elements are $W_i = X_i/d_i$, where $d_i = \frac{1}{2}(USL_{X_i} - LSL_{X_i})$, for $i = 1, 2, \dots, p$.

In general, we need to estimate Σ_W , thus obtaining estimates of λ_1 and MP_{PC} .

[26] show, through an interesting example, that MP_{PC} may mislead the measurements of process capability. They also note that this index is applicable only for a rectangular specification region.

- Mp_1 and Mp_2 [26]:

$$Mp_1 = P \left\{ Y = (Y_1, \dots, Y_v)' \in V \mid Y \sim N_v(\mu_Y = T_Y, \Sigma_Y = \text{diag}(\lambda_1, \dots, \lambda_v)) \right\}$$

and

$$Mp_2 = P \left\{ Y = (Y_1, \dots, Y_v)' \in V \mid Y \sim N_v(\mu_Y, \Sigma_Y = \text{diag}(\lambda_1, \dots, \lambda_v)) \right\},$$

where Y_1, \dots, Y_v denote the first v PCs (explaining approximately 90% of the process variation), $\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_p$ are the eigenvalues of Σ , T_Y is the target vector for Y , and

$$V = \{(y_1, \dots, y_v)' \mid LSL \leq Uy \leq USL, \text{ where } y = (y_1, \dots, y_p)'\} \text{ such that} \\ y_r = E[Y_r], r = v + 1, \dots, p\}$$

is the specification region for Y , where $U = (u_1, \dots, u_p)$ is the matrix of eigenvectors of Σ .

As noted by [26], Mp_1 is analogous to MC_p and Mp_2 is analogous to MC_{pk} . Therefore, if $Mp_1 \geq 0.9973$, the process is **potentially capable**, and if $Mp_2 \geq 0.9973$, the process is **actually capable**.

Since the computation of these two MPCIs involve the evaluation of multiple integrals on complicated regions, the authors suggest to assess them based on the empirical probability distribution of PCs. Their empirical approach is so described as follows. First, generate two random samples of a large size N (e.g., $N \geq 20,000$) from the distribution of the first v PCs with the following mean vector:

- Sample I: $\mu_Y = T_Y$,
- Sample II: μ_Y .

Then, estimate Mp_1 and Mp_2 as

$$\hat{M}p_1 = \frac{\text{Number of observations } y = (y_1, \dots, y_v)' \text{ from sample I } \in V}{N},$$

$$\hat{M}p_2 = \frac{\text{Number of observations } y = (y_1, \dots, y_v)' \text{ from sample II } \in V}{N}.$$

According to [26], the empirical approach enables us to consider non-hyper rectangular specification region (not considered in other existing MPCIs).

3 SYSTEM ARCHITECTURE

In this section, we present the main modules of the CEP Online system and the web-based approach. The online system is based on four main modules and the web-based approach uses only free softwares.

3.1 Main modules

The main modules of the proposed system are shown in Figure 1. The quality measurements (variables or attributes) are used as a source of information for the production process. Univariate or multivariate methods are applied to the inserted data, allowing an overall assessment of the process quality. To perform this, we use the R software [19], which brings back the information stored in the online database and dynamically calculates descriptive measures, control charts and capability indices. The results are stored again in a database and displayed in several system reports.

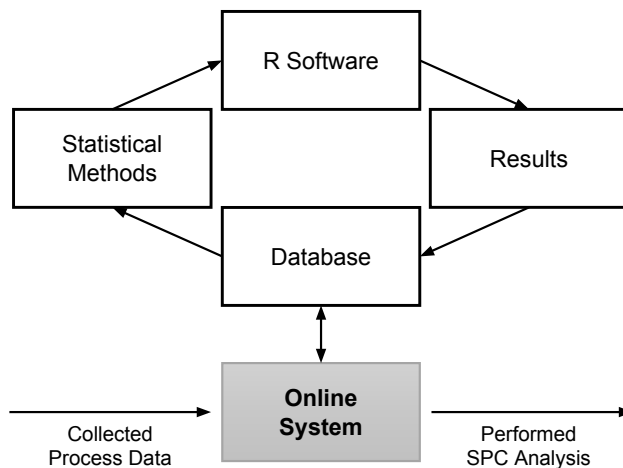


Figure 1 – Main modules of the CEP Online system.

