Products Test time Improvement: Statistical Process Control and Process Capability Analysis with Non-normal Data

Adisorn K.[#1], Somsak Ch.[*2]

[#]*College of Data Storage Technology and Applications,* [*]*Department of Electronic Engineering, Faculty of Engineering, King Mongkut's Institute of Technology Ladkrabang, Chalongkrung Rd., Ladkrabang, Bangkok 10520., Thailand.*
[1]`adisorn44@hotmail.com`, [2]`kchsomsa@kmitl.ac.th`

Abstract— **This paper proposes a product test time reduction where a hard disk manufacturing is taken as a case. Our method employs** *statistical process control (SPC)* **and** *process capability analysis (PCA)* **as our tools. Algorithm verification was made upon the collected data of a specific product type. Box-Cox transformation has been applied to those data sets that hold non-normal distribution. By applying the method to major time-consuming sub-processes, the test times could be reduced at a considerable level.** *Minitab program* **was used as a mathematical analysis tool. The obtained result has convinced us the benefit of non-normal distribution to normal distribution data transformation.**

Keywords— **Hard disk testing, SPC, PCA, Non-normal distribution, normality test**

I. INTRODUCTION

Under the highly competitive environment of many commercial products especially in the IT market, the statistic process control is now widely used for the purpose of process improvement. In conventional case, data with normal distribution are considered. For the case that data with non-normal distribution are to be analyzed correctly and the obtained results should consistently hold their true representation, a special technique must be deployed.

At the present days, hard disk testing is generally a time consuming process with high inventory cost. Clearly, improving the test time is not only reducing the manufacturing cost but also increasing the productivity. The challenge of the work is that "Can we improve the test time while maintaining the customer's satisfaction with minimum unit reworking?" In some particular cases when the pre-recorded data have shown their non-normal distribution: -what is the best way to analyze data with such features? The rest of this paper is organized as follows. A general process of hard disk testing will be given in the rest of this section. The concerned mathematics is outlined in section 2 whilst section 3 elaborates of SPC and PCA. The experiment set up is given in section 4 together with some obtained results. Finally, the work is concluded in section 5.

To get the understanding of applications of SPC and PCA in hard disk manufacturing, let's first have some understanding of general hard disk testing process. Some industries perform the hard disk testing based on the following steps

1) *Functional test,* microcode installation, and parameter adjustment for read-write optimization, take approximately 10% of the total testing time.
2) *SRST test;* i.e. inspection of media surface and record defect locations under squeeze condition, takes approximately 76% of the total testing time.
3) *Final test;* test the performance function based on customer requirements, takes approximately 14% of the total testing time.

Each main process mentioned above actually includes 10-15 sub-processes. One may notice that SRST test is the most time consuming step that takes 3/4 of the total testing time. Thus, we are focusing at this step to see how well SPC and PCA can help.

II. MATHEMATICS

A. Basic Statistics

On one hand, statistics is the method to collect the minority of data or samples to predict the majority of data or populations that one cannot work out all.
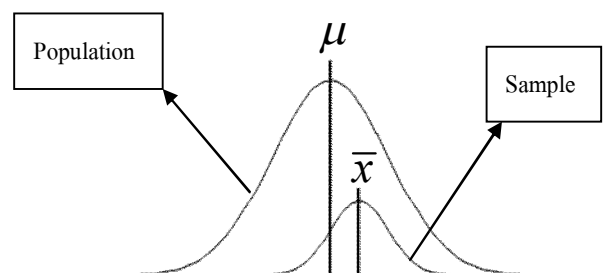

Fig. 1 Statistics data analyzing process

Fig. 2 Population and Sample

On another hand, statistics is another type of tools that interprets raw data to be important meanings and allows people who are in charge making decisions which are

unnecessary to be always according to the meanings [1], [2]. Fig. 2 shows the close population data distributions compared to sample data under correct data collection conditions. The parameters explained populations called "Population Parameters" include, μ; population average, and σ; population standard deviation. Parameters explain samples called "Sample Parameters" include, $\overline{x}$; sample average, and S; sample standard deviation.

Available of data groups are conventionally used to explain "Sample" or "Populations" to achieve the most points of view. One technique that generally applied to explain data is a term called "central tendency". There are three basic statistics parameters available to explain this term; namely mean, median, and mode.

