International Journal Of Computational Engineering Research (ijceronline.com) Vol. 2 Issue. 8

Hotellings T-square & Principal Component Analysis Approaches to Quality Control Sustainability.

Onwuka, Gerald. I.

Department of Mathematics & Statistics, Kebbi State University of Science and Technology, Aliero

Abstract:

In this paper, Multivariate Techniques are applied to a Production/Manufacturing industry to examine various measurable characteristics of a product since no two items produced are exactly alike. Although, natural or assignable causes of variation possibly must be found in every production process, attempt has been made using this technique to examine correlation and tests for equality of means on collected multivariate data. Principal Component Analysis and Hotelling's T^2 tests were used, with the 3-characteristics measured showing negligible/low correlation with nearly all the correlation coefficients small. Only one variable possess on the average about 70% of the total variation. The result of Hotelling T^2 test shows that the average daily means are not equal and provide estimate of the interval to use. The purpose of this study is to monitor the in-control condition for data on pipes production, which clearly concludes that data for the production process was really obtained from a statistically monitored and controlled process.

Keywords: Quality Control, Multivariate Analysis, PCA & Hotelling T^2.

I Introduction

Statistics technique is applied to a manufacturing industry for the determination of quality raw materials and finished products. It is also applied in modeling manufacturing processes to determine when component could be replaced or the process turned off for turnaround maintenance. Using statistics can help increase not only the quality of the products that are being manufactured but also the quantity. The reason is that manufacturing companies can use the statistics to create a plan of action that will work more efficiently. In order to be able to effectively forecast the future productively with statistics, there would need to be a program setup that can do the following: Forecast production, when there is a stable demand and uncertain demand, Quantify the risk associated within the operations and financial costs. Predict when there will be operational bottleneck, in sufficient time before they occur. Pinpoint when and which specific model will be the cause of uncertainty. Calculate given information to show the statistical outcome. Calculate summary statistics in order to setup sample data.

If there is not a visible quality issue, then there is no way to fix it. When statistics are used to increase the quality of production and products, it is easy to track and make appropriate changes to improve this overall quality. Statistics can also help maintain the quality in the areas of business process, the mechanical and engineering process, in addition to the part production process.

With the possibility of as much up to date, and real time feedback, the quality can be increased almost instantly. In order to setup the right process for statistical tracking and predicting in quality improvement, there would need be a support process and information gathered. This is a step that many manufacturers have had a hard time completing. However, once this is done, the overall payoff of being able to improve quality through statistics has had a huge benefit, increase productivity at a lower cost. And that is what it is all about.

II Back Ground

The quality of the manufactured goods depend on a variety of factors beginning from the quality of raw materials, the process of production, the conditions of the machines and other equipments, the skills of the labour force and the inspection techniques, adopted at every stage of production. The goods can be sold in the market only if they conform to pre-determined quality standards about which the prospective buyers have been briefed. The statistical quality control helps the producer to achieve this objective by keeping the various steps in the production process within statistical control.

III Statistical Quality Control

The field of statistical quality control can be broadly defined as those statistical and engineering methods that are used in measuring, monitoring, controlling, and improving quality. Statistical quality control is a field that dates back to the 1920s. Dr. Walter A. Shewhart of the Bell laboratory was one of the early pioneers of the field. In 1924 he wrote a memorandum showing a modern control chart, one of the basic tools of statistical process control. Harold F.Dodge and Harry G. Romig, two other Bell system employees provided much of the leadership in the development of statistically based sampling and inspection methods. The work of these three men forms much of the basis of the modern field of statistical quality control. World War II saw the widespread introduction of these methods to U.S. industry. Dr W. Edwards Deming and Dr. Joseph M. Juran have been instrumental in spreading statistical quality control methods since World War II. Quality Control is regarded as the most powerful trademark behind market. (Montgomery, 2001)

Statistical quality control is based on the theory of probability and sampling and is extensively used in all types of industries in fact in all repetitive processes; the statistical quality control plays a very important and significant role. (Sharaf Eldin et al, 2006).

