Portuguese Study Groups' Reports

Report on "Understanding correlations between raw materials inputs"

Problem presented by PRIOBIOCOMBUSTÍVEIS at the
127th European Study Group with Industry
8th – 12th May 2017
University of Aveiro
Aveiro
Portugal

September 22, 2017

Problem presented by: Sónia Henriques (Prio Biocombustveis)

Study group contributors: Gnord Maypaokha, Helena Alvelos, Joaquim Correia, Marina A. P. Andrade, Mijiddorj Renchin-Ochir, Xaivang Xiongshieng

Report prepared by: Helena Alvelos (e-mail: helena.alvelos@ua.pt)
and
Marina A. P. Andrade (e-mail: marina.andrade@iscte.pt)

1 Introduction

1.1 Challenge and aims

Prio Biocombustveis is a company that has a biodiesel plant in Porto de Aveiro. The production capacity of this plant is 113 880 ton/year. The plant has one lab to control the Biodiesel and Raw Materials quality. Every day Prio improves its process to assure the production of the best world-class biodiesel that any costumer can buy.

Prio's raw materials include several types of oils used in different percentages in the production process. Oils quality has the denomination of characteristic X, which varies with the different types of oils. Even in the same batch of the same oil it's possible to observe quality oscillations.

Oil's neutralization is a dynamic process. The process recipe and operational parameters determine the final quality of the oil. All production shifts collect the operational parameters and analyze the quality in the lab for both intermediate and final oils. A diagram showing the main process operations is presented in Figure 1.

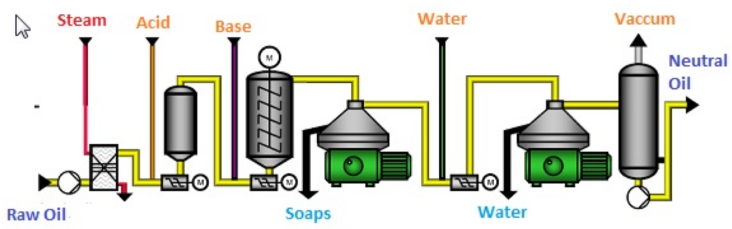

Figure 1: Neutralization Process.

The raw oil is heated to achieve the required process temperature and treated with acid and base using agitation. The produced soaps are separated from the Neutral Oil using centrifuges, which requires the control of several operational parameters. In neutralization process the main reaction is:

$$\text{Oil Mix (Raw Oil)} + \text{Base} \rightarrow \text{Neutral Oil} + \text{Soaps}.$$

The aim of this challenge is to understand correlations between raw materials inputs (in particular characteristic X) that define the recipe and highly influence the operational parameters that allow obtaining a high quality neutral oil (outputs).

1.2 Methodology

In a first phase, the process was detailed and the database made available by Prio was analysed. It contained data from 1-Jan-2016 until 30-Apr-2017, which consisted in 1334 cases and 32 variables. It was found that the combination of some raw materials were different, depending on the time of the year, so the database was split in two, according to the referred combination. The subsequent phases of the methodology, described next, were then applied to both sets of data. The second phase consisted in the descriptive analyses of each variable separately, using descriptive statistics, histograms and box plots (Ott and Longnecker, 2015; Guimarães and Cabral, 2011). It was possible to remove from the database some cases that were identified as outliers, and to correct some values of the variables that were identified as erroneous. The variables were grouped as Input Variables (6 for each database set), Output Variables (4) and Process Variables (18). There were also the variables related to the date and time of the collection of the oil sample analysed.

The third phase of this study consisted in the analysis of the correlations between all the variables in the databases. The correlation matrix was calculated and the values were analyzed (Ott and Longnecker, 2015; Guimarães and Cabral, 2011; Hair et al., 2009). There were identified three groups of pairs of variables, according to the values of the correlations. There were also performed some factor analyses, one with all the variables and others with part of the variables that it was thought could be more appropriate for the use of this multivariate technique, but the results were discarded as the assumptions related to the use of factor analysis were not satisfied at all (Hair et al., 2009).

The next step consisted in trying to relate one of the input variables with some others using a multiple linear regression, as well as to relate each output variable with the input variables, using the same technique. The statistical analysis were performed using the IBM SPSS Statistics 24 package and Excel 2016.

1.3 Report's structure

This report has five main sections. The first one is the present introduction, followed by the specification of the problem, where it is included the definition of the variables considered in the study and the databases used. The third section refers to the exploratory data analysis and the forth one addresses the inferential statistical approach to the problem. Finally, in section 5, the main conclusions and some recommendations are presented.

2 Specification of the problem

The main raw material of the oil neutralization process (raw oil) is a mix of oils that will be called (for confidential reasons) Oil A, Oil B and Oil C. In fact, the mixture can only be of two types: (Oil A + Oil C) or (Oil B + Oil C), depending on the time of the year. The quantity of the each oil in the mixture is expressed in percentage, and is represented by QA, QB and QC, respectively. The quality of the oils (input, intermediate and output) is assessed, among others, by a chemical parameter which will be called X and is expressed in percentage. It is measured in the raw materials Oil A – XMPA, Oil B – XMPB and Oil C – XMPC and also in the mixture after agitation – XM. The variable X is also estimated by a company responsible that has a lot of experience with the process and takes into account the percentage of each type of oil in the mixture, (QA, QB, or QC) the values of X measured for each of the input oils (XMPA, XMPB or XMPC) and the result of the XM measurement. The value of the estimated X will be represented by XMix A+C or XMix B+C depending on the type of the mixed oils, or Xlinha, when there is no doubt which oils are being studied.

The variables just mentioned are considered the input variables of the study that, together with the process (or control) variables are going to influence the output. There are 17 process variables that are of very different nature and so are measured in different measurement units. They can be parameters from machines, like pressure or temperatures or characteristics of some products added in the process like acid or base. For confidential reasons, those variables are not going to be specified and will be represented by V1, V2, . . . , V18. In fact, the variables were initially 18 but variable 16 had such a lower number of observations such that it was decided not to consider it.

Finally, the output variables considered in this study were the X characteristic measured in the final product the neutral oil (XFinal), two more characteristics measured in the final step of the process Yfinal and ZFinal, and another one measured during the process, which is ZIntermediate or ZInterm.

In the next sections the variables just described are going to be analyzed and related to each other. It should be emphasized that the original database was divided in two, one containing the information for the cases where Oil A and Oil C were used, and another for the cases of Oil B and Oil C.

3 Exploratory Analysis

In this section the descriptive statistics, histograms and box-plots for the two database sets earlier referred are presented. In order to make it easier to follow, those are exhibited in separate subsections whether it is database for oils A and C or database for oils B and C. Also, since there are different types of variables along the process – input variables, output variables and control variables – the different cases are presented in different subsections.

3.1 Oil A and Oil C mixture

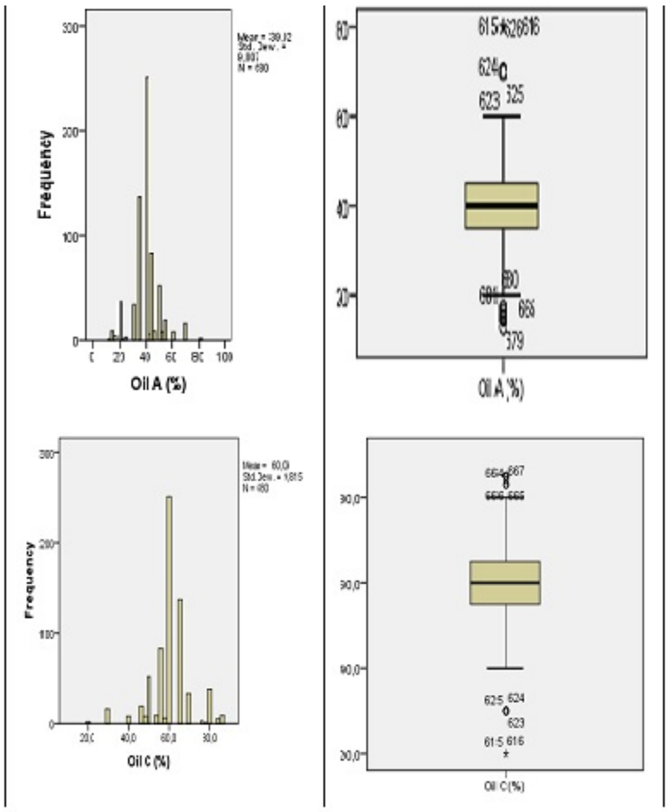

Figure 2: Histograms and box-plot for QA and QC.

