Journal of Industrial Engineering International *Islamic Azad University, Tehran South Branch*
September 2005, Vol. 1, No. 1, 20 - 27

Estimating process capability indices using ridge regression

S. Raissi*

Department of Industrial Engineering, Islamic Azad University, Tehran South Branch, Tehran, Iran

R. Noorossana

Department of Industrial Engineering, Iran University of Science and Technology, Tehran, Iran

Abstract

Process capability indices show the ability of a process to produce products according to the pre-specified requirements. Since final quality characteristics of a product are usually interrelated to its previous amounts in earlier workstations, one need to model and consider the relationship among them to assess the process capability properly. Hence, conducting process capability analysis in multivariate environment is inevitable; unfortunately, the analysis in multivariate environment is usually complex and requires extensive calculations. Sometimes it is preferable to simplify the analysis by assuming independency among quality characteristics and evaluating process performance with respect to each individual quality characteristic using univariate process capability indices such as C_P, C_{PK}, C_{PM}, and C_{PMK}. However, this simplification introduces some error in the analysis leading to under or overestimation of the process capability index. This paper models the interrelationship among quality characteristics that are produced in different workstations to provide an overall process capability index. Ridge residual regression is used as a vehicle to evaluate process capability and helps quality engineers to provide a reasonable quality policy for controlling and reducing variation in quality characteristics.

Keywords: Process capability indices; Variance propagation; Ridge regression; C_P; C_{PK}; C_{PM}; C_{PMK}

1. Introduction

There are several well-known methods for estimating the potential and actual process capability indices. Reliable estimation of these indices could be obtained through the reliable estimates of the process mean and variance along with careful selection of tolerance limits and target values. Process capability analysis could be conducted both in univariate and multivariate environments. In univariate environment, a single quality characteristic of a product is considered and process performance is evaluated with respect to that quality characteristic. Many researchers including Kane [10], Marcucci and Beazley [12], Chan et. al. [3], Choi and Owen [6], Spiring [23], Koons [11], Wheeler and Chamber [28], Pearn et. al. [20], Bissel [1], Pearn W. L. and Chen K. S. [17] and Chen J.P. and Chen K.S. [4] have contributed to the development of univariate process capability indices when quality characteristic of interest follows a normal distribution. Some authors including Munechika [14], Clemets [8], Wright [29], Somerville and Montgomery [22], Chen and Ding [5] and Chou et. al. [7] have discussed and developed indices when distribution of the quality characteristic under study is non-normal. However, there are many situations in which the interrelationship among quality characteristics of a product must be considered in order to evaluate the process performance properly. This fact is sometimes overlooked when a set of correlated quality characteristics are evaluated individually using separate univariate capability indices. Such an approach ignores the correlation that exists between the quality characteristics and results in frequent process adjustments and eventually will lead to an unstable process. Many authors including Hubele

* *Corresponding author. E-mail: raissi@azad.ac.ir*

et. al. [9], Chan et. al. [3], Taam et. al. [24], Nickerson [16], Niverthi and Dey [17], Shahriari et. al. [22], Wang et. al. [27], Wang and Du [25], Wang and Hubele [26] and Pearn et. al. [19] have contributed to the development of multivariate process capability indices. However, most of the proposed procedures involve complex calculations which make them less attractive to practitioners.

The purpose of this paper is to develop a procedure for evaluating process capability when quality characteristics of a product are the outputs of different workstations. The proposed procedure is based on ridge regression and leads to proper estimation of the process capability indices such as C_P, C_{PK}, C_{PM}, and C_{PMK}.

2. Model description

Consider a product with k quality characteristics and each quality characteristic having its own technical tolerance (TT). In general,

$$TT_i=[LSL_i, T_i, USL_i] \text{ for } i=1,2,\ldots,k, \tag{1}$$

Where LSL_i, USL_i, and T_i are the lower specification limit, the upper specification limit, and the target value (if it exists) for the i^{th} quality characteristic, respectively. Suppose these quality characteristics are the outputs of various workstations or sub-processes. It is obvious that the variability in the final product is the result of the variability induced in each quality characteristic in different workstations. This point is discussed in the following example.

