

Development of a software module for quality control and statistical data processing in production

Olesya Golubeva^{1,*}, Alina Pogorelova¹

¹Don State Technical University, 1, Gagarin Square, 344003, Rostov-on-Don, Russia

Abstract: the article considers the algorithmic and mathematical justification for the development of a new software module for product quality control and statistical data processing at the Auditor enterprises. The software is designed for automated calculation of product quality control and evaluation indicators. The program has a wide range of functions for statistical data processing and product quality control tools at the enterprise. The functional characteristics of the software include: comparing risks for a given set of factors for two cases, building a histogram of the distribution of cases over a given numerical interval, comparing objects for a given set of values using a generalized estimate, calculating global priorities for a given set of factors for several cases, calculating the I-th predicted value for a linear regression model, calculating the i-th predicted value for an exponential regression model, calculating the i-th predicted value for a parabolic regression model, calculating the consistency of experts for a given set of factors, calculating the n+1 value of a time series, and evaluating the level of employee satisfaction.

1 Introduction

In the current global environment, there is an urgent need to improve the quality and automation of production. Almost everywhere we use outdated technological maps for manufacturing products based on old databases of materials, equipment, and product specifications. Manual indicators of accounting and quality control of manufactured goods are more common than automatic ones.

Now there are information systems that can be used to manage enterprises for the production of any product. But these systems are far from perfect and require constant improvement.

Almost all enterprises have established and operate quality control departments for products, but they can not always work perfectly – this affects the quality and quantity of finished products.

To improve and ensure the quality of products, as well as the convenience of statistical processing of data obtained in production, a software module was developed as part of the information system being developed. Its goal is to simplify the solution of some of the

*Corresponding author: 1354565@mail.ru

issues related to the quality management of the production process by automatically processing input data based on various quality tools.

The General structure of the program includes 9 sections: calculation of the organization's risks related to the quality of products; basic concepts of decision-making theory; construction of a generalized assessment; decision-making based on the hierarchy analysis method; construction of a regression model; decision justification under many performance criteria; expert assessment method; methods of time series analysis; motivation management.

2 Calculating the organization's risks related to product quality

Consider the sequence of statistical assessment of the degree of risks associated with the quality of production.

First of all, we will determine the weighted average value of the analyzed parameter, which is a causal risk factor:

$$\Pi_{cp} = \Sigma \Pi_i \times P_i \quad (1)$$

where: Π_i – is the i-th value of the analyzed parameter;

P_i - probability of occurrence of the i-th value of this parameter.

Probability:

$$P_i = \frac{K_i}{n} \quad (2)$$

where: K_i – is the number of times the I-th parameter value is repeated;

n – the total number of analyzed parameter values.

Next, we determine the variance of the values of the analyzed parameter:

$$D = \Sigma(\Pi_i - \Pi_{cp})^2 \times P_i \quad (3)$$

using the value of the found variance, we determine the standard deviation of the expected value of the analyzed parameter from its average value:

$$\sigma = \sqrt{D} \quad (4)$$

where D is the variance of the values of the analyzed parameter.

The larger the value of the standard deviation, the higher the uncertainty of the appearance of a particular parameter value and, consequently, the higher the risk associated with the fact that the necessary (desired) parameter value will not be obtained.

To compare different solutions with different expected results and different risks, you can use the coefficient of variation:

$$\gamma = \frac{\sigma}{\Pi_{cp}} \quad (5)$$

The value of the coefficient of variation characterizes the size of the risk per unit of the expected result. Therefore, the option that has the lowest value should be chosen as the least risky (with a lower relative risk).

3 Basic concepts of decision theory

Calculation algorithm in this section:

- 1) Building a histogram of the output distribution, the x_i intervals are plotted on the abscissa axis, and the number of observations m_i is plotted on the ordinate axis.
- 2) The calculation of the value of the empirical average:

$$\bar{X} = \sum_{i=1}^n x_i \times \frac{m_i}{N} \quad (6)$$

- 3) Calculation of variance:

$$D(x) = \frac{\sum_{i=1}^n (x_i - \bar{X})^2}{n} \quad (7)$$

- 4) Calculation of the mean square deviation:

$$\sigma_x = \sqrt{D(x)} \quad (8)$$

5) Determination of the coefficient of variation:

$$V = \frac{\sigma_x}{\bar{x}} \quad (9)$$

The calculated numerical characteristics provide certain information about the probabilistic parameters of the process under consideration.

4 Building a generalized estimate

The theory of this section is based on the cybernetic principle of deviation control using target control methods. The method focuses on achieving the final results of the subsystems functioning and on the resulting mismatches.

Calculation algorithm:

- 1) In order to match the real numerical values of the x factors with the numerical values of the OX scale, you must enter the scale factor (M), which is determined by the formula:

$$M_i = \frac{x_{imax} - x_{imin}}{X_B - X_A} \quad (10)$$

- 2) Next, you need to translate the natural (real) values of the factors into dimensionless ones.

If the influence of the factor on the result is direct ($NAP = 1$), then the formula is used:

$$x_i = X_A + \frac{x_i^p - x_{imin}}{M_i} \quad (11)$$

If the influence of the factor is the opposite ($NAP=0$), the formula is used

$$x_i = X_B - \frac{x_i^p - x_{imin}}{M_i} \quad (12)$$

- 3) To find the partial desirability coefficients d_i substitute the obtained values of X in the equation.:

$$d = e^{-e^{4-x}} \quad (13)$$

- 4) We determine the generalized value of the preference coefficient for each technology D by the formula:

$$D = \sqrt[n]{\prod_{i=1}^n d_i} \quad (14)$$

5 Decision making based on the hierarchy analysis method

This sequence of steps is recommended for solving the problem.

1. Outline the problem and determine what you want to know.
2. Build a hierarchy starting from the top (goals-from the management point of view), through intermediate levels (criteria that determine subsequent levels) to the lowest level (which is usually a list of alternatives).
3. Build a set of matrices of paired comparisons for each of the lower levels - one matrix for each element of the level adjacent to the top.
4. using hierarchical synthesis to weigh the eigenvectors with the weights of the criteria, the sum of all the corresponding weighted components of the eigenvectors of the hierarchy level below is calculated.
5. Interpretation of calculation results. Conclusion.

6 Building a regression model

Building a factor regression model consists of several stages:

- 1) Preliminary data analysis (descriptive statistics, distribution histograms, "boxes with moustaches").

- 2) Search for, exclude anomalous observations and normalize data in order to exclude the influence of different units of measurement and scales, or logarithm without excluding anomalous ones.
- 3) analysis of variable spread graphs.
- 4) Correlation analysis. Building a list of model specifications.
- 5) Estimation of parameters of regression model variants from the list.
- 6) Search for and eliminate abnormal residues.
- 7) analysis of model residuals.
- 8) correction of model errors.
- 9) Compare the quality of models.
- 10) Interpretation of constructed models.
- 11) Forecast for the best model.

7 Justification of the decision under many performance criteria

Decision-making under many criteria is based on the principle of Pareto consistent optimism and is a multi-step process that begins with the appearance of a problem and ends with the implementation of solutions.

- 1) collecting source data and structuring the problem:
 - a) limiting complexity
 - b) displaying the situation
 - d) evaluating the resource
 - e) identifying relationships
- 2) Identification and systematization of potential possible solutions "Sifting" solutions, identifying a variety of competing solutions
- 3) Justification of particular criteria for evaluating competing solutions
- 4) Building a logical-mathematical system model and its verification
- 5) Choosing a rational solution that should be:
 - a) unique
 - b) timely
 - c) implementable
 - d) sustainable
 - e) promising

Despite the fact that decision-making methods are universal, their successful application largely depends on the professional training of a specialist who must be able to correctly set the task.

The algorithm for one of the correct ways to set the problem is shown in figure 1.

