Journal of Modern Applied Statistical Methods

Volume 7 | Issue 2 Article 23

11-1-2008

A Methodology to Improve PCI Use in Industry

Milind A. Phadnis
The University of Alabama at Birmingham, phadnismilind@gmail.com

Matthew E. Elam
Texas A&M University, Matthew_Elam@tamu-commerce.edu

Journal of Modern Applied Statistical Methods
November, 2008, Vol. 7, No. 2, 588-596

1538 – 9472/08/$95.00

A Methodology to Improve PCI Use in Industry

Milind A. Phadnis
The University of Alabama at Birmingham

Matthew E. Elam
Texas A&M University-Commerce

This article presents the development of a methodology using decision trees to resolve issues in industry with using process capability indices (PCIs). The methodology forms the structure of a prototype decision support system (PDSS) for PCI selection, calculation, and interpretation. Download instructions for the PDSS are available at http://program.20m.com.

Key words: Process capability index; decision tree; control chart; normality check; decision support system.

Introduction

Process capability may be defined as the ability of a process to achieve a certain objective. Process capability indices (PCIs) have been used for some time to provide a quantitative measure of this ability. Many PCIs have been developed in the literature for different situations encountered by industry. However, industry has not been able to achieve the full benefit from using PCIs for the following reasons:

- Abuse of PCIs by violating their underlying statistical assumptions;
- Lack of practical usage of multivariate PCIs and their interpretations;
- Unavailability of PCIs for data limited (short-run) situations;
- Shortcomings in software packages capable of calculating PCIs; and
- Lack of appropriate usage of PCIs in data with asymmetric specifications.

Milind A. Phadnis is a Research Assistant and is pursuing a PhD. in Biostatistics. Email him at phadnismilind@gmail.com. Matthew E. Elam is an Associate Professor of Industrial Engineering at Texas A&M University-Commerce and is an ASQ Certified Quality Engineer. Email him at Matthew_Elam@tamu-commerce.edu.

This article details a methodology for resolving the above mentioned issues. It makes use of a top-down decision making approach to select the appropriate PCI(s) regarding particular kinds of data. It also makes use of the latest theory available in the statistical literature pertaining to the definitions and properties of various PCIs. The methodology was developed by considering the situations in which industry needs PCI results, determining the PCIs available for these situations, and determining the decision-making process for handling these situations simultaneously.

The methodology forms the structure of a prototype decision support system (PDSS) built in order to facilitate easy usage in industry (Phadnis, Elam, Fonseca, Batson, & Adams, 2005). The PDSS analyzes the process data, verifies the statistical assumptions necessary for handling different types of process data, selects the most appropriate PCI(s) depending on the process parameters, calculates the PCI(s), provides a practical interpretation of the PCI(s), and guides the user towards the source of corrective action needed, if any. Visual Basic 6.0 and Microsoft Excel 2002 were used to design the PDSS so that it has a user-friendly graphical interface, portability, and ease of use for industry. The PDSS requires the user to enter only elementary characteristics of the collected process data, the process data itself, and the process's engineering specifications. Instructions for downloading the PDSS are available at http://program.20m.com.

Methodology

After considering the situations in which industry needs PCI results and studying the properties of the various PCIs available in the literature, the decision tree shown in Figure 1 was constructed as the backbone of the complete structure of the methodology. This decision tree presents a basic overview of the formulations used in constructing the methodology and can be further expanded into various branches and sub-branches. Thus, whenever branching is possible, a series of asterisks "*" is placed in the corresponding block to denote the same, and this particular block has been further expanded in subsequent figures in the Appendix.

As shown in Figure 1, the constructed methodology is equipped to handle the following types of data collected by the user:

- Type 1: univariate sufficient data (total number of observations $\geq$ 50), which also involves Appendix Figure 2;
- Type 2: univariate short-run data (total number of observations < 50), which also involves Appendix Figure 3; and
- Type 3: multivariate sufficient data (total number of observations $\geq$ 100), which also involves Appendix Figure 4.

The methodology adopted for selecting and evaluating PCIs is different for each of the above mentioned data types.