3.2 Web-based approach

The current number of Internet users is up from 3.5 billion, as estimated and announced by Internet Live Stats (www.internetlivestats.com), which means that approximately 40% of the world population has an Internet connection today. Due to the amplitude and existing convenience of the Internet (from a simple connection, an individual can visit pages that are stored on servers anywhere in the world and view many types of content), among other features, we propose an innovative computer system, called CEP Online, as an online statistical tool directed to monitor and control process quality, as well as to assess process capability.

The system has a cloud structure, so it is not necessary to install on local use, requiring only a connection to the Internet. Therefore, a key characteristic of this approach is that CEP Online can be used in large scale and accessed from different locations.

The CEP Online system is essentially built upon free softwares, which can be easily found on the web. Basically, we use an online server with operating system Linux Ubuntu Server 9.10 and the R software (version 3.3.1) as computational environment for carrying out statistical analysis. The HTML and PHP web languages were considered for the pages' design and connections to the database. For data management, we employ MySQL 5.0, which uses the SQL language as interface, and phpMyAdmin version 2.7.0-pl2, which allows the administration of MySQL over the Internet. Using phpMyAdmin, we can create and remove databases; create, delete and change tables; insert, remove, and edit fields; execute SQL codes and manipulate key fields.

Figure 2 exhibits the general structure of the system, starting from the access by users until the structure of the database. The following steps show the connection between the user and the system.

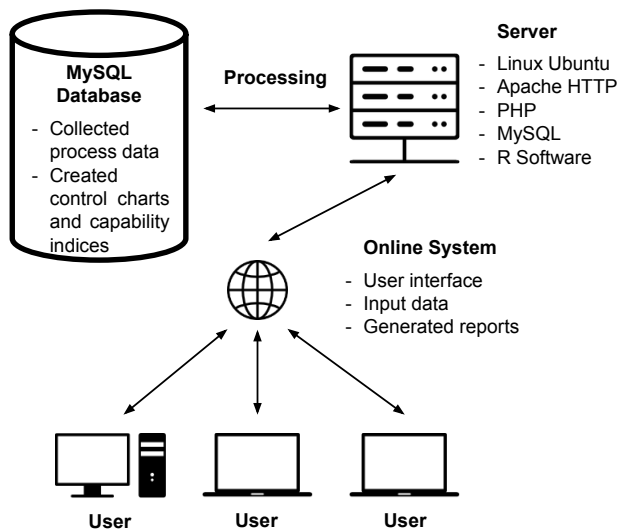


Figure 2 – General structure of the CEP Online system, as well as the languages used.

1. The user accesses the website and enters the login and password in the specified fields;
2. The system receives the information and checks user name and password on database. If all is in order (i.e. the login and password are authenticated), the system allows the access; otherwise, it returns a message saying that the provided information is invalid;
3. With access allowed, the system asks user about which SPC module he/she wants to use: univariate or multivariate, depending on the number of quality characteristics considered. Then, the system redirects the user to a page with the available options.

Once logged into the system, the user can insert new process data and evaluate the results related to process quality control and capability. Available options for each SPC module of CEP Online are shown in Figure 3.

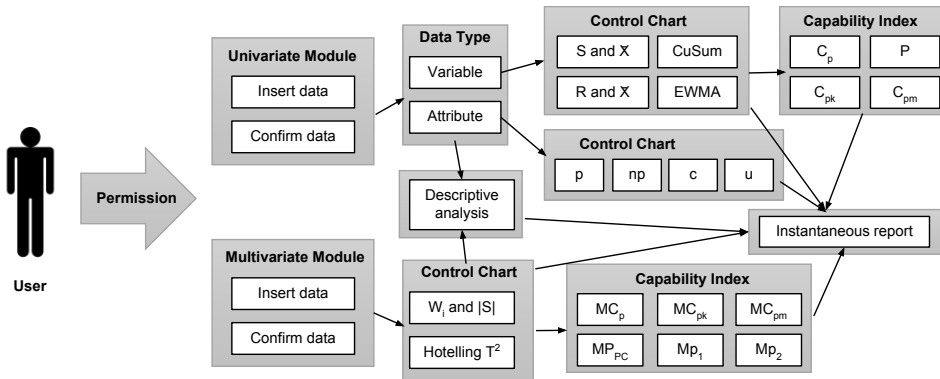


Figure 3 – Available tools in each SPC module of the CEP Online system.

