

Effect of the number of non-conforming samples on the Kappa indicator values

*Pavel Klaput*¹ and *David Vykydal*²

¹VSB-TU Ostrava, Department of quality management, 17. listopadu 15/2172, Ostrava, Czech Republic

Abstract. Evaluate the quality of this measurement system is possible by using multiple methods, which are described in the methodology guides for evaluation of the measurement system quality. Within these methods, values of different indicators are evaluated. This paper deals with the cross tabulation method and foremost Kappa indicator. More specifically is examined the effect of the number of used nonconforming samples on explanatory power of this indicator. The effect on the kappa values is examined from several aspects that must be taken into account during evaluation of the quality of measurement system for attributive quality characteristics.

1 Introduction

From practical experience with the implementation of measurement system analyses (MSA) and from more and more frequent practice requirements, several problem areas can be defined in assessing the quality of measurement systems. Where a measurement system is used to control attribute quality characteristic, there are problem areas in performing and evaluating analyses of measurement systems used to products control. Obstacles to greater use of cross tabulation method in practice can be seen in particular in the lack of information about right sampling and in the analysis evaluation itself. For these reasons, a detailed examination of this issue can lead to a better understanding and wider application of MSAs in practice. For this reason, in Excel prepared and conducted simulations aimed at the calculation of kappa indicator. The influence of different numbers of nonconforming samples on the difficulty of the measurement system analysis evaluated by the cross-table method was evaluated. Consideration was also given to the explanatory power of the calculated values of the kappa indicator and also the fact, whether the conditions of the analysis correspond to the real conditions of the process [1].

2 Cross tabulation method

Evaluation of the measurement system for attribute quality characteristic is most often done by the cross-table method, which assesses the agreement between the different operators and the agreement between the operators and the reference value. An example of a cross table for conformity assessment between operator A and reference is shown in Table 1 [2].

Table 1. Setting Word’s margins.

			Reference		Sum
			0	1	
Operator A	0	Result	X ₀₀	X ₀₁	X _{0.}
		Expected result	x ₀₀	x ₀₁	x _{0.}
	1	Result	X ₁₀	X ₁₁	X _{1.}
		Expected result	x ₁₀	x ₁₁	x _{1.}
Sum		Result	X _{.0}	X _{.1}	N
		Expected result	x _{.0}	x _{.1}	n

The values in the result cells of the table labelled of capital letter X (X₀₀, X₀₁, X₁₀, X₁₁) indicate how many times the possible combinations of operator A results with the reference rank occurred. Values in cells in total (X_{0.}, X_{1.}, X_{.0}, X_{.1}) summarize the overall a decision on acceptance and rejection [1]. Calculation of expected results is performed in the following manner:

$$x_{00} = \frac{X_{0.}}{N} \cdot X_{.0} \tag{1}$$

$$x_{01} = \frac{X_{0.}}{N} \cdot X_{.1} \tag{2}$$

$$x_{10} = \frac{X_{1.}}{N} \cdot X_{.0} \tag{3}$$

$$x_{11} = \frac{X_{1.}}{N} \cdot X_{.1} \tag{4}$$

For the calculation of kappa indicators is needed to calculate values p_o and p_e value is p_o share results when operator agree with the reference on a decision on acceptance or rejection of samples. Value p_o should ideally approximate to one and is calculated as follows:

$$p_o = \frac{X_{11} + X_{00}}{N} \tag{5}$$

The p_e value represents the share of expected results when operator agree with the reference on the decision to accept or reject the samples. It is calculated as follows:

$$p_e = \frac{x_{11} + x_{00}}{n} \tag{6}$$

The k (kappa) indicator is used to assess the agreement between the operator and the reference and is calculated according to the formula:

$$k = \frac{p_o - p_e}{1 - p_e} \tag{7}$$

The values of the kappa indicator can be in the range from -1 to 1. Kappa is a measure rather than a test. A general rule of thumb is that values of kappa greater than 0,9 indicate excellent agreement (with a maximum kappa = 1). Values less than 0,4 indicate poor agreement. Based on this indicator, decisions and interventions are made to improve the quality of the attributive measurement system. All these decisions are made on the basis of faith in the flawless information contained in the value of kappa indicator. However, on the basis of earlier analyzes, it can be concluded that the used values of the kappa indicators may be affected by setting the input parameters of the evaluation used methods or other restrictive conditions [3, 4].

