Six Sigma: Main Metrics and R Based Software for Training Purposes and Practical Industrial Quality Control

Ana Rita Costa, Carla Barbosa, Gilberto Santos, M. Rui Alves

ABSTRACT

Purpose: To clarify the different types of data likely to occur in any service or industrial process, the main applicable statistics for each type of data and the Six Sigma metrics that allow characterising and benchmarking organisational processes.

Methodology/Approach: A short reference to the statistical process control is carried out, from Shewhart's works to Motorola's achievements, followed by a short discussion of the use of Six Sigma tools as a part of today's total quality approaches, and by a discussion of the continuous, attribute and counting data worlds and their main applications in process analysis. Because many quality professionals may have difficulties dealing with engineering perspectives, a review of main classic and Six Sigma process metrics is done with examples. Complementing discussions, four functions written in the R language are presented, which can deal with real organisational data, or can be used for training purposes.

Findings: The functions developed provide useful graphical displays and calculate all necessary metrics, having the ability to let the user provide theoretical values for training activities. Real and simulated case studies help understanding data worlds and respective Six Sigma metrics.

Research Limitation/implication: This paper reports an intentionally simple theoretical perspective of Six Sigma metrics and friendly software which is available to all interested professionals on request to the authors.

Originality/Value of paper: The paper presents clear definitions of main data types and metrics and is supported by a set of four new functions that can be used by any researcher with a minimum knowledge of the R software.

Category: Technical paper

Keywords: six-sigma; data worlds; process capability; sigma level; Six Sigma software

1   INTRODUCTION

Six Sigma is essentially a methodology developed by Motorola, to address manufacturing or service problems. It is related to improvement projects following the define, measure, analyse, improve and control methodology (known as DMAIC), a well-defined project approach, utilisation of selected quality tools, and, in many circumstances, applying lean methodologies and many other specific quality tools. Theory, applications and developments can be found in many references, like Aboelmaged (2010), Antony and Banuelas (2002), Firat et al. (2017), Ward, Poling and Clipp (2008), Thomas, Barton and Chuke-Okafor (2009), Tjahjono et al. (2010).

However, it started to be a well-defined methodology to reduce variability, whose origins can be traced back to the beginning of the twentieth century after the works of Walter Shewhart (Mitra, 2016; Wadsworth, Stephens and Godfrey, 2007) dealing with the control of the variability of high yield machines. He proposed the use of statistics to control high throughput processes, and the quality control charts devised by Shewhart, which can be faced as the beginning of "statistical process control" (SPC) were the main tool to follow and control processes' performances and are still referred to as one of the seven basic quality tools (Mitra, 2016; Pyzdek and Keller, 2018).

The works of Shewhart are still valid, but the advent of computers and efficient software enabled the production of high yield machines and high throughput processes, forcing professionals to be increasingly concerned with precision and control, maintaining process parameters within very narrow specification limits, to avoid production of defectives (waste) and re-work (Antony and Banuelas, 2002; Marques et al., 2018; Pyzdek and Keller, 2018; Santos et al., 2006; Tjahjono et al., 2010).

In present times, companies without quality do not survive (Araújo et al., 2019; Santos and Milán, 2013; Santos, Murmura and Bravi, 2019). There are many companies where process control is integrated in Management Systems (quality, environmental, safety and others), seeking to optimize human and material resources (Carvalho, Santos and Gonçalves, 2018; Zgodavová et al., 2019; Santos, Rebelo and Santos, 2017; Ribeiro et al., 2017) promoting the emergence of new businesses (Bravi, Murmura and Santos, 2017; Santos et al., 2018; Doiro et al., 2017; Santos, Bravi and Murmura, 2018). In this context, tools like Six Sigma and lean are becoming more and more popular.

Six Sigma, as it is now formulated, tries to address all problems referred above, but uses procedures and metrics that are sometimes difficult to understand. Also, possibly because too many people with very different backgrounds are working in the field, a lot of confusion in critical aspects of this methodology is evident.

For example, there is unanimity in the definition and calculation of the traditional "process capability" metric (*Cp*), but in what concerns metrics defined within Six Sigma, definitions are not unanimous. The acronym *Pp* is used to refer "overall capability" or "process performance", and some authors do not provide clear definitions. The same happens with acronyms *DPMO* and *PPM*, crucial Six Sigma metrics, that can represent "defects per million opportunities", "defectives per million", or both (Antony and Banuelas, 2002; Barsalou, 2015; Brook, 2006; Moosa and Sajid, 2010). The same is true for the definition of data types, where virtually all authors agree in what concerns the "continuous data world", but with different opinions arising in relation to the "discrete data world" and the meaning of "counting", leading to difficulties in the establishment of a clear difference between "defectives" and "defects", and respective metrics.

It is essential to realise that Six Sigma is very important and is a popular quality tool, highly accepted in all types of organisations, hence the interest by the research community, the considerable amount of published papers and textbooks, and the significant number of software packages dedicated or including routines, for six-sigma. For those interested in this subject, the books of Pyzdek and Keller (2018) and Stamatis (2003) are highly recommended. But to understand that there are objections to the validity of Six Sigma as a quality tool/model by itself, reading the interview of Dr. Juran conducted by Paton (2002) is also highly recommended.

The purposes of this paper are three:

1. To try to clarify some crucial aspects of the Six Sigma methodology in what concerns important metrics, following closely the terminology used by the American Society for Quality (ASQ) (Barsalou, 2015);

2. To address three types of data worlds (continuous, attribute and counting data worlds) following the initial (not actual) Minitab terminology (Brook, 2006);

3. To present a set of useful functions, written by the authors using the R language (Alves, 2011; R Core Team, 2019) specially designed for Six Sigma lecturers and for Six Sigma practitioners, enabling the users to analyse several types of process data, producing useful graphs and automatically calculating the most important metrics.