4 IMPLEMENTATION AND EVALUATION

In this section, we perform a SPC analysis on (univariate) data extracted from [14], using the CEP Online system. Due to space limitations, we are unable to provide a SPC analysis on multivariate data too. The (univariate) data consist of the weights (in grams, g) of chocolate bars made by certain (fictitious) company. Some other relevant information on the data used for this study are: $m = 24$ and $n = 15$ (total of 360 sampling units). The required weights are contained in the interval [90g ; 110g].

The following subsections present the (step-by-step) SPC analysis on these data through CEP Online. Note that the graphical analyses are made with R software and displayed in the system as (flexible) interactive JavaScript charts (examples, including source code and live charts, are available at www.highcharts.com).

4.1 Logon to CEP Online

The user could access the system using logins and passwords provided by CEP Online, as shown in Figure 4 at the right area.

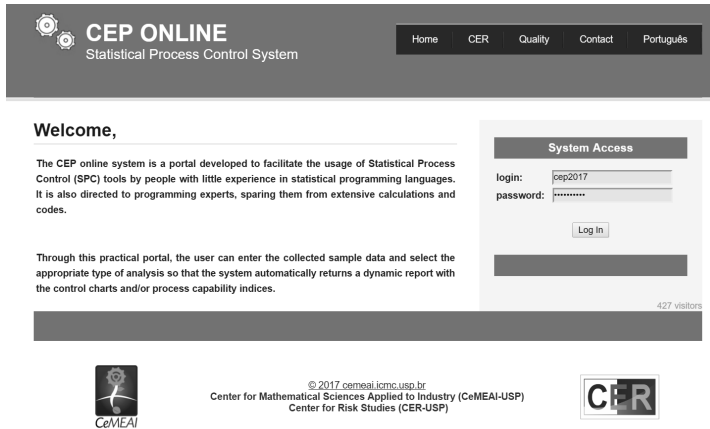


Figure 4 – Login area (www.mwstat.com/novocep).

4.2 SPC module selection

In the next step (page), the user must select the appropriate SPC module (univariate or multivariate) to perform the process quality analysis.

For the present example, which involves only one variable (chocolate bar's weight), choose the univariate SPC module, as shown in Figure 5, and then click on “>>” button.

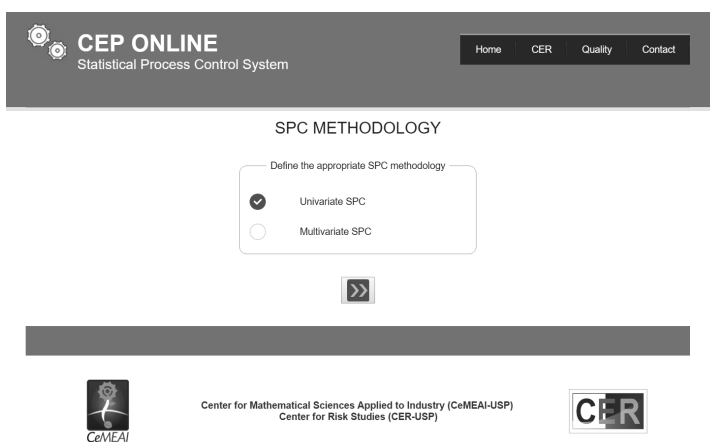


Figure 5 – SPC module selection.

4.3 Data insertion

Data can be inserted into CEP Online in a direct way, i.e. by simply typing (“Manual data entry”), or by reading from external text (.txt) or Excel (.xlsx, .xls or .csv) files. In the blank field, the user must enter the variable name (accents and spacings are not allowed) in the first row and use “enter” to separate data. The data are then stored in the online database, and the system shows some information regarding them in the history of uploaded data sets, such as the file name and extension, total number of observations, upload date and time, as well as buttons that allow the user to view (“Visualize”) and delete (“Delete”) entire data sets. Moreover, in this system page, the user must inform the number of samples to be used in the analysis, as well as the sample size value (equal sample sizes) or vector (unequal sample sizes). The user must also select the type of inserted data (“Attribute” or “Variable”). Finally, click on “Analyze” button to go to the next page.

