Comparison of individual charts to monitor peroxide index of olive oil

Luís M. Grilo, and Helena L. Grilo

***Abstract* —** The peroxide index is an important variable measured in order to assess the quality of olive oil. However, the normality assumption of its distribution fails the test in this case study. The control limits of individual chart for peroxide index of virgin olive oil are obtained considering the average and also the median moving range. A non-parametric method based on empirical quantiles, quite robust against deviations from normality, is also applied with the bootstrap procedure to estimate control limits for the individual chart. The conclusions obtained with these approaches are relatively different.

***Keywords* —** Bootstrap, Empirical quantiles, Non-normality, Non-parametric method.

I. Introduction

Olive oil is a food product with growing importance in health due to its nutritional richness and therapeutic properties. The variable "peroxide index (mEq O_2/Kg) of virgin olive oil" (*PI*), which can be affected by several factors (such as climate, soil and latitude), measures its initial oxidation and subsequent deterioration to assess the quality as well as to characterize the type of oil (e.g., virgin, extra virgin). Control charts have been used effectively for years to monitor processes and to detect abnormal behaviors so, in this study, we obtain a Shewhart individual (*X*) chart in order to analyse central location and to monitor eventual changes in the process. The empirical distribution of *PI* could not be considered symmetric nor mesokurtic and the normality assumption is violated, which may seriously affect the performance of the corresponding control chart, in what regards the false alarm rate and the time required to detect the occurred changes ([1]-[3]). According to [4] and [5] these type of charts are a robust tool and work well under non-normality. Despite the controversy, we obtain and compare *X* chart for *PI* computing the control limits based on the average moving range (AMR) and on the median moving range (MMR), as in [6] and [7]. Some simulation studies show the excellent performance of the empirical quantile (EQ) control charts, which are a special case of the bootstrapping control charts, for a broad range of (non)normal distributions. Thus, with individual observations stem from a process which is statistically in-control, we also estimate robust control limits based on EQ to make the *X* control chart more sensitive to persistent assignable causes ([8] and [9]).

In Section II we analyse the sample data with some descriptive statistics and test the normality of the variable *PI*. In Section III we introduce three methods used to compute the control limits for the *X* chart. In Section IV we obtain and compare the statistical performance of control limits for the *X* chart considering different approaches. In Section V we present some final considerations and conclusions.

This work was partially supported by the Fundação para a Ciência e a Tecnologia (Portuguese Foundation for Science and Technology) through the project UID/MAT/00297/2013 (Centro de Matemática e Aplicações - CMA).

L. M. Grilo is with the Unidade Departamental de Matemática e Física, Instituto Politécnico de Tomar, Quinta do Contador – Estrada da Serra, 2300-313 Tomar, and Centro de Matemática e Aplicações (CMA), FCT, UNL, Portugal (+351 249 328 166; e-mail: lgrilo@ipt.pt).

H. L. Grilo is with the Centro de Sondagens e Estudos Estatísticos, Instituto Politécnico de Tomar, Quinta do Contador – Estrada da Serra, 2300-313 Tomar, Portugal. (e-mail: helenagrilo56@gmail.com).

II. Data Analysis

The individual measurements of *PI* of one olive oil supplier of a Portuguese brand leader, obtained in laboratory, are given in the Table I.

Table I: The 29 individual values (observations) of *PI*.

No. of observ.	***PI***	**No. of observ.**	***PI***	**No. of observ.**	***PI***	**No. of observ.**	***PI***
1	6.6	**9**	6.9	**17**	6.4	**25**	8.4
2	6.5	**10**	6.8	**18**	8.5	**26**	8.4
3	6.4	**11**	6.8	**19**	8.3	**27**	7.9
4	9.2	**12**	7.4	**20**	7.5	**28**	6.9
5	9.2	**13**	7.0	**21**	7.6	**29**	7.7
6	6.6	**14**	8.6	**22**	7.2		
7	6.6	**15**	6.5	**23**	9.0		
8	6.3	**16**	6.4	**24**	8.1		

Based on a small data sample (size n = 29) we compute some descriptive statistics which are presented in Table II. The location measures: mean, trimmed mean and median are close; only the mode is relatively smaller (almost one unit). The variation coefficient, i.e. the ratio between the standard deviation and the mean, is approximately 12.6%, which allows us to consider a relatively low dispersion (also the small value of standard error of mean indicates a good stability or a small sampling error). The shape of the empirical distribution (see also Fig. 1) is slight skewed to the right and platykurtic (the skewness and kurtosis coefficients have positive and negative signals, respectively). Furthermore, we do not identify outliers.

Table II: Results of some descriptive statistics of *PI*.

PI	**Statistic**	**Std. Error**
Mean	7.438	0.174
5% Trimmed Mean	7.402	
Median	7.200	
Mode[a]	6.400	
Std. Deviation	0.936	
Minimum	6.300	
Maximum	9.200	
Skewness	0.524	0.434
Kurtosis	-1.065	0.845

[a] Multiple modes exist. The smallest value is shown.

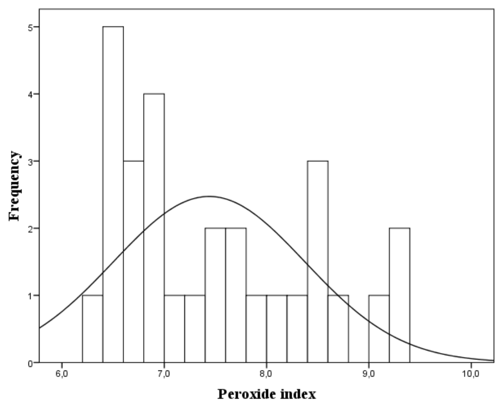

Fig. 1: Frequency histogram of *PI* with normal curve.

The results of the Shapiro-Wilk normality test with original data set (Table III) led us to reject the null hypothesis of normality (p - value = 0.011 < 0.05), for a significance level of 5%. However, the decision will be different if we consider a significance level of 1%.

Table III: Results of the normality Shapiro-Wilk test of *PI*.

PI	**Shapiro-Wilk**		
	Statistic	df	p - value
original data set	0.902	29	0.011

III. Control Limits Methods

To monitor *PI* for shifts that modify its mean, we use the X control chart based on the AMR and on the MMR. As an alternative we also obtain and compare non-parametric control chart based on EQ, which use the bootstrap method ([9]). To construct this control chart we consider the central line as the average value and we have to obtain both a lower control limit (LCL) and an upper control limit (UCL). When an individual measurement of *PI* falls outside of these control limits the process is called out-of-control.

Here we only consider a X control chart without transforming the data (instead of using Box-Cox transformations as we did in [7] and [10]).

When the cumulative distribution function (c.d.f.) F is assumed to be normal/gaussian (in this case commonly represented by Φ), with mean μ and standard deviation σ, the control limits of the Shewhart X chart are,

$$\begin{aligned} UCL &= \mu + \Phi^{-1}(1-\tfrac{\alpha}{2})\sigma, \\ LCL &= \mu - \Phi^{-1}(\tfrac{\alpha}{2})\sigma, \end{aligned} \tag{1}$$

where Φ^{-1} is the standard normal quantile function and the α level represents the false alarm rate. The parameters μ and σ in (1) are usually unknown because in practice the normality assumption is questionable. However, if an independent and identically distributed (i.i.d.) random sample (X_1, X_2, …, X_n) is available, we can estimate these parameters using the classical estimators, which are respectively, the sample mean

$$\hat{\mu} = \bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$$

and the sample standard deviation

$$\hat{\sigma} = S' = \sqrt{\frac{1}{n-1}\sum_{i=1}^{n}(X_i - \bar{X})^2}.$$

The sample standard deviation, S', is asymptotically efficient for an i.i.d. normal random sample, but it is also sensitive to trends and oscillations, which is a disadvantages. An estimator less sensitive to these deviations is AMR, defined by

$$\overline{MR} = \frac{1}{n-1}\sum_{i=2}^{n}\left|X_i - X_{i-1}\right|, \tag{2}$$

which can be scaled by $d_2(2) = 2/\sqrt{\pi}$, in order to obtain an unbiased estimator for σ under normality, i.e.

$$\hat{\sigma} = \frac{\overline{MR}}{d_2(2)} = \frac{1}{d_2(2)}\frac{\sum_{i=2}^{n}\left|X_i - X_{i-1}\right|}{n-1} = \frac{\sqrt{\pi}}{2}\overline{MR}. \tag{3}$$

Control limits based on Average Moving Range

The control limits based on the AMR, in (2) and (3), for the traditional X control charts are ([11]),

$$\begin{aligned} UCL_{AMR} &= \bar{X} + \Phi^{-1}(1-\tfrac{\alpha}{2})\tfrac{\sqrt{\pi}}{2}\overline{MR}, \\ LCL_{AMR} &= \bar{X} - \Phi^{-1}(\tfrac{\alpha}{2})\tfrac{\sqrt{\pi}}{2}\overline{MR}. \end{aligned} \tag{4}$$

For an α level of 0.0027 we obtain Φ^{-1} (1 - α/2) = 3 and considering $d_2(2) = 2/\sqrt{\pi} \approx 1.128$ we can rewrite the control limits in (4) as

$$\begin{aligned} UCL_{AMR} &= \bar{X} + 3\frac{\overline{MR}_n}{d_2(2)} \approx \bar{X} + 2.66\overline{MR}, \\ LCL_{AMR} &= \bar{X} - 3\frac{\overline{MR}_n}{d_2(2)} \approx \bar{X} - 2.66\overline{MR}. \end{aligned} \tag{5}$$

The AMR control charts tend to perform reasonably well for moderate sample sizes, more or less independently of the observations probability distribution ([4]-[6]).

ISBN: 978-1-61804-360-3

Control limits based on Median Moving Range

According to [4] and [6], when some large moving ranges have inflated the AMR, we may use the MMR as an alternative way to compute more robust limits for these control charts. This approach allows the computation of narrower limits, because they are less severely inflated by the large ranges. The control limits based on the MMR ([4], [6] and [7]) are

$$\begin{aligned} UCL_{MMR} &= \bar{X} + 3.145MR, \\ LCL_{MMR} &= \bar{X} - 3.145MR. \end{aligned} \tag{6}$$

Control limits based on Empirical Quantile

The control limits of the EQ for X chart will be defined according to [9], where a natural estimator of the q-quantile of the unimodal unknown c.d.f. F is the empirical quantile $F^{-1}(q)$, which is defined as

$$F^{-1}(q) = \inf\{x \mid F(x) \geq q\}, \quad 0 < q < 1$$

where F is the empirical c.d.f. that puts mass $1/k$ at each X_i, $1 \leq i \leq k$, i.e.

$$F(x) = \frac{1}{k}\sum_{i=1}^{n} I_{\{X_i \leq x\}}, \quad -\infty < x < +\infty,$$

where I represent the indicator function, i.e. $I_{\{x \leq y\}}$ equals 1 if $x \leq y$ and 0 otherwise. Thus, the obvious estimators of the upper and lower control limits based on the EQ are

$$\begin{aligned} UCL_{EQ} &= \hat{F}^{-1}(1-\tfrac{\alpha}{2}) = X_{\lceil (1-\frac{\alpha}{2})n \rceil}, \\ LCL_{EQ} &= \hat{F}^{-1}(\tfrac{\alpha}{2}) = X_{\lfloor \frac{\alpha}{2}n+1 \rfloor}. \end{aligned} \tag{7}$$

Here $X_{(1)} \leq X_{(2)} \leq \cdots \leq X_{(n)}$ denotes the order statistics of the initial sample $X_1, X_2, \ldots, X_n$, and $\lceil \cdot \rceil$ denotes the ceiling of the argument, that is, the smallest integer not less than the argument, and $\lfloor \cdot \rfloor$ denotes the floor of the argument, that is, the largest integer that does not exceed the argument.

Here we should note that the non-parametric control charts become attractive if large data sets are available, i.e. we need at least 1,000 observations in order to attain reasonably performance. Nevertheless, this may be overcome with the bootstrap approach which is a computational intensive technique based on the philosophy that the unknown c.d.f. F of a random variable will be replaced by an empirical c.d.f. F. Thus, we apply the bootstrap procedure to obtain, for example, $UCL_{EQ} = \hat{F}_n^{-1}(1-\frac{\alpha}{2})$ as an estimate of the $UCL_{EQ} = F_n^{-1}(1-\frac{\alpha}{2})$ in control charts for individual observations ([9]).

IV. Comparison of Control Limits for X Charts

In Fig. 2 an X control chart is drawn (i.e. we put the data on a process behaviour chart) to create a picture of how the process is running over time which helps to separate the routine variation from the exceptional variation. The bilateral control chart for PI displays the individual measurements and the estimates of LCL_{AMR} and UCL_{AMR} (also in Table V).

Although no obvious patterns were identified in the plot, some of the first observations present an unusual behaviour of PI and there are three points that violate the control rules (Fig. 2 and Table IV). The eight consecutive points on one side of the average line could be interpreted as a potential signal of some change in the process.

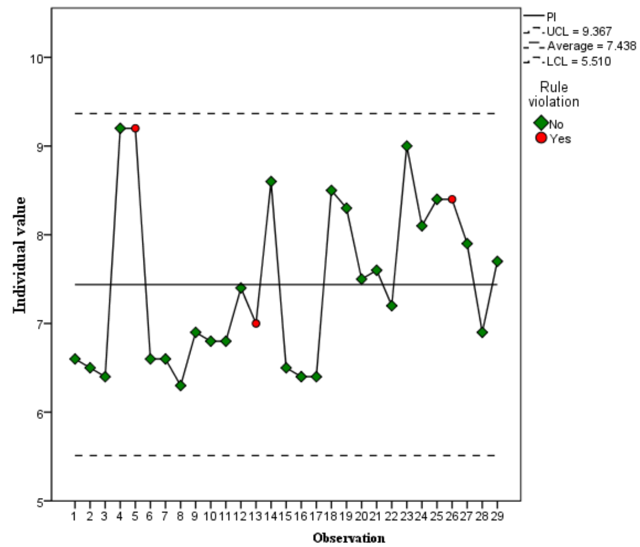

Fig. 2: Individual control chart of PI.

Table IV: Rule violations for run of PI.

Rule Violations for Run	
Case Number	**Violations for Points**
5	2 points out of the last 3 above +2 sigma
13	8 consecutive points below the centre line
26	4 points out of the last 5 above +1 sigma

3 points violate control rules.

In Table V we have the estimates of the control limits (LCL and UCL) for X chart of PI, considering the usual α level of 0.0027.

To calculate the control limits with AMR method we use (5), the mean value in Table II and $\overline{MR} = 0.725$.

For control limits obtained with MMR method we use (6), mean value in Table II and $MR = 0.4$.

For the X chart based on EQ method we simply draw 1,000 bootstrapped samples, with the same size as the data sample, with replacement from a population made up of the data sample. Then, we determine the mean of each sample, which create the sampling distribution of the mean, and using (7) we obtain the control limits.

Table V: Control limits for individual charts of PI.