3. Typical Example

Consider the product in Figure 1 which has four quality characteristics X_1, X_2, X_3 and X_4 representing part length, step length, head diameter, and step diameter, respectively. All the characteristics are variable types with specific technical tolerances.

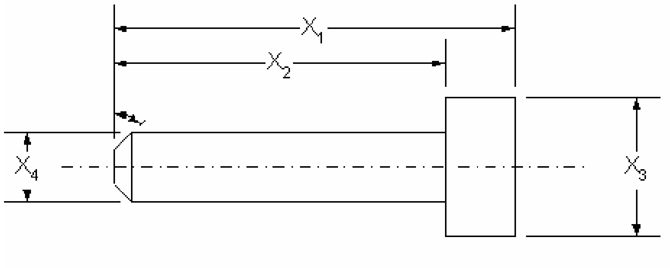


Figure 1. Quality characteristics of a product.

Appendix 1 presents the detailed operational sequences and workstations required to produce the above product. The major steps involved in workstations (WS) are as follows:

a) Selecting a free length rod with the external diameter X_{31} (in WS # 1 as denoted in Appendix 1).
b) Cutting the rod to length X_{11} (in WS # 2).
c) Surface machining to final diameter X_{32} (in WS # 3).
d) Machining the step in length X_{21} (in WS # 2) with proper cutting depth so that the final step diameter will be X_{41} (in WS # 3).
e) Face machining and rounding edge such that X_{22} (in WS # 3) will be the step length. The quality characteristic X_{12} (in WS # 3) denotes the part length.
f) Improving the mechanical properties through heat treatment. The final quality characteristics, X_1, X_2, X_3 and X_4 are the outputs of these sub-processes.

Since heat treatment can cause dimensional changes in the part, quality characteristics of the product will be a function of the relevant characteristics in the previous workstations. However, this function is generally unknown and process designers can introduce the variables involved in the function by the means of logical relations and cause and effect analysis in production technology. Furthermore, by getting appropriate samples and analyzing the correlation coefficients matrix, the analyst can get some information about the impact and contribution of each workstation on the quality characteristics of interest. In the illustrated example, the quality characteristics can be considered to be related to X_{ij}'s as follows:

$$\begin{aligned} X_1 &= f_1(X_{11}, X_{12}) \\ X_2 &= f_2(X_{21}, X_{22}) \\ X_3 &= f_3(X_{31}, X_{32}) \\ X_4 &= f_4(X_{41}, X_{31}, X_{32}) \end{aligned} \tag{2}$$

Dispersion analysis of the quality characteristics shows that the final value of the i^{th} quality characteristic that is shown by Y_i is influenced by the operations done in the m previous workstations. If the value of the i^{th} quality characteristic in workstation j is shown by X_{ij} then Y_i for each quality characteristic is a function of X_{ij}s in m previous workstations, i.e.:

$$Y_i=f_i(X_{i1}, X_{i2}, \ldots, X_{im}) \quad i=1,2,\ldots,k \tag{3}$$

The parameters of the model can be estimated by applying Ordinary Least Square (OLS), when no correlation exists among independent variables, or ridge regression method, when independent variables are correlated. According to Neter et. al. [15] the results from the regression analysis are still applied if the following conditions hold:

1. The conditional distributions of Y_i, given X_i, are normal and independent.
2. The X_i's are independent random variables, whose probability distribution does not involve the regression parameters.