The process of formulation of the problem

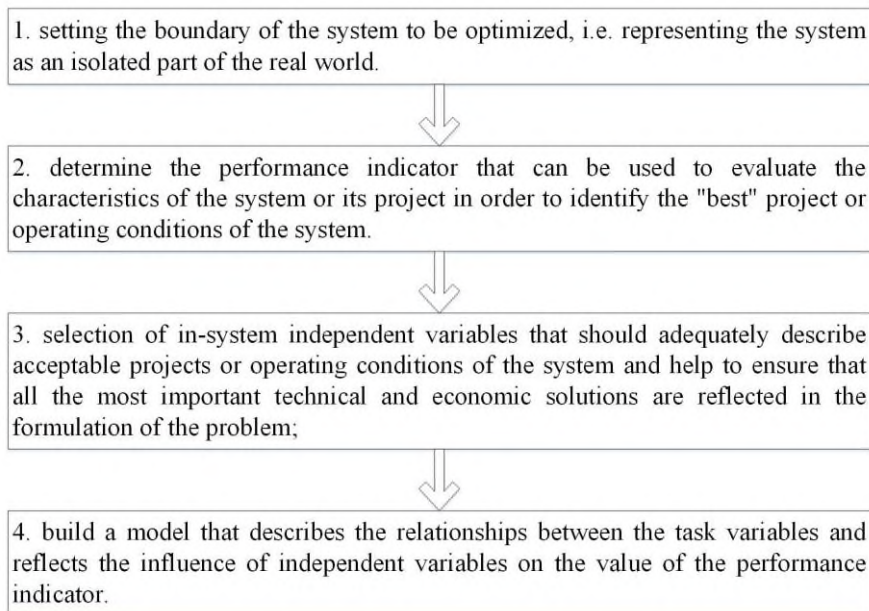


Figure 1. The process of formulation of the problem.

Following this algorithm, the employee will be able to complete the task much more efficiently and less labor-intensive.

8 The method of expert evaluation

THE SEQUENCE OF STEPS OF THE ANALYSIS

Stage 1. Experts are offered a questionnaire in which each of the n experts is asked to write a list of parameters (quality indicators) that characterize, in their opinion, a specific object (product, technological process, etc.). as a result, a table is compiled with a list of indicators (m is the number of parameters) and their symbols.

Stage 2. Each expert individually fills out the second questionnaire, where the indicators should be ranked (ordered) by their significance. If, according to the expert, the indicators do not differ in the strength of their influence (significance) on the object under study, then they are assigned the same ordinal number (rank). Indicators are ranked in descending order from most significant (rank 1) to least significant (rank m).

Stage 3. On the basis of individual questionnaires, a General matrix of ranks is formed: X_{ij} – the rank of the j -th indicator for the i -th expert; n – the number of experts; m – the number of indicators.

Stage 4. Based on the data, the degree of consistency of the opinions of the interviewed experts is assessed. The need for this procedure is due to the fact that: firstly, due to the individual characteristics of each expert, the level of their knowledge and ideas about the object, the parameters can be interpreted ambiguously; secondly, the ranking of parameters can be made unqualified due to insufficient knowledge of the object.

The degree of consistency of experts' opinions is assessed using the concordance coefficient. If there are equal rank values in the i -th row of the rank matrix, then the concordance coefficient is defined as

$$W = \frac{S}{\frac{1}{12}n^2(m^3 - m) - n \sum_{i=1}^n T_i}, \quad (15)$$

where S and T are defined by the formulas:

$$S = \sum_{j=1}^m \left(\sum_{i=1}^n X_{ij} - \frac{1}{m} \sum_{j=1}^m \sum_{i=1}^n X_{ij} \right)^2 \quad (15)$$

$$T_i = \frac{1}{12} \sum_{j=1}^m (t_j^3 - t_j) \quad (16)$$

where t_j is the number of ranks of the j -th index in each row of the rank matrix. If the rank matrix does not have equal rank values in each row, the assessment of the degree of consistency of expert opinions is determined by the formula

$$W = \frac{12S}{n^2(m^3 - m)} \quad (17)$$

The value $W = 1$ indicates a complete agreement of experts' opinions. The value $W = 0$ indicates a complete discrepancy between the opinions of experts. In practical situations, $0 < W < 1$. If the coordination coefficient is close to zero ($W = 0.1 \div 0.5$), it should be concluded that the composition of experts was selected unsuccessfully or the object was not studied fully enough. If the value of W is close to one ($W \geq 0.9$), it may turn out that the examination was carried out formally without proper study of the object. In both of these cases, the examination should be repeated.

Stage 5. The significance (difference) of the concordance coefficient W from zero is checked using the Fisher z-test:

$$z = \frac{1}{2} \ln \frac{(n-1)W}{1-W} \quad (18)$$

If $z \geq z_\alpha$, then with probability $P = (1-\alpha)$ (where α is the significance level of the criterion), it can be argued that there is non-random consistency in the opinions of experts. In this case, z is found by the formula (5), and z_α is found using statistical tables. The initial data for obtaining z_α are the significance level of the z-criterion α , which is usually taken to be 0.01-0.10, and the degrees of freedom ν_1 and ν_2 :

$$\nu_1 = (m-1) - \frac{2}{n} \quad (19)$$

$$\nu_2 = (n-1) \nu_1 \quad (20)$$

If $z < z_\alpha$, then with probability $P \geq (1-\alpha)$, it can be argued that there is no consistency between experts and it is necessary to conduct a new survey or identify a group of experts who have a high consistency of opinions. For this purpose, one expert is excluded from the population and the W_1 coefficient is calculated for the remaining experts. If $W_1 > W$, this expert Advisor is excluded from the aggregate. Such calculations are performed for each expert. As a result of calculations, the degree of consistency of the opinions of the experts remaining in the aggregate increases.