Mean (average or arithmetic mean) is depicted as shown in eqn. (1) above. It is widely used in most cases.

$$\overline{x} = \frac{1}{n}\sum_{i=1}^{n} x_i \qquad (1)$$

Median and mode are also applied when data are needed to be analyzed in order to verify that they have natural deviation because these will affect the central tendency value. Tendency can be considered as data quality respected to the size of deviation or "Dispersion" by measuring various statistics parameters include "Range: R", "Standard deviation: S" and "Variance: S^2" which are given below in eqn.(2), (3) and (4) respectively [3].

$$R = x_{\max} - x_{\min} \qquad (2)$$

$$S = \sqrt{\frac{1}{(n-1)}\sum_{i=1}^{n}\left(x_i - \overline{x}\right)^2} \qquad (3)$$

$$S^2 = \frac{1}{(n-1)}\sum_{i=1}^{n}\left(x_i - \overline{x}\right)^2 \qquad (4)$$

Finally, to confirm the data dispersion we need to test for the normal distribution. Using probability $p(x)$ plot when x is the sample value, n is the number samples, and i is an index. Probability value can be calculated as following steps. First, the data are ranked from least to large, also insert the number of step for each sample. Probability can be computed as given in eqn. (5). Scattering plot can be performed between probability and data value (for instance; x-axis for data value and y-axis for the corresponding probability). This method is suitable for data with small to moderate sample size. When the sample space is getting large, the process is quite tedious and time consuming. Moreover, the distribution moves toward normal distribution.

$$p(x_i) = \frac{i - 0.5}{n} , \quad i = 1,2,3,.....n \qquad (5)$$

B. Normality Test

As mention in the previous sub-section, the plot should be imposed on randomly selected samples. According to the plot, parameter that defines the normal distribution is the P-value. This value simply indicates how big the number of sample points lines up around (within an amount of discrepancy) the straight line of normal distribution. P-value > 0.05, implies normal distribution of tested data whilst P-value < 0.05 implies non-normal distribution of tested data.
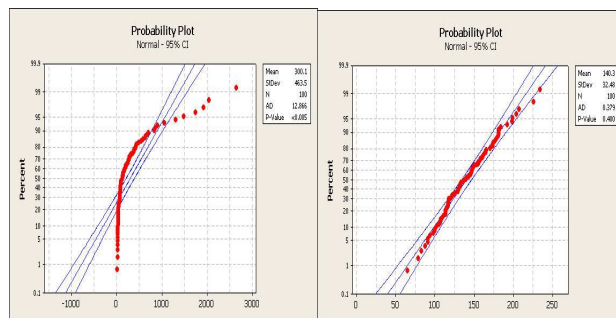
a) Non-normal distribution   b) Normal distribution

Fig. 3 Probability plot of non-normal data (P-value <0.005) and normal distribution data (P-value = 0.400)

C. Non-normality with Graphical Summary

Quite frequent that some particular processes may hold data with a non-normal distribution, but with well-understood probability distribution. One good example shown in Fig. 4 represents a data set of sub-process V073. The graphical summary shows the data that highly skewed to the left. Apparently, the distribution is non-normal.
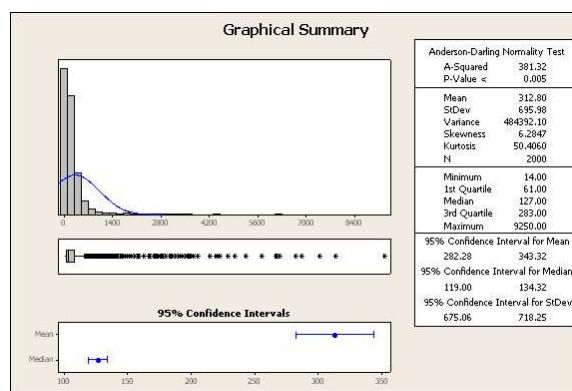

Fig. 4 Graphical summary of non-normal data

III. STATISTICS PROCESS CONTROL (SPC)

The objective of applying statistics to control the production process are to improve and to monitor the processes when stability and/or controllability are taken into account. A controlled specification line can be defined if the measured data show normal distribution property. Let CL, USL, and LSL be control line, upper control line, and lower control line respectively, specification limits can be termed as:

$$USL = \bar{x} + 3S \qquad (6)$$

$$CL = \bar{x} \qquad (7)$$
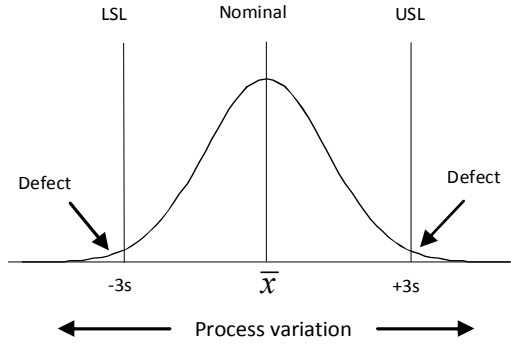
$$LSL = \bar{x} - 3S \qquad (8)$$

Fig. 5 process control standard at 6σ level

The bound of $\bar{x} \pm 3\sigma$ or 6σ control is widely used as standard one. It offers confidence interval of 99.73%.