As long as these conditions are met, all results on estimation still hold even though the X_i's are now random variables. Let μ_{ij} and σ_{ij} be the mean and variance of the i^{th} quality characteristic in workstation j with technical tolerances given by $[LSL_{ij}, T_i, USL_{ij}]$. In general, one of the two following cases may occur:

a) Quality characteristics in different workstations are independent

In this case, the i^{th} quality characteristic of the final product can be represented using the following general linear regression model:

$$Y_i = a_0 + \sum_{j=1}^{m} a_j X_{ij} + U_i \quad ; i = 1, 2, 3, ..., k \tag{4}$$

where U_i is white noise having normal distribution with mean zero and a given variance. The regression coefficients can be estimated by ordinary least squares estimators. If X_{ij}s follow normal distribution, then Y_i also has a normal distribution with the following parameters:

$$\mu_{\hat{Y}_i} = \sum_{j=1}^{m} a_j \mu_{ij} \quad ; i = 1, 2, 3, ..., k \tag{5}$$

$$\sigma_{\hat{Y}_i} = \sqrt{\sum_{j=1}^{m} a_j^2 \sigma_{ij}^2} \quad ; i = 1, 2, 3, ..., k \tag{6}$$

Proper estimates of $\mu_{\hat{Y}_i}$ and $\sigma_{\hat{Y}_i}$ can be obtained through reliable estimations of μ_{ij} and σ_{ij}. When X_{ij}s are assumed to be normal random variables, the most typical process capability indices namely C_P, C_{PK}, C_{PM}, and C_{PMK} can be calculated using the following equations:

$$C_{P_i} = \frac{USL_i - LSL_i}{6\sigma_{Y_i}} \tag{7}$$

$$C_{PK_i} = Min\left\{\frac{USL_i - \mu_{Y_i}}{3\sigma_{Y_i}}, \frac{\mu_{Y_i} - LSL_i}{3\sigma_{Y_i}}\right\} \tag{8}$$

$$C_{PM_i} = \frac{USL_i - LSL_i}{6\sqrt{\sigma_{Y_i}^2 + (\mu_{Y_i} - T_i)^2}} \tag{9}$$

$$C_{PMK_i} = Min\left\{\frac{USL_i - \mu_{Y_i}}{3\sqrt{\sigma_{Y_i}^2 + (\mu_{Y_i} - T_i)^2}}, \frac{\mu_{Y_i} - LSL_i}{3\sqrt{\sigma_{Y_i}^2 + (\mu_{Y_i} - T_i)^2}}\right\} \tag{10}$$

If the distributions of X_{ij}s are non-normal, different approaches such as data transformation and quantile estimation can be used to modify process capability indices. However, many researchers prefer quantile estimation over other methods recommended for non-normal situations. In quantile estimation method, appropriate quantiles of the distribution of Y_i are selected such that the desired portion of the distribution lies between the selected upper and lower quantiles. For illustration purposes, suppose $Y_{i,\,99.865}$ and $Y_{i,\,0.135}$ are used as the upper and lower quantiles of the distribution of Y_i such that 99.73 percent of the distribution lies between them. When these quantities are used in the analysis, the process capability indices would be modified to:

$$C_{P_i} = \frac{USL_i - LSL_i}{Y_{i,99.865} - Y_{i,0.135}} \tag{11}$$

$$C_{PK_i} = Min\left\{\frac{USL_i - Y_{i,50}}{Y_{i,99.865} - Y_{i,50}}, \frac{Y_{i,50} - LSL_i}{Y_{i,50} - Y_{i,0.135}}\right\} \tag{12}$$

$$C_{PM_i} = \frac{USL_i - LSL_i}{6\sqrt{\left[\frac{(Y_{i,99.865} - Y_{i,0.135})}{6}\right]^2 + (Y_{i,50} - \frac{USL_i + LSL_i}{2})^2}} \tag{13}$$

$$C_{PMK_i} = Min\left\{ \frac{USL_i - \frac{USL_i + LSL_i}{2}}{3\sqrt{\left[\frac{(Y_{i,99.865} - Y_{i,50})}{3}\right]^2 + (Y_{i,50} - \frac{USL_i + LSL_i}{2})^2}}, \frac{\frac{USL_i + LSL_i}{2} - LSL_i}{3\sqrt{\left[\frac{(Y_{i,50} - Y_{i,0.135})}{3}\right]^2 + (Y_{i,50} - \frac{USL_i + LSL_i}{2})^2}} \right\} \quad (14)$$

In the non-normal case, if we can find a distribution form for the data, one that provides a reasonable fit, then we can obtain more accurate measures of the quantiles leading to better estimation of the process capability indices.

