ADJUSTMENT OF SIX SIGMA TOOLS FOR A BETTER CONTROL OF PROCESS TIME

FILIP TOŠENOVSKÝ

1 INTRODUCTION

The article discusses an appropriate adjustment of standard statistical tools used by Six Sigma methodology to limit problems that occur in organization processes. The adjustment is specifically related to the problem of process time control the unique characteristics of which require a modification of some of the standard statistical methods generally used in Six Sigma.

2 STANDARD SIX SIGMA APPROACH

If processes in a company last too long, there is a good reason to make an effort at curtailing the time because the more time it takes to run the process, the costlier it usually becomes, let alone the fact that customers awaiting the process output become impatient and later dissatisfied too. There are several ways how to approach the problem with process time. Given the current trend, however, it is quite likely that companies facing such a problem will choose Six Sigma as a solution.

Six Sigma relies heavily on the define, measure, analyse, improve and control methodology known as the DMAIC cycle. The cycle is a step-by-step procedure to reveal major causes of problems occuring in a process, and to eliminate the causes. After a Six Sigma project is defined from the organizational point of view (phase D of the cycle), data from the problematic process are gathered and checked for their accuracy, stability of the process is determined based on the data and the actual performance of the process is estimated (M phase). If the data are correct and reflect the process in question, i.e. the process is stable and its behaviour is not contaminated by unplanned and unexpected external events, the analytical phase A follows, trying to detect the mechanisms by which problems in the process arise. The A phase states as its result a group of major factors that create the problems in the process, and also a relation that expresses the problem as a function of the behaviour of the major factors. Based on the known relation or relations, a favourable configuration of the factors is proposed so that the pre-set behaviour of the factors caused as few problems in the process as possible (I phase). In the final C phase of the DMAIC cycle, a surveillance over the improved process is designed to make sure the improvements really take place and the process stays improved for a longer period of time.

The measure and analyse phases are the key to solving the problem in the process as they gather information on the process and decide what and how causes the problems. Statistical tools that are often used in the two phases involve control charts to determine the stability of the process and regression or correlation analysis to define the relations between factors and the problem they create in the process.

3 THE STANDARD APPROACH AND THE PROBLEM OF PROCESS TIME

Effective elimination of problems in processes requires effective implementation of the measure and analyse phases of the DMAIC cycle. In order for this to happen, right conditions must be met so that control charts and regression worked properly in the two phases.

In order for control charts to give a valuable and trustworthy result, data collected from the process should be at least approximately normally distributed and their collection should have the form of a random sampling, i.e. the data ought to be statistically independent. Although some say too much of the probability theory is put into control charts, and the charts are rather an intuitive tool, the above-mentioned conditions were in place when control charts were invented and successfully tried out, so the conditions are justified. None the less, the truth probably lies somewhere between the two intellectual streams.

When a process lasts too long, and its time is therefore the focus of a Six Sigma project, it may easily happen that none of the two conditions is met and control charts will then be a too rough tool for making a judgement on the stability of the process – on the stability of time as an observed process characteristic. Situations when queueing occurs is an example. Process time as a random variable is typically modelled by Gamma distribution which not even remotedly resembles the normal distribution. Moreover, process time t_n may depend on the previous process time t_{n-1} when queueing takes place. The time it takes to serve a customer in the bank, for instance, depends on the time it took to serve a previous customer if the two customers joined in the same queue. We may also add that the fact the data on process time are not independent in such cases is more inappropriate than the fact that time is not normally distributed as independence might in many cases ensure at least approximate normality thanks to the central limit theorem provided the data samples gathered about the process time are large enough. If the independence is not in place, the theorem doesn't apply.

Success of regression modelling also largely depends on the situation, of course. The standard approach works mainly with linear modelling, which gives good results if the modelled variable – the variable that represents the problem in the process – is normally distributed and the coefficients of the model are estimated via the least squares method. Under normality and least squares, linear model is the best for the description of such variable. The word „best" means that the analytical form of the model is the right one (Anděl 1984), and if we had a chance to get an infinite number of series of data samples on the process, we would – in average - estimate the unknown theoretical coefficients in the model precisely, that is we would get unbiased estimates. In addition to that, if we don't estimate the coefficients precisely with an individual data sample, the imprecision is the smallest we can achieve. Unfortunately, this is not the case with process time which is not normally distributed. A different approach must then be considered.