Type 1: Univariate Sufficient Data ($\geq$ 50 Observations)

The classifications of sufficient data as that with at least 50 observations, and a short-run situation as that with less than 50 observations, are based on the fact that the statistical properties of the commonly used PCIs do not permit calculation of an index when less than 50 observations are available as noted by Deleryd & Vannman (1998). Univariate data may further be classified into data collected in subgroups and data collected as individual observations. Each of these cases is discussed below. (See Figure 1 and Appendix Figure 2.)

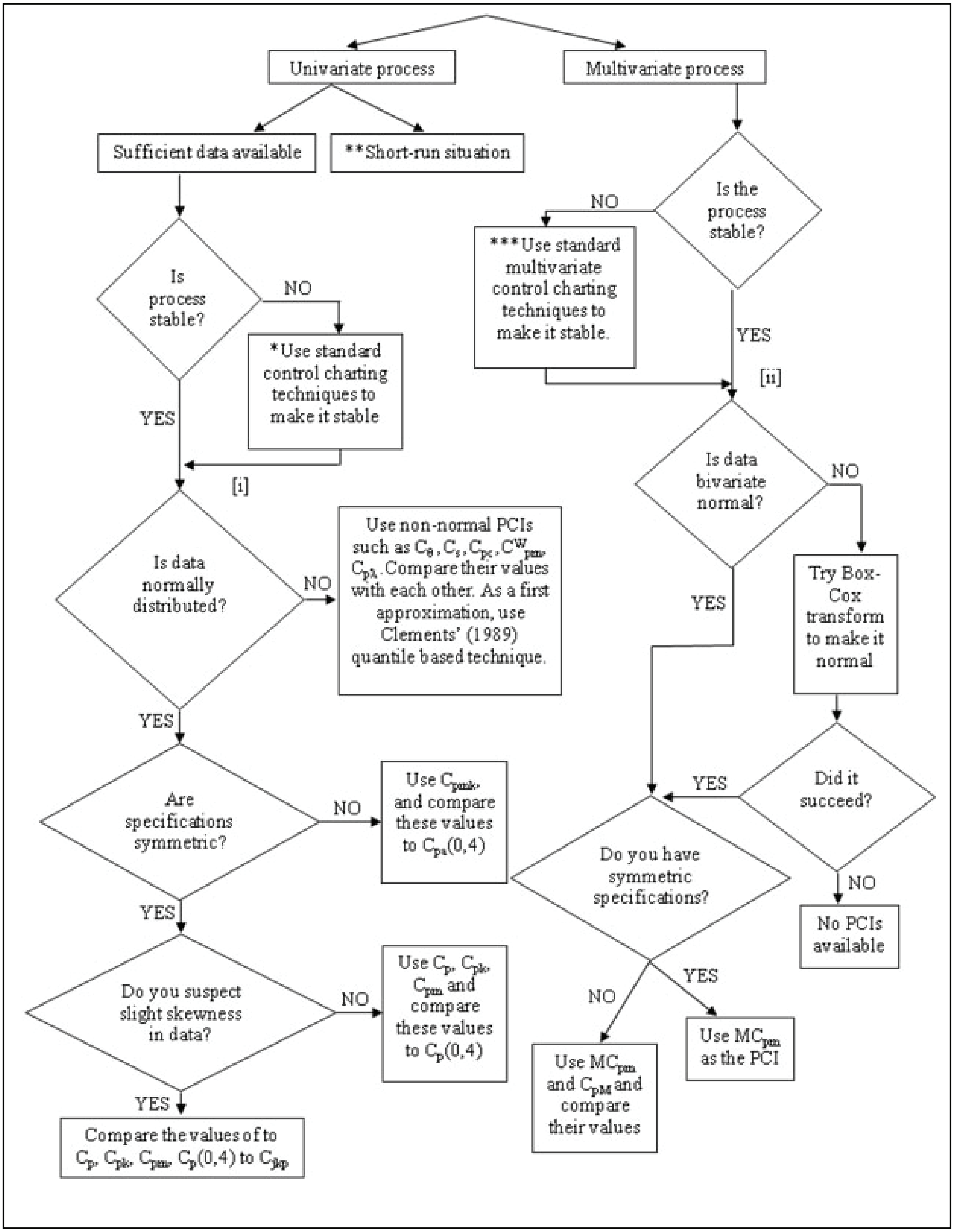
m Subgroups of Equal Size n

The data used to calculate any PCI must come from a stable process (i.e., a process governed by a single probability distribution). Statistical control charting with a delete and revise (D&R) procedure is one way to ensure this. In a D&R procedure, the data used to construct the control charts is also plotted on the charts to retrospectively test if the process was in control while the initial data was being obtained. Any points that plot outside the control limits are deleted and the remaining data is used to construct revised control charts. One of the several variations of the D&R procedure repeats this process until no points plot beyond the control limits, at which time the remaining data would be considered stable or in control.

For $2 \leq n \leq 10$, the usual $\overline{X}$ and R control charts (Montgomery, 2001) are used to perform control charting in order to establish control of the data. For n > 10, the usual $\overline{X}$ and S charts are used to perform control charting as the range method for estimating σ loses statistical efficiency for moderate to large subgroup sizes, as mentioned in Montgomery (2001).

Once the above procedure is completed, the remaining data is subjected to a normality check via the Kolmogorov-Smirnov (K-S) test, the procedure for which can be found in any standard statistical text, such as Ebeling (2000). If the normality assumption is satisfied, the decision tree approach makes use of the PCIs as shown in Figure 1 for this situation in order to evaluate process capability. PCIs like C_p, C_{pk}, C_{pm} (Kotz & Lovelace, 1998), and $C_p(0,4)$ (Vannman, 1993) are used when the target value is equal