Figure 6 shows how to fill this system page for the current example.

<= Multivariate Module

UNIVARIATE MODULE

Input data

Choose the format (extension) of the file to upload, or manually enter the data:

Excel file
(XLSX ou XLS)

Excel file
(CSV)

Notepad
(TXT)

Manual data entry
(TXT)

	Name	Extension	Number of observations	Analyze	Upload date	Upload time	Visualize	Delete
1	Example	txt	360	<input checked="" type="checkbox"/>	22/12/2016	20:02:44	Visualize	Delete

Analysis

First, select a file for analysis in the table above.

Enter the number of samples that will be used for the analysis:

Is the sample size constant? Yes No

Which value?

Choose the data type

Attribute

Variable

Figure 6 – Data insertion.

4.4 Data confirmation

The user must check the information after inserting data. The system presents a table with some information regarding the data inserted in the previous page, such as the data set (file) name; number of variables (always one if the user has selected the univariate SPC module), observations and samples; sample size value (or “Different” in the case of unequal sample sizes) and data type (“Attributes” or “Variables”). If something goes wrong, e.g. the total number of observations in the data set does not match the number of samples and sample size(s), the system will send an alert message to the user, who should return to the previous page by clicking on “Return” button and correct the misspecified information. Observe that, if desired, users can perform a descriptive analysis by clicking on “>>” button. Otherwise, they can go directly to the SPC analysis by clicking on “Next” button.

Figure 7 shows this system page for the chocolate bars example.

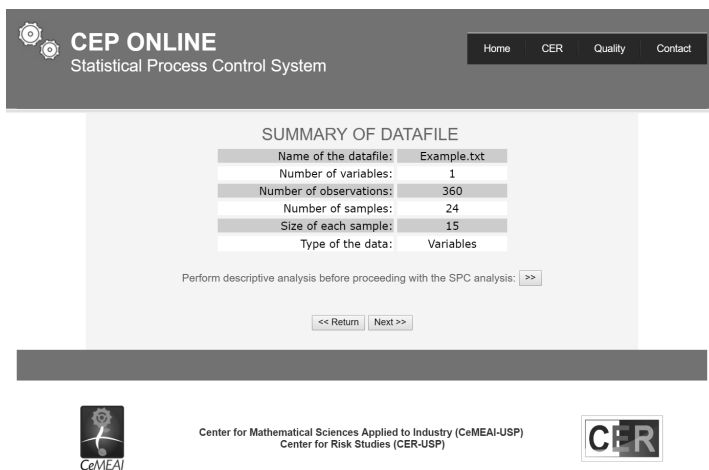


Figure 7 – Data confirmation.

4.5 Descriptive analysis

The optional descriptive analysis displays several summary statistics (total number of observations, minimum and maximum values, first and third quartiles, median, mean, variance, standard deviation and interquartile range), graphics (boxplot, histogram and normal probability plot) and normality tests (Shapiro-Wilk, Anderson-Darling, Lilliefors, Shapiro-Francia and Cramér-von Mises). These summary statistics are also displayed (for each quality characteristic) in the optional descriptive analysis of the multivariate SPC module, in addition to the correlation matrix, graphical tools (scatter plot matrix, chi-square Q-Q plot) and multivariate normality tests (the generalization of the Shapiro-Wilk test by [32], the E-statistic (energy) test by [28], and the [16]’s test).

The descriptive analysis for the current example is presented in Figures 8 and 9. From these figures, we can observe, among others, that the chocolate bars' weights are normally distributed.