It is must be highlighted that some functions were written in R by Cano, Moguerza and Redchuk (2012) and Cano et al. (2018) to deal with Six Sigma. However, although very comprehensive and freely accessible to the community, they require a good knowledge of the R language. Consequently, there is still an opportunity to produce simple, easy to handle functions, helpful for young practitioners and for people from backgrounds other than engineering, with low computational competencies.

2   METHODOLOGY

Four functions were written by the authors using the freely available R project software (R Core Team, 2019). The first purpose of these functions is to provide two main possibilities: (i) "simulate" and analyse data using argument "simulate=TRUE": the user supplies specification and process parameters and the functions simulate data and analyse it, so that they become quite friendly for teaching/learning purposes; (ii) "read" and analyse real examples supplied in simple text files (e.g., Notepad running in MS Windows environment), using argument "read=TRUE": the user supplies specification parameters and a data file, which is then automatically analysed.

In both "simulating" and "reading" modes, these functions produce two frames: (i) a frame with information on data, Six Sigma metrics and statistical tests; (ii) a frame with graphical aspects, including histograms or barplots and time plots.

All functions were designed in order to be confined to specific data worlds, requiring minimum knowledge of computation and Six Sigma. These functions are available to all interested readers on request.

3   RESULTS

In this section, the definition of data worlds follows that used in previous Minitab approaches (Brook, 2006), and metrics follow the definitions of the ASQ (Barsalou, 2015) closely.

3.1   The Continuous Data World

Many processes can be described by "key process indicators" (KPI) that are continuous variables following normal distributions, or that can be transformed in variables following that important probability distribution. This implies that a KPI can be characterized by a mean value (μ) and an uncertainty (3σ), i.e., in the form $\mu \pm 3\sigma$, where σ is the process standard deviation and value ± 3, which multiplies σ, is a z value from the standard normal distribution inducing a 99.74% confidence interval. Then, the "lower process limit" (LPL) is given by $LPL = \mu - 3\sigma$ and the "upper process limit" (UPL) is $UPL = \mu + 3\sigma$, with the "process range" (PR), also called precision, given by $PR = UPL - LPL = 6\sigma$.

To analyse these processes, they must be compared with specifications. Any specification (required by customers, laws or imposed internally) can be written in a way similar to a process, i.e., considering a "target value" (TV) and an "allowable error" (AE), and be written in the form $TV \pm AE$. Than the "lower specification limit" (LSL) is $LSL = TV - AE$ and the "upper specification limit" (USL) is $USL = TV + AE$, and the "specification range" (SR), also called tolerance, is $SR = USL - LSL$.

3.1.1   Capability Metrics

Taking these definitions into consideration, the main metrics for process evaluation are the "process capability" (Cp) and the "process capability index" (Cp_K) (Brook, 2006; Bartalou, 2015) given by the following equations where σ_w represents an estimate of the process σ based on several samples collected from production over time:

$$Cp = \frac{SR}{PR} = \frac{tolerance}{precision} = \frac{specification\ range}{process\ range} = \frac{USL - LSL}{6 \times \sigma_w} \qquad (1)$$

$$Cp_L = \frac{LSL - \mu}{3 \times \sigma_w} \qquad Cp_U = \frac{\mu - USL}{3 \times \sigma_w} \qquad Cp_K = \begin{cases} Cp_L & if\ |Cp_L| < |Cp_U| \\ Cp_U & if\ |Cp_L| \geq |Cp_U| \end{cases}$$

It is also important to evaluate the number of items falling out of the specification limits, expressed in "parts per million" (PPM): the PPM_L or $PPM < LSL$, PPM_U or $PPM > USL$ and $PPM_T = PPM_L + PPM_U$. These PPM metrics refer to the process "defective yield". The "sigma level" (Z_σ or Z_{bench}) is the value of the standard normal variable corresponding to PPM_T, imagining all defective items as exceeding the UPL. PPM is calculated as areas under the normal curve exceeding the specification limits and are therefore the "expected" or "probable" fraction of defective units that will be produced, multiplied by 10^6.

Because there is confusion on several references, it must be emphasised that the total area under any normal curve always equals 1, so that tables of the normal distribution (or appropriate software) will give results (probabilities) as fractions of 1 and not as percentages. Consequently, percentages (%) are fractions of 1 multiplied by 10^2 and parts per million (PPM) are fractions multiplied by 10^6.

Dispersion (D) is a new metric, herein introduced, calculated as $D = AE/\sigma = 3 \times Cp$, and is used to make a distinction between the sigma level and the common sense of 3σ. If a process mean is centred in the specification target value, D is the number of standard deviations between the process mean and the specification limits.

For the estimation of the process σ, referred to as σ_{within} or just σ_w, processed units can be collected in three different ways: (i) samples collected one at a time, therefore collecting N samples of size $n=1$ and estimating σ_w based on moving ranges; (ii) groups of units always with the same number of units, i.e., N samples of size n, with $n > 1$, and estimating σ_w based on samples' variances; (iii) groups of units with different numbers of units per group, i.e., N samples of different sizes, and estimating σ_w based on samples' sums of squares. These estimates of the process σ based on σ_w reflect the common causes of variation, inherent to the process, and therefore represent the best that the process can do, and are used for the calculation of capability metrics.

3.1.2   Function *SS.Cp*

In order to understand these definitions, function *SS.Cp* was built (mainly for teachers and students). The user supplies the values for a specification (*TV* and *AE*) and the process parameters (μ and σ). The function draws the specification and the normal curve corresponding to the process parameters and calculates all metrics referred above.

Fig. 1 was produced with function *SS.Cp* with arguments $TV = μ = 500$, $AE = 15$ and $σ = AE/3 = 5$. In this situation, the process is fully adjusted to the specification. In Figure 1, the specification values, process characteristics and the yield metrics just discussed are calculated and shown on the left-hand side, and on the right-hand side, the specification is plotted together with the normal curve corresponding to the process. Because the process is centred in the specification, *Cp* and *CpK* are equal, and because the process limits and specification limits are also equal, $Cp = 1$ and $D = 3$. Nevertheless, some defective units are produced and $Z_{bench} ≈ 2.8$: hence the difference between "dispersion" and "sigma level".

Fig. 2 was produced with function *SS.Cp* with the same arguments for the specification, $TV = 500$ and $AE = 15$, but with $σ = AE/4 = 3.75$ and $μ = 500 – 1.5 × σ = 494.375$, reflecting a 1.5 sigma-shift. In this situation, although the process is capable, as seen by $Cp = 1.33$ and $D = 4$, the sigma-shift displaced the process to the left, resulting in a *CpK* which is negative and smaller than *Cp* in absolute value. Consequently, more defectives are produced exceeding the *LSL*. The sigma level is now only $Z_{bench} = 2.5$.
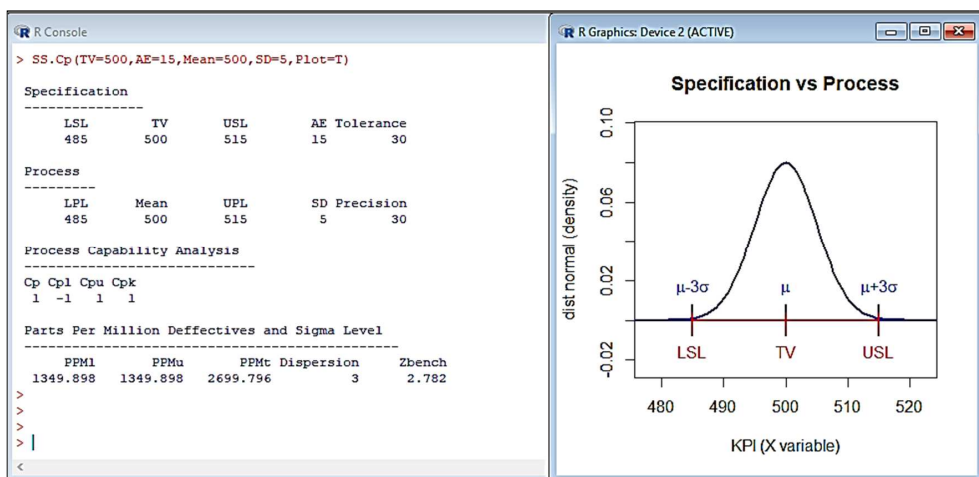

Figure 1 – Output of Function SS.Cp, Comparing a Process with μ Adjusted to TV, and with σ = AE/3. The Dispersion is 3, Cp = CpK = 1 and 2700 PPM are Produced, Corresponding to a Z_{bench} = 2.78. The Process Parameters μ and σ are not Calculated, but Stipulated by the User
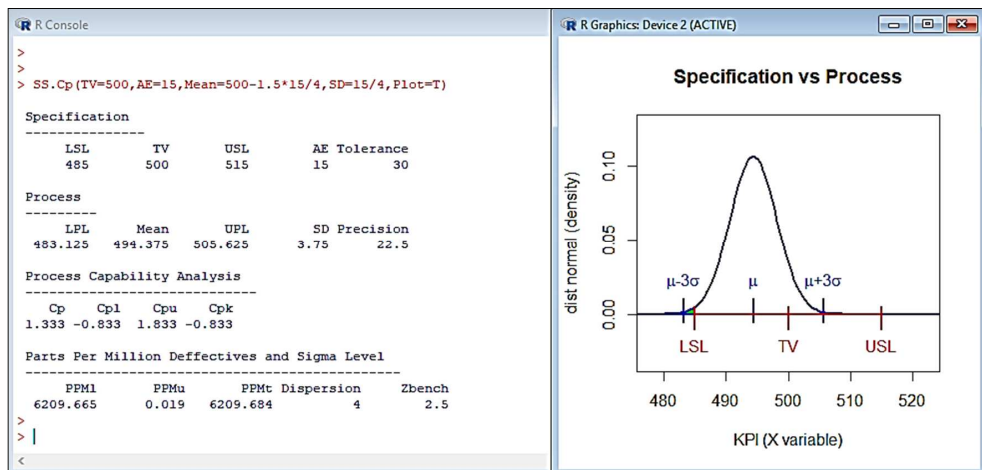
Figure 2 – Output of Function SS.Cp, Showing a Process with a 1.5σ Shift to the Left of the TV, with σ = AE/4. The Dispersion is 4, Cp = 1.33, Cp_K = −0.833 and 6200 PPM are Produced, leading to a Z_{bench} = 2.5

3.1.3   Performance Metrics

In Six Sigma projects, besides capability metrics, performance metrics are also important (Barsalou, 2015): these are the "process performance" (*Pp*), and the "process performance index" (*Pp_K*). The difference between capability and performance metrics resides in the estimate of the process standard deviation:

$$Pp = \frac{SR}{PR} = \frac{tolerance}{precision} = \frac{specification\ range}{process\ range} = \frac{USL - LSL}{6 \times \sigma_o} \tag{2}$$

$$Pp_L = \frac{LSL - \mu}{3 \times \sigma_o} \qquad Pp_U = \frac{\mu - USL}{3 \times \sigma_o} \qquad Pp_K = \begin{cases} Pp_L \ if\ |Pp_L| < |Pp_U| \\ Pp_U \ if\ |Pp_L| \geq |Pp_U| \end{cases}$$

As it can be seen comparing equations (1) and (2), the performance metrics are equal to capability metrics, but use a different estimate of the process σ, referred as $\sigma_{overall}$ or just σ_o. The calculation of σ_o uses all units collected, irrespectively of time or samples, i.e., as if it was just one big sample. Such an estimate is called the "overall sigma" and incorporates all common causes of variation as well as the special causes of variation arising along time.

Following the definitions for capability and performance metrics, it can be concluded that these metrics will be different if: (i) the process is affected by special causes of variation; (ii) if the process variability increases with time; (iii) if there is a sigma shift, a term that is used in Six Sigma (Pyzdek and Keller, 2018) to refer to a displacement of the project mean in relation to the specification target value, measured in standard deviation units. In all these three cases, σ_o will be higher than σ_w, and all performance metrics will be significantly higher than the corresponding capability metrics.

3.1.4 Function *SS.Norm*

Once capability, performance and defective yield metrics are understood, processes can be analysed in business practice with function *SS.Norm*. This function needs to be supplied with the specification values (*TV* and *AE*), a file name with practical data, and argument "samples". The latter argument is just the number of samples actually analysed, as described in section 3.1.1. Based on these arguments, function *SS.Norm* will automatically calculate estimates for σ_o and σ_w and all metrics just discussed.

Fig.3 presents a real industrial example: the analysis of daily losses expressed as the fraction of wasted fruit mass per day in a fruit juice processing plant. On the left-hand side of the function output, all metrics are presented: observed data and the estimates of capability, performance and yield, together with some statistical tests. On the right-hand side, the specification, a histogram of observed data and the estimates of process behaviour as two normal curves (the narrower, blue curve representing the process with only common causes of variation, and the wider, yellow curve representing the process in the long term including the observed special causes of variation), and shaded areas under the normal curves showing the fraction of defective units. Also, on the same side, a plot of data values over time enables the visual observation of the special causes of variation.

Fig. 3 is an example of a very bad process: *Cp* is lower than 1, meaning that the process range is higher than the specification range; the positive and small values of Cp_K and Pp_K demonstrate that the process mean is displaced towards the right, producing higher daily losses; the sigma level is very low and even if all special causes of variation were removed, Z_{bench} would only be slightly higher than 1.2, which is a very low value.

Fig. 4 shows the analysis of a real industrial example relative to coffee packaging. The specification limits were derived from Portuguese legal aspects related to the metrological control of pre-packaged foods (Portaria 1198/91) and internal specifications. As it can be seen on the left-hand side of Fig. 4, the short-term sigma level is extremely high ($Z_{bench(ST)}$ = 8.5), reflecting a process with very high precision. The long-term sigma level is also high, and curiously reflects a sigma-shift of 1.5 ($Z_{bench(LT)}$ = 7.0), as predicted in the Six Sigma theory (Pyzdek and Keller, 2018). This shift shows that the process is very accurate, but there is a margin for improvement just by the removal of the special causes of variation. It is worth noting the short-term metrics $Cp \approx 3$ and a dispersion $D \approx 9$.
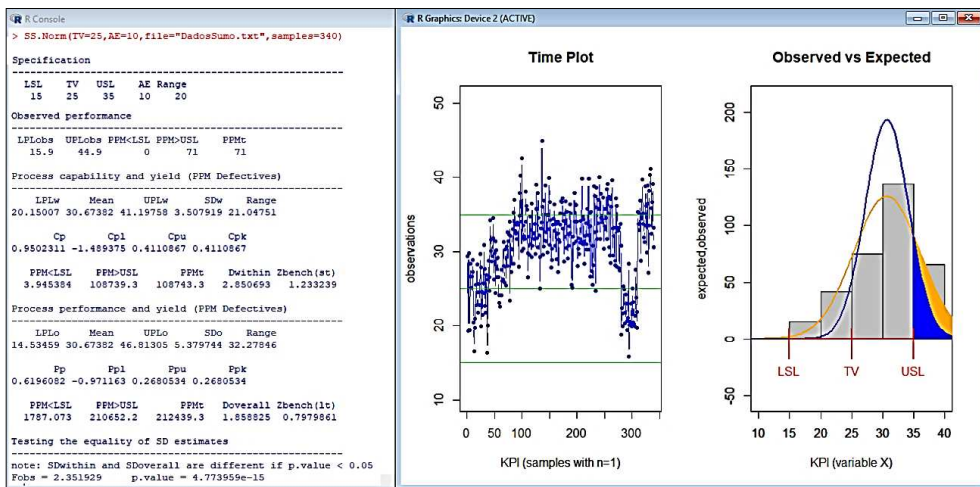
Figure 3 – Output of Function SS.Norm, Showing a Real Industrial process dealing with Daily Losses in Juice Production. The Process has Very Low Performance (Pp = 0.62) and is Displaced towards the Right (Pp_K = 0.268), Producing 108743 PPM Corresponding to a Long Term Sigma Level of 0.798. There are Evident Special Causes of Variation (p = 4.773959×10⁻¹⁵), and a Cp = 0.95 Shows a Need for Improvement
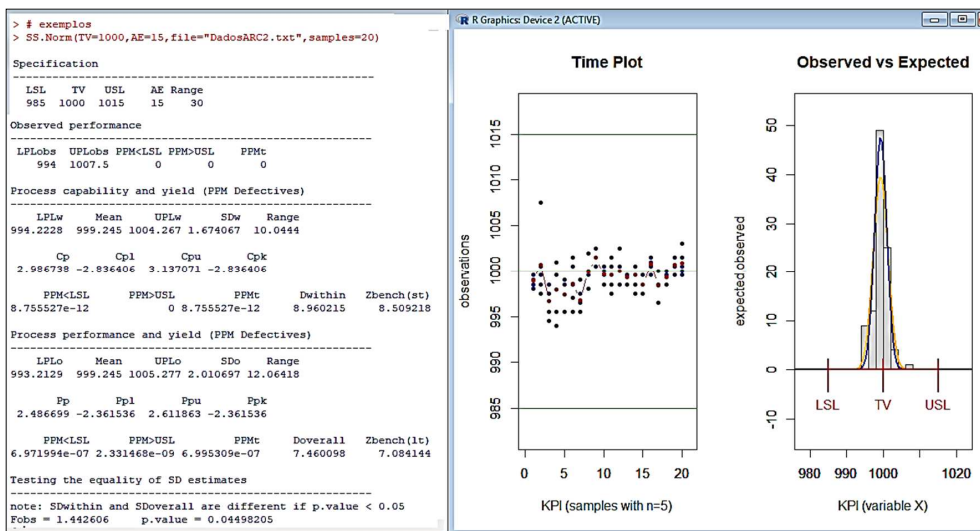

Figure 4 – Output of Function SS.Norm, Showing a Real Industrial Process of Coffee Packaging. The Process has a Very High Capability Expressed as a Short-Term Sigma Level Equal to 8.5. Metrics Show that with Time this Level Will Be Reduced to 7, Reflecting a True 1.5σ Shift.

3.2 The Attribute Data World

Some processes can be described by *KPI*s that classify units as "defective" or "non-defective", i.e., as "1" or "0". In this situation, data follows a binomial distribution, and the critical process parameter is the fraction of defective items produced, denoted by π. These *KPI*s that correspond to counting defective units are expressed as only two possible quality outcomes and are referred to as "attributes" (Brook, 2006).

3.2.1 Main Metrics for Attributes

In the case of attributes, one has to collect and inspect *N* production units, classifying each unit as a "1" (defective) or a "0" (non-defective). Units can be collected and inspected in different ways, leading to different aspects.

One can collect and inspect *N* units, one at a time, leading to a series of *N* "zeros" and "ones". In this case, two data displays are useful: (i) a barplot with two bars, one for the number of defective units, the other for the amount of non-defective units; (ii) a plot of results (of "zeros" and "ones") over time, to try to investigate possible special causes of variation.

If one collects *N* samples of equal size *n*, this leads to a series of natural values ranging between "0" and "*n*" representing the number of defective units observed in each sample. Therefore, a barplot with *n* columns, with the height of each column representing the number of samples with 0, 1, …, *n* defectives will be informative. Also, a plot of the evolution of the number of defective units per sample along time will be necessary.

Finally, one may have a series of *N* samples of different sizes. In this situation it is essential to convert the number of defective units per sample to the fraction of defective units per sample, producing a plot of the fraction of defective units observed over time and a histogram showing the distribution of the fraction of defectives per sample.

Whatever the procedure applied, all data is used to estimate the process defective fraction, π. Such an estimate is calculated as the sum of all values (the sum of all defective units) divided by the total number of inspected units (*M*), which is referred by *P*.

In mathematical notation, one collects a series of *M* values, which can be represented as x_1, x_2, x_3, …, x_i, …, x_M, and *P*, the estimate of π, is:

$$\pi \sim P = \frac{1}{M} \sum_{i=1}^{M} x_i \qquad\qquad PPM_T = P \times 10^6 \qquad\qquad (3)$$

It is worth noting that it is irrelevant if one is talking about samples of size 1, or equal or unequal sizes since the average of sample's defective fractions are always equal to the overall defective fraction.

As shown in equations (3), an estimate of the long-term defective production yield is just the part per million defectives, calculated as $PPM_T = P \times 10^6$. Consequently, a sigma level, $Z_{bench(LT)}$, can be calculated using P as the extreme right-hand side area under the standard normal distribution and determining the value of the corresponding z variable.

However, with this type of data, unless one is talking about very high numbers of inspected units, it is virtually impossible to distinguish between short- and long-term defective fractions, as well as special causes of variation. Hence, all data is faced as long-term, and if a short-term sigma-level, $Z_{bench(ST)}$, is desired, it can be approximated by $Z_{bench(ST)} = Z_{bench(LT)} + 1.5$.

3.2.2 Function *SS.Defectives*

In order to deal with attribute data, function *SS.Defectives* was built. This function has two possibilities: (i) using argument "Simulate = TRUE", intended for teaching and studying purposes, leading to the simulation of production outputs based on values supplied by the user for π (the true production defective fraction), N (number of samples) and n (sample size); (ii) indicating a data file, which overrides all other arguments, leading to the automatic analysis of real data.
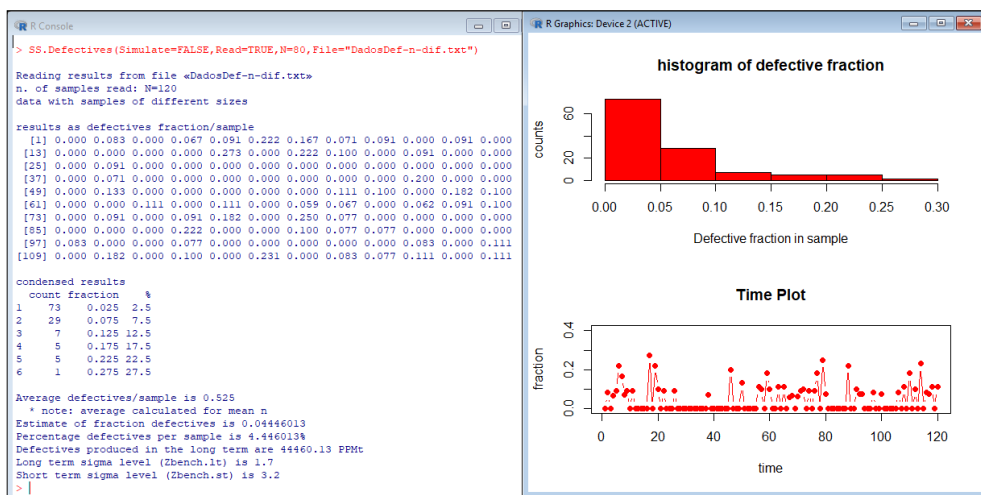

Figure 5 – Output of Function SS.Defectives, Showing a Real Industrial Process, Analysing Samples of Different Sizes of Pre-Packaged Coffee Beans. Units with Contents Lower than the Legal AE or Higher than the Internal AE, Are Defective. $Z_{bench\ (LT)}$ Shows Possibilities for Improvement

Fig. 5 shows an example of the use of function *SS.Defectives* applied to real data consisting of N=120 samples of different sizes. All raw values in the file are converted to sample defective fractions, which are shown on the left-hand figure frame, together with a division in classes for analysis of dispersion and as the basis for the histogram. Several metrics are also presented, mainly the estimate of

the defective production fraction, total *PPM*, and the sigma-levels $Z_{bench(LT)}$ and $Z_{bench(ST)}$. In this example, 4.45% of defective units are estimated to be produced over time, which corresponds to a long-term sigma-level of 1.7.

3.3   Counting Data World

The last type of *KPI*s of interest belongs to the "counting data world". To work with this type of data, each production unit is faced as an "area", or "quantity", where some "types of defects" are possible. Each type of defect is called an "opportunity", and the total number of opportunities is referred to as "*NO*". The number of units inspected is *N* (Barsalou, 2015; Brook, 2006).

The purpose of this quality control practice is to count the number of defects observed for each opportunity and for each unit, for all units, and calculate an estimate of the population parameter of interest: the average number of defects per opportunity in the long-term, λ, expressed in parts per million, usually referred to as "defects per million opportunities" (*DMPO*) (Moosa and Sajid, 2010).

3.3.1   Main Metrics for Counts

In what concerns the actual work that has to be carried out to calculate main metrics, the following aspects are important: (i) a template is prepared with a matrix-like structure, with columns representing opportunities (types of defects) and rows representing sample units (see left hand-side frame of Fig. 6); (ii) a unit is collected and inspected, counting the number of defects of each type, filling a matrix row; (iii) this practice is repeated for all *N* units, filling all *N* matrix rows; (iv) the last matrix column is the automatic row sum, i.e., the total "number of defects per unit" (*NDU*); (v) the last matrix row is the automatic column sum, i.e., the total "number of defects per opportunity" (*NDO*).

Based on these initial metrics (*NDU* and *NDO*), two crucial intermediate metrics are derived: (i) the average "number of defects per unit" (*DPU*); (ii) the average "number of defects per opportunity" (*DPO*):

$$DPU = \frac{1}{N}\sum_{i=1}^{N} NDU_i \quad DPO = \frac{1}{N \times NO}\sum_{i=1}^{N}\sum_{j=i}^{NO} NDO_{ij} = \frac{DPU}{NO} \quad (4)$$

Because any unit with one or more defects is faced as a defective unit, *DPU* is an estimate of the process parameter λ, the long-term number of defects per unit. Because this parameter follows a *Poisson* distribution, the probability of "0" defects per unit, i.e., the probability of *NDU*=0, can be calculated and the process yield, in terms of fraction defective units, in the long-term, is expected to be the quantity $1-e^{-DPU}$ (Brook, 2006):

$$P(NDU = 0) = \frac{DPU^0 \times e^{-DPU}}{0!} = e^{-DPU} \quad (5)$$

$$P = fraction_{(defectives)} = 1 - e^{-DPU} \quad PPM_t = 10^6 \times P \qquad (6)$$

The long-term sigma-level, $Z_{bench(LT)}$, is the z value corresponding to the extreme right-hand side area under the standard normal curve equivalent to P (fraction defective), and the short-term sigma-level, $Z_{bench(ST)}$, is:

$$Z_{bench(ST)} = Z_{bench(LT)} + 1.5 \qquad (7)$$

Finally, if instead of PPM_T the interest is in the study of "defects per million opportunities", referred as $DPMO$, instead of defective units per million (PPM_T), then metric $DPMO$ is the important metric, calculated as (see Fig. 6):

$$DPMO = 10^6 \times \frac{DPU}{NO} \qquad (8)$$

This metric, although famous in the area of Six Sigma projects, should always be avoided, unless one can be sure that all opportunities for defects are unambiguously defined and that no new opportunities will arise in the course of a practical industrial work (Brook, 2006).
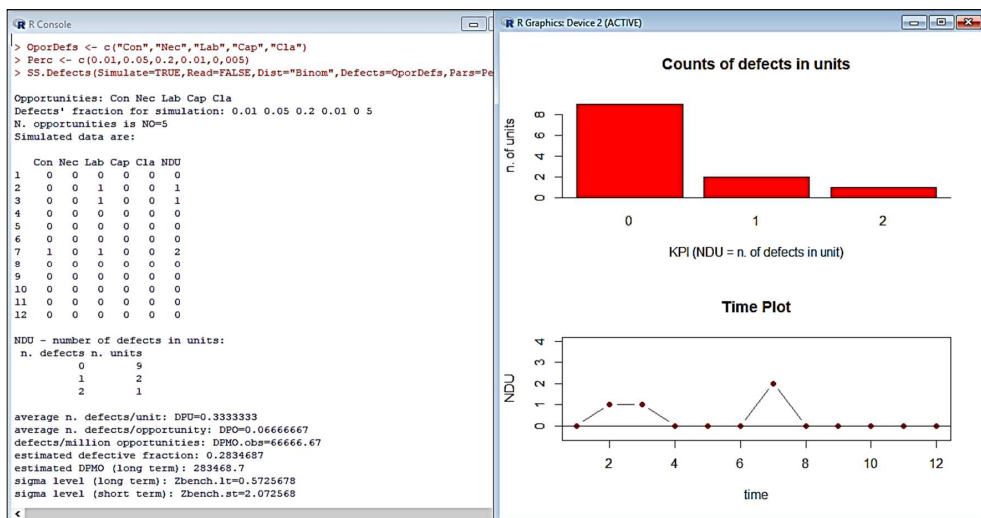

Figure 6 – Output of Function SS.Defects, in the Simulation Mode. Simulated Data is Presented on the Left-Hand Side, Together with the Main Metrics: DPU, DPO, DPMO and Sigma-Levels. On the Right-Hand Side, Barplot with Main Observed Counts of Samples per Number of Defects, and Evolution with Time. $Z_{bench(LT)}$ is Very Small and $Z_{bench(ST)}$ Must be Seen with Caution Since Its Meaning is Doubtful

3.3.2 Function *SS.Defects*

Function *SS.Defects* was built to address counting data problems as described in section 3.3.1, and is similar to function *SS.Defectives*, enabling simulations for teaching/studying purposes, but also the treatment of practical organizational problems.

Figure 6 provides an example of the use of this function in the simulation mode. The user supplies the function with a list of opportunities (a list of names of defects that can occur in bottles of mineral water) and a list of defective fractions for each opportunity (a list of π values). In this simulation, the opportunities were "content" (*Con*), "bottleneck" (*Nec*), "label" (*Lab*), "capsule" (*Cap*) and "clamping" (*Cla*), with π values equal to 0.01, 0.05, 0.2, 0.01, 0.5, respectively. The user also tells the function how many units are to be generated (*N*). Then, using the binomial distribution, the *SS.Defects* function, generates random data for each opportunity and, afterwards, provides two frames: a frame with reports and metrics (left-hand side), and a frame with graphical aspects (right-hand side).

4   CONCLUSION

Although Six Sigma is a very comprehensive methodology, involving a lot of quality tools and ways to address projects, Six Sigma metrics may be a key question in the methodology.

In this paper, the main metrics and the main graphical displays were presented, and discussions were followed with practical cases treated with functions *SS.Cp*, *SS.Norm*, *SS.Defectives* and *SS.Defects*, which were designed using the R language, to work within the "continuous", "attribute" and "counting" data worlds. These functions are very versatile because they can be used in the "simulation mode" for teaching/learning purposes, and on the "reading mode" for the analysis of real data, enabling users to quickly carry out any process analysis with the correct Six Sigma metrics.

The examples show that with very little knowledge of computing or statistics, any interested professional can analyse real cases, producing illustrative graphs and all relevant Six Sigma metrics. Furthermore, these functions can be used to simulate data, treating simulated data as real and produce the same graphs and metrics, being very useful for teaching Six Sigma metrics and for students to compare specifications with process parameters and understand the meaning of Six Sigma metrics. These functions are available to any interested reader.

All functions presented in this paper accept only one *KPI* and single stage processes. In order to make them more versatile, work is being carried out in order to treat data arising from multistage processes and it is envisaged to extend them to deal with more than one KPI.

ACKNOWLEDGEMENTS

REFERENCES

Aboelmaged, M., 2010. Six Sigma quality: a structured review and implications for future research. *International Journal of Quality &Reliability Management*, [e-journal] 27(3), pp.268-317. http://dx.doi.org/10.1108/02656711011023294.

Alves, M., 2011. *Introdução à Programação em R*. Porto: Publindústria, Edições Técnicas.

Antony, J. and Banuelas, R., 2002. Key ingredients for the effective implementation of Six Sigma program. *Measuring Business Excellence*, 6(4), pp.20-27.

Araújo, R., Santos, G., Costa, J. and Sá, J.C., 2019. The Quality Management System as a Driver of Organizational Culture: An Empirical Study in the Portuguese Textile Industry. *Quality Innovation Prosperity Journal,* [e-journal] 23(1), pp.1-24. http://dx.doi.org/10.12776/qip.v23i1.1132.

Barsalou, M., 2015. *The ASQ Pochet Guide to Statistics for Six Sigma Black Belts*. Milwaukee: ASQ Quality Press.

Bravi, L., Murmura, F. and Santos, G., 2017. Attitudes and Behaviours of Italian 3D Prosumer in the Era of Additive Manufacturing. *Procédia Manufacturing,* [e-journal] 13, pp.980-986. https://doi.org/10.1016/j.promfg.2017.09.095.

Brook, Q., 2006. *Six Sigma and Minitab – A Complete Guide for all Six Sigma Practitioners*. London: QSB Consulting, Ltd.

Cano, E., Moguerza, J. and Redchuk, A., 2012. *Six Sigma with R: Statistical Engineering for Process Improvement*. New York: Springer.

Cano, E., Moguerza, J., Prieto, M. and Redchuk, A., 2018. *Six Sigma Tools for Quality Control and Improvement - Vignette for Package 'SixSigma'*. [pdf] The R Foundation. Available at: <https://cran.r-project.org/web/packages/SixSigma/SixSigma.pdf> [Accessed February 2019].

Carvalho, F., Santos, G. and Gonçalves, J., 2018. The disclosure of information on Sustainable Development on the corporate website of the certified Portuguese organizations. *International Journal of Quality Research*, 12(1), pp.253-276.

Doiro, M., Fernández, J.F., Félix, M.J. and Santos, G., 2017. ERP - machining centre integration: a modular kitchen production case study. *Procédia Manufacturing*, [e-journal] 13, pp.1159-1166. https://doi.org/10.1016/j.promfg.2017.09.178.

Firat, S., Arioglu, M., Ersoy, E., Gök, S. and Ünal, E., 2017. A Six Sigma DMAIC Process for Supplier Performance Evaluation using AHP and Kano's Model. *International Journal of Business Analytics*, [e-journal] 4(2), pp.37-61. https://doi.org/10.4018/IJBAN.2017040103.

Marques, C., Lopes, N., Santos, G., Delgado, I. and Delgado, P., 2018. Improving operator evaluation skills for defect classification using training strategy supported by attribute agreement analysis. *Measurement*, 119, pp. 129-141. https://doi.org/10.1016/j.measurement.2018.01.034.

Mitra, A., 2016. *Fundamentals of Quality Control and Improvement*. New Jersey: John Wiley & Sons.

Moosa, K. and Sajid, A., 2010. Critical analysis of Six Sigma implementation. *Total Quality Management*, 21(7), pp.745-759.

Paton, S., 2002. Juran: A Lifetime of Quality. *Quality Digest*, [online] Available at: <https://www.qualitydigest.com/aug02/articles/01_article.shtml> [Accessed May 2019].

Portaria 1198/91, 1991. *Diário da República I – Série – B*, *291*, pp 6682-6684.

Pyzdek, T. and Keller, P., 2018. *The Six Sigma Handbook*. New York: McGraw-Hill.

R Core Team, 2019. *R: A Language and Environment for Statistical Computing*. Vienna: R Foundation for Statistical Computing.

Ribeiro, F., Santos, G., Rebelo, M. and Silva, R., 2017. Integrated Management Systems: trends for Portugal in the 2025 horizon. *Procédia Manufacturing*, [e-journal] 13, pp.1191-1198. https://doi.org/10.1016/j.promfg.2017.09.194.

Santos, D., Rebelo, M. and Santos, G., 2017. The Integration of certified Management Systems. Case Study – Organizations located at the district of Braga, Portugal. *Procédia Manufacturing,* [e-journal] 13, pp.964-971. https://doi.org/10.1016/j.promfg.2017.09.168.

Santos, G., Afonseca, J., Lopes, N., Félix, M. J., and Murmura, F., 2018. Critical success factors in the management of ideas as an essential component of innovation and business excellence. *International Journal of Quality and Service Sciences*, [e-journal] 3(3), pp.214-232. https://doi.org/10.1108/IJQSS-05-2017-0051.

Santos, G. and Barbosa, J., 2006. QUALIFOUND - a modular tool developed for Quality Improvement in Foundries. *Journal of Manufacturing Technology Management*, [e-journal] 17(3), pp.351-362. https://doi.org/10.1108/17410380610648308.

Santos, G. Bravi L. and Murmura, F., 2018. Fabrication laboratories: The development of new business models with new digital technologies. *Journal of Manufacturing Technology Management,* [e-journal] 29(8), pp.1332-1357. https://doi.org/10.1108/JMTM-03-2018-0072.

Santos, G. and Milán, A.L., 2013. Motivation and benefits of implementation and certification according ISO 9001 – The Portuguese experience. *International Journal for Quality Research,* [e-journal] 7(1), pp.71-86. https://doi.org/10.4314/ijest.v6i5.1.

Santos. G., Murmura, F. and Bravi, L., 2019. Developing a model of vendor rating to manage quality in the supply chain. *International Journal of Quality and Service Sciences*, [e-journal] 11(1), pp.34-52. https://doi.org/10.1108/IJQSS-06-2017-0058.

Stamatis, D., 2003. *Six Sigma and beyond: statistics and probability.* Boca Raton: St. Lucie Press (CRC Press LLC).

Thomas, A., Barton, R. and Chuke-Okafor, C., 2009. Applying lean Six Sigma in a small engineering company - A model for change. *Journal of Manufacturing Technology Management*, [e-journal] 20(1), pp.113-129. https://doi.org/10.1108/17410380910925433.

Tjahjono, B., Ball, P., Vitanov, V., Scorzafave, C., Nogueira, J., Calleja, J., Minguet, M., Narasimha, L., Rivas, A., Srivastava, A., Srivastava, S. and Yadav, A., 2010. Six Sigma: a literature review. *International Journal of Lean Six Sigma*, [e-journal] 1(3), pp.216-233. http://doi.org/10.1108/20401461011075017.

Wadsworth, H., Stephens, K. and Godfrey, A., 2007. *Modern Methods for Quality Control and Improvement.* Chichester: Wiley.

Ward, S., Poling, S. and Clipp, P., 2008. Selecting Successful Six Sigma Projects. *Quality*, 47(10), pp.50-51.

Zgodavová, K., Bober, P., Sütöová, A., Lengyelová, K., 2019. Supporting sustainable entrepreneurship in injection molding of plastic parts by optimizing material consumption. *Przemysł Chemiczny,* [e-journal] 98(3), pp.1001-1009. http://doi.org/10.15199/62.2019.3.9.

ABOUT AUTHORS

Ana Rita Costa – is a BSc in Food Engineering and is now a student carrying out a one-year training period in industry as a partial fulfilment of the Master Degree in Food Engineering. She is applying statistical process control technics included in lean/Six Sigma projects. E-mail: anaritarcosta@hotmail.com, Author's ORCID: https://orcid.org/0000-0002-1756-7789.

Carla Barbosa – is Prof. at Polytechnic Institute of Viana do Castelo (IPVC), Viana do Castelo, Portugal. She is a member of LAQV/REQUIMTE and CISAS (investigation centres accredited by FCT – the Portuguese Foundation for

Science and Technology), holds a MSc (Food Engineering), and a PhD (Nutrition and Food Chemistry). Her current research interests include Quality Management and Quality Tools, Statistical Process Control and Product Development. Special focus on physical properties and their relationships to Consumer Acceptabilities and Preferences. E-mail: cbarbosa@estg.ipvc.pt, Author's ORCID: https://orcid.org/0000-0002-8042-8642.

Gilberto Santos – is Prof. at Polytechnic Institute Cávado Ave (IPCA) Barcelos, Portugal. He holds a PhD degree in Mechanical Engineering from Minho University, Portugal. His current research interests include industrial and mechanical engineering, namely, innovation, quality systems and integrated management systems. Gilberto participates as a speaker in several national and international conferences, and he is currently author of several publications in integrated management systems (IMS) – QES (quality, environment and safety) (books, papers in international and national journals). He is member of the research team of several IMS – QES research projects and supervises Master and PhD thesis, as well as reviewer of international journals. He founded a master's course on "Integrated Management Systems QES" at the College of Technology of Polytechnic Institute Cávado, where he was Director of the MSc during the first six editions. E-mail: gsantos@ipca.pt, Author's ORCID: https://orcid.org/0000-0001-9268-3272.

M. Rui Alves – is this paper's corresponding author. He is Full Prof. at Polytechnic Institute of Viana do Castelo (IPVC), Viana do Castelo, Portugal. He is a member of LAQV/REQUIMTE and CISAS (investigation centres accredited by FCT – the Portuguese Foundation for Science and Technology), holds an MPhil (Food Science & Technology), a PhD (Nutrition and Food Chemistry) and a Post-Graduation (Six-Sigma Black Belt). His current research interests include Quality Management and Quality Tools, Statistical Process Control, Product Development and Consumer Studies. Special focus on R based friendly algorithms for non-statisticians. E-mail: mruialves@estg.ipvc.pt, Author's ORCID: http://orcid.org/0000-0003-1844-485X.