to the midpoint of the specifications (target = midpoint). These values are compared to C_{jkp} (Kotz & Lovelace, 1998) if doubt of slight skewness exists in the data. If not, C_p, C_{pk}, and C_{pm} are compared to C_p (0,4). If the target value is not equal to the midpoint of the specifications, PCIs such as C_{pmk} (Kotz & Lovelace, 1998), C'_{pm} (Perakis & Xekalaki, 2003), and C_{pa} (0,4) (Vannman, 1997) are used to evaluate process capability.

Figure 1: Main Decision Tree



If the normality assumption is not satisfied, non-normal PCIs such as C_θ , C_s , C_{pc} , C^W_{pm}, and $C_{p\lambda}$ (Kotz & Lovelace, 1998) are used to evaluate process capability. Because there is no evidence in the statistical literature as to which of these indices is better for a particular situation, the values of these indices are compared with each other as per the methodology.

m Subgroups of Variable Sizes with Maximum Subgroup Size n

In this case, the usual $\overline{X}$ and S control charts for variable subgroup sizes (Montgomery, 2001) are used to perform control charting in order to establish control of the data. Once the process data is stable, the methodology proceeds with normality, symmetric specification, and skewness checks as described previously. The appropriate PCI(s) are then selected.

m Individual Observations

In this case, the usual Individuals (X) and Moving Range (MR) control charts (Montgomery, 2001) are used in order to establish control of the data. The moving range used here is defined by the equation:

$$MR_i = \left| x_i - x_{i-1} \right| \tag{1}$$

where x_i and x_{i-1} are two successive observations collected as individual process data.

The PCI selection procedures for the data remaining after the D&R procedure are performed in the same manner discussed above. However, it is necessary to ascertain that individual observations obtained are normally distributed even before control limits for these charts are calculated, because even for moderate departures from normality the use of the X and MR charts is not appropriate. Hence, if the data collected is not normally distributed, it should be transformed to another variable that is approximately normally distributed (this was not an issue in previous descriptions because the Central Limit Theorem could be invoked subgrouped data).

If the normality assumption is satisfied, the methodology suggests the continuation of the PCI selection procedure as mentioned earlier. However, if the normality assumption is not satisfied, the data should undergo a Box-Cox transformation of the type in equation:

$$Y = \left(X^{\lambda} - 1\right)/\lambda \tag{2}$$

where the optimal value of λ is determined by an iterative procedure using the following steps as mentioned by Johnson & Wichern (2003):

1. Construct a normal probability plot of the individual observations and determine the correlation coefficient, r.
2. For different values of λ ranging from -2 to 2, determine the value of r. Determine r_{max}, the maximum value of r among all the values calculated.
3. The value of λ which gives r_{max} is used for the transformation in accordance with the following values of λ: 2 (square transformation), 1 (use the original data), 1/2 (square root transformation), 0 (logarithm transformation), -1/2 (reciprocal square root transformation), and -1 (reciprocal transformation).

The transformed data is again checked for normality. If the transformed data is found to be normally distributed, the PCI selection procedure is conducted using the methods explained previously. However, if that is not the case, the data is considered to be strongly non-normal. As a result, control charting cannot be done and PCIs cannot be selected.

Type 2: Univariate Short Run Data (< 50 Observations)

In this case, the data may have been collected either in m subgroups each of size n or as individual observations. The following procedure is adopted for evaluating PCIs in this situation. (See Figure 1 and Appendix Figure 3.)

m Subgroups of Equal Size n

The control charting procedure adopted in this case for establishing control of the data is the short run $\overline{X}$ and S control charts from Elam & Case (2005a, 2005b). Once this procedure is completed, the remaining data is checked for normality via the Kolmogorov-Smirnov (K-S) test and the correlation coefficient test (Johnson

& Wichern, 2003) for normality at the specified level of significance. The underlying reason for using both tests is that, for a small number of observations, the correlation coefficient test is considered to be a very powerful test for normality. If the remaining data are found to be normally distributed, short-run PCIs such as C_{sp}, C_{spk}, and C_{spm} are used to evaluate process capability as mentioned by Balamurali (2003). According to this procedure, the remaining data are bootstrapped into 1,000 resamples, each of which are equal to the total number of observations in the remaining data. These are then used to calculate the short-run PCIs, and the standard bootstrap method is used to construct a 95% confidence interval for each index.

If the remaining data is found to be non-normal at the specified level of significance, the Box-Cox transformation is used to transform the original non-normal data to normal data. If the transformation is successful (the transformed data is subjected to the K-S test and the correlation coefficient test for normality), short-run PCIs as discussed above are evaluated. If the transformation is unsuccessful, the short-run PCIs are still evaluated. It should be noted, however, that the results obtained from PCI calculations may be inaccurate, as for a non-normal process, the coverage percentage points for 95% confidence limits might indicate a high proportion of values that are significantly different from the expected value of the index at the specified level of significance.

m Individual Observations

The control charting procedure adopted in this case for establishing control of the data is the short run X and MR control charts from Elam & Case (2008, 2006). Once this procedure is completed, the remaining data is subjected to the same procedures as related earlier in the m Individual Observations, starting with the normality check. The short-run PCIs discussed previously are used to evaluate process capability.

Type 3. Multivariate Sufficient Data (Observations ≥ 50) (See Figure 1 and Appendix Figure 4.)

m Subgroups of Size n

In this case, the usual Hotelling T^2 control chart (Montgomery, 2001) is used along with the usual bivariate control chart for dispersion (Johnson & Wichern, 2003) to conduct control charting for establishing control of the data. The remaining data are subjected to a bivariate normality check because the PCIs to be calculated are strictly based on the assumption of bivariate normality. This bivariate normality check is performed by:

$$(X-\mu)' S^{-1}(X-\mu) \le \chi_2^2(0.5) \tag{3}$$

The average μ and variance-covariance matrix S are for the remaining data grouped together. If approximately 50% of the remaining data grouped together satisfies equation (3) the data is considered to be bivariate normal as per Johnson & Wichern (2003).

If the bivariate normality assumption is satisfied, the bivariate PCIs C_{pM} and MC_{pm} (for bivariate process data with asymmetric specifications) and MC_{pm} (for bivariate process data with symmetric specifications) are evaluated as shown in Wang, Hubele, Lawrence, Miskulin & Shahriari (2000). If the bivariate normality assumption is not satisfied, the Box-Cox transformation of the data is performed. The optimal value of λ is the one that maximizes the following equation:

$$l(\lambda) = (-n/2)\ln\left((1/n)\sum_{j=1}^{n}\left[x^{(\lambda)}{}_j - \bar{x}_j^{(\lambda)}\right]^2\right) + (\lambda-1)n\sum_{j=1}^{n}\ln\left[x_j\right] \tag{4}$$

where n is the total number of filtered observations, $x^{(\lambda)} = \left(x^{\lambda}-1\right)/\lambda$ if $\lambda \neq 0$, and $x^{(\lambda)} = \ln(x)$ if $\lambda = 0$. If, after the above procedure, bivariate normality is not satisfied, then it is not possible to calculate a bivariate PCI.

m Individual Observations

In the case of individual observations of bivariate data, the usual T^2 control chart for individual observations (Johnson & Wichern, 2003) is used to establish control of the data.

Once this has been accomplished, the bivariate data is subject to a bivariate normality check in accordance with the procedure discussed herein. The PCI selection procedure continues similarly to the case for bivariate data collected in subgroups.

Results and Conclusion

The methodology used in formulating a decision tree approach in order to aid industry practitioners regarding the selection of a PCI has been discussed; the main advantage of this methodology that it offers a structured approach for programming the same into a decision support system for easy usage in industry. By incorporating such a methodology into a computer program with the capability to select, calculate, and interpret the appropriate PCI(s) for the situation under consideration, the problems industry experiences with PCIs, as noted in the Introduction, are alleviated. As all statistical assumptions have been taken into consideration while developing this methodology, a robust structure to the application of PCI usage in industry has been accomplished.

References

Balamurali, S. (2003). Bootstrap confidence limits for short-run capability indices. *Quality Engineering*, *15(4)*, 643-648.

Clements, J. (1989). Process capability calculations for non-normal distributions. *Quality Progress*, September, 95-100.

Deleryd, M. & Vannman, K. (1998). Process capability studies for short production runs. *International Journal of Reliability, Quality, and Safety Engineering*, *5(4)*, 383-401.

Ebeling, C. E. (2000). *Reliability and maintainability engineering*. New Delhi: Tata McGraw-Hill.

Elam, M. E. & Case, K. E. (2005a). Two-stage short-run ($\overline{X},s$) control charts. *Quality Engineering*, *17(1)*, 95-107.

Elam, M. E. & Case, K. E. (2005b). A computer program to calculate two-stage short-run control chart factors for ($\overline{X},s$) charts. *Quality Engineering*, *17(2)*, 259-277.

Elam, M. E. & Case, K. E. (2008). Two-stage short-run (X,MR) control charts. *Journal of Modern Applied Statistical Methods, 7(1)*, 275-285.

Elam, M. E. & Case, K. E. (2006). A computer program to calculate two-stage short-run control chart factors for (X,MR) charts. *Journal of Statistical Software*, *15(11)*, http://www.jstatsoft.org/.

Johnson, R. A. & Wichern, D. W. (2003). *Applied multivariate statistical analysis*. New York: Pearson Education.

Kotz, S. & Lovelace, C. (1998). *Process capability indices in theory and practice*. New York: Oxford University Press Inc.

Montgomery, D. C. (2001). *Introduction to statistical quality control*. 4th ed. New York: John Wiley and Sons.

Perakis, M. & Xekalaki, E. (2003). On a process capability index for asymmetric specifications. *Communication in Statistics – Theory and Methods*, *32(7)*, 1459-1492.

Phadnis, M. A., Elam, M. E., Fonseca, D. J., Batson, R. G., & Adams, B. M. (2005). A prototype DSS for PCI selection, calculation, and interpretation. *Proceedings of International Conference, Institute of Industrial Engineers*, Atlanta, USA.

Vannman, K. (1993). A unified approach to capability indices. *Statistica Sinica*, *22(2)*, 537-560.