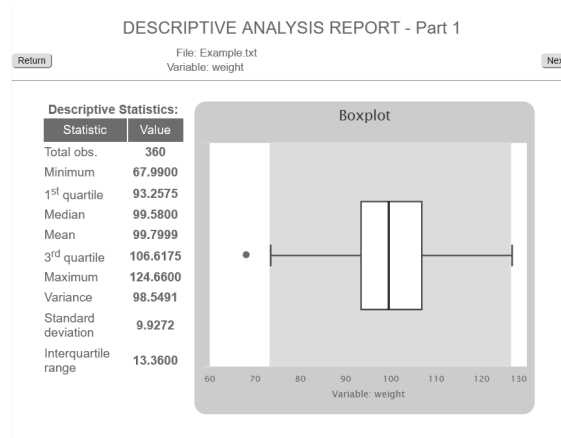


Figure 8 – Descriptive analysis - part 1.

4.6 Control charts selection

In this system page, the user must select the appropriate control charts for the data. If the data are of variable type, the system will show the options: “ S and \bar{x} ” or “ R and \bar{x} ”, “CuSum” (CuSum chart) and “EWMA” (EWMA chart). It is important to note that the system does not allow user to select both options, “ S and \bar{x} ” and “ R and \bar{x} ”, since the S and \bar{x} charts are recommended for $n \geq 10$ or unequal sample sizes; otherwise, the R and \bar{x} charts are the recommended ones (as commented before in Section 2.1.1). If the data are of attribute type, the system will present the options: p , np , c and u . Furthermore, the user must define the process condition, i.e. if there is parameter specification (“with parameter specification”) or not (“without parameter specification”).

Figure 10 shows this system page filled for the current example, with no parameter specification for the production process. Note that the S and \bar{x} charts are more adequate than the R and \bar{x} ones, since $n = 15 > 10$.

4.7 SPC analysis

Here, the CEP Online system generates the (interactive) control chart(s) chosen in the previous page. If the data are of variable type, there will be some blank fields at the left area, that are filled with the necessary information (lower specification limit - “LSL”, upper specification limit - “USL”, and confidence level - “ α ”) to perform a capability study of the process (only in the case where the process is under control). At this area, there is also the confirmation of the type of data and control chart(s) selected in the previous steps.

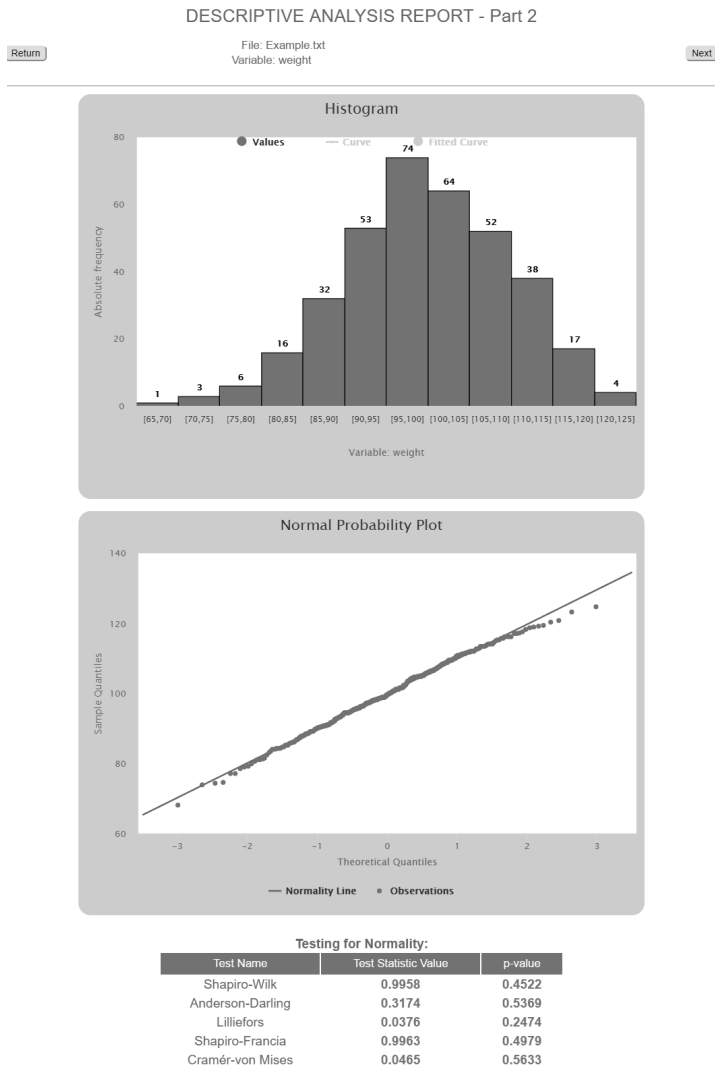


Figure 9 – Descriptive analysis - part 2.