4 PROCESS TIME AND SIX SIGMA TOOLS ADJUSTMENT

If process time is the problem and the data on process times indicate the condition of their independence is not met, control charts might be an inappropriate tool to assess the process stability in terms of the process characteristic observed. However, since the process times are necessarily recorded at a certain point in time, they can be represented as a time series, and the fact that a dependence exists among the data then starts to be conversely an advantageous feature of the data. It is advantageous because the Box-Jenkins representation of time series, which is based on an existing dependence among the data, allows us to analyse such series properly (Box, Jenkins 1969). What is even more important is that the definition of the so called „stationarity of a process“, as presented in the theory of time series analysis, may be considered a synonymum for the term „stability of a process“, as presented in the theory of control charts. Thus, if control charts are not available due to existing dependence in the data on process times, we may check the stability of the process by forming a time series of process times and validating stationarity of the series according to Box and Jenkins theory.

As regards regression, there is more than one way how to handle non-normally distributed variables like process time that we want to model as a function of major factors involved. We may stay with linear regression model, benefiting from the simplicity of the model, but the full appropriateness of the model for the given situation will not be guaranteed. Or we may also try to find a model that suits the situation, utilizing Box-Cox transformation, for instance. The transformation can bring the variable in question, e.g. the process time in our case, closer to normality, for which we know how the good model looks like – it is linear. Thus, the following procedure may lead to a good, generally nonlinear, regression model describing the behaviour of process time as a function of several conceivable factors:

1) Find a transformation T that brings the modelled process time y closer to normality.
2) Model the average behaviour of T(y) with the best model, i.e. with linear model:

$$E[T(y)] = \beta_0 + \beta_1 x_1 + \ldots + \beta_\kappa x_k,$$

where the operator E stands for expected value (average), β_i, i=1,…,k, are coefficients in the model to be estimated by least squares and x_i, i=1,…,k, are major factors influencing the process time.

3) Some transformations have the property $E[T(y)] \cong T[E(y)]$, for instance the logarithmic transformation, which turns out to be in many cases a good transformation in step one (McCullagh, Nelder 1990). If this is the case, then $T[E(y)] \cong E[T(y)] = \beta_0 + \beta_1 x_1 + \ldots + \beta_k x_k$, in other words $E(y) \cong T^{-1}(\beta_0 + \beta_1 x_1 + \ldots + \beta_k x_k)$ provided the inverse function T^{-1} is also available.

This way, it is possible to get a regression model $E(y) \cong T^{-1}(\beta_0 + \beta_1 x_1 + \ldots + \beta_k x_k)$ which suits the situation better than a linear model.

5 CONCLUSION

Six Sigma has proven to be a good approach to solving problems in organization processes. However, while the DMAIC cycle Six Sigma consists of methodologically is fine-tuned, statistical tools used to analyse the data gathered from the process must be selected ad hoc, depending on the nature of the problem occuring in the process. Problems of specific nature show up when it is the process time that must be shortened by Six Sigma. The specific nature of process time requires us to alter some of the standardly used Six Sigma statistical methods, particularly those with a major influence on the result of a Six Sigma project. Those are control charts used in the M phase of the DMAIC cycle, and the modelling of relations among variables used in the A phase of the cycle. Box-Jenkins theory on time series enables to check stability of a process when control charts are insufficient due to dependence in analysed data, while the Box-Cox transformation can lead to a nonlinear regression model which describes relations in the data better than linear model if the modelled variable does not follow normal distribution.

REFERENCES

Anděl, J.: *Matematická statistika*, Státní nakladatelství technické literatury (1984).

Box, G.E.P., Jenkins, G.M.: *Time Series Analysis, Holden-Day*, 1969.

McCullagh, P., Nelder, J.: *Generalized Linear Models*, Chapman & Hall/CRC (1990).

ABOUT THE AUTHOR

Ing. Filip Tošenovský, Ph.D., The School of Business Administration in Karvina, Silesian University, Univerzitní náměstí 1934/3, 733 40 Karviná, Czech Republic tosenovskyfilip@opf.slu.cz.