2 Influence of the number of nonconformities

The most important area using the value of the kappa indicator as a measure of the quality of the measurement system is the evaluation of the compliance between the operator and the reference assessment. The basic parameter of the cross-table method is the number of samples (parts) used for application of this method. As mentioned above, a sufficient number of samples (50) for which their reference values are known have to be used for the analysis of the cross-tabulation method. Among the selected samples should be a certain proportion of nonconforming samples to achieve the most relevant results of the analysis. None of the methodological manuals does not say the recommended number of nonconforming samples. Due to the method of calculating the value of the kappa (formula 5, 6 and 7), it can be assumed that the number of used non-conforming samples will affect the possible range of kappa value. At the half representation of non-conforming samples used for the analysis of attributive measurement systems, the value of the kappa indicator ranges from -1 (operator took no right decision) to a value of 1 (the operator made the right decision in all cases). The range of possible kappa values, depending on the number of correct decisions made by the operator for half number of the non-conforming samples (25 from 50), is clearly shown in Figure 1.

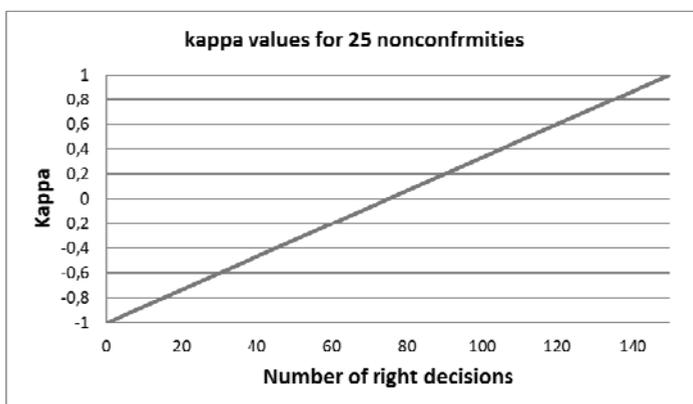


Fig. 1. Kappa values for 25 nonconformities.

Constructed graph shows that in the case of uniform proportion of conforming and nonconforming samples values of the kappa indicator changes uniformly. In practice, however, it is often possible to encounter a situation where for a variety of reasons it is not possible to ensure a uniform proportion of nonconforming samples. The changes of kappa values in the case of less than half number of nonconforming samples are clearly presented in Figure 2. Analysing the displayed data, it can be noticed that with low numbers of nonconforming samples used to analyse the attributive measurement systems, it is not possible to reach the full range of kappa values. For example, if ten nonconforming samples is used, the value of the kappa in the case of the zero match with reference is equal to -0.47. From the course of the constructed curves it is clear that the increase in the values of the kappa indicator with the increasing number of correct decisions taken by the operator is not linear. A special case is a situation where no one non-conforming sample was used to evaluate the quality of the attributive measurement system. In this case, the value of the kappa cannot be different from zero, so in this case the kappa value cannot be evaluated and the analysis of the attributive measurement system is in this case meaningless (see Figure 2).

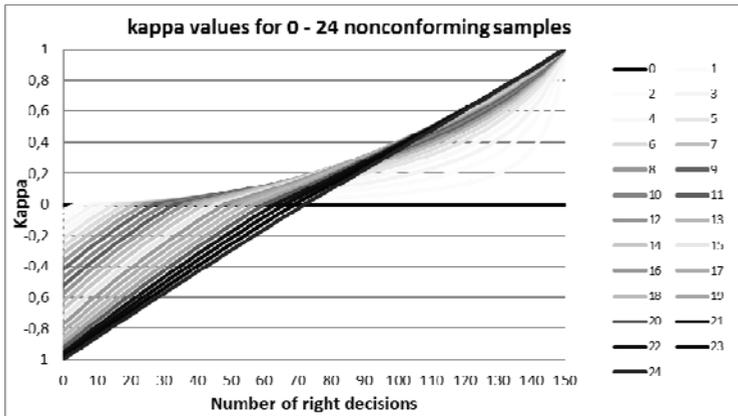


Fig. 2. Kappa values for 0 - 24 nonconformities.