Figure 11 presents the S and \bar{x} Shewhart control charts for the current example. Note that both the variability and mean of the quality characteristic (chocolate bars’ weights) are under control. We can thus perform a capability study of this production process.

4.8 Process capability analysis

The last page of the univariate SPC module of CEP Online shows a capability report for the process. This report includes a histogram with shadowed specification region, several capability indices (the ones described in Section 2.1.3) and their $100(\alpha)\%$ confidence intervals (only for

CONFIGURATION OF THE SPC ANALYSIS REPORT

Choose the appropriate charts:

S and \bar{x}

R and \bar{x}

CuSum

EWMA

Define the chart condition:

with parameter specification

without parameter specification

Figure 10 – Control charts selection and process condition.

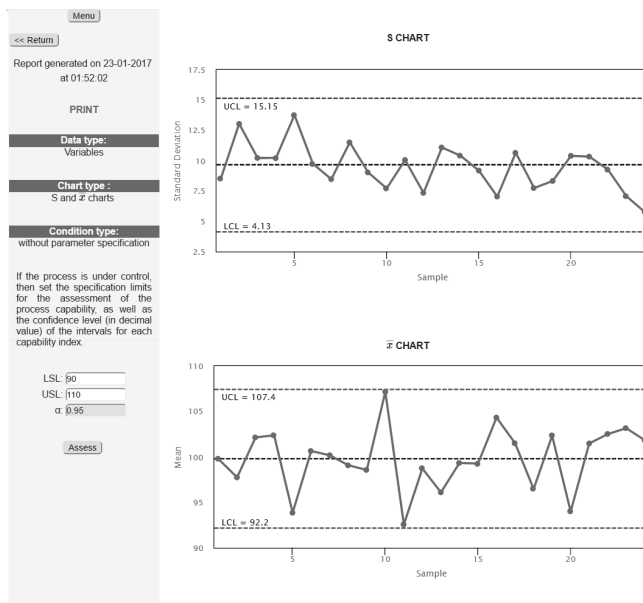


Figure 11 – SPC analysis.

some of them). At the left area, there is again the confirmation of the information (type of data, control chart(s), process condition, lower and upper specification limits, and confidence level) inserted in the previous pages.

From Figure 12, we can observe that the process considered here (chocolate bars' weights) is not able to meet its specifications, since the C_p , C_{pk} and C_{pm} values are much lower than 1 (at the 95% confidence level). Moreover, the P value tells us that this process uses approximately 300% of the specification band.

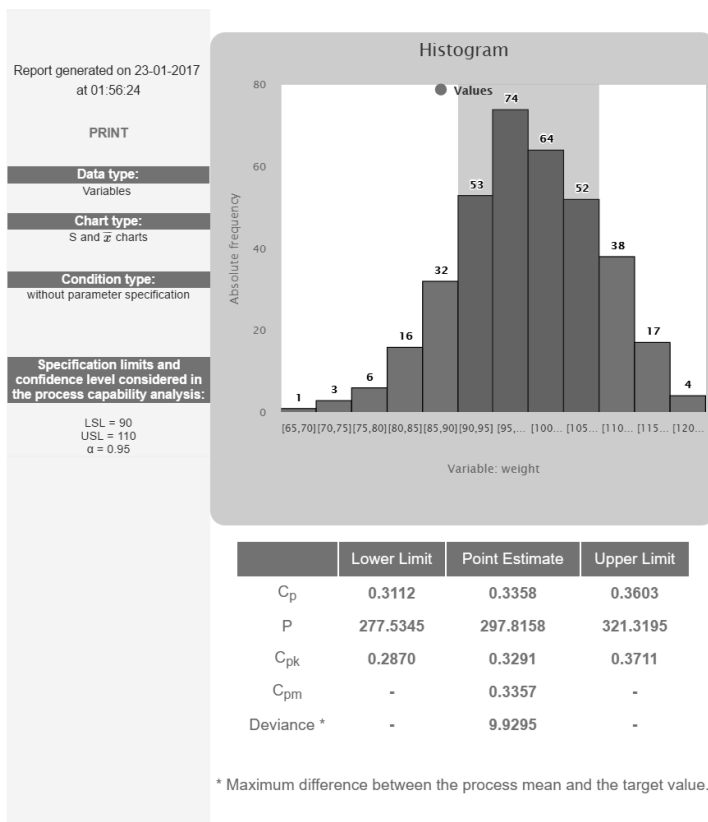


Figure 12 – Process capability analysis.

5 FINAL REMARKS

In this paper, a web-based expert system was developed to perform SPC analysis. It presents two statistical modules, one univariate and the other multivariate, which create different types of reports that show the process quality performance. The CEP Online system has a structure composed by some web languages, free softwares and the use of R software to make instantaneous analysis.

To the best of our knowledge, there is no web-based system that was developed to perform on-line SPC analysis. As we have seen here, the proposed methods are simple, though considering several well-known univariate and multivariate analysis. Besides, our online system can be considered an innovative software, that was built to help mainly small and medium-sized companies to perform SPC and capability analyses at low cost, i.e. with no need of using expensive well-known softwares, such as SAS, SPSS, Minitab and STATISTICA.

The CEP Online system allows the continuous monitoring of the process in a simple and efficient way, taking into account the data typed directly into the system or read from external files. It also allows the study of the ability of the process to meet its specifications/requirements.

In order to promote and popularize the access of information and the statistical science applied to process control and monitoring, the CEP Online system can be used in any company of the country, as well as in SPC training courses.

Since the proposed system has a cloud structure, it is not necessary to install it on local use, requiring only a connection to the Internet (low cost). This indicates that CEP Online can be used in large scale and accessed from different locations and different types of device.

As future prospects, we have the development and implementation of new statistical methods in both SPC modules, including modules that can be built by other researchers/practitioners, besides the continuous improvement of the presented modules.

Hence, we visualize the CEP Online system being used worldwide, giving equal opportunities to companies of all sizes and types of performing SPC analysis, even the ones that can not pay for statistical softwares.

ACKNOWLEDGEMENTS

The research is partially funded by the Brazilian organizations, CNPq and FAPESP.

REFERENCES

- [1] ALT FB. 1985. Multivariate Quality Control. In: KOTZ S, JOHNSON NL & READ CR (Eds.), *The Encyclopedia of Statistical Sciences*, John Wiley & Sons, New York, vol. 6, pp. 110–122.
- [2] BARROS F, BARBOSA E, GONCALVES E & RECCHIA D. 2017. *IQCC: Improved Quality Control Charts*. <http://CRAN.R-project.org/package=IQCC>, r package version 0.7.
- [3] BOYLES RA. 1991. The Taguchi Capability Index. *Journal of Quality Technology*, **23**(2): 107–126.
- [4] BUSABABODHIN P & AMPHANTHONG P. 2016. Copula modelling for multivariate statistical process control: a review. *Communications for Statistical Applications and Methods*, **23**(6): 497–515.
- [5] CANO EL, MOGUERZA JM & REDCHUK A. 2012. *Six Sigma with R*. Springer, New York.
- [6] CHEN N, ZI X & ZOU C. 2016. A Distribution-Free Multivariate Control Chart. *Technometrics*, **58**(4): 448–459.
- [7] FRAGA GR, PEIXOTO TA & RANGEL JJA. 2018. Simulation optimization in dosing process control system in real time in a free and open-source software. *Pesquisa Operacional*, **38**(2): 273–289.
- [8] GANDY A & KVALOY JT. 2013. Guaranteed Conditional Performance of Control Charts via Bootstrap Methods. *Scandinavian Journal of Statistics*, **40**(4): 647–668.
- [9] HOLMES DS & MERGEN AE. 1993. Improving the performance of the T² control chart. *Quality Engineering*, **5**(4): 619–625.
- [10] JOHNSON RA & WICHERN DW. 2007. *Applied Multivariate Statistical Analysis*. Pearson, New Jersey, 6 ed., 800 pp.