A. Process Capability Analyze

Capability index (C_p) is a parameter indicating the control range. It implies data deviation of populations compared with the allowed deviation. Mathematical expression of the capability index is given in eqn. (9) below.

$$C_p = \frac{USL\text{-}LSL}{6S} \qquad (9)$$

When samples are considered instead of population, $\bar{x}$ is used instead of μ. C_p is now termed as C_{pk}. Its lower and upper limits are given with respected to mean $\bar{x}$ as follows.

$$C_{pu} = \frac{USL\text{-}\bar{x}}{3S} \qquad (10)$$

$$C_{pl} = \frac{\bar{x}\text{-}LSL}{3S} \qquad (11)$$

In practice, the lower one is preferable. Therefore,

$$C_{pk} = \min(C_{pu}, C_{pl}) \qquad (12)$$

B. Process Capability Analysis with Non-normal data

Some to be controlled processes may hold data with non-normal feature. When data follows a well-known distribution type, but non-normal distribution, such as a Weibull or log-normal distribution, calculation of defect rates is accomplished using the properties of the distribution. Parameters of the distribution and the specification limits must be input to the procedure. Alternatively, Those non-normal data can be mathematically transformed to the approximated normal distribution and calculate the process capability using the assumption of normality under the specification limits. Commercial package such as *Minitab* do provide this function as its standard. It also offers the calculation of the defective part per million (DPPM) of the transformed data. In the background *Minitab* uses the transformation tool so called "*Box-Cox Transformation*".

Box and Cox [4] proposed a family of power transformations on necessarily positive response variable x given by

$$x_i^{(\lambda)} = \begin{cases} \dfrac{(x_i^{\lambda}-1)}{\lambda}, & if\ \lambda \neq 0 \\ \ln(x_i), & if\ \lambda = 0 \end{cases} \text{ for } 1 \leq i \leq n \quad (13)$$

This transformation depends on a single parameter λ that is estimated using Maximum likelihood estimation (MLE) [5]. The transformation of non-normal data to normal data using Box-Cox transformation is available in most statistical software packages. Apparently, we can deploy this technique directly to evaluate PCA.

IV. EXPERIMENT

Data from 320 Gigabyte SRST hard disk tester have been studied. 4,000 hard drive units of which 2,000 are pass drives and the rest are fail drives. The "pass" and "fail" criteria are not classified by our model. Instead, they are classified by the current manufacturing criteria that are less related to SPC or PCA. Fig. 6 shows experiment steps of this research. *Minitab* package is used for data analysis. Upon the analyzed result the control parameters can be decided. The control model then can be verified to term out the process capability and performance.
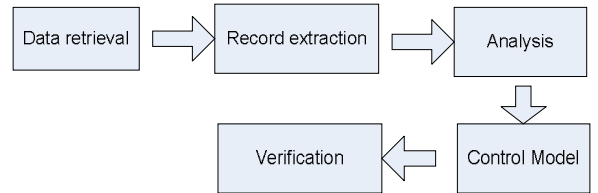

Fig. 6 Experiment procedure

In this study, we are working backward. For the process data we do analysis those data to see whether the process is stable and under control. We use PCA to check if we can cut down the test time. And if PCA is applicable –what effect it produces upon its application; i.e. FAR and FRR. Therefore, no action is taken to the real process. However, this study should demonstrate the possibility of test time improvement.