Vannman, K. (1997). A general class of capability indices in the case of asymmetric tolerances. *Communication in Statistics – Theory and Methods*, *26(10)*, 2381-2396.

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

Figure 2: Decision Tree for Univariate Data

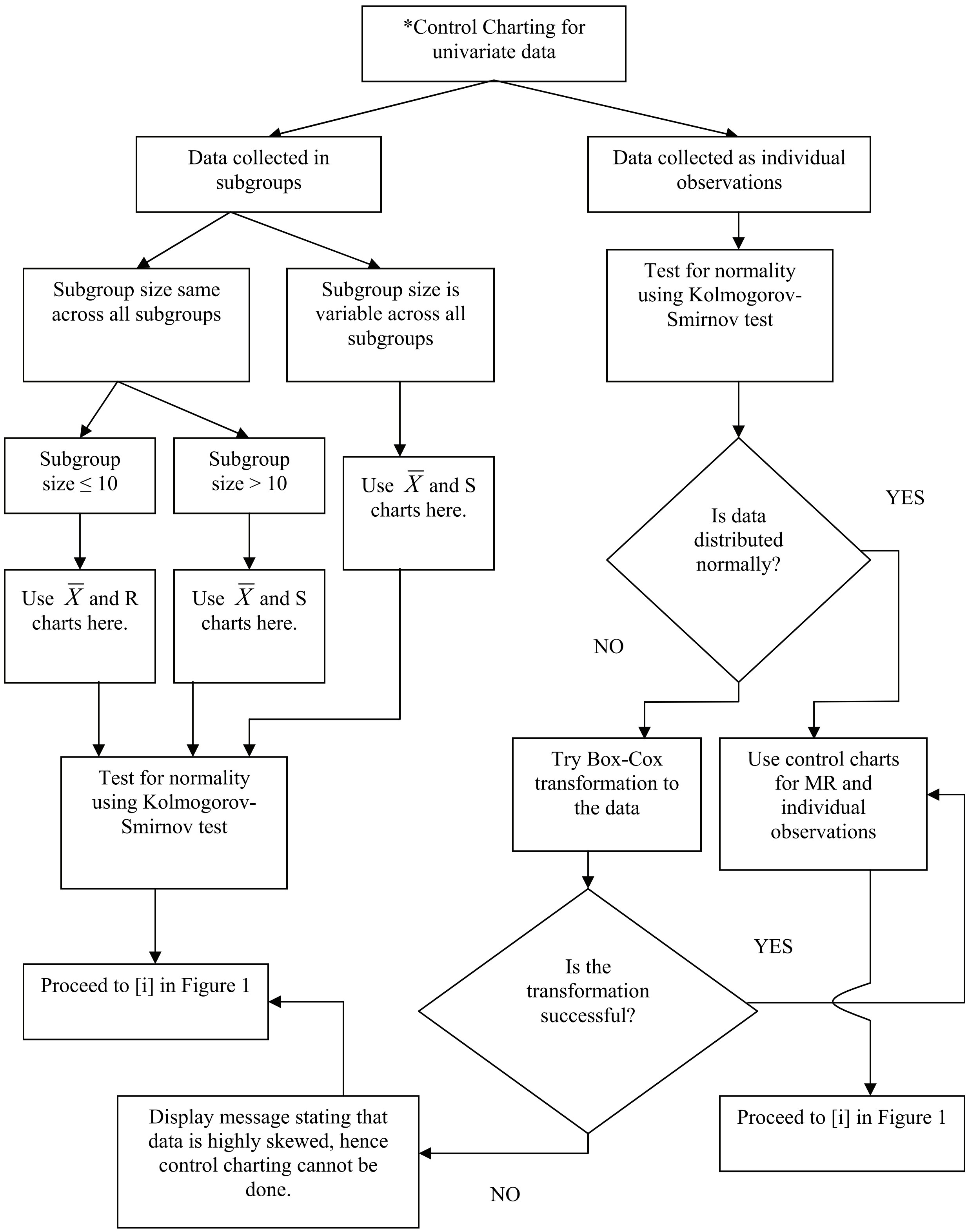


Figure 3: Decision Tree for the Short-Run Situation

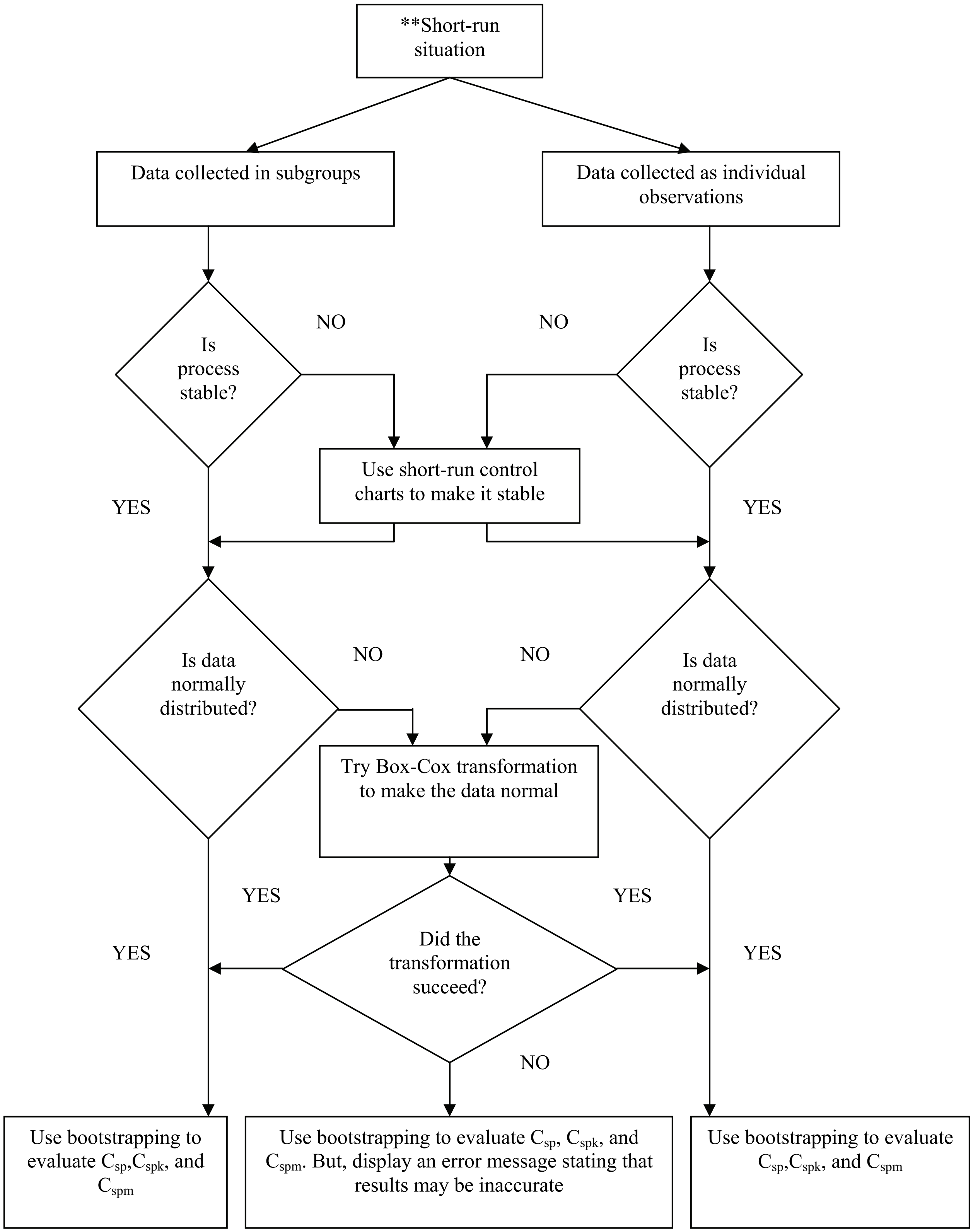


Figure 4: Decision Tree for Bivariate Data