Research on the Use of Statistical Methods in Quality Control in Selected Czech Manufacturing Companies

MARTIN KOVÁŘÍK, PETR KLÍMEK
Department of Statistics and Quantitative Methods
Tomas Bata University in Zlín
Mostní 5139, 760 01 Zlín
CZECH REPUBLIC
m1kovarik@fame.utb.cz, klimek@fame.utb.cz http://www.fame.utb.cz

Abstract: The article deals with the issue of research in statistical methods and tools in this paper. The main two aims will be to determine whether the selected statistical methods reported in scientific publications related to SPC are actually used in practice and if does the degree of difficulty of used statistical methods in the SPC depend on the firm size. We decided to use a questionnaire to map the state in selected companies in the use of statistical methods. 80 Czech companies from different sectors were addressed in which was expected use of statistical quality control tools. The results of the survey will be evaluated using the methods of mathematical statistics, concretely by analysing the dependence of two categorical and two numerical variables. The result presentation of this survey questionnaire will be mentioned below.

Key Words: Questionnaire Survey, SPC (statistical process control), \( \chi^2 \)-test of Independence, Association Coefficients, Spearman Rank Correlation Coefficient, Contingency Coefficients.

1 Introduction
Current research concerning the use of exact methods in quality control indicates a growing trend in the application on these methods.

This paper will take the issue of statistical methods and tools research where the goal will be to determine whether the selected statistical methods reported in scientific publications related to the issue of SPC (statistical process control) are actually used in practice. To perform research, data were collected and this was obtained through working with managers of companies and managers in production. It was used a web interface and database for collecting and storing data. The last step was the evaluation of these data with adequate statistical methods including interpretation of the conclusions provided by them.

At this point, we must point out that similar surveys were already made in Japan in the 70 years. Using mathematical and statistical methods for process control is a preventive quality management tool since the early detection of significant deviation in the process from a predefined level allows the process to implement interventions in order to maintain it at acceptable and long-term stable level or eventually enable process improvement. In addition to classical methods such as descriptive statistics, ANOVA, hypothesis testing, regression analysis, reliability analysis, there are still some special statistical methods to the needs of quality management. These are especially different types of control charts, the analysis of discrete variables, statistical acceptance, capability analysis, sampling, reliability analysis, Pareto analysis and others. It is obvious that a worker without adequate education or good training cannot effectively and properly use these tools.

2 Literature Review
The result of the implementation and active use of statistical methods in the process data analysis is more accurate image of the process or even construction of a model of interdependencies and influencing the various elements and phases of technology.

This model can be used to find the causes of variability and instability and to real increase in quality and stability. The aim of introducing statistical analysis is mainly to improve quality. Absolutely dominant tool for these tasks is a computer statistical analysis. The vast majority of programs only replace the routine work of practitioners.

This state is often considered to be satisfactory. But problems arise when methods assumptions are not met (e.g. normality, absence of gross errors, sufficient sample size, statistical process stability,
etc.) From this perspective, some statistical programs or their combination is better. They allow complex statistical data analysis and also control chart construction and other additional tools for quality management [1].

Integration of statistical analysis in economic activities is strictly required by many regulations and standards. The best known is a series of ISO 9000. Generally, we can name ČSN and ČSN ISO from the field of applied statistics, standards from statistical regulation, standards from statistical acceptance, standards ČSN and ČSN ISO from reliability and a new standard ISO 22514 series (Statistical Methods in Process Management – Capability and Performance – Part 1: General Principles and Concepts) [7].

In this era of strains on the resources and rising costs of manufacturing, it becomes increasingly apparent that decisions must be made on facts, not just opinions. Consequently, data must be gathered and analysed. This is where statistical process control (SPC) comes in. For over 70 years, the manufacturing arena has benefited from the tools of SPC that have helped guide the decision-making process.