9 Time series analysis methods

A number of methods are used to analyze time series. Within the created module, the moving average method is used.

This method is based on the fact that a more accurate forecast for the future can be obtained if recent observations were used, and the "newer" the data, the greater their weight for the forecast should be.

To calculate a simple moving average, the average of the last observations of a fixed number N is used to estimate the next value of the time series.

$$\bar{x}_k = \frac{1}{N} \sum_{i=1}^N x_{k-1} \quad (21)$$

To assess the accuracy of forecasts, the average of absolute deviations (CAO) and the average of relative errors (COOP) are used, calculated using the formulas::

$$CAO = \frac{\sum_N |x_i - \bar{x}_i|}{N} \quad (22)$$
$$COOP = \frac{\sum_N \frac{|x_i - \bar{x}_i|}{x_i} \times 100\%}{N} \quad (23)$$

Where N is the number of forecasts.

10 Motivation management

To assess the level of staff motivation, the motivation factors are given. Select the indicators that need to be evaluated for each employee of the Department where the motivation analysis is performed, or members of the sample group.

Thus, the degree of satisfaction (motivation) of an individual employee of an enterprise can be calculated using the formula:

$$SU = \frac{(\sum U_i^+ m_i + \sum U_i^- q_i)}{100\%} \quad (24)$$

Where SU – is an assessment of the employee's satisfaction level (motivation level);

U_i^+ – employee's positive assessment of the motivation factor;

U_i^- – employee's negative assessment of the motivation factor;

m_i и q_i – are the weights of the factor as a motivator and as a hygiene factor, respectively.

The motivation factor is usually evaluated based on the results of an employee's survey.

The level of motivation of employees directly determines what restructuring tasks can be solved with the participation of these employees.

Conclusion

The program includes a wide range of statistical data processing and product quality control tools at the enterprise, including building various types of charts, graphs and diagrams based on them, which are widely used in statistical methods, which allow you to identify the most important factors that affect production processes.

Based on the described functionality of the program, we can conclude that the object of software implementation can be the educational process in such disciplines as "System analysis", "Decision Theory" and "Quality Management". This software can also be recommended for implementation in the enterprise information system.

References

- 1) K. Linz, G. Muller-Stievens, A. Zimmerman, *Radical change in Business Model, Adaptation and survival in a competitive environment* (Alpina Publisher, 2019)
- 2) K.V. Rochev, *Information Technology. Analysis and Design of Information Systems* (2019)

- 3) Halim, A. 2013. *Pengaruh Kompetensi dan Independensi Auditor Terhadap Kualitas Audit dengan Anggaran Waktu Audit dan Komitmen Profesional Sebagai Variabel Moderasi*. Disertasi. Universitas Brawijaya: Malang.
- 4) O.L. Golitsina, I.I. Popov, N.V. Maksimov, *Information Systems and Technologies. Tutorial* (2019)
- 5) D.Y. Katalievskiy, A.Y. Ivanov, *Modern Agrotechnologies, Economic and Legal and Regular Aspects* (2018)
- 6) A.G. Skhirtladze, *Technological Process Automation* (2016)
- 7) I.F. Borodin, *Automation of Technological Processes and Automatic Control Systems* (2018)
- 8) L. A. Ginis, *Statistical Methods of Quality Control and Management, Applied Software* (2019)
- 9) A.K. Ershov, *Quality Management* (2017)
- 10) S.I. Solonin, *Control Charts Methods* (2014)
- 11) I.V. Maruseva, *Modern Management, Classic and Applied Aspects* (2018)
- 12) V.N. Kraev, *Management Decision Making Methods* (2014)
- 13) J.C. Choi, C. Kim, J. Mater. Process Technol., 110 (2013)
- 14) G. Colombo, D. Ferretti, U. Cugini, *Proceedings of international symposium on advanced geometric modelling for engineering applications*, 2-15 (2014)
- 15) G. La Rocca, L. Krakers, M.J.L. van Tooren, *Proceedings 9th symposium on multidisciplinary analysis and optimization*, 2-13 (2014)
- 16) Luximon, A. *Handbook of Footwear Design and Manufacture* — 2013. — Vol. 64, № 5. — P. 416.