In addition to lower fraction of nonconforming samples used for the analysis may occur converse situation, where (especially in the case of processes that produce a high percentage of non-conforming products) will be analysed majority proportion of nonconforming samples (26 to 50). This situation is illustrated in Figure 3.

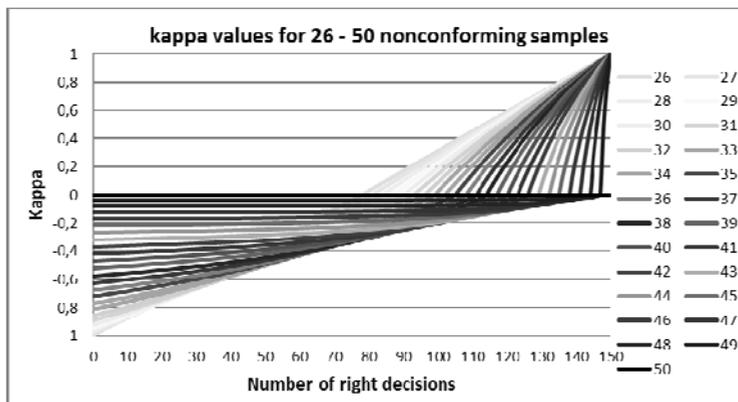


Fig. 3. Kappa values for 26 - 50 nonconformities.

Also, in this case, it can be noticed that with higher numbers of nonconforming samples used to analyse the attributive measurement systems, it is not possible to reach the full range of kappa values. For example, if forty nonconforming samples is used, the value of the kappa in the case of the zero match with reference is equal to -0.47. This is the same value as in the case of the ten nonconforming samples. From the above mentioned findings, it is clear, that to achieve the limit values of the kappa indicator for the conditionally acceptable (0.7) resp. acceptable measurement system (0.9), it will be necessary to achieve at a different number of correct decisions according to the reference for a different number of nonconforming samples used for analysis. Changes in the number of correct decisions in dependence on the number of nonconforming samples that must be taken to achieve the required values of the kappa indicator is shown in Figure 4. From the constructed curves, it can be observed that with the decreasing proportion of nonconforming samples used to analyse of attributive measurement systems, increases the number of correct decisions that the operator has to make to assess acceptable compliance with the reference. The same finding is also valid for the increasing proportion of non-conforming samples. Using too small or too large numbers of non-conforming samples, thus placing greater requirements on the accuracy of operator evaluation.

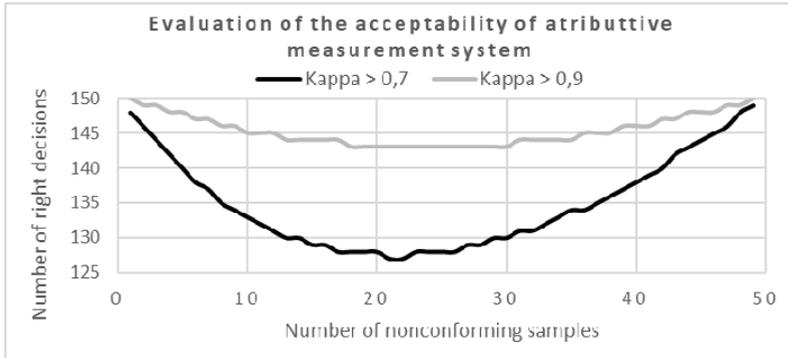


Fig. 4. Dependence of the number of correct decisions on the number of non-conforming samples.