In particular, the control chart has helped determine whether special-cause variation is present implying that action needs to be taken to either eliminate that cause if it has a detrimental effect on the process or to make it standard operating procedure if that cause has a beneficial effect on the process. If no special-cause variation is found to be present, SPC helps define the capability of the stable process to judge whether it is operating at an acceptable level.

Accountability with hard data, not fuzzy opinions, is being demanded. Existing processes must be examined and new ones discovered. The good news is that improved quality inherently lowers costs as it provides a better product and/or service. Statistical Process Control provides accountability and is an essential ingredient in this quality effort.

Statistical Process Control is not an abstract theoretical exercise for mathematicians. It is a hands-on endeavour by people who care about their work and strive to improve themselves and their productivity every day. SPC charts are a tool to assist in the management of this endeavour. The decisions about what needs to be improved, the possible methods to improve it, and the steps to take after getting results from the charts are all made by humans and based on wisdom and experience [6].

3 Problem Formulation

The first survey was conducted for the issue of statistical methods and tools. It consisted in studying the available literature on these instruments. The aim of this research was to analyse which methods and tools are recommended with a focus on statistical process control. It was studied 25 (of which 4 Czech and 21 English) professional publications on the theme of the use of statistical methods in quality control. Each has its own unique approach to characterize this strategy as well as different approaches to methods and tools in the SPC issue.

The list of most frequently reported methods and tools according to the literature is shown in Figure 8. It was subsequently conducted second survey which consisted in questioning the applicability of statistical methods in quality management in selected enterprises. 26 companies (32,5 %) involved to a questionnaire survey of all 80 surveyed companies. The aim was to determine whether the selected statistical methods presented in professional publications related to the SPC are really used in practice.

Two research hypotheses were defined:

1. Does the degree of difficulty of used statistical methods in the SPC depend on the firm size?
2. Are the selected statistical methods reported in scientific publications related to the SPC really used in practice?

3.1 Research Method Conducted

Qualitative and also quantitative research was used for the issue solution.

When solving these research problems methods are adequate to issues and objectives – a combination of qualitative and quantitative methods of research. Quantitative research method is used for summarizing the results of comparative survey and a questionnaire. Statistical methods (descriptive and mathematical statistics – statistical induction) used in processing the results are based on the experience of the authors.

3.2 Processing the Results of Research by Analysing the Dependence of Two Categorical and Two Numerical Variables

This section defines the statistical methods used in processing the results of our research. When investigating the possible dependence of two
nominal variables defined in the first research hypothesis will be used χ²-test in contingency table. The results of observations are recorded for easy reference in the contingency (pivot) table. Pivot table is created when we sort the file according to two variants of qualitative features A and B where A has r variants and B has s variants. The null hypothesis is: A variable and B variable are independent. Test statistic is displayed by following formula:

\[ \chi^2 = \sum_{i=1}^{r} \sum_{j=1}^{s} \frac{(n_{ij} - n_{ij}^*)^2}{n_{ij}^*}, \]  

(1)

where \( n_{ij}^* \) are the theoretical frequencies [2], [5].

The rejection region is defined by: \( \chi^2 \geq \chi_{1-\alpha}^2((r-1)(s-1)) \). If the value of the test statistic falls into rejection region we can reject the null hypothesis. It means that the dependence between A and B was proved [3].

The Spearman rank correlation coefficient (similar to the Pearson paired correlation coefficient) is used in investigating if there a dependence between two numeric (ordinal) variables. It is defined by the following formula

\[ r_s = 1 - \frac{6 \sum (i_x - i_y)^2}{n(n^2 - 1)}, \]

(2)

where \( i_x \) and \( i_y \) are the ranks of values of variables \( x \) and \( y \), \( n \) – sample size [4], [10].

Values of the Spearman rank correlation coefficient are from the interval \((-1; 1)\), while values around 0 indicate independence, values close to 1 or \(-1\) indicate direct or indirect dependence.