After the histogram and boxplot of QA and QC variables, the descriptive statistics were computed and are presented in figure 3.

Descriptive Statistics							
	N	Minimum	Maximum	Mean	Std. Deviation	Skewness	
	Statistic	Statistic	Statistic	Statistic	Statistic	Statistic	Std. Error
Oil A (%)	680	13	80	39.92	9.807	0.334	0.094
X M.P. A (%)	680	0.338	1.919	1.05047	0.535923	0.419	0.094
Oil C (%)	680	20.0	85.0	60.087	9.8148	-0.336	0.094
X M.P. C (%)	675	2.390	4.675	3.42208	0.349251	0.036	0.094
X Oils A + B + C	681	1.430	3.658	2.48652	0.470477	-0.026	0.094
X' Mix A + B + C	681	1.380	3.500	2.47324	0.463735	-0.014	0.094
Z Intermédiate Product	681	239.44	1565.93	835.7206	194.61117	0.627	0.094
Z Final Product	680	28.43	864.85	176.3370	66.09917	3.271	0.094
Y Final Product	680	230.10	1289.40	604.7376	147.06144	1.113	0.094
X Final (%) Product	680	0.049	0.257	0.10543	0.028788	2.160	0.094
Valid N (listwise)	672						

Figure 3: Descriptive statistics for Oil A and Oil C input and output variables.

The histograms for the variables XMPA, XMPC, the mixture and the estimated function during the process were also computed and are presented in figure 4:

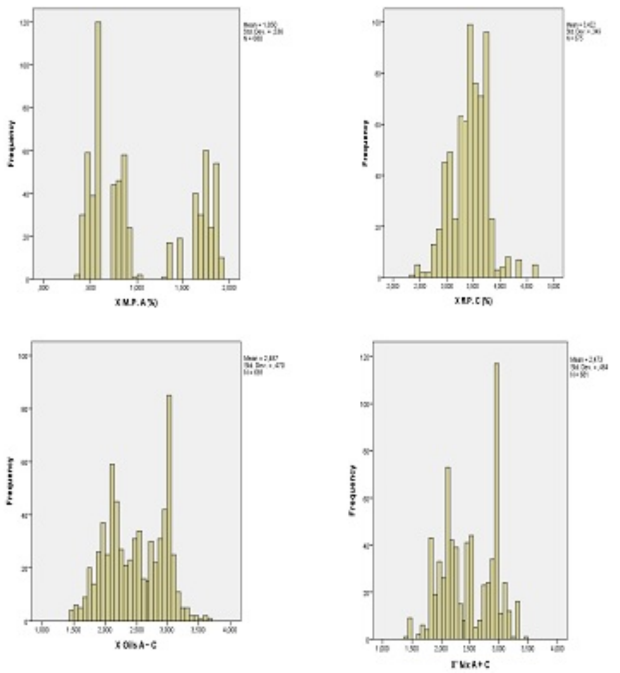

Figure 4: XMPA, XMPC, XM (A+C) and estimated function histograms.

The output variables histograms and box plots for the mixture oil A and oil C:

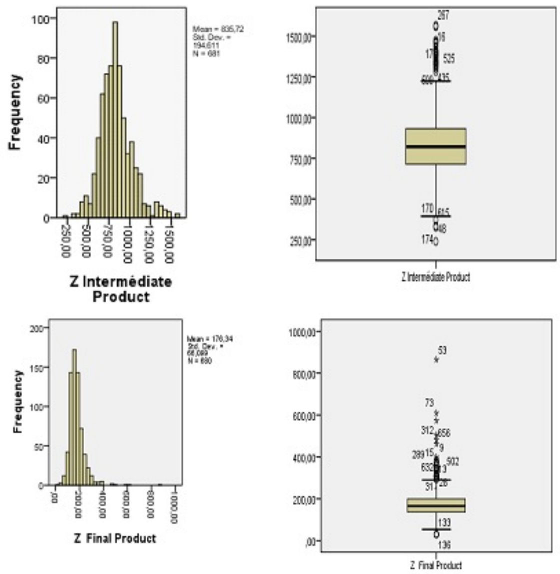

Figure 5: Output variables for Oil A plus Oil C.

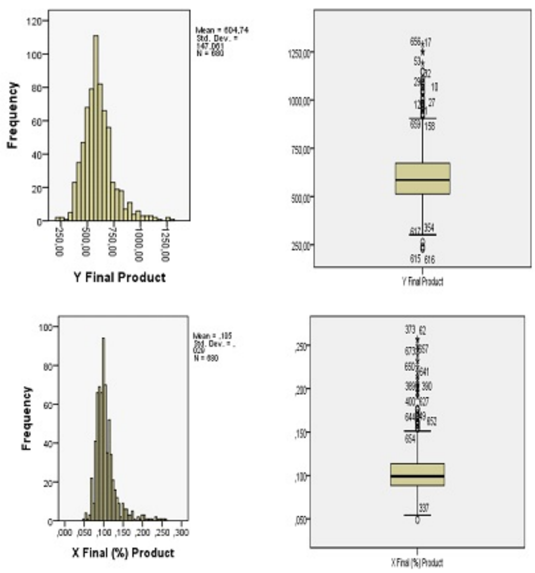

Figure 6: Output variables for Oil A plus Oil C.

As it was mentioned an important objective was to consider and measure the correlations between the variables present in the process. In that sense for each mixture the correlations were computed. For Oil A and Oil C those are presented in figure 7:

	QA (%)	XA (%)	QC (%)	XC (%)	XM (A+C)	Xlinha	Zinterm	Zfinal	Yfinal	Xfinal
QA (%)	1.000									
XA (%)	-0.211	1.000								
QC (%)	-0.999	0.210	1.000							
XC (%)	-0.107	0.586	0.108	1.000						
XM (A+C)	-0.631	0.771	0.631	0.638	1.000					
Xlinha	-0.602	0.782	0.602	0.673	0.953	1.000				
Zinterm	-0.289	0.330	0.290	0.149	0.322	0.321	1.000			
Zfinal	-0.269	0.071	0.268	0.007	0.120	0.106	0.570	1.000		
Yfinal	-0.308	-0.132	0.306	-0.149	-0.035	-0.052	0.341	0.635	1.000	
Xfinal	-0.129	0.025	0.129	0.070	0.138	0.047	0.104	0.174	0.129	1.000
V 1	0.120	-0.058	-0.119	-0.124	-0.169	-0.137	0.315	0.290	0.277	-0.054
V 2	-0.005	-0.213	0.004	-0.191	-0.110	-0.225	-0.063	-0.019	-0.022	-0.198
V 3	-0.110	0.082	0.110	-0.055	0.078	0.071	0.071	0.011	0.178	-0.075
V 4	-0.581	0.678	0.581	0.571	0.873	0.896	0.287	0.092	-0.037	0.001
V 5	-0.593	0.629	0.593	0.548	0.855	0.862	0.149	-0.021	-0.118	0.026
V 6	0.036	0.069	-0.037	-0.050	0.000	0.020	-0.047	-0.106	-0.026	-0.242
V 7	0.007	-0.409	-0.011	-0.248	-0.327	-0.304	-0.098	0.087	0.344	-0.214
V 8	-0.002	-0.209	0.007	-0.121	-0.141	-0.160	-0.198	-0.022	0.022	0.042
V 9	-0.031	-0.027	0.028	-0.079	-0.052	-0.038	-0.004	-0.010	0.096	-0.153
V 10	0.168	-0.189	-0.173	-0.171	-0.250	-0.228	-0.155	-0.089	0.072	-0.070
V 11	-0.073	-0.060	0.075	-0.002	0.014	0.006	-0.001	0.029	0.064	0.082
V 12	-0.201	-0.325	0.202	-0.389	-0.215	-0.254	0.059	0.220	0.408	0.125
V 13	-0.354	-0.231	0.350	-0.175	-0.053	-0.046	0.206	0.350	0.631	-0.035
V 14	-0.371	0.428	0.372	0.242	0.440	0.425	0.243	0.133	0.037	0.160
V 15	-0.159	0.153	0.170	0.069	0.188	0.166	0.088	0.074	0.023	0.098
V 17	0.025	-0.131	-0.025	-0.141	-0.147	-0.134	-0.106	-0.027	0.023	-0.085
V 18	-0.276	-0.277	0.274	-0.247	-0.149	-0.154	0.153	0.317	0.579	-0.028

Figure 7: Correlations between input and output variables and the intermediate ones (considering Oil A and Oil C).

Some of the correlations shown in figure 7 are highlighted according to their values. The ones that have absolute values between 0.9 and 1 are highlighted in red, between 0.7 and 0.9 in yellow and between 0.5 and 0.7 in green.

It can be observed that Xlinha is strongly correlated with XM(A+C). This fact means that the estimation made by the company responsible (Xlinha) is in line with the value that is measured in the process (XM(A+C)). Xlinha and XM(A+C) are also correlated with the percentage of the quantity of each oil in the mixture (QA and QC) and with their own values of X characteristic (XA and XC). Concerning the input and output variables, it can be observed that the correlations are weak, meaning that there is no linear relationship between those variables. Considering the process variables, V4 and V5 show important correlation values with all the input variables, the same happening between V13 and V18 and YFinal.

3.2 Oil B and Oil C mixture

Input variables histograms and box plot for mixture Oil B and Oil C:

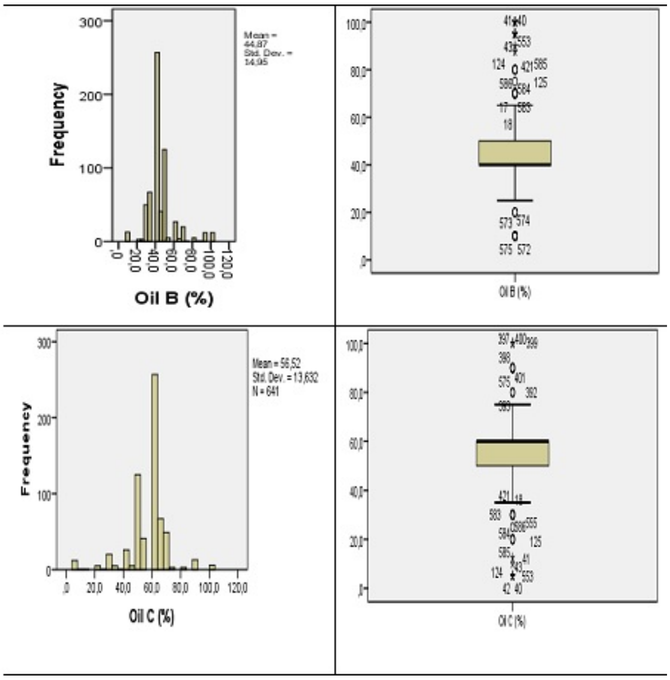

Figure 8: Histograms and box-plot for QB and QC.

The descriptive statistics for the mixture Oil B and Oil C:

Descriptive Statistics							
	N	Minimum	Maximum	Mean	Std. Deviation	Skewness	
	Statistic	Statistic	Statistic	Statistic	Statistic	Statistic	Std. Error
Oil B (%)	647	10.0	100.0	44.875	14.9500	1.637	0.096
X M.P. B (%)	647	0.330	1.579	0.86422	0.218220	0.272	0.096
Oil C (%)	641	5.0	100.0	56.523	13.6320	-0.916	0.097
X M.P. C (%)	639	2.672	4.305	3.18034	0.360425	0.963	0.097
X Oils B + C	653	0.379	3.500	2.16670	0.475297	-0.674	0.096
X' Mix B + C	653	0.390	3.170	2.18524	0.444662	-0.878	0.096
Z Intermediate Product	653	102.83	1874.16	819.7248	244.11576	0.171	0.096
Z Final Product	652	6.85	2100.57	189.2545	132.48151	7.707	0.096
Y Final Product	653	10.65	4419.60	707.9951	338.51041	2.977	0.096
X Final (%) Product	652	0.025	0.890	0.09946	0.046569	12.197	0.096

Figure 9: Descriptive statistics for Oil B and Oil C input and output variables..

Subsequently the same procedure, histograms and box plots for the database with the mixture of oil B and Oil C were computed and are presented in what follows

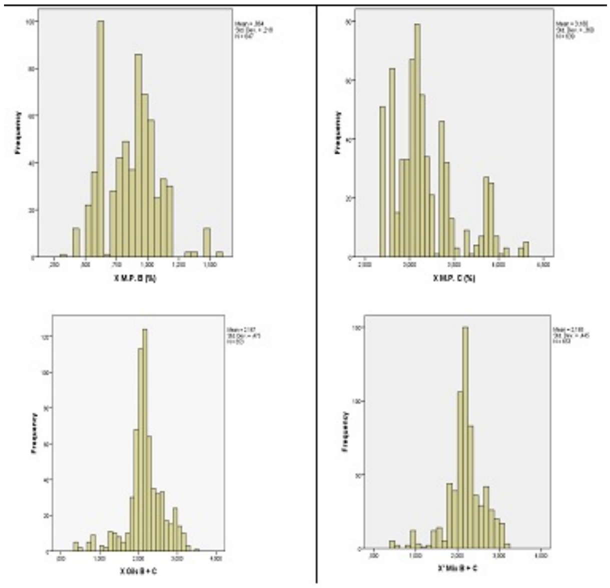

Figure 10: XMPB, XMPC, XM (B+C) and estimated function histograms.

The output variables histograms and box plots for the mixture oil B and oil C:

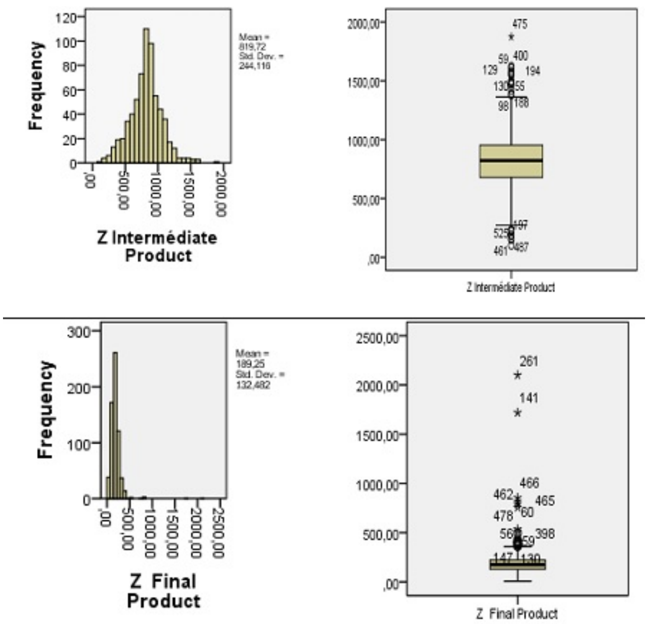

Figure 11: Output variables for Oil B plus Oil C.

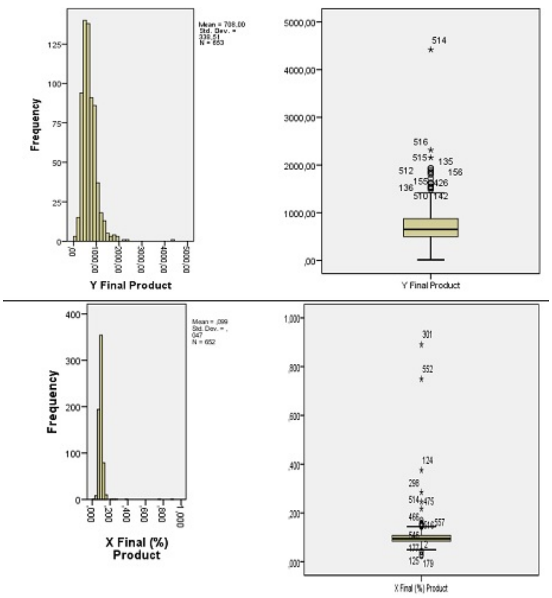

Figure 12: Output variables for Oil B plus Oil C.

For Oil A and Oil C the correlations were also computed and are presented in figure 13:

	QB(%)	XB(%)	QC(%)	XC(%)	XM(%)	Xlinha(%)	ZInterm	ZFinal	YFinal	XFinal
QB(%)	1									
XB(%)	-0.14662	1								
QC(%)	-0.9957	0.077676	1							
XC(%)	-0.08892	-0.0103	0.045174	1						
XM(%)	-0.79218	0.28628	0.731595	0.490858	1					
Xlinha(%)	-0.79167	0.271847	0.724185	0.539453	0.935021	1				
ZInterm	-0.52607	-0.03498	0.481796	-0.01538	0.434345	0.462939	1			
ZFinal	-0.25793	-0.06404	0.236016	-0.08155	0.20089	0.16496	0.414354	1		
YFinal	-0.28511	-0.03882	0.254899	-0.25176	0.084576	0.060807	0.342322	0.340702	1	
XFinal	0.010144	0.020685	0.08768	0.01884	0.083477	-0.01464	0.017694	0.000509	0.036517	1
V1	-0.0737	-0.1559	0.071088	-0.15829	-0.05252	0.009653	0.458343	0.28853	0.238691	-0.05867
V2	0.241654	-0.23797	-0.26411	-0.10759	-0.22241	-0.23011	0.124424	0.103971	0.047664	-0.17617
V3	-0.09868	0.158128	0.144351	-0.01087	0.140842	0.114713	0.097431	0.06826	0.10865	0.006809
V4	-0.61492	0.093765	0.531399	0.376809	0.73414	0.771499	0.502764	0.227471	0.122473	-0.05803
V5	-0.52683	0.161397	0.408576	0.45148	0.717087	0.731365	0.277541	0.081291	-0.01166	-0.05025
V6	-0.03662	0.117819	0.045589	0.165614	0.173632	0.172234	0.0658	0.018223	0.057403	-0.02313
V7	0.11741	-0.46138	-0.17598	-0.1953	-0.25195	-0.22509	0.206087	0.201072	0.181472	-0.08894
V8	-0.02659	0.036101	0.052917	-0.13068	-0.03579	-0.05311	-0.09201	0.010778	0.091327	0.0315
V9	0.063151	0.176754	-0.06866	-0.12448	-0.11459	-0.10609	0.117776	0.211505	0.262096	-0.10937
V10	-0.1244	0.156355	0.149043	-0.07349	0.102183	0.088785	0.153828	0.11922	0.283811	-0.00489
V11	-0.17647	-0.02716	0.204432	-0.22966	-0.00026	0.008366	0.049669	-0.01518	0.214522	0.026279
V12	-0.22329	-0.09218	0.271609	-0.58856	-0.11768	-0.14941	0.190041	0.21	0.395123	0.007118
V13	-0.23888	-0.2368	0.201088	-0.16199	0.05819	0.052083	0.368252	0.357009	0.706148	-0.03124
V14	0.03985	-0.03894	-0.05156	0.002373	-0.07345	-0.0619	-0.02812	0.023595	0.063038	-0.02346
V15	-0.0142	0.302864	0.013387	-0.19986	-0.05379	-0.05123	-0.00329	0.147319	0.139164	-0.01808
V17	0.170068	-0.01213	-0.18938	-0.05198	-0.14185	-0.12302	0.006886	0.158494	0.210944	-0.05837
V18	-0.36915	-0.17853	0.357997	-0.17163	0.165984	0.167288	0.366	0.243744	0.54542	0.012069

Figure 13: correlations considering Oil B and Oil C.

As happened with figure 7, some of the correlations shown in figure 13 are highlighted according to their values. The ones that have absolute values between 0.9 and 1 are highlighted in red, between 0.7 and 0.9 in yellow, between 0.5 and 0.7 in green, and between 0.3 and 0.5 in magenta.

It can be observed that, as with Oil A + Oil C mixture, Xlinha is strongly correlated with XM(B+C). This fact means that, again the estimation made by the company responsible (Xlinha) is in line with the value that is measured in the process (XM(B+C)). Xlinha and XM(B+C) are also correlated with the percentage of the quantity of each oil in the mixture (QB and QC) and with the value of X characteristic in oil C (XC), but not with XB, as happened in the previous case. Concerning the input and output variables, it can be observed that the correlations are weak, meaning that there is no linear relationship between those variables, except in the case of variables XC and V12. Considering the process variables, V4 and V5 show important correlation values with some of the input variables, the same happening between V13 and V18 and YFinal.

3.3 Process Variables

In what concerns the set of process variables, as it was already explained, it is large. Therefore, here are presented only four histograms and box plots to illustrate the variety of what was observed. The remaining histograms and box plots are presented in the appendix.

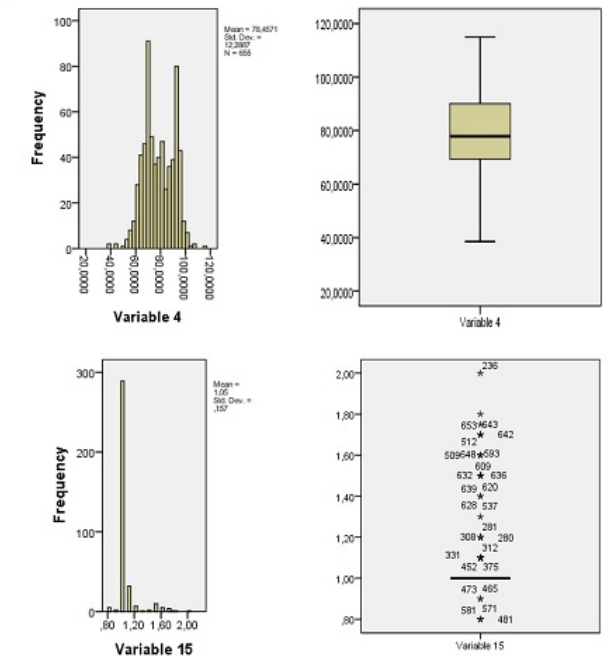

Figure 14: Process variables – Var 4 and Var 15.

As it can be observed variable 4 and variable 15 show a huge difference in terms of existence and absence of outliers. In the two other process variables here presented, figure 15, there can be also noticed important differences in the graphical representation below:

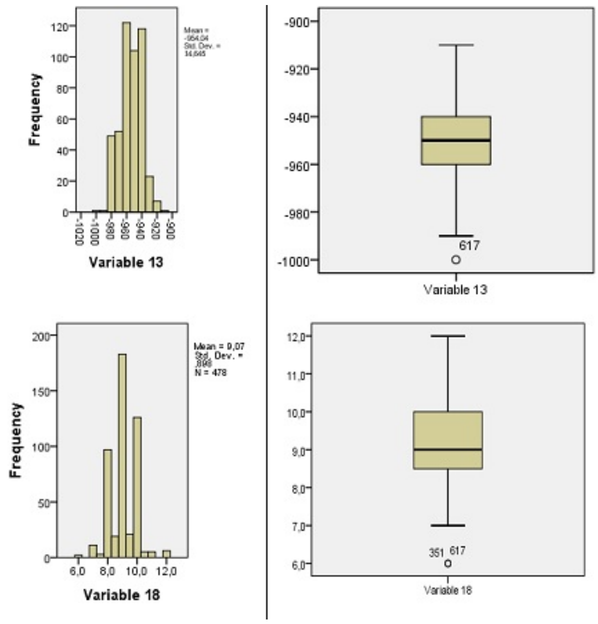

Figure 15: Process variables – Var 13 and Var 18.

The correlations between the process variables in the two database sets considered was another issue worth to consider. In the next figures, 7 and 13 , those correlations are presented:

	V1	V2	V3	V4	V5	V6	V7	V8	V9	V10	V11	V12	V13	V14	V15	V17	V18
V1	1																
V2	0.054964	1															
V3	-0.08467	-0.12376	1														
V4	-0.11946	-0.08722	0.347807	1													
V5	-0.30838	-0.09429	0.348143	0.953859	1												
V6	-0.14752	0.016848	0.754956	0.338452	0.355153	1											
V7	-0.08931	0.256439	0.232659	-0.2209	-0.1783	0.307445	1										
V8	0.077116	0.069916	-0.2694	-0.16066	-0.18801	-0.26048	-0.09045	1									
V9	-0.127	0.094908	0.048937	-0.06565	-0.00971	0.159396	0.26539	-0.33447	1								
V10	-0.02937	-0.05136	0.20689	-0.23456	-0.22644	0.182994	0.197299	0.076055	0.186925	1							
V11	0.063674	-0.02754	-0.11911	-0.0077	-0.03576	-0.21035	-0.10616	0.312891	-0.18604	0.130119	1						
V12	0.272002	0.262409	-0.16629	-0.21926	-0.26153	-0.41924	0.134646	0.203761	-0.06877	0.002627	0.232232	1					
V13	0.14242	0.064266	0.172036	-0.03261	-0.05878	0.07197	0.431439	-0.04935	0.11925	0.099614	0.022791	0.217726	1				
V14	-0.04617	-0.1547	-0.05433	0.379488	0.367952	-0.17351	-0.32828	-0.04579	0.110042	-0.18424	0.058823	0.100566	-0.0265	1			
V15	0.007434	-0.19776	-0.01946	0.142659	0.129624	-0.19134	-0.31611	-0.03115	-0.06617	-0.10477	0.140717	0.110712	-0.06183	0.83688	1		
V17	0.02904	0.028142	0.035613	-0.10886	-0.10684	0.069054	0.102791	-0.00156	0.050038	0.10373	-0.07856	0.000267	0.025151	0.085035	0.197887	1	
V18	0.186693	0.07126	0.128747	-0.14146	-0.18292	-0.00165	0.412383	0.067178	0.1903	0.136372	0.106014	0.258376	0.606696	-0.04339	-0.05267	0.013946	1

Figure 16: Process variables correlations for oil A and Oil C.

	V1	V2	V3	V4	V5	V6	V7	V8	V9	V10	V11	V12	V13	V14	V15	V17	V18
V1	1																
V2	0.270571	1															
V3	-0.15117	-0.15844	1														
V4	0.117902	0.060304	0.448908	1													
V5	-0.2825	-0.04615	0.485047	0.900838	1												
V6	-0.21629	-0.0789	0.796624	0.558277	0.594416	1											
V7	0.300655	0.533439	-0.03807	0.121289	-0.02538	0.043191	1										
V8	0.020593	-0.0737	-0.08775	-0.14288	-0.14496	-0.20639	-0.10105	1									
V9	0.24251	0.120982	0.27058	0.05365	-0.05721	0.190842	0.114185	0.060654	1								
V10	-0.07046	-0.11901	0.666438	0.351534	0.366887	0.56849	-0.0363	-0.01019	0.326087	1							
V11	0.004425	-0.08029	0.044251	0.047355	0.04472	-0.07251	-0.01701	0.256477	-0.08665	0.154398	1						
V12	0.230606	-0.01273	0.001014	-0.05618	-0.20027	-0.21581	0.178544	0.229004	0.125187	0.067378	0.417029	1					
V13	0.39903	0.126382	0.129824	0.245407	0.049555	0.085729	0.318881	0.006321	0.286616	0.279686	0.085384	0.28853	1				
V14	-0.01335	0.008453	0.011059	-0.04737	-0.04262	0.018626	0.022337	0.120858	0.062936	0.067144	-0.01152	-0.02597	0.086549	1			
V15	0.072787	-0.09957	0.079303	-0.11912	-0.16265	-0.036	-0.17765	0.195396	0.681894	0.124291	0.019197	0.231587	0.036474	0.008082	1		
V17	0.106231	0.163553	0.090279	0.043631	0.014354	0.125603	0.164993	0.120336	0.19929	0.287521	0.090907	-0.06113	0.259471	0.096042	0.040349	1	
V18	0.211753	0.05523	0.216893	0.303388	0.155104	0.166627	0.267833	0.07659	0.252014	0.269626	0.14409	0.272327	0.538715	0.036529	0.103107	0.056307	1

Figure 17: Process variables correlations for oil B and Oil C.

4 Inferential Statistical Approach

In addition to the exploratory descriptive analysis, for both database sets, two approaches were considered in inferential terms: linear regression and factorial analysis in order to reduce the number of variables. The preliminary results are presented below.

4.1 Linear Approach for the mixture oil A and Oil C

For the database formed by the mixture of oil A and oil C the results are presented in figure 18:

$X' = f(X_A, Q_A, X_C, X_M)$

Model Summary

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	.939[a]	.881	.881	.143452
2	.946[b]	.894	.894	.135597
3	.950[c]	.903	.902	.129911
4	.953[d]	.908	.907	.126736

a. Predictors: (Constant), X Oils A + B + C

b. Predictors: (Constant), X Oils A + B + C , X M.P. C (%)

c. Predictors: (Constant), X Oils A + B + C , X M.P. C (%), X M.P. A (%)

d. Predictors: (Constant), X Oils A + B + C , X M.P. C (%), X M.P. A (%), Oil A (%)

Coefficients[a]

Model		Unstandardized Coefficients B	Unstandardized Coefficients Std. Error	Standardized Coefficients Beta	t	Sig.
1	(Constant)	.130	.037		3.525	.000
	X Oils A + B + C	.944	.015	.939	64.764	.000
2	(Constant)	-.229	.056		-4.125	.000
	X Oils A + B + C	.818	.021	.814	39.883	.000
	X M.P. C (%)	.195	.024	.169	8.268	.000
3	(Constant)	-.025	.060		-.421	.674
	X Oils A + B + C	.675	.028	.671	24.092	.000
	X M.P. C (%)	.202	.023	.174	8.913	.000
	X M.P. A (%)	.123	.017	.167	7.173	.000
4	(Constant)	.336	.089		3.782	.000
	X Oils A + B + C	.566	.034	.563	16.687	.000
	X M.P. C (%)	.245	.023	.212	10.434	.000
	X M.P. A (%)	.168	.019	.228	8.994	.000
	Oil A (%)	-.007	.001	-.087	-5.437	.000

a. Dependent Variable: X' Mix A + B + C

Figure 18: Linear regressions (stepwise) – Oil A and Oil C.

As it can be noticed from figure 18 the best model e model 4 and can be written as:

$$Xlinha = 0.563 * XM(A+C) + 0.212 * XMPC + 0.228 * XMPA - 0.087QA.$$

4.2 Linear Approach for the mixture oil B and oil C

For the database formed by the mixture of oil B and oil C the results are presented in figure 19:

$$X' = f(X_B, Q_B, X_C, X_M)$$

Model Summary

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	.922[a]	.850	.850	.151911
2	.929[b]	.863	.862	.145604
3	.939[c]	.882	.881	.135288
4	.941[d]	.885	.884	.133664

a. Predictors: (Constant), X Oils B + C
b. Predictors: (Constant), X Oils B + C, X M.P. C (%)
c. Predictors: (Constant), X Oils B + C, X M.P. C (%), Oil B (%)
d. Predictors: (Constant), X Oils B + C, X M.P. C (%), Oil B (%), X M.P. B (%)

Coefficients[a]

Model		Unstandardized Coefficients B	Std. Error	Standardized Coefficients Beta	t	Sig.
1	(Constant)	.384	.031		12.370	.000
	X Oils B + C	.833	.014	.922	59.891	.000
2	(Constant)	.064	.052		1.231	.219
	X Oils B + C	.774	.015	.856	49.952	.000
	X M.P. C (%)	.141	.019	.129	7.539	.000
3	(Constant)	.504	.065		7.732	.000
	X Oils B + C	.568	.025	.628	22.630	.000
	X M.P. C (%)	.245	.020	.224	12.097	.000
	Oil B (%)	-.007	.001	-.241	-10.037	.000
4	(Constant)	.461	.065		7.076	.000
	X Oils B + C	.527	.027	.583	19.666	.000
	X M.P. C (%)	.268	.021	.245	12.886	.000
	Oil B (%)	-.008	.001	-.268	-10.879	.000
	X M.P. B (%)	.108	.027	.059	4.047	.000

a. Dependent Variable: X' Mix B + C

Figure 19: Linear regressions (stepwise) – Oil B and Oil C.

As it can be noticed from figure 19 the best model e model 4 and can be written as:

$$Xlinha = 0.583 * XM(B+C) + 0.245 * XMPC + 0.054 * XMPB - 0.268 * QB.$$

5 Conclusions and recommendations

The correlations that can explain the values of the final variables are:

- Zinterm is influenced by XA, XM and Xlinha;
- Yfinal is influenced by QA, QC, Zinterm an Zfinal (only Zinterm and Zfinal in the case of B+C);
- ZFinal is influenced by Zinterm.

Concerning XFinal, no correlations were found neither with input variables nor with the other output variables or control variables.

Some sugestions were considered like analyzing the relationship between Xlinha and V5 with XFinal, but no significant results were obtained. There are some correlations between control variables, that are somewhat different concerning the mix used (A+C or B+C).

Concerning the estimation of variable X, it is more correlated with the values X of the input oils and the X of the intermediate sample in the case of (A+C) than in the case of (A+B).

As future work it is suggested to analyze the concentrations of variables A, B and C separately, as well as some periods of time that correspond probably to different batches of raw materials. Additionally, other types of relationships (nonlinear) between the variables should be explored, considering eventual autocorrelations present the variables.

References

[1] Ott, R. L. and Longnecker, M., 2015. *An Introduction to Statistical Methods and Data Analysis (7th Ed.)*. (Stamford, Connecticut: Cengage Learning).

[2] Guimarães, R.C. and Cabral, J.S., 2011. *Estatstica (2nd Ed.)*. (Lisboa: Dashfer Holding Ltd. e Verlag Dashfer Lda).

[3] Hair, Joseph F., Black, William C., Babin, Barry J. and Anderson, Rolph E., 2009. *Multivariate Data Analysis (7th Ed.)*. (New Jersey, Prentice Hall).

Appendix

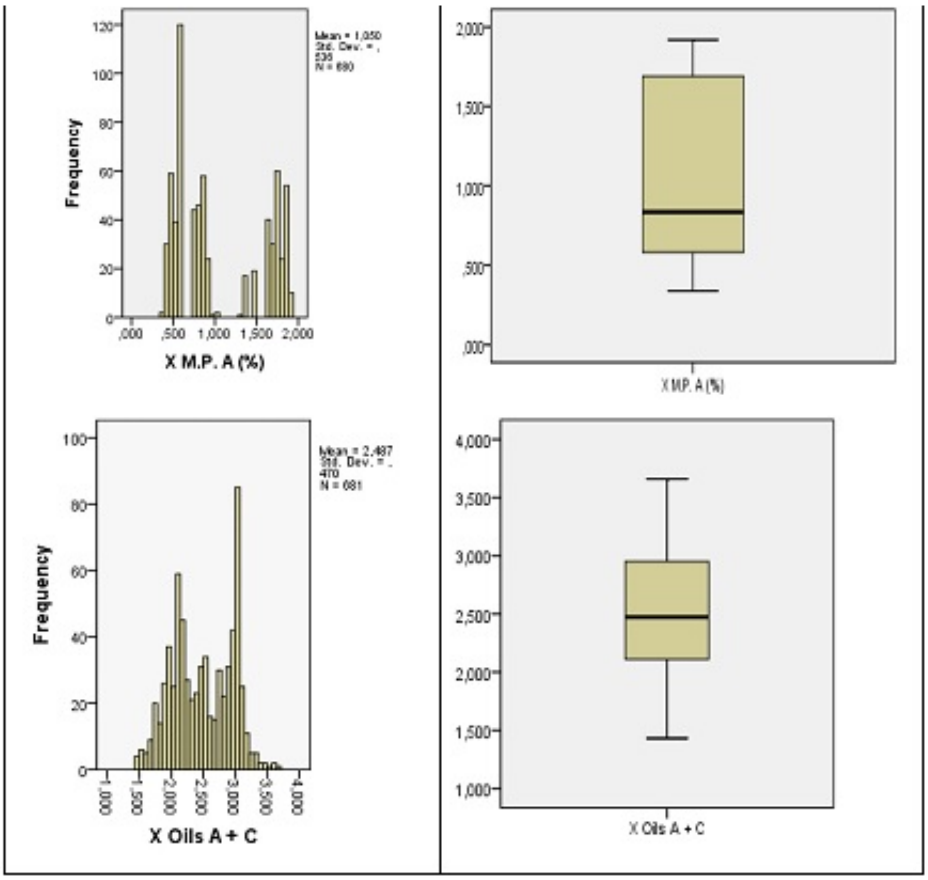

Figure 20: Input variables.

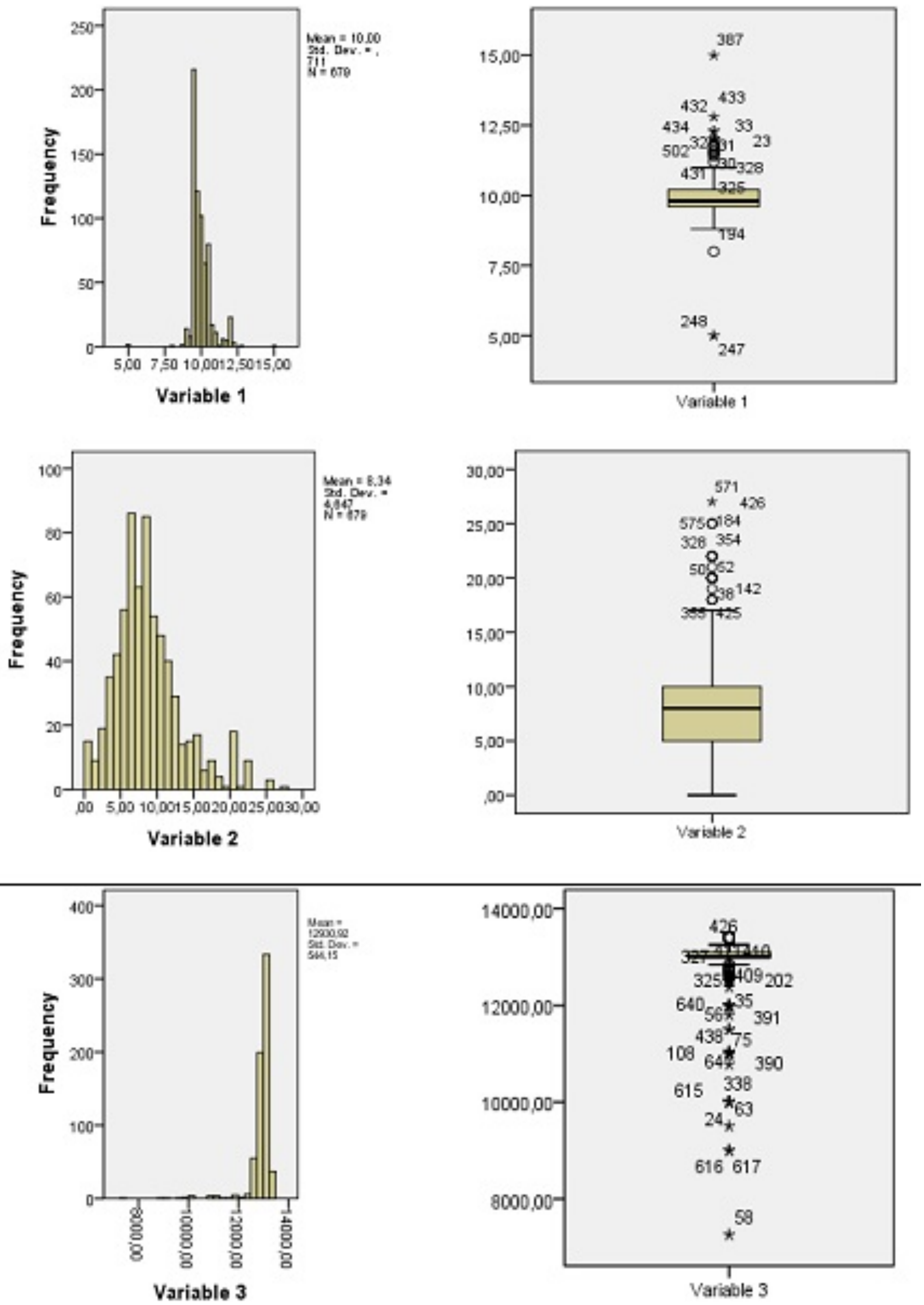

Figure 21: Process variables – Var 1, Var 2 and Var 3.

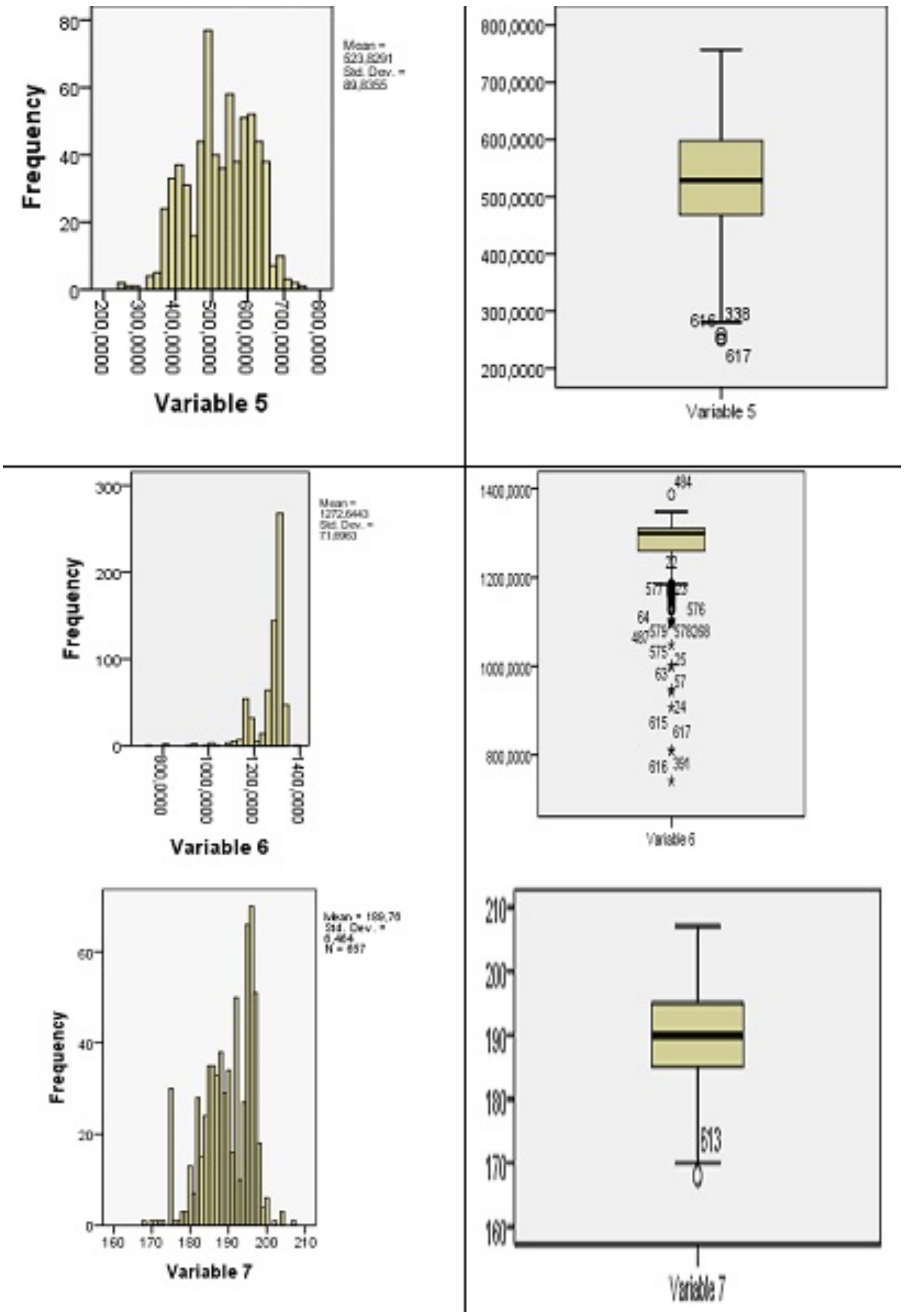

Figure 22: Process variables – Var 5, Var 6 and Var 7.

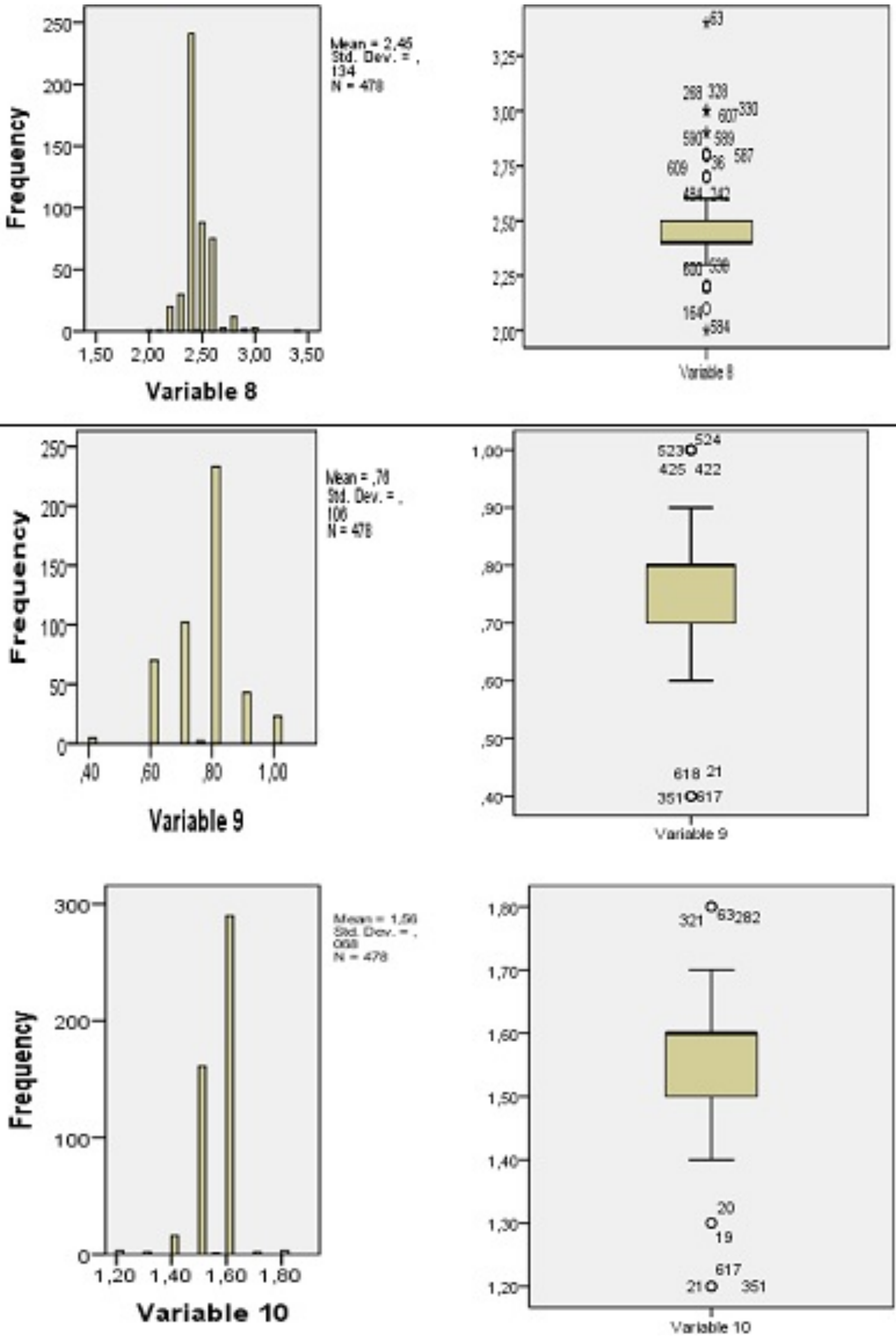

Figure 23: Process variables – Var 8, Var 9 and Var 10.

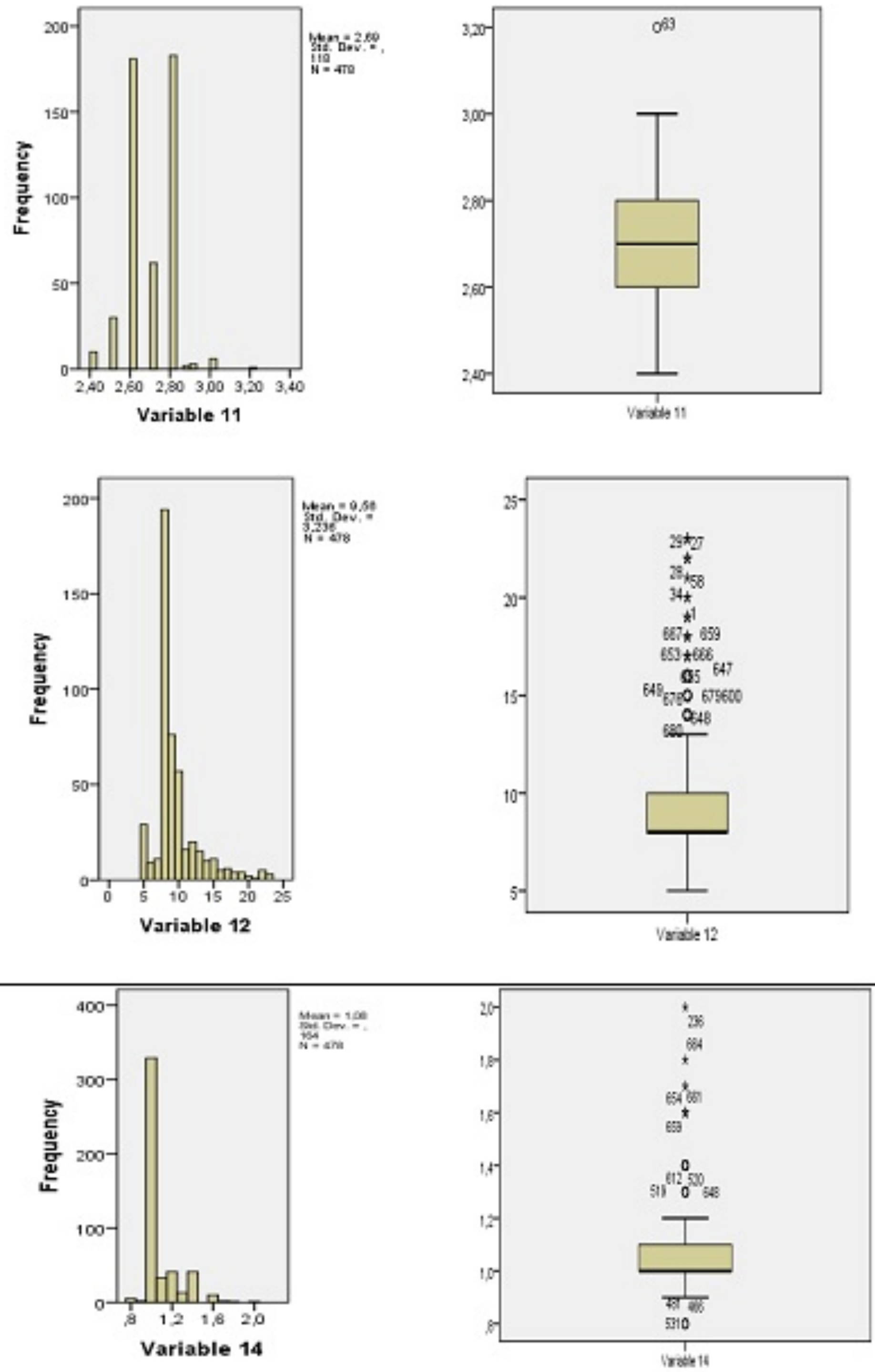

Figure 24: Process variables – Var 8, Var 9 and Var 10.