b) Quality characteristics in different workstations are not independent

When the independent quality characteristics are correlated, intercorrelation or multicollinearity is said to exist. Multicollinearity leads to high variation in the estimated regression coefficients. There are different remedial measures available to lesser the effects of multicollinearity. However, according to Neter et. al. [15] ridge regression seems to be the most effective remedy. Ridge regression is modified general version of regular least square method by using the biasing constant *K* which is a number grater than zero. When *K* is zero, Ridge regression acts as OLS and elsewhere, the regression coefficients obtained are biased but more stable. It is vital that the value of *K* remains as low as possible and still be great enough to provide a good estimate. Refer to Montgomery and Freidman [13] for more details. It is noted that the biased estimates have much less mean square error compared to the unbiased estimates. Generally speaking, an estimator that has only a small bias but is substantially more precise is the preferred estimator because it will be closest to the true value of the parameter. By using appropriate standard statistical software the method of ridge regression could easily be applied to equation 4 to estimate the regression coefficients. Due to the presence of multicollinearity, equation 6 needs to be modified to the following equation:

$$\sigma_{Y_i} = \sqrt{\sum_{j=1}^{m} a_{ji}^2 \sigma_{ij}^2 + 2a_j a_k Cov(X_{ij}, X_{ik})} \quad (15)$$

However, the rest of the relationships presented in the previous section remain unchanged.

4. Numerical Example

To illustrate the steps involved in the calculation of the proposed procedure, consider the example discussed earlier. For the sake of simplicity, we only consider the step length (called X_2) of the rod. Based on the engineering analysis, it seems that two of the workstations have profound effect on the quality of the step length. As illustrated in appendix 1, the final step length X_2 with technical tolerance given by USL= 80.4, T=80.2 and LSL=79.9 may be modeled as a function of its previous amounts, X_{21} and X_{22}, produced in the two previous workstations. This is generally shown as $X_2 = f(X_{21}, X_{22})$.

Suppose each of the mentioned quality characteristics follows a normal distribution with the parameters given in table 1. Further assume that these quality characteristics are statistically under control and it is desired to perform process capability analysis.

Table 1. Parameters of the quality characteristics.

Quality Characteristic	X_2	X_{22}	X_{21}
Mean	80	81.4	82.5
Standard Deviation	0.1	0.3	0.3

In order to assess the overall capability of this process, 40 samples where each sample consists of three observations is collected. Table A-1 in the 2nd appendix shows the collected data.

Note that in this example, since the quality characteristics have the same dimensional scale, there is no need to standardize the data. The correlation coefficients presented in table 2 is indicative of some correlation among the three quality characteristics.

Table 2. Correlation matrix for the quality characteristics.

	X_2	X_{22}	X_{21}
X_2	1.00	0.16	0.24
X_{22}	0.16	1.0	0.41
X_{21}	0.24	0.41	1.00

Pass 2000 statistical software was applied to the data presented in Table A2 and the following ridge regression equation with a biasing factor of *K=0.005* which helped the regression coefficients to remain constant was obtained:

$$\hat{X}_2 = 64.108 + 0.063X_{22} + 0.130X_{21}$$

Figure 2 presents the normal probability plot for the residuals. This figure confirms the normality assumption.

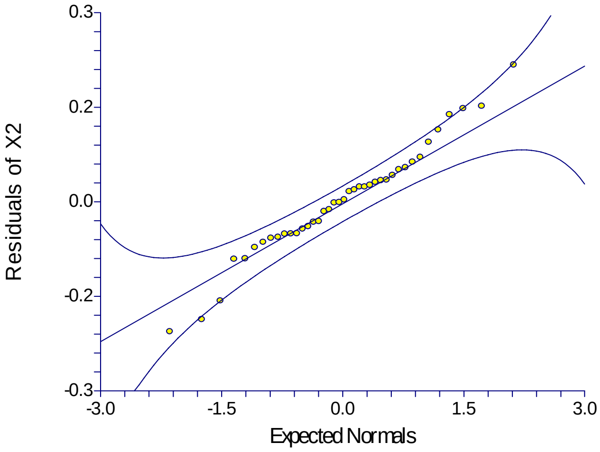


Figure 2. Normal probability plot of residuals with 95% confidence interval.

