UNIVERSIDAD AUTÓNOMA DE BAJA CALIFORNIA

LATIN AMERICAN JOURNAL OF APPLIED ENGINEERING

Increasing Productivity by Principal Component Analysis: a Case Study

Incremento de la Productividad por Análisis de Components Principales: un Caso de Estudio

Julian Aguilar-Duque [a,*], Juan Hernández-Arellano [b], Jesus Salinas-Coronado [a], Guillermo Amaya-Parra [a]

[a]*Facultad de Ingeniería, Arquitectura y Diseño, Universidad Autónoma de Baja California, Ensenada, México*
[b]*Instituto de Arquitectura, Diseño y Arte, Universidad Autónoma de Ciudad Juárez, México*
**Corresponding author: Julian.aguilar@uabc.edu.mx*

***Abstract*—**The aim of this research is to expose the use of the technique of Principal Component Analysis to increase the productivity of production system whit a high level of downtimes. For the develop of the research, a manufacturer system integrated by 20 cells of manufacturing was considered under the restriction of the same conditions for each cell. Ten variables classified as a cause of the death times. The methodology implemented based on the develops of analysis of component principal to determine the effect of the principal variables on the system and their correlation. With the result achieved it was possible to reduce the number of variables from ten to five; they explain the 80% of the variation of the system. Finally, the implementation of lean manufacturing tools and the integration of the chain supply make an achievement of an increase in the productivity of 23%.

***Keywords*—**Manufacture, Principal Analysis Component, Productivity.

***Resumen*—**La presente investigación tiene como finalidad exponer el uso de la técnica de análisis de componentes principales para el incremento en la productividad de un sistema con un nivel de tiempos muertos elevados. Para el desarrollo de la investigación se realizó el análisis en un sistema de manufactura integrado por 20 células de manufactura con las mismas condiciones de operación y de productos. Se clasificaron 10 variables generadoras de tiempos muertos. La metodología implementada se basa en el desarrollo de análisis de componentes principales para determinar el efecto de las principales variables en la productividad del sistema y su correlación entre sí. Con los resultados obtenidos se logró reducir el número de variables a 5, las cuales explican en conjunto el 80% de la variabilidad del proceso. Finalmente se logró implementar el uso de herramientas lean y de cadena de suministros para incrementar el nivel de productividad en un 23%.

***Palabras Claves*—**Análisis de Componentes Principales, Productividad, Manufactura.

I. INTRODUCTION

Historicaly, the continuous improvement in the manufacturing systems has been the goal of many companies around the world. When manufacturer talks about subsisting, they refer to be the best in their area, it doesn't matter if they are the focus in the manufacturing process or the industry of services.

There is an special interesting in renew the methodology of been successfully manufacturing process, as it could be seen in the reconfiguration of chain supply across the understanding the attitudes of buy of costumers [1] or think about the human factor, as an analysis of multi-objective methodology and outline of a mechanism to analyze the relevance in the selection of economic projects [2].

Productivity generally express the relationship between the quantity of goods and services produced (output) and the quantity of labor, capital, land, energy and other resources to produce it (input) [3]. Before describe the importance of the productiveness and the success in the global competition is necessary to describe the principal elements of productivity. In 2001 Smith publicized the 5I's, considered the "key elements" that impact productivity: *Inventions*, even though there are relatively few inventions, they can have an enormous impact on productivity. *Investment*, this particular element suggests that making the right investments is paramount to improve productivity. *Integrations*, it refers to integrate elements in the organization "*no organization can produce with only a single resource*". *Information*, this is easy to describe if the requirements are not made known, all concerned, then is likely they will not met. Finally, *Innovations*, is in many cases de origin of the successful because the product is the creation of new products or services.

Wherever, the "*I*" used by the organization, there is an interest to know the right use for the manufacturers around the world, that means the importance that Americans, Europeans, and Asians considerate for the increase their productivity.

In a 2010 survey jointly undertaken by Deloitte LLP and the Manufacturing Institute (2010), 78 percent of Americans believe that the manufacturing sector is "*crucial*" to US economic prosperity, and 76 percent believe that is "very important" to our standard of living. 55 percent of survey respondents believe the long-term outlook for manufacturing will waken while only 8 percent of respondents believe that the US manufacturing sector will strengthen over the long term. [4] According to this information is easy to describe the different necessities of the industry for two things, the first one is to increase their capital and consolidate their market and the second, to support the perception of the citizens, especially in US where the manufacturing industry is the motor of the nation.