### 4 Problem Solution

First, we focus on the distribution of companies by sector in the questionnaire. The following Figure 1 shows the relative proportions of respondents by the industry sector. Respondents answered the following question: “In which industry sector your company operates?”

The previous graph shows that most of the companies surveyed were from rubber and plastics industries. The relatively high proportion was also observed in the area of chemical and engineering industries. Section Other includes companies operating in the food and service sector. Construction companies have the lowest proportion (7.69%).

Another part of the questionnaire survey focused on the distribution of companies by the number of employees. Interviewed respondents answered the question: “How many employees work in your company?” Results are shown in Figure 2.

Most respondents are companies that employ more than 250 employees and companies from 50 to 249 employees.

The question arises at this point whether the degree of statistical methods in the SPC difficulty depends on the company size. This question is answered below. The following Figure 3 shows respondents’ answers to the question: “What does your company turnover in millions of Euros?”

![Fig. 1 Distribution of Surveyed Companies by Sector [Source: Own Processing]](image)

![Fig. 2 Distribution of Surveyed Companies by Number of Employees [Source: Own Processing]](image)

![Fig. 3 Distribution of Surveyed Companies According to Annual Sales in Millions Euros [Source: Own Processing]](image)

The question: “Is there legislation in your company (instructions, directives, manuals, etc.) that would describe the statistical methods used?” replied 84.62% of surveyed respondents in the affirmative suggesting that the companies actually use the statistical methods in quality control.
22 respondents answered in the affirmative in the absolute frequencies.

Last but one question concerning the companies was: “Who carries out the application and evaluation of statistical methods?” As shown in the following Figure 5, application and evaluation of statistical methods is most performed at the level of middle management and specialist. The statistical analysis is carried out by trained Black Belt at the level of top management in the large companies.

4.1 Dependence of Degree of Difficulty of the Statistical Methods in Quality Control on the Company Size

We try to verify or reject this research hypothesis on the basis of statistical nonparametric $\chi^2$-test of independence in the contingency table. Companies were divided into categories according to the number of employees and turnover according to [9]. Difficulty of the used statistical methods was divided as follows including their abbreviations in parentheses:

- **Elementary methods**: Pareto analysis (PA), cause and effect diagram (CED), histogram (HI), scatter plot (SP) and flowcharts (PM).
- **Medium intensive methods**: hypotheses testing (HT), analysis of variance (ANOVA), time series (TS), design of experiments (DOE), regression and correlation (RaCA).
- **Intensive methods**: statistical process control (SPC), investigation of process capability (ICP), measurement system analysis (MSA), method FMEA (FMEA), statistical acceptance (AS), multivariate methods SPC (MMSPC).

Tab. 1 $\chi^2$ Test of Independence in Contingency Table, NCSS 2007 Output [Source: Own Processing]

<table>
<thead>
<tr>
<th>Chi-Square Statistics Section</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Chi-Square</td>
<td>17.0093</td>
</tr>
<tr>
<td>Degrees of Freedom</td>
<td>6</td>
</tr>
<tr>
<td>Probability Level</td>
<td>0.00924</td>
</tr>
<tr>
<td>Cramer's V</td>
<td>0.29018</td>
</tr>
<tr>
<td>Pearson's Contingency Coefficient</td>
<td>0.37965</td>
</tr>
<tr>
<td>Tschuprow's T</td>
<td>0.26220</td>
</tr>
<tr>
<td>Kendall's tau-B</td>
<td>0.01980</td>
</tr>
</tbody>
</table>

Probability Level $\leq \alpha$ (0.05), we can reject the null hypothesis about the independence of two categorical variables. It follows that the complexity of the methods depends on the size of the company. The contingency coefficients are measures of intensity of dependence. Following coefficients can be calculated in statistical software: Pearson's, Cramer's contingency coefficients or Kendall's $\tau_b$ and Tschuprow's contingency coefficients.