Using equations 5 and 15, the parameters of the final quality characteristic X_2 can be calculated as follows:

$$\mu_{\hat{X}_2} = 64.108 + 0.063(81.4) + 0.130(82.5) = 79.961$$

$$\sigma_{\hat{X}_2} = \sqrt{(0.063)^2.(0.3)^2 + (0.13)^2.(0.3)^2 + 2(0.063)(0.130)}$$
$$= 0.045$$

Once these values are replaced in equations 7 through 10, the true process capability indices for the rod's step length are computed as C_P*=1.83,* C_{PK}*=0.45,* C_{PM}*=0.34* and C_{PMK}*=0.14*. Now the process engineer can use the appropriate index or indices to make the required adjustments to the process.

5. Conclusions

Univariate process capability indices are well known to quality engineers in different industries. They could be easily applied to estimate potential and actual process performance when process is under statistical control. Due to existence of correlation among quality characteristics, process capability analysis in multivariate environments is not usually an easy task to perform. In this paper, by applying regression method, final quality characteristic was modeled as a function of its previous amounts in earlier workstations and a relatively easy method based on multivariate ridge regression analysis was developed to estimate four typical common process capability indices known as C_P, C_{PK}, C_{PM}, and C_{PMK}.

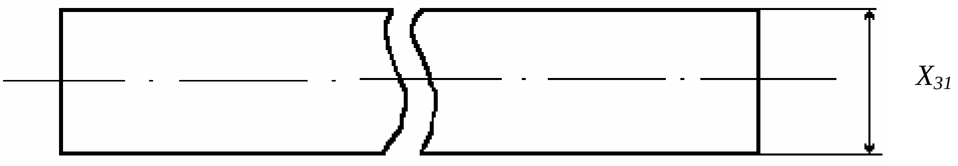
This new approach has the following advantages:

1. It is constructed based on a well-known multiple regression method which can be easily calculated using statistical packages.
2. It can be applied to estimate many well-known process capability indices for both cases of normal and non-normal distributions.
3. The critical workstations can be identified using the regression coefficients.

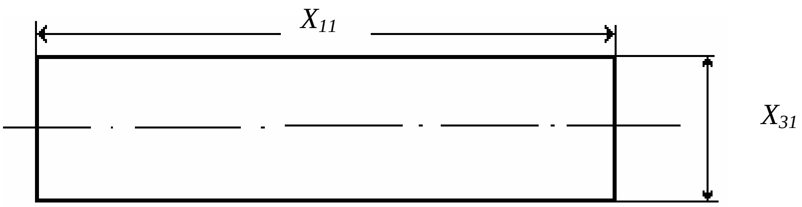
A numerical example was also considered to model the interrelationship among quality characteristics in various workstations and to show the efficiency of the proposed method in terms of the amount of computations involved.

Appendix 1

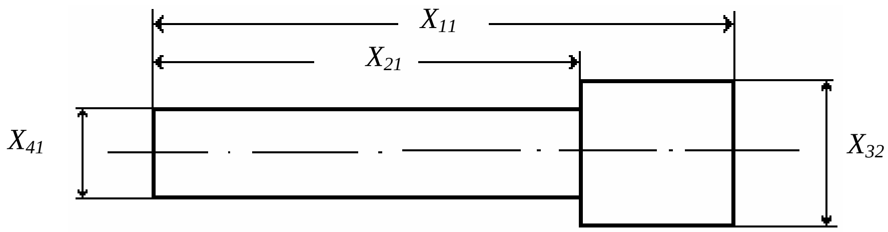
The operational sequence in each workstation for the illustrated example is as follows:
Starting point: Selecting a free length rod with the external diameter X_{31} .