A. Data Retrieval and Extraction

A database file generated by a tester usually contains several information reveal the testing parameters and results. Common to all and of our interest are testing steps and corresponding times spent for such steps. Although the actual process involves many step sequences, only few time-

consuming steps are selected. These are shown in Fig. 7 where the process step and its corresponding average test times are plotted.
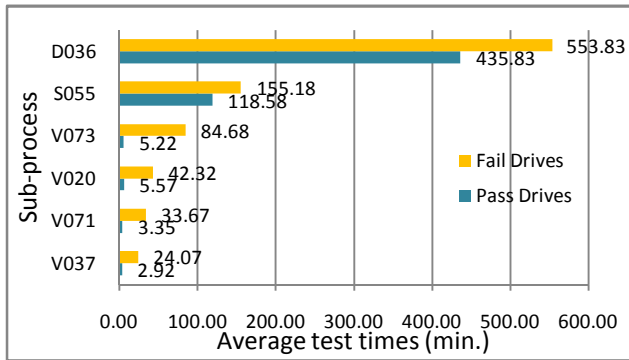

Fig. 7 Pareto chart of major sub-processes

Drives are classed to Pass and Fail according to the time limited set for the tester. For a particular testing step "D036", as a case, individual unit test times is shown in Fig. 8 below.
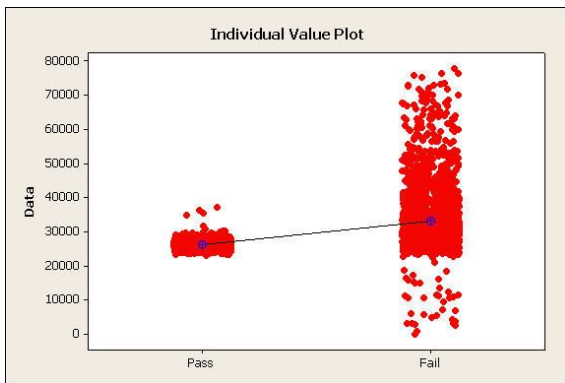

Fig. 8 D036 individual unit test times (sec.)

Refer to the said figure, we can clearly note the variance and range of data in each groups. The fail group holds larger variation. Regarding this observation we can use the data of the pass group to define the statistic parameters.

The extracted data were tested for their normality by examining the P-value. The obtained results are shown in table I. Those with P-value less than 0.005 are assumed to have non-normality distribution and must be transformed for SPC and PCA applications.

TABLE I NORMALITY TEST OF THE DISTRIBUTION (P-VALUE)

V020	D036	V037	S055	V071	V073
<0.005	0.246	<0.005	<0.005	<0.005	<0.005

B. Simulation of Process Capability with Non-normal Data and Result

Distribution of the raw-collected data (V073) is shown in Fig. 4 in the previous section. The non-normal distribution can be clearly observed. With Box and Cox transformation defined

by eqn. (13), the new distribution is shown in Fig. 9 where the transform parameter was set to $\lambda = 0$.

This transformation estimates the DPPM to be 68,000 pieces or 6.8%. The process capability c_{pk} of 0.54 is obtained as USL was set to 6σ (USL=773 sec.). The control is set to one side specification (using only upper specification limit).
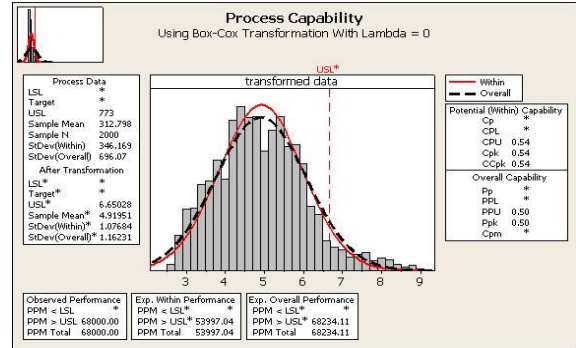

Fig. 9 Process capability of Box-Cox transformed data

Box-Cox transformation parameter can be made varied as shown in Table II. By comparing the process capability indices and defective part per million (DPPM) for each lambda, the optimal one can be selected.

TABLE II
Cpk. AND DPPM. CHANGES ACCORDING TO TRANSFOMATION PARAMETER, LAMBDA

Lambda (λ)	Process capability (V073) according to Box-Cox Transformation			
	Cpk Indices	DPPM (Total)		
		Observed	Within	Overall
-2.0	0.26	6.80%	21.93%	31.56%
-1.5	0.29	6.80%	19.03%	26.69%
-1.0	0.35	6.80%	14.70%	19.65%
-0.5	0.44	6.80%	9.4%	11.60%
0	0.54	6.80%	5.53%	6.82%
0.5	0.57	6.80%	4.42%	9.95%
1.0	0.44	6.80%	9.18%	25.42%
1.5	0.21	6.80%	26.56%	41.85%
2.0	0.01	6.80%	49.33%	49.84%

A main objective of applying specification limit to each sub-process is to reduce the test times. With defined control level, the obtained time-saving of each sub-process is shown in Table III below. Many sub-processes hold the time-saving of about 10% or more. For example, V073 can save 25% and D036 can save 14%. However when we look into details, V073 have much less impact. This is because the average test time of V073 is only 5.2 minutes whilst that of D036 is 435.8 minutes. Outcomes (yield) of each sub-process can be computed based on the fault rejected drives and the total number of pass drives.

