Journal of Asian Scientific Research, 2014, 4(4):174-181

Journal of Asian Scientific Research

journal homepage: http://aessweb.com/journal-detail.php?id=5003

OUTLIERS DETECTION USING CONTROL CHARTS FOR OIL WELLS

Daniel Francisco Maranhão Evangelista

Depart of Statistics and Actuarial Science – Federal University of Sergipe

José Augusto Andrade Filho

Program of Pos-Graduate in Intellectual Property Science – Federal University of Sergipe

Glaucio José Couri Machado

Program of Pos-Graduate in Intellectual Property Science – Federal University of Sergipe

Gabriel Francisco da Silva

Program of Pos-Graduate in Intellectual Property Science – Federal University of Sergipe

Suzana Leitão Russo

Program of Post-Graduate in Intellectual Property Science – Federal University of Sergipe

ABSTRACT

The presence of moderate values in a normal population is more likely than the presence of extreme values. Within this context, the assumption of normality of any population is due to the high probability of data to be normally distributed [1, 2]. The definition of outliers is subject to analysis and interpretation of results. Decisions regarding the identification of outliers should be taken individually and depend on a specific experiment [1]. Control charts are records of observations in statistical process built in a Cartesian coordinate system. The measurements obtained are represented in a time/space order and compared with the control limits. If any measurement exceeds the control limits, the process is considered to be out of bounds of statistical control and the value identified is defined as an outlier. Thus, this work aims at identifying outliers by control charts using data from drilling oil wells in order to improve the generation of synthetic sonic profile. This work was supported by FAPITEC and CNPq.

Keywords: Quality control, Control charts, Outliers, Sonic profile, Wells.

1. INTRODUCTION

Currently, oil is the fossil fuel with the highest economic importance worldwide and a major industrial propellant. Therefore, its value is rising, and presenting a gradual increase in the value of oil barrels over the years. It is important to note that the demand for oil deposits is increasing and it is accompanied by concerns to decrease extraction costs [3].

The sonic profile in oil wells, was a study initiated in the 50's, with the objective of assisting the seismic prospect; then, it became extensively used for studies on the porosity of rocks crossed by the well [4]

The sonic profile data uses the variables DT, PROF, DCALI, and the outliers present in each stratigraphic level of the well 2 will be identified by means of Box-Plot graphics. Later, apply Shewhart Control Charts to identify outliers in the measurements performed during the drilling of oil wells. Thus, defining limits which will indicate whether the data is "under control". If there is data out of the defined bounds, it will be considered as an anomaly, i.e., a possible outlier. After identifying the outliers, they are removed

2. LITERATURE REVIEW

2.1. Profile Sonic

The sonic profile is typically used on the early drills of each area of interest, in order to collect data and evaluate the hydrocarbon potential of the area. Besides, due to the high cost, the use of probes in all drilled wells would be prohibitive [3].

The seismogram (graphic records) is based on the concept that the sound propagation rate in a rock is given by the inverse of the time taken for a wave to go through a rock for a given distance measure. The acoustic impedance curve is obtained multiplying the density by the sonic speed, with this curve, we find the reflectivity coefficient that will merge with the signature of a seismic source and, finally, generate a synthetic seismogram [4].

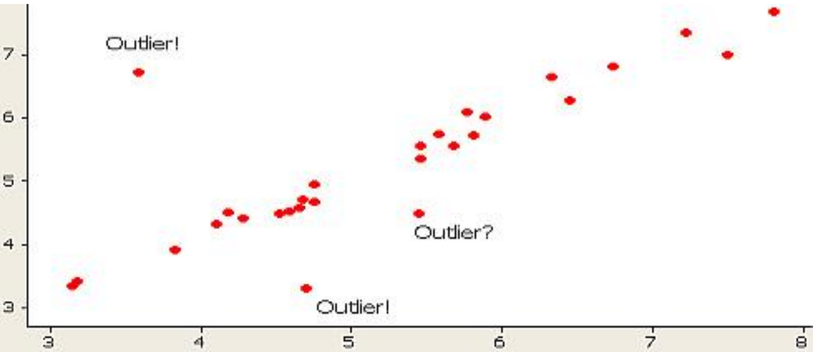
2.2. Outliers

In most databases analyzed, it is often found structural changes (outliers) that may be associated with unexpected events, as measurement and sample record errors. These changes are known as anomalies, distortions, aberrations and can interfere in the results of the statistical analyzes of the data [5].