Control Limits			
Method	***LCL***	***UCL***	**Limits range**
AMR	5.510	9.367	3.857
MMR	6.180	8.696	2.516
EQ	6.938	7.921	0.983

To compare the control limits of the three methods we take into account the limits range, where the smallest limit range corresponds to the EQ control limits. These control limits could be superimposed in Fig. 2 to visualize how narrower the different limits are. The MMR method produce narrower limits than AMR method, because they are less severely inflated by the large ranges and consequently the number of individual values outside the limits increase. For the EQ control limits, more sensitive to small shifts in the mean, we have a considerable number of individual values that violate the control rules.

The different results obtained here show how important is to use robust methods under non-normal distributions, given that in this case we also have a small sample size. In result, the empirical quantile control chart is preferable in order to attain reasonable performance and it also performs well under the normality of the observations ([9]). This non-parametric method, that use bootstrap procedure, has also the advantage of being easy to compute (for example using Excel) and distribution-free when we have an in-control situation.

V. Conclusion

In this case study all the original measurements of peroxide index of olive oil are within the legal specification limits for human consumption. Nevertheless, these results are considered very interesting by the members of the enterprise (typically well-trained in chemical engineering), once they allow to evaluate the variable peroxide index of olive oil itself and to compare its behaviour among other brand suppliers.

The different precision of the control limits obtained, with the three statistical methods (AMR, MMR and EQ), also show how important is to take into account the deviations from normality, the presence of some large moving ranges and the sample size.

The approaches used here are relatively simple and easy to interpret, constituting an important tool to help non-statisticians to judge whether or not a process is considered statistically in-control.

References

[1] W. A. Jensen, L. A. Jones-Farmer, C. H. Champ, and W. H. Woodall, "Effects of Parameter Estimation on Control Chart Properties: A Literature Review". *Journal of Quality Technology*, vol. 38, no. 4, pp. 349-364, 2006.

[2] C. M. Borror, D. C. Montgomery, and G. C. Runger, "Robustness of the EWMA control chart to non-normality". *Journal of Quality Technology,* vol. 31, no. 3, pp. 309-316, 1999.

[3] Z. G. Stoumbos, and M. R. Jr. Reynolds, "Robustness to no-normality and autocorrelation of individuals control charts". *Journal of Statistical Computation and Simulation*, vol. 66, pp. 145-187, 2000.

[4] D. J. Wheeler, *Advanced Topics in Statistical Process Control*, SPC Press, Knoxville, Tennessee, 1995.

[5] D. J. Wheeler. (2009). Do you have leptokurtophobia?, *Quality Digest*. Available: http://www.qualitydigest.com.

[6] D. J. Wheeler. (2010). Individuals Charts Done Right and Wrong, *Quality Digest*. Available: http://www.qualitydigest.com.

[7] L. M. Grilo, and H. L. Grilo, "Individual and moving range control charts in the production of olive oil". *12th International Conference of Numerical Analysis and Applied Mathematics* (*ICNAAM 2014*), AIP Conf. Proc. 1648, 840011-1–840011-4; 2015; doi: 10.1063/1.4913051.

[8] T. R. Willemain, and G. C. Runger, "Designing control charts using an empirical reference distribution". *Journal of Quality Technology*, vol. 28, no. 1, pp. 31-38, 1996.

[9] M. B. Vermaat, R. J. M. M. Does, and C. A. J. Klaassen, "A comparison of Shewhart individuals control charts based on normal, non-parametric, and extreme-value theory". *Quality and Reliability Engineering International*, vol. 19, no. 4, pp. 337-353, 2003.

[10] L. M. Grilo, D. M. R. Mateus, A. C. Alves, and H. L. Grilo, Robust control charts in industrial production of olive oil. *10th International Conference of Computational Methods in Sciences and Engineering* (*ICCMSE 2014*), AIP Conf. Proc. Vol. 1618, pp. 539-542, 2014, doi: 10.1063/1.4897794.

[11] A. J. Duncan, *Quality Control and Industrial Statistics*, 5th ed. Irwin Homewood, Illinois, 1986.