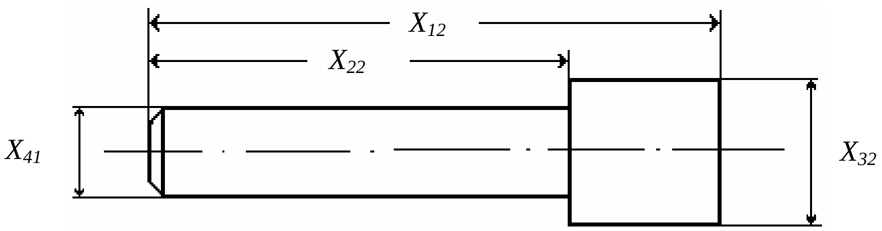
1st workstation: Cutting the rod to length X_{11} .



2nd workstation: Surface machining to final diameter X_{32} and machining the step in length X_{21} with proper cutting depth so that the final step diameter will be X_{41} .



3rd workstation: Face machining and rounding edge such that X_{22} will be the step length and X_{12} denotes the part length.



4th workstation: Heat treating the final quality characteristics such as X_1, X_2, X_3, and X_4 are the final outputs of all sub-processes (Refer to Fig. 1).

Appendix 2

TableA1. 40 samples on the step length of the rod in three workstations (WS) related to numerical example.

Sample No.	Step length in			Sample No.	Step length in		
	Final WS (X_2)	WS # 3 (X_{22})	WS#2 (X_{21})		Final WS (X_2)	WS # 3 (X_{22})	WS#2 (X_{21})
1	80.0	81.6	82.8	**21**	79.9	81.4	82.5
2	80.0	81.7	82.1	**22**	80.1	81.7	82.6
3	80.1	81.3	82.2	**23**	80.0	81.6	82.2
4	80.0	81.0	82.3	**24**	79.9	81.4	82.3
5	80.2	81.4	82.4	**25**	79.9	81.8	82.8
6	80.0	81.5	82.3	**26**	80.0	81.1	82.3
7	80.0	81.5	82.6	**27**	79.8	81.1	82.6
8	80.1	81.1	82.6	**28**	79.9	81.1	82.4
9	80.0	81.0	82.5	**29**	80.1	81.6	82.5
10	80.1	81.2	82.9	**30**	80.0	81.3	82.9
11	80.0	81.4	82.5	**31**	80.1	81.3	83.1
12	80.0	81.6	82.3	**32**	80.0	81.8	82.7
13	79.9	80.9	82.0	**33**	80.1	81.5	83.0
14	80.0	81.5	82.3	**34**	79.9	81.3	82.0
15	80.2	81.8	82.7	**35**	80.0	81.2	82.6
16	80.2	81.5	82.9	**36**	79.9	81.0	82.7
17	79.9	81.2	82.4	**37**	79.9	81.8	82.0
18	79.8	81.7	82.4	**38**	80.0	81.9	82.7
19	79.9	81.2	82.0	**39**	79.9	80.9	82.3
20	80.0	81.6	82.3	**40**	80.0	81.7	82.6

References

[1] Bissel, Derek, 1994, *Statistical Methods for SPC and TQM.* Chapman & Hall, New York, NY, p. 240.

[2] Chan, L. K., Cheng, S. W. and Spiring F. A., 1991, A multivariate measure of process capability. *International Journal of Modeling and Simulation*, 11(1).

[3] Chan, L. K., Cheng, S. W. and Spiring, F.A., 1988, *The Robustness of the Process Capability Index to Departures from Normality.* Statistical Theory and Data Analysis, Proceeding of Second Pacific Area Statistical Conference, Tokyo, 223-239.

[4] Chen, J. P. and Chen, K. S., 2004, Comparison of two process capabilities by using indices C_{PM}: an application to a color STN display, *International Journal of Quality & Reliability Management,* 21(1), 90-101.

[5] Chen, J. P. and Ding, C. J., 2001, A new process capability index for non-normal distributions. *International Journal of Quality & Reliability Management,* 18(7), 762-770.

[6] Choi B. C. and Owen, D. B., 1990, A study of a new process capability index. *Communications in Statistics, Theory and Methods,* 19(4), 1231-1245.