2.0 MATERIALS AND METHODS

Data for this study was collected from Tower galvanized products Nigeria Limited Kaduna. The company was incorporated in May 1975 and is managed by CATISA Genera. CATISA is currently managing more than 60 manufacturing operations in Nigeria and more than 400 operations in 55 countries in the world, out of which 65% of these are in Africa. Some of the well-known companies of the group in Nigeria are: Borno aluminum company Ltd. Maiduguri, Queens way aluminum company Ltd. Kaduna., Tower aluminum Nigeria PLC. Ikeja, Midland galvanizing products Ltd. Asaba. among others, the company has pioneered the development of furniture manufacturing industries in the northern states of Nigeria, pipe produced by the company are extensively used in the distribution of water for domestic and industrial use. The company's products enjoy good reputation among other products available in the market. Most of the products are distributed through the facilities available in Kaduna and Kano units. The products include: Head-pan, Cut-sheets, Long-span circular profile roofing (Aluminium, G.I. &Aluzinc) etc The scope of this study covers production data for four(4) different products produced daily by the company. The data consists of the length, circumference, and outside diameter for the four different pipes produced by the company. The pipes are: 1X1" square pipe, 7/8" round pipe, 3" round pipe, 2X2" square pipe,

The data was drawn from sample of size 5 taken from each day's production for 25 days (working days), with the aim of determining the degree of dependency between the various components of each particular product. So as to check the contribution of each component to the total variation (PCA), performing, multivariate test on the data set using Hotelling T^2 approach and to estimate confidence interval using Roy-Bose approach.

This study covers all aspects of production –raw materials, labour, equipment and management. In production industries, data on updated information about the status of the process are collected a long time. These data are frequently served to control charts, in order to see and decide whether production is operating under statistical control or if some special cause has interfered with it. (The Process).Normal operation conditions are set using the control charts, analyze the data from period of normal operations and as long as the process rests within the normal operations conditions, as soon as it moves outside such boundaries, the root of that cause is identified and corrective measures is taken to bring back to normal operation. This is because in any production process, regardless of how well designed, certain amount of inherent or natural variability would always exist.(Montgomery & Lowry, 1995). These "background noise" is the cumulative effect of many small, while a process that is operating with only chance causes of variation present is said to be in statistical control while any variability that are not part of the chance cause pattern are called assignable causes. It may result due to operator errors, improper adjusted machines and sub-standard raw materials. A process that operates in the presence of assignable cause is said to be out of control (Montgomery, 1990).

3.0 RESULTS AND DICUSSION USING THE VARIOUS TECHNIQUES MENTIONED ABOVE

3.1 PRINCIPAL COMPONENT ANALYSIS (PCA)

This is a mathematical procedure which does not require user to specify an underlying statistical model to explain the 'error' structure. (Jackson,1985) In particular, no assumption is made about the probability distribution of the original variables. It is a technique to use to get a 'feel' for a set of data. Hopefully, it assist user to a better understanding of correlation structure and may generate hypothesis regarding the relationship between the variables. Its two main objectives are;

(1) identification of new meaningful variables
(2) reduction of dimensionality of the problem as a prelude to further analysis of the data.

3.1.1 PROCEDURE IN PCA.

(a) Decide if it is worth including all the variables recorded in the original data matrix. And whether any of the variable need to be transformed.
(b) Calculate correlation (or covariance) matrix bearing in mind that a correlation co-efficient should not be calculated for pairs of variables whose relationship is non-linear.
(c) Examine the correlation matrix and observe any natural grouping of variables with 'high' correlation. However, if nearly all the correlation coefficients are "small", there is probably not much point to do principal component analysis because of the results obtained from the analysis (see appendix).. Because the variables i.e diameter, circumference and length of the 4 type of pipe i.e

- 1X1" square pipe
- 7/8" round pipe
- 3" round pipe
- 2X2" square pipe

Used in the analysis have very low correlation coefficient in the various matrix obtained.

These result force the research to the following conclusion on each of the 4 type of pipe used in the analysis. Meaning there was no need to obtain a linear combination of the variables but conduct hypothesis test using Hotellings T^2 and estimate Roy-Bose Confidence interval for each characteristics of each particular type of pipe.(Hotelling, 1947)

(1) From matrix or Table 3.11,3.12,3.13 and 3.14, (see appendix) all the correlation matrix shows a very low relationship between the variables i.e diameter, circumference and length which indicate that there is no further need to conduct principal component analysis.

(2) Also, matrix or table 3.11,3.12,3.13 and 3.14,shows that in each of the type of pipe produced, the diameter contribute higher variation to the total variation in the production process. This is clearly demonstrated by the eigen values, and from the graphs m, n, o and p prove that by the sharp drop from the first eigen value in each graph.

3.2 MULTIVARIATE TESTS

A multivariate statistical tests specifies conditions about the parameters of the population from which sample was drawn. This means that multivariate model considers that the population distribution should have a particular form (e.g a normal distribution) and also involves hypotheses about population parameters.(Morrison,2005) The reliability of the results of multivariate tests depends on the validity of these assumptions:

It is paramount to note that, the most powerful tests are those having more extensive assumptions. The T^2 test, for it to be applied the underlying conditions must be at least 70% satisfied, (Chatfield & Collins, 1980),

1. Observations must be independent (uncorrelated)
2. Observations must be drawn from normally distributed populations
3. Population must have variance-covariance matrix and mean vector.
4. Variables must have been measured in at least interval or ratio scale.

If these conditions are satisfied in the data under analysis, one can choose T^2 or wilks lambda test. When these set of assumptions are met, the tests are most likely to reject H_O when it is false.

3.3 HOTELLING T^2 DISTRIBUTION

Harold Hotelling (1931), propose a multivariate generalization of the student t-distribution and T^2 is used in multivariate hypothesis testing.

If x, x, ---,x_n ~N (μ, $\square^2$) with μ and $\square$ unknown, we can test the hypothesis; Ho: $\mu=\mu_o$ using

$t = (\bar{X} - \mu)/\square /\sqrt{n}$

So that $t^2 = (x- \mu_o)^2/(\square^2/n)$

$t^2 = n\,(x- \mu_o)^1\,(\square^2\,)^{-1}\,(x- \mu_o)$.

the generalization for x_1, x_2 --, x_n,~ $N_p(\mu,\Sigma)$ with $P\geq 1$, is the Hotelling T^2 statistic

$T^2 = n(x- \mu_o)^1\, s^{-1}(x- \mu_o)$

Where mean $\bar{X} = 1/n\ \Sigma x$, and s^{-1} is the inverse of the variance covariance matrix. n is the sample size. The diagonal elements of S are the variances of the x_i and the off diagonal are the covariance for p variables.

The multivariate techniques should posses' three important properties:

1. They produce a single answer to the question: is the process in control?
2. Has the specified type I error been maintained?.
3. These techniques must take in to account the relationship between the variables.

3.3.1 HOTELLING'S T^2 TEST

Below is the covariance result of the 3`` round pipe data.

```
             diam          circum          length
diam    0.00119900    0.00025500   -0.00004317
circum  0.00025500    0.01726667   -0.02476333
length -0.00004317   -0.02476333    0.41491233
```

Inverse of covariance matrix

```
      843.11         -12.89 0.00
      -12.89          63.17 2.58
        0.00           2.58 2.58
```

T^2= 47.3684.

T=14.47.

F(0.05)(3,22)=3.05.

From the result, the null hypothesis is rejected at $\alpha = 0.05$, that the means are not equal i.e. the mean of diameter, circumference and length for the individual 3``round pipe differs. And that the following intervals should be considered for the acceptance of 3`` round pipes produced.

3.4 CONFIDENCE INTERVAL (Roy – Bose confidence interval) for 3`` round pipe data.

C.I= $\overline{X} \pm Ka/2,(n-1)\ S/\sqrt{n}$

$Ka/2 = \sqrt{}\ [p(n-1)/(n-p)\,F\alpha,\ ,\ (_{p,\,n-p})]$

C.I(diameter)= (8.99,9.03).
C.I(circumference)= (30.09,30.25).
C.I(length)= (599.22,600.04).

Also we can use, C.I= $\overline{X} \pm ta/2,(n-1)\ S/\sqrt{n}$
C.I(diameter)= (9,9.02).
C.I(circumference)= (30.12,30.22).
C.I(length)= (599.36,599.9).

Note that this test was done for all 3 types of pipe with interpretation of each before the **(ROY- BOSE C.I)**

4.0 SUMMARY

In this research Quality Control using multivariate techniques tests for independence and equality of means was conducted using multivariate data. From the result of each tests using principal component analysis, the variables are found to be reasonably uncorrelated that the degree of relationship(correlation) is not strong enough. The hypothesis conducted using Hotelling's T^2 , each tests rejected the null hypothesis(H_o : means are the the same).

4.1 CONCLUSION

The multivariate tools used, principal component analysis and Hotelling T^2 are both statistical method of inference. Multivariate methods usually depend upon the assumption of a specific distribution form, for example an approximate multivariate normal distribution. And data for these methods will be in interval or ratio scale. The analysis of the data from pipe Production Company using the multivariate techniques produced a fairly reasonable result. From the analysis using principal component analysis to have a "feel" for the set of data and understanding of their correlation structure, the data set for diameter shows higher variation contribution compared to that for circumference and length. And the correlation that exists from each product variable is "substantial". Meaning nearly all the correlation coefficients are "small", there is probably not much point to further conduct complete principal component analysis. The Hotelling T^2 analysis result for the different product rejected the null hypothesis in favor of the alternative hypothesis (H_o; the means are different). The Roy-Bose confidence interval techniques tend to estimate interval for the production department to consider for the acceptance of a pipe that would be produced when the process is in statistical control with a confidence that items produced reached 95% standard quality. Based on the findings, it is therefore concluded that the process that generates the data is in statistical control.

Numerous quality control related problems are multivariate in nature, and using univariate statistical process control charts is not effective when dealing with multivariate data exhibiting correlated behavior. Therefore, multivariate statistical process control procedure provides reliable result.

APPENDIX

3.2 PRINCIPAL COMPONENT ANALYSIS

Table 3.11 shows the result of the 3`` round pipe data using PCA.

Table 3.11 :

———— **10/10/2012 9:08:34 AM** ————————————————————Welcome to Minitab, press F1 for help.

CORRELATION

```
       D       C          L
D     1      0.056      -0.002
C   0.056      1        -0.269
 L  -0.002  -0.269        1
```

Covariances: diameter(cm), circumf(cm), length(cm)

```
               diam         circum        length
diam     0.00119900   0.00025500  -0.00004317
circum   0.00025500   0.01726667  -0.02476333
length  -0.00004317  -0.02476333   0.41491233
```

Eigen analysis of the Covariance Matrix

```
Eigenvalue  0.41645  0.01573  0.00119
Proportion    0.961    0.036    0.003
Cumulative    0.961    0.997    1.000
Variable            PC1       PC2       PC3
diameter(cm)        -0.000    -0.017    1.000
circumf(cm)         -0.062    -0.998    -0.017
length(cm)          0.998     -0.062    -0.001
```

Table 3.12 shows the result of the 7/8`` round pipe data using PCA.

Table 3.12 :

———— **10/10/2012 11:38:11 AM** ————————————————————Welcome to Minitab, press F1 for help.

CORRELATION

	D	C	L
D	1	0.166	0.331
C	0.166	1	0.067
L	0.0331	0.067	1

Covariances: diameter(cm), circumf(cm), length(cm)

```
                 diam          circm          length
      diam   0.00258933   0.00133333   0.00872900
      circm  0.00133333   0.02500000   0.00550000
      length 0.00872900   0.00550000   0.26826933
```

Eigenanalysis of the Covariance Matrix

```
Eigenvalue  0.41645  0.01573  0.00119
Proportion    0.961    0.036    0.003
Cumulative    0.961    0.997    1.000
Variable            PC1       PC2       PC3
diameter(cm)        -0.000    -0.017    1.000
circumf(cm)         -0.062    -0.998    -0.017
length(cm)          0.998     -0.062    -0.001
```

Table 3.13 shows the result of the 2 X 2 round pipe data using PCA.

Table 3.13 :**CORRELATION**

	D	C	L
D	1	0.043	0.074
C	0.043	1	0.42
L	0.074	0.42	1

Covariances: diameter(cm), circumf(cm), length(cm)

```
             diam           circum             length
   diam     0.00251433   0.00419833   0.00105117
   circum   0.00419833   0.03793333   0.02292167
   length   0.00105117   0.02292167   0.07944733
```

Eigenanalysis of the Covariance Matrix

```
Eigenvalue  0.089695  0.028218  0.001982
Proportion    0.748     0.235     0.017
Cumulative    0.748     0.983     1.000
Variable            PC1       PC2       PC3
diameter(cm)        0.031     -0.131    0.991
circumf(cm)         0.407     -0.904    -0.132
length(cm)          0.913     0.407     0.026
```

Table 3.14 shows the result of the 1 X 1 round pipe data using PCA.

Table 3.14 :**CORRELATION**

	D	C	L
D	1	0.044	-0.283
C	0.044	1	0.034
L	-0.283	0.034	1

Covariances: diameter(cm), circumf(cm), length(cm)

	diam	circum	length
diam	0.00325100	0.00035167	-0.00335833
circum	0.00035167	0.01943333	0.00100000
length	-0.00335833	0.00100000	0.04333333

Eigenanalysis of the Covariance Matrix

Eigenvalue	0.043651	0.019406	0.002960
Proportion	0.661	0.294	0.045
Cumulative	0.661	0.955	1.000

Variable	PC1	PC2	PC3
diameter(cm)	-0.082	0.030	0.996
circumf(cm)	0.040	0.999	-0.026
length(cm)	0.996	-0.038	0.084

From Table 3.11,3.12,3.13 and 3.14,the correlation matrix shows a very low relationship(in appendices Table*) between the variables which indicate that there is no further need to conduct principal component analysis.

Also, Table 3.11,3.12,3.13 and 3.14,shows that in each of the type of pipe produced, the diameter contribute higher variation to the total variation in the production process. This is clearly demonstrated by the eigen values and in the Appendix section, Figures m, n, o and p prove that by the sharp drop from the first eigen value in each graph.

(3)

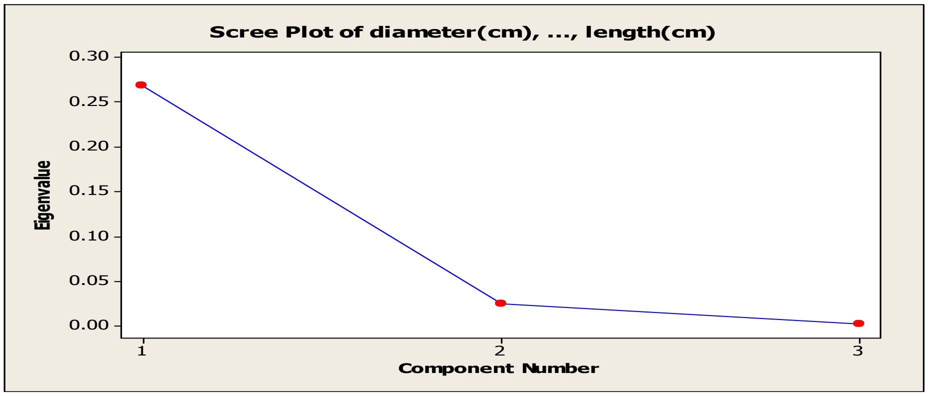


Fig 3-1

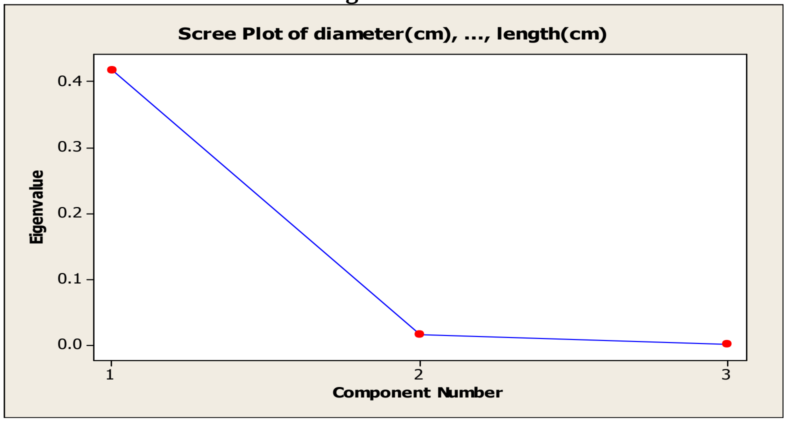


Fig 3.2

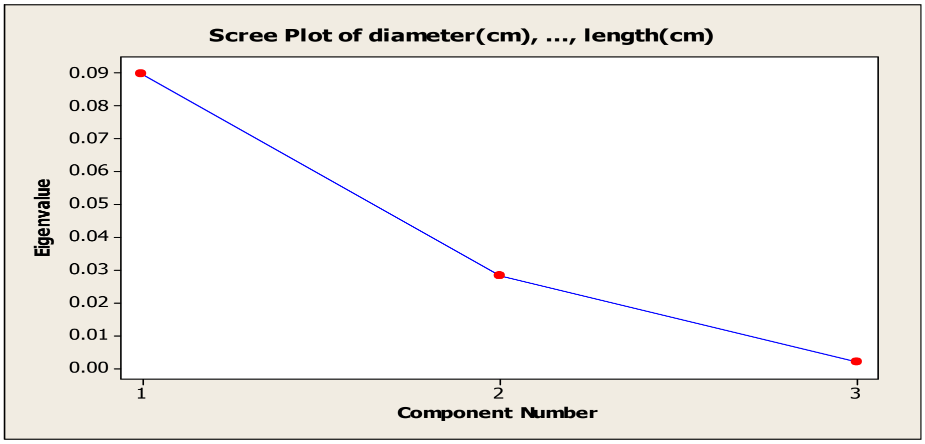
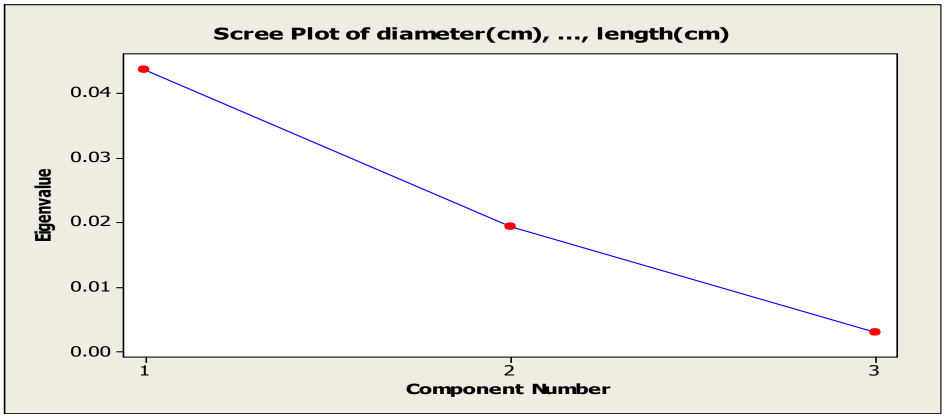


Fig 3.3



Fig, 3.4

References

[1.] Chatfield C. and Collins A.J (1980) Introduction to Multivariate Analysis
[2.] Chapman and Hall, Cambridge
[3.] Hotelling, H. (1931), The Generalisation of Students' ratio. Annals of
[4.] Mathematical Statistics No 2 pp 360-378
[5.] Hotelling, H. (1947), Multivariate Quality Control, in Techniques of Statistical analysis, edited by Eisenhart, Hastay, and Wallis. McGraw- Hill, New York, NY.
[6.] Jackson, J. E. (1985). Principle Component and Factor Analysis: Part 1– principal Component, Journal of Quality Technology, 12, pp. 201- 213.
[7.] Montgomery D, C and Lowry, C.D and (1995), "The performance of
[8.] Control Charts for Monitoring Process Variation," Communication in
[9.] Statistics computation and simulation vol. 2.
[10.] Montgomery, D.C (1990), Statistical Quality Control, 3rd edition John Wiley
[11.] Publish Ltd. New York.
[12.] Montgomery, D.C (2001), Introduction to Statistical Quality Control, 4th
[13.] edition John Wiley , USA
[14.] Morrison, D.F. (2005). Multivariate Statistical Methods. Belmont. CA.
[15.] Brooks /Cole
[16.] Sharaf Eldin, H. J Rashed and M El Khabeery(2006). Statistical Process
[17.] Control Charts applied to Steelmaking Quality improvement. Quality
[18.] Technology and Quantitative management Vol 3.
[19.] Publish Ltd. New York.