Proc.	Quality Level	Spec. limit min.	Cpk	Time saving min.	Time saving %	Yield %
V020	6σ	7.65	1.17	19.19	6.73	94.30
D036	3σ	491.2	1.01	60.00	13.77	99.35
V037	6σ	4.02	1.20	11.00	7.56	90.80
S055	3σ	132.48	1.00	14.77	0.21	98.25
V071	6σ	5.70	1.52	15.72	9.41	95.80
V073	6σ	12.48	1.30	41.90	24.57	93.20

Applying only specification limit by mean of test times can result in pass drives to be rejected and fail drives to be accepted. In principle, we should reject only the fail drives and accept only the pass drives. However, the situation of rejection of good drives (termed as fault rejection rate: FRR) and acceptance of bad drives (termed as fault acceptance rate: FAR) cannot be totally avoided. We have investigated the issue with the merit of test times saving. The result is shown in Fig. 10 below. Be kept in mind that in this experiment all 4,000 units are input to each testing step. If the method proposed in this study is brought to practice, the genuine rejected drives and the fault rejected drives will be reworked upon their failure symptoms. Obviously, the lower FRR is the better. Genuine pass drives and fault accepted drives will be passed to the subsequent process step. Repeatedly tested for many step sequences, the FAR should be reduced to minimum number. Of course, we should not deliver any fail drives to the customer at all. Therefore, the FAR should be zero at the final step.
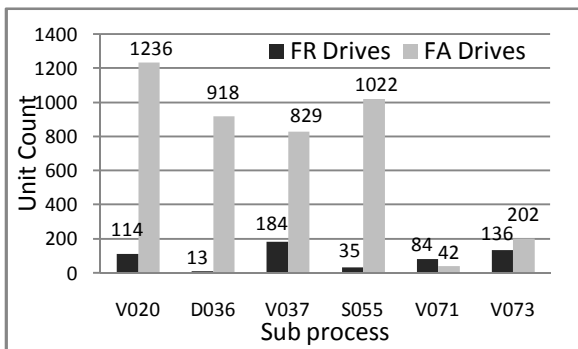

Fig. 10 Fault rejected drives and Fault accepted drives after the application of PCA to each sub-process

V. CONCLUSION

In this study we have demonstrated the application of SPC and PCA to a product manufacturing where hard disk drives test time was taken as a case. To overcome the data with nor-normality distribution, Box-Cox transformation with $\lambda = 0$ was applied. Only sub-processes that contribute large test times were selected. Clearly shown in table III, the test times can be reduced to some certain amount. The proposed method can be either applied to particular selected sub-processes or applied to all sub-process. A drawback of the method is that the control still produce fault accepted units that resulting in the fail unit to be fed to the subsequent sub-process. Moreover, the resulted fault rejected units can pile up the unit cost because of its reworking.

REFERENCES

[1] K. Poypanitchararenpol "Process Capability Analysis (PCA)", Bangkok, Technology Promotion Association (TPA) (Thailand-Japan), 2008
[2] Ch. Srikeawsew "Statistics Learning and Application Website", www.statistics.ob.tc
[3] G. E. P. Box and D. R. Cox, "An analysis of transformation", J. Roy, Stat Soc. B. 26: 211-252, 1964
[4] D. Montgromery, "Introduction of statistical quality control", 5th ed., Wiley, New York, 1996
[5] L. C. Tang, S. E. Than, "Computing process capability indices for non-normal data", 1999
[6] K. Poypanitchararenpol "Statistics for Engineering", Bangkok, Technology Promotion Association (TPA) (Thailand-Japan), 1999
[7] T. Sincharu "Researching and Analysis Statistics data by SPSS", Ed. 9, Bangkok, Business R&D, 2008
[8] K. Rezaie, B. Ostadi and M.R. Taghizadeh. "Application of Process Capability and Process Performance Indices" Journal of Applied Sciences 6(5): 1186-1191, 2006.