Aspects related to competitiveness and productivity at a level has increasingly been focused on scholarly activity for some time, as evidenced, by the large, theoretical and empirical literature. In Europe, this interest has been significantly stimulated by policy concerns regarding productivity growth and the reduction in spatial differences interregional competitiveness [5]. [5] made a description of the importance of the education inside the manufacturing process, this is an important focus of the increase de productivity in the European sectors, the review made by this author emphatics the importance of use the resources properly, specifically the resources used for education and innovation.

Productivity growth is an indication of optimal resource allocation, effective resource utilization and it has been termed as an ultimate source of economic prosperity. Entrepreneurs, policymakers as well as academic analysts are all interested in strategies for productivity improvement to enhance competitiveness. [6] Focus in the necessities of the Asian manufacturers [6] makes emphasis in the importance of use new strategies for productivity, this means develop new techniques, new philosophies, and new theories, are the most important for the successful in the Asian industry, it doesn't matter if is manufacturer or services.

As it has been mentioned, the productivity in the manufacturing process and services is many times affected by internal parameters of the organization impeding the proper functioning or overall performance of the company. In some cases there exist to many variables that affect the process, in consequence is hard to decide or determinate whit exactitude which of them are really important ant they have a high weighting in the productivity of the process. [7]

One of the multiple tools used for the multivariate analysis is the *Principal Component Analysis* (*PCA*). This tool will be described next due to the present investigation is supported by the application of the *PCA* to determinate which of the different component that affect a process are the principal [8].

As a reference, the main purposes of a principal component analysis are the analysis of data to identify patterns and finding patterns to reduce the dimensions of the dataset with minimal loss of information [9].

Follow the methodology established by the Multivariate Analysis, in the next section as a macro part of the research, the methodology of the *PCA* is described as a micro part focused in reduce the dimensions or the number of variables.

The First Principal Component is the linear combination of the original variables whit maximum variance [9]. Values in this first component of the *n* data are represented by a vector z_1, given by

$$z_1 = Xa_1 \tag{2}$$

Original variables have average zero by consequence z_1 will have average 0 and then their variance will be:

$$\frac{1}{n}z_1'z_1 = \frac{1}{n}a_1'X'Xa_1 = a_1'Sa_1 \tag{3}$$

Where *S* is the matrix of variances and covariance of the data. According to [9] the variance can be maximized increasing the a_1. If we want to get the solution of the maximization for (3), it's necessary to establish a restriction to the vector a_1, and, without loss the generality, it'll use $a_1'a_1 = 1$. Introducing the restriction, using LaGrange multiplication the result is:

$$M = a_1'Sa_1 - \lambda(a_1'a_1 - 1) \tag{4}$$

Maximizing (4) and deriving respect to the components of a_1 and matching to 0. Then

$$\frac{\partial M}{\partial a_1} = 2Sa_1 - 2\lambda a_1' = 0 \quad (5)$$

The solution is:

$$Sa_1 = \lambda a_1 \quad (6)$$

In conclusion, by (3), λ is the variance of Z_1, the interest is to maximize this quantity, λ will be the eigenvalue of the matrix S and their associated vector a_1, define the coefficients of each variable in the first principal component.

Considering the last equations and focusing in apply this methodology to increase the production, statistics are use to focus to describe *p* dimensional points whit the minimum loss of information in one space with dimension one, is equal to substitute *p* variables original's by one new, Z_1, with the capacity of resume the information optimally. In other words, this new variable should have the maximum correlation whit the original's or, should be prevented or calculate the original's variables whit the maximum precision [9].

The case of study, the present article considers the utility of the *PCA* focused in increase the productivity of a manufacturing process. 20 cells of manufacture integrates the processes and the aim is to determinate which of them are the most important for the process, in other words, determinate which of they affect the productivity.

The aim is to apply the *PCA* on a group of 20 entities (manufacturing cells) and 10 variables (the causes of the death times) and determinate which parameters are de most important and which of them are significant for productivity in the process.

II. METHODOLOGY

The methodology used in this case study was focused on the six basic steps used by every multivariate analysis technique. The first step is to determinate de aim of the analysis, in this step is very important delineate all the possible objectives, in other words, is necessary specify, why is necessary to use determinate technique or tool.

The second step is to design the analysis, while this step is used to integrate all the possible variables. During this integration the process of classification is very important, the focus is to duplicate the reality, make a diagnostic of the problem and use the most recent information.

In the third step consist in establish the hypothesis, in this part of the study is necessary to determinate what is the explanation of the phenomenon.

After establish the hypothesis, the next step is to make the analysis. During this part of the process, is necessary to take all the considerations for the study, integrating all the techniques necessary, and if it is possible, try to integrate new analysis tools.

Interpretation results are the step number five; this is the most important and fine parts of the process due the importance of the information get it by the proofs. During the interpretations of the resuls is necessary consider all the factors used for the researcher, including methods and techniques.

Finally, the validation is the last part of the process. The validation includes the decision of accept the results get the proof, and accept that the proceedings used by the researcher were correctly established, and they were appropriate for the research. Figure 1 presents a diagram of the steps described in the last paragraphs.

As in all research, the objective of this methodology is determinate all the factors and discover the new possible factors, verify and test important facts, analyze the event focused to determinate the cause and effect relationship. If a process is conducted with success, it's possible to find solutions to scientific, non-scientific and social problems and overcome or solve the corresponding situations.

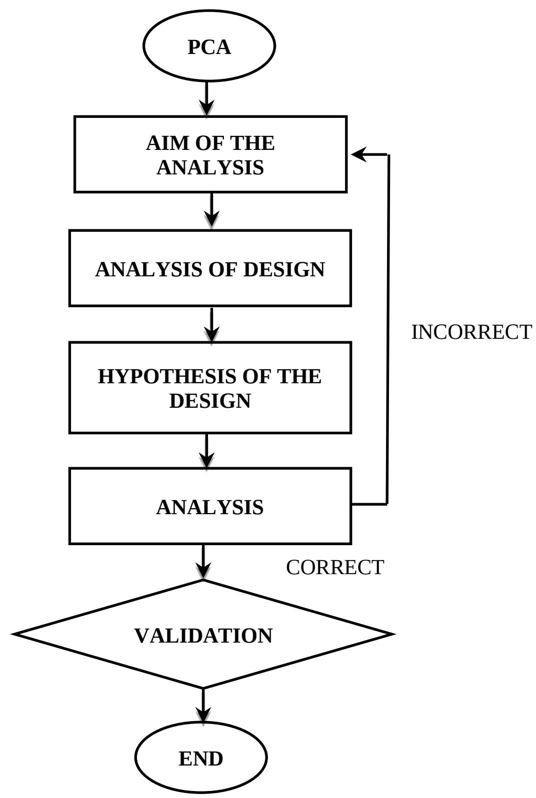

Fig. 1. Methodology used for Multivariate Analysis.

In general de Figure 1, present a macro methodology used for a Multivariate Analysis. As it has been described, in this article the micro methodology used is PCA, this one is integrated by six general steps listed below. [10]

1. Take the whole dataset consisting of d-dimensional samples ignoring the class labels.
2. Compute the d-dimensional mean vector.
3. Compute the scattering matrix of the whole data set.
4. Compute eigenvectors and corresponding eigenvalues.
5. Sort the eigenvector by decreasing eigenvalues and chose k eigenvectors whit largest eigenvalues to form a *d x k* dimensional matrix w.
6. Use this *d x k* eigenvector matrix to transform the samples onto the new subspace, this can be summarized by the mathematical equation (7).

$$y = W^T \times x \qquad (7)$$

III. DESIGN

For this analysis, it has been determinate to work whit a group of data got it from an accumulative sample of one month of production. The data correspond to the death times generated by the 20 manufacturing cells originated by the next variables:

1. Sub ensemble. - This variable considers the time used by make a subensemble that doesn't be delivered on time in the production area by the chain supply.
2. Deliver Material. - This variable is used for determinate the time that the equipment stop due to the material wasn't deliver on time in the production area programmed using the chain supply.
3. New Products and Fixture. - The fixtures are mechanism designed to increase the productivity, making easier the operation inside the process for manufacture a new product or other that exists. The death time considered in this variable is the death time generated by broken fixtures, adjusts or reparations.
4. Lack of material. – In this variable it has been considered the time generated by the cell when it stops producing due the lack of material, or elemental component for the manufacturing of the product, provided that is available in the factory.
5. Organization. – This variable describes the time that the cell stops the production due to external situations, no planned with anticipation.
6. Quality. – During the search, all the death times generated for a defect in row material or material in the process where classified in this variable. In other words, this variable considers the time generated by defects of products during the process or raw material out of specifications.
7. Micro Failure. – Al the death times minor to 5 minutes, were classified in this variable, it doesn't matter the cause.
8. Machinery failure. – Is the time generated by the equipment when a stop occurs, and it is greater than 5 minutes. Commonly these failures are electrical issues, by design, programming or state of components of the machines; in most cases it requires the assistance of a technician or specialist.
9. Set up by Model. – This variable considers the total time generated by set up the equipment. This time is considered since the last good piece of a model was produced until the first good piece of a new model is produced, this time include the stop inside the production line due to the necessary adjusts of machinery, human resources, material, fixtures, and special tools.
10. Downtimes programmed. – Downtimes programme are the time that stops the fluency of production to make planned activities, for example, launch time, workshop, meetings, 5's, etcetera.

According to the last information and for the design of this document, the variables were labeled with numbers: 1 (Subensemble), 2 (Deliver Material), 3 (New Products and Fixtures), 4 (Lack of material), 5 (Organization), 6 (Quality), 7 (Micro Failure), 8 (Failure of Machine), 9 (Set up by model) and 10 (Downtimes programmed). The table I shows the death times generated by the variables 1 to 5 in the 20 manufacturing cells, the unit of time is a minute.

TABLE I

Cell	1	2	3	4	5
1	58	85	95	475	664
2	24	103	214	10	283
3	60	44	10	275	725
4	41	82	67	6	775
5	8	182	95	0	342
6	17	50	332	20	120
7	60	204	315	892	196
8	44	64	259	25	251
9	22	10	162	120	149
10	58	9	24	475	266
11	53	96	68	465	703
12	34	19	10	0	224
13	56	129	37	200	358
14	5	55	52	625	320
15	15	233	0	365	105
16	51	140	66	1250	349
17	46	185	51	50	290
18	27	64	72	207	232
19	17	172	65	380	243
20	34	216	83	274	200

The Table II shows the death times generated by the variables 6 to 10 in the 20 manufacturing cells, the unit of time is a minute.

TABLE II

Cell	6	7	8	9	10
1	105	16	251	599	10
2	565	0	266	596	20
3	63	0	224	809	0
4	58	0	320	461	0
5	43	0	128	444	0
6	100	0	161	837	15
7	335	77	157	562	0
8	75	0	243	420	0
9	270	0	232	783	20
10	32	10	161	430	0
11	32	0	299	619	0
12	128	0	593	791	12
13	412	0	168	690	0
14	20	0	254	450	13
15	96	0	155	512	17
16	170	20	100	809	0
17	320	0	368	512	11
18	128	30	270	697	10
19	59	15	540	791	0
20	15	0	95	430	12

IV. RESULTS

According to the aim of this article and executing the software Minitab® for the analysis of PCA, the table III and IV present the values of the analysis and the vectors for the matrix of covariance. The matrix of covariance was used due to the necessity of calculate the principal components under the restriction of don't standardize the variables.

Tables III and IV show the eigenvalue of each principal component, this value specifies the proportion of each component explain. For the CP_1, the eigenvalue it's equal to 2.4096 this describes the 24.1% of the variance in all the process. The CP_2, has an eigenvalue equal to 2.1144 this describes the 21.1% of the variance in all the process. Describing the CP_3, and according to his value, this variable has an eigenvalue of 1.5097; this one represents the 15.1 % of the variance in all the process. For the CP_4, the eigenvalue it's equal to 1.0312 this describes the 10.3% of the variance in the process. The CP_5, has an eigenvalue equal to 0.9016 this describes the 9.0% of the variance in the process.

TABLE III

Cell	CP_1	CP_2	CP_3	CP_4	CP_5
Eigenvalue	2.4096	2.1144	1.5097	1.0312	0.9016
Proportion	0.241	0.211	0.151	0.103	0.090
Cumulative	0.241	0.452	0.603	0.706	0.797

Continuing with the description in table number IV, de results show and 0.7642 eigenvalue for the variable number 6 (CP_6) this variable represents the 7.6 of the variance whit respect at all the process. For the variable number 7, the eigenvalue has a value of 0.4831; this one represents the 4.8% of the variance in the process. Describing the *CP* number 8, the eigenvalue determinate for this variable has a measure of 0.3734 which it represents a 3.7% of the variability in the process. The last two variables can be described using the CP_9 and CP_{10}, for the variable number 9 the eigenvalue is 0.2336 which it represents a 2.3% of the total variance in the process and finally the variable number 10 has an eigenvalue of 0.1793, this one represents the 1.8 of the variance in all the process.

TABLE IV

Cell	CP_6	CP_7	CP_8	CP_9	CP_{10}
Eigenvalue	0.7642	0.4831	0.3734	0.2336	0.1793
Proportion	0.076	0.048	0.037	0.023	0.018
Cumulative	0.873	0.921	0.959	0.982	1.000

As it has been showed in the table III and IV, the proportion represented by each eigenvalue respect whit their total it's equal to the proportion explained by the principal component [7].

Using as a reference table I and II, the component's number one, two, three, four, and five can be selected due to in addition they explain the 79.7% of the total variation in the process, and they represent the combination of 5 parameters, if it has to consider another component, it's possible to explicate the 87.3% of the variation of the process nevertheless it's necessary to include an extra variable, but the aim of use the PCA is to reduce de variable's, for this reason, it's convenient use only the first five component's because together explain a significant proportion of the variables that affect the process.

Table V presents the matrix of coefficients of the principal components numbers one, two, three, four and five. The effect and their relation it will be explained in the tables numbers VII, VIII, IX, X, and XI.

TABLE V

Variable	CP_1	CP_2	CP_3	CP_4	CP_5
Subensemble	-0.483	0.116	0.276	0.326	0.182
Deliver Material	-0.148	-0.305	-0.430	-0.237	0.558
New Products and Fixture	-0-040	-0.455	0.190	0.419	-0.364
Lack of Material	-0.461	-0.187	0.009	-0.439	-0.188
Organization	-0.289	0.475	0.075	0.276	0.104
Quality	0.074	-0.365	0.364	0.308	0.584
Micro Failure	-0.384	-0.400	0.192	-0.188	-0.037
Failure of Machine	0.211	0.290	0.350	-0.339	0.384
Set up by model	0.067	0.008	0.635	-0.385	-0.118
Downtimes programmed	0.495	-0.219	0.036	-0.030	0.011

Table VI, presents the matrix of coefficients of the principal components numbers six, seven, eight, nine and ten. The effect and their relation it will be explained in the tables numbers VII, VIII, IX, X, and XI.

TABLE VI

Variable	CP_6	CP_7	CP_8	CP_9	CP_{10}
Subensemble	-0.131	-0.202	-0.307	0.580	-0.223
Deliver Material	0.080	0.444	0.147	0.304	-0.122
New Products and Fixture	0.394	0.260	0.241	0.064	-0.403
Lack of Material	-0.321	-0-228	0.209	-0.279	-0.498
Organization	-0.077	0.178	0.734	-0.077	0.128
Quality	-0.240	-0.055	-0.035	-0.481	0.017
Micro Failure	0.347	-0.284	0.172	0.093	0.616
Failure of Machine	0.569	-0.247	0.117	0.006	-0.339
Set up by model	-0.252	0.568	-0.051	0.165	0.118
Downtimes programmed	-0.383	-0.379	0.445	0.466	-0.007

Now it's necessary to make a description of the coefficients for each principal component. Table VII describes the coefficients or weights that determine which variables represent each principal component. In this case Subensemble, lack of material and down times programmed have high values in contrast whit the others, this means that these variables are correlated and they represent in some way the "*same*". For this case of the study, the productivity is affected by lack of subassembly, lack of material and downtimes programmed. Probably the death times due to downtimes programmed are a consequence of lack of material and lack of subassembly.

Table VII

Variable	CP_1
Subensemble	**-0.483**
Deliver Material	-0.148
New Products and Fixture	-0-040
Lack of Material	**-0.461**
Organization	-0.289
Quality	0.074
Micro Failure	-0.384
Failure of Machine	0.211
Set up by model	0.067
Downtimes programmed	**0.495**

Table VIII describes the coefficients or weights that determine which variables represent each principal component. In this case New products and fixture, organization, and micro failure have high values in contrast whit the others, this means that these variables are correlated, and they represent in some way the "*same*". In this case of the study, the productivity is affected by these variables. Probably the death times due to new products and fixture are a consequence of organization and micro failure.

Table VIII

Variable	CP_2
Subensemble	0.116
Deliver Material	-0.305
New Products and Fixture	**-0.455**
Lack of Material	-0.187
Organization	**0.475**
Quality	-0.365
Micro Failure	**-0.400**
Failure of Machine	0.290
Set up by model	0.008
Downtimes programmed	-0.219

The Table showed bellow (Table IX) describes the coefficients or weights that determine which variables represent each principal component. In this table, the variable's Deliver Material and Set up by the model have the highest values in contrast with the others by consequence they are correlated. Probably the death times generated by this variable are due to the set up by model, and the consequence probably is the delivery of the material.

Table IX

Variable	CP_3
Subensemble	0.276
Deliver Material	**-0.430**
New Products and Fixture	0.190
Lack of Material	0.009
Organization	0.075
Quality	0.364
Micro Failure	0.192
Failure of Machine	0.350
Set up by model	**0.635**
Downtimes programmed	0.036

In table X, present the coefficients that determine which variables represent by each principal component. In this table the variable's New Products and fixture and Lack of Material have the highest values in contrast with the others, by consequence they are correlated. Probably the death times generated by this variable are due to new products and fixture generated by the lack of material.

Table X

Variable	CP_4
Subensemble	0.326
Deliver Material	-0.237
New Products and Fixture	**0.419**
Lack of Material	**-0.439**
Organization	0.276
Quality	0.308
Micro Failure	-0.188
Failure of Machine	-0.339
Set up by model	-0.385
Downtimes programmed	-0.030

Finally, Table XI describes the coefficients or weights that determine which variables represent each principal component. In this table, the variable's delivered material, and quality have the highest values in contrast with the others by consequence they are correlated. Probably the death times generated by this variable are due to the quality, and the consequence probably is the delivery of the material.

Table XI

Variable	CP_5
Subensemble	0.182
Deliver Material	**0.558**
New Products and Fixture	-0.364
Lack of Material	-0.188
Organization	0.104
Quality	**0.584**
Micro Failure	-0.037
Failure of Machine	0.384
Set up by model	-0.118
Downtimes programmed	0.011

The information presented in the tables described below could be verified in the graphic of the figure 2. This one presents the variance of each principal component.

The description to this graphic is in function of the variances of each component. In the top of the chart, the component number on is the most representative as it was described in the component. The productivity is affected by the lack of subassembly, lack of material and downtimes programmed. With a less high, the second component has the capacity of describes with tow variables the variance of 21%, just three percentage points less than the first component.

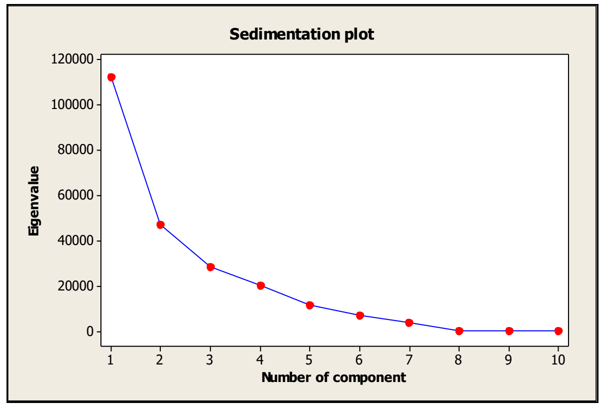

Fig. 2. Graphic of variances for each component.

In the figure, 3 shows the graphic of score for the first two components. Before to describe the graphic, is necessary to describe that the second component for the variable of delivery material is integrated the eigenvalue who defines the new variable: $z_2 = 0.116x_1 - 0.305x_2 - 0.455x_3 - 0.187x_4 + 0.475x_5 - 0.365x_6 - 0.400x_7 + 0.290x_8 + 0.0080x_9 - 0.219x_{10}$.

This variable is approximate the difference between two weighted averages. The first average considerate most important two variables, Subensemble (x_1) and lack of material (x_2). The second average considers most important the downtimes programmed. After the description of the integration of the second component, it´s less complicated describe that the figure 3, represent each variable in a plane of the first two principal components. Each point appears represented by their coordinates with respect to the axis defined by the principal components and can be interpreted as a projection of the points in a space of dimension 10.

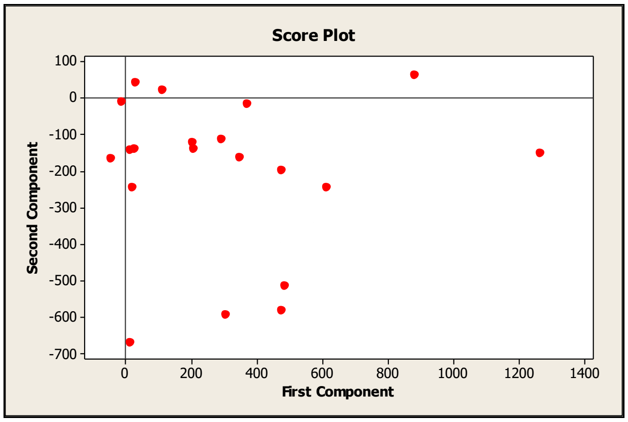

Fig. 3. Graphic of variance for each component.

In the figure number 4, the projection plot graphic is used to represents the points for the second principal component (axis y) versus the points for the first principal component (axis x). About the figure, it's possible to describe that the variable lack of material has a big dispersion in horizontal axis compared whit the others. In the axis vertical, the variables Organization, quality and new products show a dispersion compared whit the others variables, this means that the variables organization, lack of material and new products are important factors for the efficiency of the system.

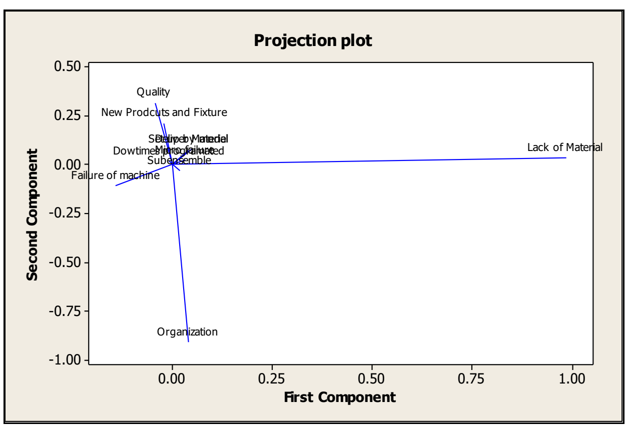

Fig. 4. Graphic of variances for each component.

Using the double projection plot as a tool for analysis, it has been considered that the position of the graphics of score and weight of the two principal components could generate more evidence about the most important variable. In figure 5, confirms that the lack of material, quality, and organization still been the most important variables for this case.

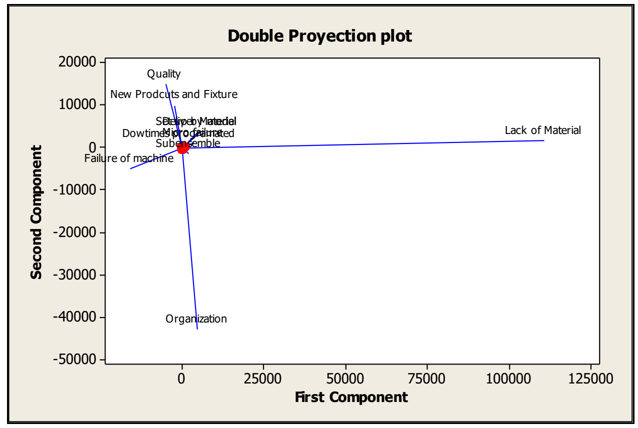

Fig. 5. Graphic of variances for each component.

After the analysis of the variables, is elemental to make a "student t" whit the finality of verify the difference between the variables Lack of material, Organization, Quality and New products and Fixtures. Table XII describes the results of the analysis.

Table XII

Variable	N	Average	Std. Desv.	Std. Error Aver.	CI of 95%
Subensemble	20	36.50	18.57	4.15	(27.81, 45.19)
Deliver Material	20	107.71	71.20	15.9	(73.80, 140.40)
New Products and Fixture	20	103.80	99.50	22.3	(57.30, 150.40)
Lack of material	20	305.70	330.60	73.9	(151.00, 460.40)
Organization	20	339.80	206.40	46.1	(243.20, 436.30)
Quality	20	151.30	150.70	33.7	(80.80, 221.80)
Micro Failure	20	8.40	18.35	4.10	(0.00, 16.99)
Failure of machine	20	249.30	131.10	29.3	(187.90, 310.60)
Set up by model	20	612.10	151.90	34.0	(541.00, 683.20)
Downtimes programmed	20	7.00	7.65	1.71	(3.42, 10.58)

As it has been showed, in the last figure, the variable lack of material has death times between 151 and 460 minutes on average, the variable Organization has death times between 243.20 and 436.30 minutes also the variable Quality has an IC between 80.80 and 221.80 all of the considering a 95% of confidence.

Figure 6 presents a box plot of the dispersion of the variables. There exist a group of outliners that make confuse the information, for example, the variable lack of material has an outliner of 1250 minutes, this data has an important contribution to this variable, the responsible of deliver the material mentioned that this delay sometimes occur whit more frequency. Now describing the variable organization it’s possible to describe compact phenomena but there is a group of out liner that makes difficult to discriminate this variable.

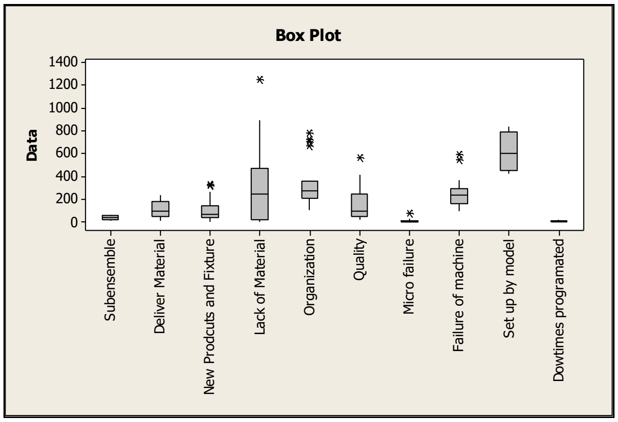

Fig. 6. Box plot for every variable.

Finally, as a part of the process of validation, the box plot showed in the figure number six confirm the impact of the variables in the death times, also the correlation demonstrated in the Tables VII to XI. It´s necessary to clear, that this analysis present the information that define the variable more important for the productivity and as a consequence of the analysis, the recommendation is to make an efforts to reduce the death times generated by the lack of material, the organization, and the quality.

V. DISCUSSION

The use of statistical tools for the development of productivity studies is one of the new modalities used in the national industry. As a part of this project, the aim was the integration of the technique PCA, not only to determine the root causes of the dead times, but to determine the effect that the variables have individually and as a whole. During the development of the research, we considered the use of the technique of discriminant analysis, however, due to the determination of the factors can develop by means of analysis of averages, we were interested in integrates the analysis of principal components through their matrix of correlations. In the future, is possible to integrate discriminant analysis as a technique for support the results generated by the Principal Component Analysis.

VI. CONCLUSION

The results obtained by the analysis of principal components, showed that the first f,ive principal components could be explained the 79.7% of the total variation, decreasing the number of variables from ten to five, that they presents more effect on the level of productivity.

According to the data obtained by First Principal Component, we could conclude that the productivity is affected by causes related to lack of material and downtimes programmed. According to the factory, this information is true, due to the 90% of the death times where the cause of the lack of material during the downtimes programmed.

Considering the second factor, the causes that originate the death times during the develop of new product is caused by the organization an the agenda generated whit the different areas integrated into the project, micro failures define one of the most common errors.

Finally, during the implementations of lean manufacturing tools, the focused on improve the production system and whit the aim of reduce the lack of material, it was implemented the technique of develop of suppliers and their integration as a part of the chain supply whit the new objective, reduce the times of delivery and maintain a minimum stock of materials.

In case of the organization and the effects due to external variables, the company decides to develop a strategy of marketing, focused on the use of enterprise resource planning and the control via the internet. For the quality, the death times generated by the quality decrease until 50% due to the manufacturing area now has the correct materials for each product in quantity and on time.

In resume, determinate the effect of variables and their correlation is very useful when the process has too many variable with to many effects, analyze this process with elemental quality tools is impossible, in this cases is recommendable to use analysis multivariate of data. Analysis multivariante is the shortest way to reduce the number of variables a determinate which of them are de most important.

REFERENCES

[1] O. Bustinza-Sanchez, G. Parry, F. Vendrell-Herrero y V. Myrthianos, «Recovering Revenue Through Coustomerization and Service Channel Linking,» *DYNA Management,* vol. 3, nº 1, pp. 1-8, 2015.

[2] M. B. Bernábe Loranca, A. D. Pinto, J. A. Ruiz-Vanoye, R. J. Espinoza y B. E. Olivares, «A multi-objective proposal for aggregation of economically inactive population,» *International Journal of Combinatorial Optimization Problems and Informatics,* vol. 3, nº 1, pp. 70-79, 2012.

[3] K. Smith, «The concept and importance of productivity,» de *Maynard's industrial engineering handbook,* New York, Mc Graw Hill, 2001, pp. 2.32-2.10.

[4] T. A. Hemphill, «Policy Debate: The US advanced manufacturing initiative: Will it be implemented as an innovation - or industrial - policy?,» *Innovation: Managmente, Policy & Practice,* vol. 16, nº 1, pp. 67-70, 2015.

[5] B. Bracalente, C. Perugini y F. Pompei, «What Sorts of Agglomerations Really Matter to Productivity?: A regional Analysys of Europe´s Manufacturing Sector,» *Review of Regional Studies,* vol. 38, nº 2, pp. 145-172, 2008.

[6] L. Jeong-Dong y A. Heshmati, «Manufacturing Sector Productivity Crowth in the Asia Pacific Region,» *Global Economic Review,* vol. 36, nº 4, pp. 301-303, 2015.

[7] J. Moreno, «Análisis Multivariante en Investigaciones de Calidad del Servicio,» *Revista Venezolana de Análisis de Coyuntura,* vol. XI, nº 1, pp. 275-307, 2005.

[8] M. Richarme, «Eleven Multivariante Analysis Techniques: Key Tools in your Marketing Research Survival Kit,» 2002. [En línea]. Available: http://www.decisionanalyst.com/Downloads/MultivariateAnalysisTechniques.pdf. [Último acceso: 01 06 2015].

[9] D. Peña, Análisis de Datos Multivariantes, Madrid: Mc Graw Hill, 2002.

[10] S. Raschka, «Implementing a Principal Component Analysis (PCA) in Python step by step,» 13 04 2014. [En línea]. Available: http://sebastianraschka.com/Articles/2014_pca_step_by_step.html. [Último acceso: 15 05 2015].

Este estudio fue financiado por los autores. Los autores declaran no tener ningún conflicto de interés.