In order to start working on data analysis, the first step is to observe the data behavior. Thus, analyzing disproportionate values, the possible outliers. These can be easily identified, but there are many exceptions due to anomalies. When expressed in a graph, for instance, outliers can occur in many different settings, from simple data collection and tabulation errors, or due to existing phenomena from which data are collected [6].

Figure 1 shows two samples easily defined as outliers and a possible one, the problem in cases like this is how define if the data can be classified as an outlier or not.

Figure-1. Example of outlier.



In sonic profile data, as in most cases, there is also outliers, and as stated before it is a given outlier or not we should go check the possible causes of its appearance, this case can be read error probe or malformation of the well at that point due.

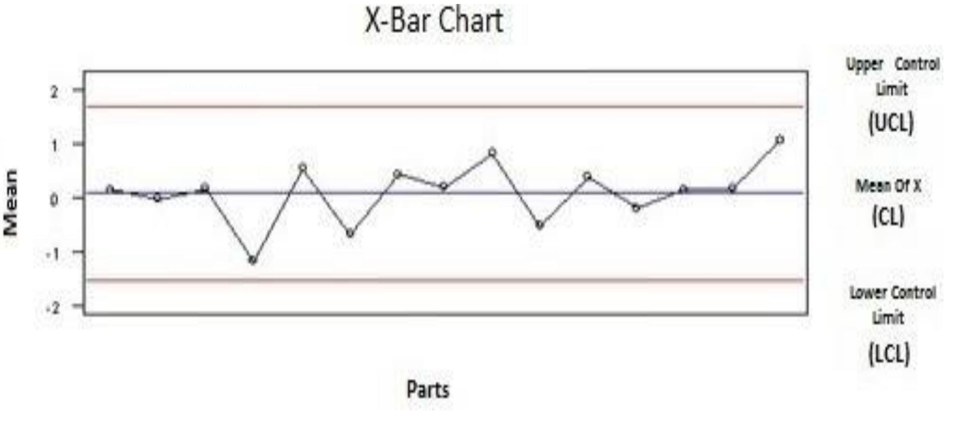
2.3. Quality Control

Quality control is an old method. It was born with the industry in order to improve the quality of products. In the beginning, the control was performed in a casual way, with the traditional visual inspection of the product status. Quality control is a procedure of high relevance in the industry, with a high demand to the company, in order to improve and ensure the production quality in economical levels, to meet their consumers requests [7].

To Montgomery [8], quality has always been a complementary part products and services. However, the awareness of its importance made the quality control has evolve and it continues to evolve over the years.

Statistical quality control is nothing more than the use of statistical tools for quality control of the process [8]. A control chart is one of the statistical tools used.

Figure-2. Example of a control chart.



This type of graph (Figure 2) plots the average variable sample. It presents a center line (CL) that shows where the variable had to be if there was no variation in the process. It also shows two boundary lines, upper control (LSC) and lower control limit (UCL), which define how far the process variability can reach. If some points exceed these limits, the process is be considered "out of control". It is clear that the control graph will not show how to put the process under control. It just shows if any action is needed [8].

Graphs X-S are preferred over the graphs X-R, when n > 10 or 12, since, for larger samples the amplitude sample R loses efficiency to estimate σ, noting that one of the requirements to implement them would be normal data [4].

Expressions for calculating the control limits of X charts - S:

Chart $\bar{X}$:

$$LSC = \bar{\bar{x}} + \frac{3\bar{s}}{c_4\sqrt{n}} = \bar{\bar{x}} + A_3\bar{s},$$

$$LM = \bar{\bar{x}},$$

$$LIC = \bar{\bar{x}} - \frac{3\bar{s}}{c_4\sqrt{n}} = \bar{\bar{x}} - A_3\bar{s}$$

where $A_3 = \frac{3}{c_4\sqrt{n}}$ is a constant tabulated as a function of sample size n.

Chart *S:*

$$LSC = \bar{s} + 3\hat{\sigma}_s = B_4\bar{s},$$

$$LM = \bar{s},$$

$$LIC = \bar{s} - 3\hat{\sigma}_s = B_3\bar{s}$$

where $\hat{\sigma}_s$ is the estimated standard deviation of the S distribution; and B3 and B4 are constants tabulated according to the sample size n [4].

2.4. Box-Cox Transformation

Having identified the lack of data normality with the Jarque-Bera test, we use the Box-Cox transformation to obtain a normal distribution. The equation in its simplified form is:

$$W = X_i^{\lambda}$$

The value of λ is chosen between 3 and -3. The idea is simple: there must be some value of λ that transforms the original variable from not Noraml to Normal. The mathematically complete Box-Cox transformation follows the equation:

$$X_i(\lambda) = \frac{X_i^{\lambda} - 1}{\lambda}$$

In practical terms, the equation is the same, but there is a different case when λ is very close to zero. Both transformation are essentially the same when λ is not close to zero and this means that the first transformation, being simpler, should be preferred. However, when λ is close to zero, the more complete transformation is preferred because the zero value means that the correct transformation would the logarithm (ln(X_i)) [9].

3. METHODOLOGY

The variable used in this study is one of several used during drilling an oil well. These series were collected at Petrobras, then put in a spreadsheet. Finally, assembled according to a careful plan; moreover, it was registered in order to plot a graph for better viewing.

The variables used in this study are: DT, PROF, DCALI [10]

- DT: Measures the time required for a sound wave to travel one foot of rock, this time is called transit time and it is inversely proportional to the sonic velocity of the rock. It is used to estimate the porosity; correlate wells; estimate the compression degree of rocks; estimate the elastic constant of rock; fracture detection; and support for seismic (synthetic seismogram). It is measured in microseconds per foot.
- PROF: It is the depth of the reading point within the well. Its measurement scale is in meters, and the reference measurement, ie, zero, is the Rotary Table (MR) of the drilling rig, which is usually less than 10 meters. To correct the Topographic Level simply subtract the MR. The depth is a variable that in order to be used properly requires knowledge of the geology of the area, i.e., be familiar with the different stratigraphic levels. The correlation of MR with other variable profiles requires geological interpretation. In general, the sonic velocity of the rock (1/DT) and density (RHOB) increase with depth due to burial and subsequent compression of the rocks.
- DCALI: It is a computed variable (CALI - DB), it measures how much the diameter of the well deviates from the nominal diameter of the drill bit that pierced a particular stretch. This value is called the "break-in of the well". When this variable presents a high frequency, it causes "roughness" on the walls of the wells and these irregularities affect the quality of the profile readings, especially the profiles running with contact shields with the walls of the pits, as the Neutron and Density profiles. This variable is measured in inches.

4. RESULTS

4.1. Descriptive Analysis

Table 1 shows a summary of the variable DT data in an oil well, in each stratigraphic level.

Table-1. Summary Description

Level	Mean	Median	Standard Error	Standard Deviation	Variance	Asymmetry	Minimum Value	Maximum Value	C.V%
1C	104,745	104,300	0,148	16,075	258,421	0,424	58,143	190,385	15,3
2M	86,582	88,539	0,671	18,940	358,737	0,079	53,214	127,899	21,9
3P	79,334	78,750	0,075	8,301	68,900	0,061	50,415	146,535	10,5
4B1	79,221	77,165	0,072	11,233	126,173	1,577	56,230	181,329	14,2
4B2	70,550	69,154	0,080	8,322	69,261	1,250	49,300	132,712	11,8
5S	66,160	65,825	0,145	4,278	18,301	0,913	53,563	86,425	6,5
6B	69,317	69,247	0,191	1,877	3,523	-0,295	63,406	73,816	2,7

Based on the stratigraphic levels, we decided to work with the level 4B2 data. Levels above the 4B2 present many distorted values, which can contaminate the analysis. On the other hand, levels below the 4B2 are more stable, what can be seen in the standard deviation and variance coefficient, which already shows a variation lower than the average in a downward trend.

Figure 3 presents the Box-Plot graph for each stratigraphic level, what shows that there is a higher occurrence of outliers in the level 4B2, implying that there is greater variability in reading the transit time (TD).

Figure-3. Box-Plot Graph

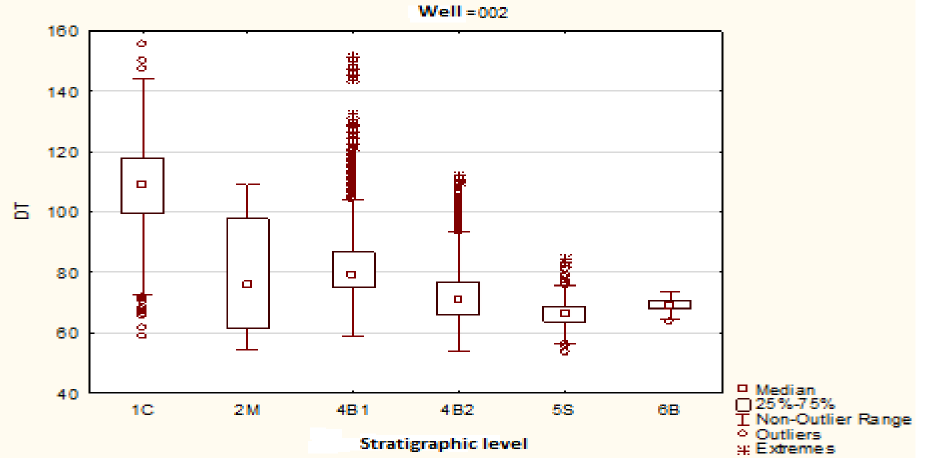
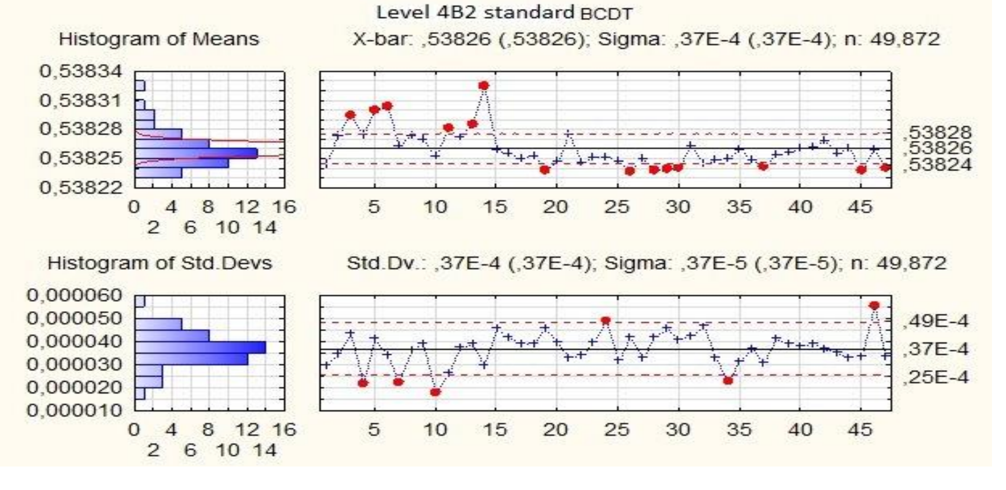


Figure 3 also shows a larger amplitude in the 1C and 2M levels, due to a high DCALI index value. The amplitudes are smaller in the 5S and 6B levels, since the well starts to taper as the DCALI approaches the actual size of the drill.

4.2. Control Charts

The Control Charts technique was applied for the level 4B2. Figure 4 shows several points outside the control limits.

Figure-4. Control Charts level 4B2.



Analyzing Figure 4, five points were considered possible outliers, as they exceed the upper and lower limits (highlighted in red), and 30% of samples in the $\bar{X}$ Chart and 13% in the S Chart were out of control (Table 2).

Table-2. Summary Chart $\bar{X}$ and Chart S

Outliers Chart $\bar{X}$	**n**	**%**	**Outliers Chart S**	**n**	**%**
Below of LCL	8	17,02	Below of LCL	4	8,51
Above of UCL	6	12,76	Above of UCL	2	4,25
Total	14	29,79	Total	6	12,76

It is also known that most samples are concentrated below the central and lower limit. The reason for this is that as the pit deepens, the variable DCALI gets closer to the actual size of the drill. One explanation may be that the mud flows down when upper levels collapses, causing a decrease in both the DT and the DCALI.

All these points are possible candidates to break the well or probe reading errors. What is not desired, because such a large number of outliers will bias features such as the average diameter and the type of rock on the well. Therefore, generating distorted information.

4.3. Revised Charts

The Revised Chart technique was applied to the variable DT data at level 4B2. Figure 6 shows that there was only one point (sample number 29) outside the bounds after applying the Revised Charts, however, this sample was not representative,since 97% of the samples were under control.

Figure-5. Chart reviewed level 4B2.

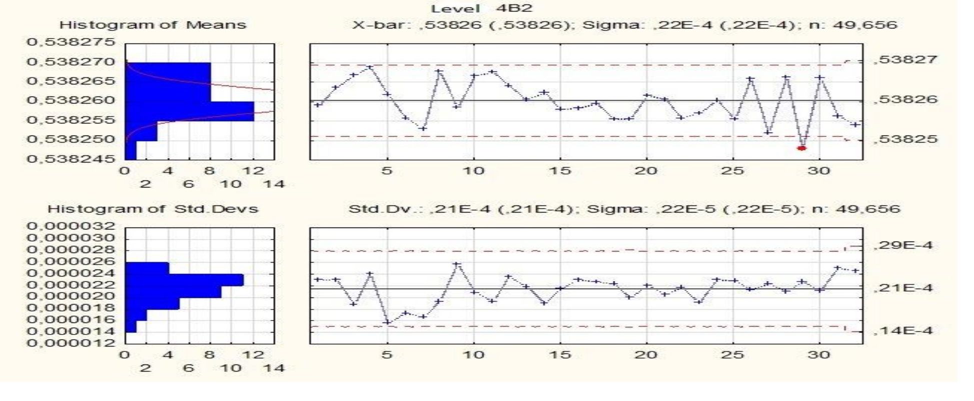


Table 3 shows the new control limits for the variable DT at Level 4B2 for the chart reviewed, which were obtained by means of the Box-Cox method in order to generate a new equation that returns the value which we were initially working.

Table-3. Summary Chart $\bar{X}$ and Chart S

Control Limits Chart $\bar{X}$		**Control Limits Chart S**	
LCL	69,18126	LCL	1,0001
$\bar{\bar{X}}$	70,89630	$\bar{\bar{X}}$	1,0002
UCL	72,73868	UCL	1,0001

5. CONCLUSION

The use of Control Charts in business are vital in order to achieve better quality levels in the production processes.

Industries are taking higher regard on what concerns the monitoring of the production process, since the market constantly demands products and services of the highest quality.

Control charts in the DT variable was applied. Several points out of the bounds of control limit were found. After detecting and removing outliers, a revised control chart was generated. Outliers may have been generated by errors in probe reading, by any irregularities in the well, among others problems.

Finally, new limits were defined on the Revised Chart, keeping the process under control in order to generate a new sonic profile for the oil well.

6. ACKNOWLEDGEMENTS

The work was supported by FAPITEC and CNPq.

REFERENCES

[1] R. A. Jhonson and D. W. Wichern, *Applied multivariate statistical analysis,* 5th ed. New Jersey: Pearson, 2002.

[2] M. R. Spiegel, *Theory and problems of probability and statistics*. New York: McGraw-Hill, 1992.

[3] E. S. Chagas, S. L. Russo, and V. H. Simon, *Geração de Perfil Sônico Sintético em Poços de Petróleo Através dos Modelos de Regressão Não Lineares Usando a Profundidade como Variável Regressora*. São Cristovão: Scintia Plena, 2010.

[4] S. L. Russo, M. S. Prata, and V. H. Simon, "Aplicação do controle estatístico de qualidade na variável DT do processo de perfilagem de um poço de petróleo," *INGEPRO,* 2010.

[5] C. F. Miranda, "Modelação Linear de Series Temporais na Presença de Outliers," *Tese (Mestrado em Estatística) – Departamento de Matemática Aplicada, Universidade do Porto,* 2001.

[6] M. M. C. Figueira, "Identificação de outliers," *MILLENIUM nº12,* 1998.

[7] F. R. C. B. Lourenco, *Controle estatístico de qualidade*. Rio De Janeiro: LTC, 1981.

[8] D. C. Montgomery, "Introdução ao controle estatístico de qualidade," *LTC,* 2009.

[9] R. W. Samohyl, *Controle estatístico de qualidade*. Rio De Janeiro: Editora Elsevier, 2009.

[10] D. A. Silva, M. J. S. Neto, V. H. Simon, and S. L. Russo, "Estudo do perfil sônico sintético: Caracterização da variáve lDT," *Natal. ERMAC: Encontro Regional de MatemáticaAplicada e ComputacionaL,* 2008.