[7] Chou, C. Y., Chen, C. H. and Liu, H. R., 2004, Effect of non-normality on the economic design of warning limit X-bar charts. *Quality Engineering,*16(4), 567–75.

[8] Clements, J. A., 1989, Process capability calculations for non–normal distributions. *Quality progress,* September, 95-100.

[9] Hubble, N. F, Shahriari, H. and Cheng, G. S., 1991, *A Bivariate Process Capability Vector in Statistics and Design in Process Control.* Marcel Dekker, NY, 299-310.

[10] Kane, VE, 1986, Process Capability Indices. *Journal of Quality Technology,* 18(1), 41-52.

[11] Koons, George F., 1992, *Indices of Capability: Classical and Six Sigma Tools.* Addison Wesley Publishing Co., Reading, MA.

[12] Marcucci, M. O. and Beazley, C. C., 1988, *Capability Indices: Process Performance Measure.* 42nd ASQC Annual Quality Congress Trans., Dallas, TX, 516-523.

[13] Montgomery, D. C. and Friedman, D. J., 1993, *Prediction using regression models with multicolinear predictor variables.* IIE Transaction, 25, 73-85.

[14] Munechika, M., 1986, *Evaluation of Process Capability for Skew Distribution.* Proceeding of the 30th EOQC Conference, Stockholm, 383-390.

[15] Neter, J., Wasserman, W. and Kutner, M. H., 1989, *Applied Linear Regression Models.* 2nd edition, Irwin.

[16] Nickerson, D. M., 1994, Construction of a conservative confidence region from projections of an exact confidence region in multiple linear regression. *The American Statistician,* 48, 120-124.

[17] Niverthi, M. and Dey, D. K., 1995, *Multivariate Process Capability: a Bayesian Perspective.* Technical Report, Department of Statistics, University of Connecticut, Storrs.

[18] Pearn, W. L. and Chen, K. S., 2002, One-sided capability indices C_{PU} and C_{PL}: decision making with sample information. *International Journal of Quality & Reliability Management,* 19(3), 221-245.

[19] Pearn, W. L., Ko, C.H. and Wang, K. H., (2002), A multi-process performance analysis chart based on the incapability index C_{pp}: An application to the chip resistors. *Microelectronics Reliability,* 42, 1121–1125.

[20] Pearn, W. L., Kotz, S. and Johnson, N. L., 1992, Distributional and inferential properties of PCI. *Journal of Quality Technology,* 24(4), 216-231.

[21] Shahriari, H., Huebele, N. F. and Lawrence, F. P., 1995, *A Multivariate Process Capability Vector.* Proceeding of The 4th Industrial Engineering Research Conference, Institute of Industrial Engineers, 304-309.

[22] Somerville, S. and Montgomery D. C., 1996, Process capability indices and non–normal distributions. *Quality Engineering,* 9(2), 305-316.

[23] Spiring F. A., 1991, Comments on J. Mirabella's letter. *Quality Progress*, p. 24.

[24] Taam, W., Subbaiah, P. and Liddy, J. W., 1993, A note on multivariate capability indices. *Journal of Applied Statistics,* 20, 339–351.

[25] Wang, F. K. and Du, T. C. T., 2000, Using principal component analysis in process performance for multivariate data. *Omega - The international Journal of Management Science,* 28, 185–194.

[26] Wang, F. K. and Hubele, N. F., 2001, Quality evaluation of geometric tolerance regions in form and location. *Quality Engineering,* 14, 203–209.

[27] Wang, F. K., Hubele, N. F., Lawrence, P., Miskulin, J. D. and Shahriari, H., 2000, Comparison of three multivariate process capability indices. *Journal of Quality Technology*, 32(3), 263–275.

[28] Wheeler, D. J. and Chambers, D. S., 1992, *Understanding Statistical Process Control.* 2nd edition, SPC Press, Knoxville, TN.

[29] Wright, P. A., 1995, A process capability index sensitive to skewness. *Journal of Statistics & Computer Simulation,* 52, 195-203.