- [11] KNOTH S. 2018. *spc: Statistical Process Control - Calculation of ARL and Other Control Chart Performance Measures*. <https://CRAN.R-project.org/package=spc>, r package version 0.6.0.
- [12] KOTZ S & LOVELACE CR. 1998. *Process Capability Indices in Theory and Practice*. Arnold, London.
- [13] LOUZADA F & ARA A. 2018. MWStat: A modulated web-based statistical system. *Pesquisa Operacional*, **38**(2): 291–306.
- [14] LOUZADA F, DINIZ C, FERREIRA P & FERREIRA E. 2013. *Controle Estatístico de Processos: Uma abordagem prática para cursos de Engenharia e Administração*. LTC, Rio de Janeiro, 282 pp.
- [15] MACHADO MAG, COSTA AFB & CLARO FAE. 2009. Monitoring bivariate processes. *Pesquisa Operacional*, **29**(3): 547–562.
- [16] MARDIA KV. 1970. Measures of multivariate skewness and kurtosis with applications. *Biometrika*, **57**(3): 519–530.
- [17] MONTGOMERY DC. 2009. *Introduction to Statistical Quality Control*. John Wiley & Sons, 7 ed., 768 pp.
- [18] MOSTAJERAN A, IRANPANAH N & NOOROSSANA R. 2018. An Explanatory Study on the Non-Parametric Multivariate T^2 Control Chart. *Journal of Modern Applied Statistical Methods*, **17**(1): 2–27.
- [19] R CORE TEAM. 2014. *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria, <http://www.R-project.org/>.
- [20] ROTH T. 2016. *qualityTools: Statistical methods for quality science*. <http://www.r-qualitytools.org>, r package version 1.55.
- [21] SAGHIR A & LIN Z. 2015. Control Charts for Dispersed Count Data: An Overview. *Quality and Reliability Engineering International*, **31**(5): 725–739.
- [22] SANTOS-FERNÁNDEZ E. 2013. *Multivariate Statistical Quality Control Using R*. Springer, New York.
- [23] SANTOS-FERNÁNDEZ E & SCAGLIARINI M. 2012. MPCl: An R Package for Computing Multivariate Process Capability Indices. *Journal of Statistical Software*, **47**(7): 1–15.
- [24] SARKAR D. 2008. *Lattice: Multivariate Data Visualization with R*. Springer, New York.
- [25] SCRULLA L. 2004. qcc: An R package for quality control charting and statistical process control. *R News*, **4**(1): 11–17, <https://cran.r-project.org/doc/Rnews/>.
- [26] SHINDE RL & KHADSE KG. 2008. Multivariate process capability using principal component analysis. *Quality and Reliability Engineering International*, **25**(1): 69–77.
- [27] SULLIVAN JH & WOODALL WH. 1995. A comparison of multivariate quality control charts for individual observations. *Journal of Quality Technology*, **28**(4): 398–408.
- [28] SZEKELY GJ & RIZZO ML. 2005. A new test for multivariate normality. *Journal of Multivariate Analysis*, **93**(1): 58–80.
- [29] TRACY ND, YOUNG JC & MASON RL. 1992. Multivariate control charts for individual observations. *Journal of Quality Technology*, **24**(2): 88–95.
- [30] VEEVERS A. 1999. Capability indices for multiresponse processes. In: PARK SH & VINING GG (Eds.), *Statistical Process Monitoring and Optimization*, Marcel Dekker, New York, vol. 28, pp. 185–194.

- [31] VERDIER G. 2013. Application of copulas to multivariate control charts. *Journal of Statistical Planning and Inference*, **143**(12): 2151–2159.
- [32] VILLASENOR-ALVA JA & GONZALEZ-ESTRADA E. 2009. A generalization of Shapiro-Wilk's test for multivariate normality. *Communications in Statistics: Theory and Methods*, **38**(11): 1870–1883.
- [33] WANG FK & CHEN JC. 1998. Capability index using principal component analysis. *Quality Engineering*, **11**(1): 21–27.
- [34] WANG FK & DU TCT. 2000. Using principal component analysis in process performance for multivariate data. *Omega*, **28**(1): 185–194.