For example, if five nonconforming samples are used, the operator will need to make 140 correct decisions so that the value of the kappa assess the compliance between the operator and the reference reaches the minimum required value of 0,7. Conversely, in the case of half-shares of non-conforming samples (25), the operator has to make only 128 correct decisions to achieve the required value of 0,7. Its error rate can be more than twice as large as 25 nonconforming samples than if 5 nonconforming samples were used. By appropriately choosing the number of non-conforming samples, the " difficulty" of the kappa indicator can be reduced. The smallest difficulty of the kappa indicator, and thus the entire cross-table method, can be achieved whit 15 to 30 non-conforming samples. Within this range, the number of necessary correct decisions varies from 127 to 130 [5, 6].

3 Conclusions

On the basis of the findings presented in the previous chapters, it is possible to define some recommendations that could reduce, respectively, eliminate the effect of different number non-conforming samples. Since methodological guides do not make clear recommendations regarding the number of non-conforming samples used for analysis, there is a need to devote increased attention to sample selection. In this selection must be taken into account a number of conflicting requirements, respectively. aspects of the composition of the samples used to measurement system analysis for attributive quality characteristics. These aspects are shown in Figure 5.

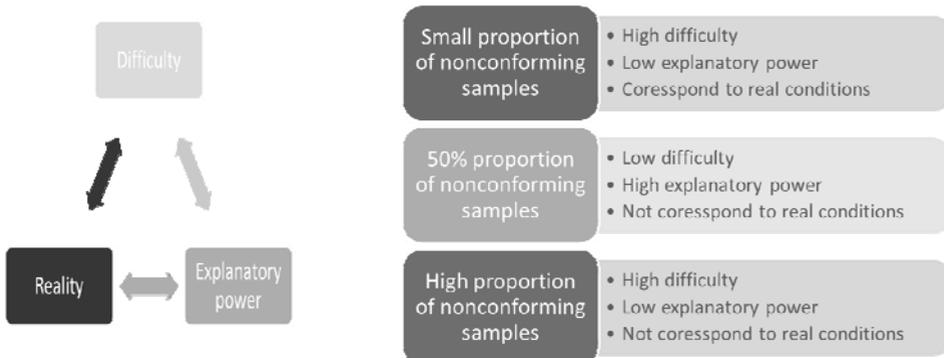


Fig. 5. Influence of aspects at different proportions of non-conforming samples

The first aspect to be taken into consideration is the difficulty or, the accuracy of operator evaluation. With a lower or higher number of non-conforming samples, the operator must make more correct decisions to achieve the required level of the kappa indicator than when the proportion of nonconforming and conforming samples is equal. Another aspect that needs to be consider is the relevancy, or the explanatory power of the whole analysis. If a small number of non-conforming samples is used, the full range of possible kinds of the defect (non-conformities) will not be covered and the obtained results may not be entirely relevant for this reason. The last considered aspect is the realism of performed analysis. Because the manufacturing process typically does not reach such a low quality level, which produces half of nonconforming products (half representation of non-conforming samples), such analysis conditions does not correspond with the reality of the production process. For certain types of attribute quality characteristic and certain types of control systems, it was observed that operators achieved less error in the case of low (realistic) proportion of non-conforming samples. [7]

This paper was elaborated in the frame of the specific research projects No. SP 2018/97 and SP 2018/109, which has been solved at the Faculty of Metallurgy and Materials Engineering, VŠB-TU Ostrava with the support of Ministry of Education, Youth and Sports, Czech Republic.

References

1. Chrysler Group LLC, Ford Company, General Motors corporation *Measurement Systems Analysis. Reference Manual. 4th ed.*, (2010)
2. D.C. Montgomery, *Statistical Quality Control: A modern introduction, 7th ed.* 768, (2013)
3. H. Pačaiová, J. Sinay, R. Turisová, Z. Hajduová, Š. Markulík, *Measurement*, **109**, 359-365 (2017)
4. J. Petřík, *Archives of Metallurgy and Materials*, **61**, 4 (2016)
5. P. Klaput, R. Macek, A. Lubojacká, *Kvalita – Quality* (2014)
6. E. Vágó, A. Kemény. *Quality and Reliability Engineering International*, **27**, 6 (2011)
7. J. Petřík, P. Palfy, M. Havlík, *International Journal of Engineering*, **11**, 1 (2013)