We can say that exist a weak dependence but it is statistically significant from results displayed in Table 1. The last question was: “What statistical methods in quality control are used in your company?” The results are given in the next Figure 6. It is obvious that most companies are satisfied only with histogram (69.23 %), capability index (65.38 %), Pareto analysis (65.38 %) and flowcharts (57.69 %).
The list of most frequently reported methods and tools according to the literature is shown in Figure 8.

It is apparent from previous graph that the most frequently mentioned methods (100%) in scientific literature are statistical process control (SPC), analysis of variance (ANOVA), and histogram (HI). Other methods and tools according to publications studied are: measurement system analysis (80%), scatter plots (80%), cause and effect diagram (96%), method FMEA (84%), Pareto analysis (84%), design of experiments (92%) and flowcharts (88%).

If the $p$-value $\leq \alpha$ (0.05) we can reject the null hypothesis about independence of variables $Y_i$ and $X_i$. In this case, the $p$-value of Spearman rank correlation coefficient is $|r_s| > r_s(\alpha)$, where $r_s(\alpha) = 0.5$, therefore we cannot reject the null hypothesis about the independence of variables $Y_i$ and $X_i$ at 5% confidence level. The above mentioned conclusion can be supported by the confidence interval for the Spearman rank correlation coefficient “95.00% CI Limits”(-0.2409; 0.7376).

10000 Monte Carlo samples with starting seed 262676

4.2 Theory and Practice Accordance in Using Selected SPC Methods

We continue with statistical data processing in program StatXact 7. Second hypothesis if the selected statistical methods reported in scientific publications related to the SPC are really used in practice. It is obvious from following Table 2 that the value of Spearman rank correlation coefficient equals to 0.2483 and the $p$-value for both sided exact test by the method Monte Carlo is 0.3454 (see Table 3).

5 Discussion

On direct questioning of respondents regarding the use of statistical methods in quality management, we often met with a faulty analysis of multiple comparisons of mean values when were carelessly used $t$-tests instead of analysis of variance.
Therefore low frequency by ANOVA method was observed. In this case, practitioners often posed a question whether to use ANOVA or \( t \)-test for two groups \((k = 2)\). It does not matter, \( p\)-value is released in both cases exactly the same because \( F \) is the square of the \( t \). We also met with nonparametric tests ignorance resulting in a low value for hypotheses testing.

Most surprising was the fact that only two respondents use multivariate control charts and multidimensional capability indices. This shows great ignorance of multivariate SPC methods among practitioners. The lowest frequencies were observed in the measurement system analysis (MSA) and design of experiments (DOE). But the majority of the scientific literature placed the greatest emphasis on these methods used in pre-production and production phases. The questionnaire survey also showed that elementary statistical methods are used most often in practice.

6 Conclusion

Two surveys were conducted to prove both research hypotheses. The first survey was conducted for the issue of statistical methods and tools. It consisted in studying the available literature on these instruments. The aim of this research was to analyze which methods and tools are recommended with a focus on statistical process control. It was studied 25 (of which 4 Czech and 21 English) professional publications on the theme of the use of statistical methods in quality control.

It was subsequently conducted second survey which consisted in questioning the applicability of statistical methods in quality management in selected enterprises. 26 companies (32.5 \%) involved to a questionnaire survey of all 80 surveyed companies. To perform research, data were collected and this was obtained through working with managers of companies and managers in production. It was used a web interface and database for collecting and storing data. The last step was the evaluation of these data with adequate statistical methods including interpretation of the conclusions provided by them.

The first research hypothesis (if the degree of difficulty of used statistical methods in the SPC depends on the firm size) was proved by the \( \chi^2 \)-test in the contingency table. Association coefficient detected only weak (but statistically significant) dependence between two nominal variables.

The nonparametric Spearman rank correlation coefficient was used for theory and practice concordance detection. In this case, the second research hypothesis (if the selected statistical methods reported in scientific publications related to the SPC are really used in practice) was not proved.

References: