https://doi.org/10.21122/2227-1031-2022-21-2-158-163

UDC 658.511

Statistical Analysis of Industrial Processes

T. I. Chepeleva[1], A. N. Chepelev[2]

[1]Belarusian National Technical University (Minsk, Republic of Belarus),
[2]Academy of Public Administration under the President of the Republic of Belarus (Minsk, Belarus)

Abstract. To analyze the synchronization of machine-building production processes, a statistical approach is used on the correctness of the choice of parameters. The accuracy of the parameters is achieved by modeling. This method allows checking how correctly the parameters are selected and whether they ensure uninterrupted operation of production. Statistical analysis of parameters gives information on failures in a particular production unit, on violations of the synchronization of production processes, technological processes. With the help of statistical characteristics, it is possible to evaluate production as a whole. Statistical multidimensional analysis of complex production data allows to analyze the work of individual units, production blocks. The cluster analysis has been carried out by the method of *K*-medium production process with minimization of the total error probability. A geometric interpretation of the results of cluster analysis of production processes is given in the paper. The influence of factors on the work of production has been determined in the paper. The index factorial method with a different comparison base and different weights has been applied. The hypothesis about the adequacy of the model of production processes has been tested. A statistical analysis of the complex data of the production process has been carried out in the search for optimal solutions in the case of uncertainty and in conditions of risk using the following methods: Bayes, Laplace and Germeyer using the simplex method. The network methods of decision-making have been used in the paper. Statistical methods with the help of mathematical modeling have substantiated the optimal sizes of both individual parts and volumes of local warehouses, so that there were no delays in the transfer of production processes, disruptions in work, downtime of working equipment. In this case, the criterion for the optimality of production volumes can be the minimum of total losses from idle time of individual units and production blocks to the possibility of disruption of the synchronous process modes due to lack of equipment (arising production pockets) or due to long-used outdated units.

Keywords: processes, statistical analysis, modeling, factors

For citation: Chepeleva T. I., Chepelev A. N. (2022) Statistical Analysis of Industrial Processes. *Science and Technique*. 21 (2), 158–163. https://doi.org/10.21122/2227-1031-2022-21-2-158-163

Статистический анализ производственных процессов

Канд. техн. наук, доц. Т. И. Чепелева[1], А. Н. Чепелев[2]

[1]Белорусский национальный технический университет (Минск, Республика Беларусь),
[2]Академия управления при Президенте Республики Беларусь (Минск, Республика Беларусь)

Реферат. Для анализа синхронизации процессов машиностроительного производства используется статистический подход о правильности выбора параметров. Точность величин достигается путем моделирования. Данный способ позволяет проверить, насколько правильно выбраны параметры и обеспечивают ли они бесперебойную работу производства. Статистический анализ величин дает информацию о сбоях в том или ином узле производства, о нарушениях синхронизации процессов производства, технологических процессов. С помощью статистических характеристик можно оценить производство в целом. Статистический многомерный анализ сложных данных о производстве позволяет анализировать работу отдельных узлов и блоков производства. Проведен кластерный анализ методом *K*-средних процессов производства с минимизацией полной вероятности ошибки. Дана геометрическая интерпретация результатов кластерного анализа процессов производства. Определено влияние факторов на работу производства. Применен индексный факторный метод с различной базой сравнения и разными весами. Проведена проверка гипотезы об адекватности модели процессов производства. Выполнен статистический анализ сложных данных процесса производства при поиске оптимальных решений в случае неопределенности и в условиях риска методами Байеса, Лапласа и Гермейера с помощью симплекс-метода. Использованы сетевые методы принятия решений. Статистическими

Адрес для переписки
Чепелева Тереса Иосифовна
Белорусский национальный технический университет
ул. Б. Хмельницкого, 9,
220013, г. Минск, Республика Беларусь
Тел.: +375 17 292-80-75
tchepeleva@gmail.com

Address for correspondence
Chepeleva Teresa I.
Belarusian National Technical University
9, B. Khmelnitsky str.,
220013, Minsk, Republic of Belarus
Tel.: +375 17 292-80-75
tchepeleva@gmail.com

методами посредством математического моделирования провели обоснование оптимальных размеров как отдельных деталей, так и объемов местных складов, чтобы не было опозданий в работе передачи процессов производства, срывов в работе, простоев оборудования. В этом случае критерием оптимальности объемов производства может служить минимум суммарных потерь от простоя отдельных узлов и блоков производства до возможности нарушения режимов синхронного процесса из-за нехватки оборудования (возникших карманов производства) или из-за давно используемых устаревших агрегатов.

Ключевые слова: процессы, статистический анализ, моделирование, факторы

Для цитирования: Чепелева, Т. И. Статистический анализ производственных процессов / Т. И. Чепелева, А. Н. Чепелев // *Наука и техника*. 2022. Т. 21, № 2. С. 158–163. https://doi.org/10.21122/2227-1031-2022-21-2-158-163

Introduction

Statistical methods are widely used in the industrial complex, which is very important at the stages of designing machines, creating them, producing them, operating them. Statistical analysis is used both for the formation of the industrial complex itself, and for the creation of separate equipment, parts, separate nodes. All technical tasks are continuously connected with the economic and social direction of the country. The presented work has a socio-economic direction. The strength and reliability of manufactured machines is a global pressing issue of our time. The work of the industrial complex should be first of all uninterrupted, it requires appropriate preliminary competent calculations, clearly reflecting the issues raised. Algorithmization of the production processes of the industrial complex was carried out, and then a statistical study was conducted on the created algorithms.

Materials and results of studies

The production process is carried out by people who own modern technology, i. e. it is a set of actions of personnel and equipment, resulting in new machines, or kits, parts to them, the necessary planned ready-made products within a certain time frame with minimal defect. Statistical analysis allows us to solve these problems even at the design stage of both individual parts, a separate machine, and the entire industrial complex [1–3]. Statistical analysis evaluates the reliability and risk of equipment failure, determines the suitability of production processes.

The work describes the mathematical apparatus, which allows to determine the efficiency of the enterprise, the effectiveness of competing processes of a multi-structural resource for the production of machines at the design stage of machines. A mathematical model of the distributed processing of competing processes of a production resource is obtained, which can be used to solve managerial issues in the manufacture of machines.

Let's look at the parallel distributed and consistent execution of production processes. An industrial complex, an enterprise must operate smoothly like a powerful computer. Let it a – a time used by a variety of distributed competing processes to organize the parallel use of the working blocks of a multi-structure production resource, $a > 0$; n – a variety of different competing processes; p – the number of intellectual performers; v – the number of work blocks; B_j – a set of work blocks ($j = \overline{1, v}$, $j = 1, n$); a_{ij} – time to execute j production blocks by i competing processes $j = \overline{1, v}$; $i = \overline{1, n}$, $2 \le v \le p$.

Let the system of competing multi-structure resource processes are equally distributed, i. e. have the same duration of production processes. Therefore, there is a coincidence by the time of each of i processes: $a_{i1} = a_{i2} = ... = a_{iv} = a_i$. Let $t^n = \sum_{i=1}^{n} a_i$ – the total run time each of the operating B_j units by all n processes; a variety $a = a\left(a_1, a_2, ..., a_n, t^n\right)$ – the characteristic set of this system; $F\left(t_a^n\right) = a$, $t_{m-1}^{\alpha} \le t_m^{\alpha} \ge t_{m+1}^{\alpha}$, $m = \overline{1, n}$, where α – the time spent on organizing the parallel use of multi-structure production resource units by a variety of distributed competing workflows, $\alpha > 0$. It proved, that then the minimum total time to perform a plurality of equally distributed competing workflows respectively in serial mode and two parallel distributed modes coincide, otherwise $t_{pr}^1 > t_{pos}^3 = t_{pr}^2$, where t_{pr}^1, t_{pr}^2 – minimum total time (it's two of them) of parallel production mode; t^3_{pos} – sequential. Indeed, if $a^\alpha_m = \max\limits_{1\le i\le n} a^\alpha_i$, then for the third sequential and the second parallel distributed production modes, $t^3_{pos} = t^2_{pr} = t^n_\alpha + (\nu-1)a^\alpha_m$ equality occurs. In this case, every work unit is continuously performed by all n processes for any characteristic set of the system a.

Within 1st parallel distributed mode with continuous execution by multi resource production operating units within each of the processes, their interaction and control subordinate algorithm: $t^1_{pr} = t^n_\alpha + (\nu-1)\left(a^\alpha_n + \sum\limits_{i=2}^{n}\max\left\{a^\alpha_{i-1} - a^\alpha_i, 0\right\}\right)$. Then $a^\alpha_n + \sum\limits_{i=2}^{n}\max\left\{a^\alpha_{i-1} - a^\alpha_i, 0\right\} = a^\alpha_m$. But $a^\alpha_m = \max\limits_{1\le i\le n} a^\alpha_i$, then when $1 \le i \le m \le n$ equality $\sum\limits_{i=2}^{m}\max\left\{a^\alpha_{i-1} - a^\alpha_i, 0\right\} = 0$ is true, and for $1 \le m \le i \le n$ – equality $\sum\limits_{i=m+1}^{n}\max\left\{a^\alpha_{i-1} - a^\alpha_i, 0\right\} = a^\alpha_m - a^\alpha_n$. $a^\alpha_n + \sum\limits_{i=2}^{n}\max \times \left\{a^\alpha_{i-1} - a^\alpha_i, 0\right\} = a^\alpha_n + a^\alpha_m - a^\alpha_n = a^\varepsilon_m$ and $a^\alpha_n + \sum\limits_{i=2}^{n}\max\left\{a^\alpha_{i-1} - a^\alpha_i, 0\right\} - \max\limits_{1\le i\le n} t^\alpha_i > 0$.

Systems, which satisfy equations of the form: $a_{i1} = a_{i2} = ... = a_{in} = a_i$, $i = \overline{1, n}$, called stationary, while they are effective at $p, \nu \ge 2$ if systems of competing processes of manufacture are equally distributed and the following relation $\nu t^n - t(p, n, \nu, \alpha) = \Delta_\alpha(n) > 0$ fulfill, where νt^n – time of the execution of the working units of manufacture B_j, $j = \overline{1, \nu}$ by all n processes in serial mode.

It is proved that if the production system operates in three modes: sequential and two parallel distributed, then for identically distributed competing production processes, the minimum total time, taking into account production delays and costs, is calculated by the formula

$$t(p, n, \nu, \alpha) = t^n_\alpha + (\nu-1)a^\alpha_{\max},$$

where $t^n_\alpha = \sum\limits_{i=1}^{n} a^\alpha_i$; $a^\alpha_{\max} = \max\limits_{1\le i\le n} a^\alpha_i$; $a^\alpha_i = a_i + \alpha$.

For an effective, equally distributed production system of competing workflows, exists a more efficient stationary equally distributed system, that satisfies the relations $\nu n \ge 2(n + \nu - 1)$, $0 < \alpha \le \min\limits_{1\le i\le n} a_i$, $n = \nu$, $p \ge \nu > 3$.

The opposite is true: an equally distributed production system of competing processes, with n, ν, p, α parameters that satisfies these relations, is effective. Indeed, if there are two effective equally distributed production systems of competing processes, it is said that the former is more effective than the second, if the value $\Delta_\alpha(n)$ of the first system is no greater than the corresponding value of the second.

Let there be an effective equally distributed production system, then, based on its effectiveness, inequality is true

$$\Delta_\alpha(n) = (\nu-1)\left(t^n - a^n_{\max}\right) - (n + \nu - 1)\alpha \ge 0,$$

where $a^n_{\max} = \max\limits_{1\le i\le n} t_i$,

and for any stationary distributed production system, it's valid:

$$\Delta^c_\alpha = (\nu-1)\left(t^n - a\right) - (n + \nu - 1)\alpha > 0,$$

where $a = t^n / n$.

It's easy to prove that $\Delta^c_\alpha(n) \le \Delta_\alpha(n)$ by substituting instead $\Delta_\alpha(n)$ and Δ^c_α corresponding values. Transforming the resulting expression, we will have: $(n-1)a \le t^n - t^n_{\max}$, since for a stationary, equally distributed system is valid $a = \min\limits_{1\le i\le n} a_i = a^n_{\min}$. Let for definiteness $a^n_{\max} = a_m$, then the following relationship is true:

$$t_n - a^n_{\max} = \sum_{i=1}^{m-1} a_i + \sum_{j=i+1}^{n} a_j \ge (n-1)a^n_{\min} = (n-1)a.$$

Let us prove the converse statement. Condition of the efficiency of production is equivalent to the inequality: $\dfrac{t^n - a^n_{\max}}{\alpha} \ge \dfrac{n + \nu - 1}{\nu - 1}$. We can verify directly that of the ratio $0 < \alpha < \min\limits_{1\le i\le n} a_i$ results a chain of inequalities: $\frac{t^n - a_{\max}^n}{\alpha} \geq \frac{(n-1)a_{\max}^n}{\alpha} \geq$ $\geq n-1,$ as by the choice α follows the inequality $\frac{a_{\min}^n}{\alpha} \geq 1.$ The inequality $n-1 \geq \frac{n+\nu-1}{\nu-1}$ follows from the expression $\nu n \geq 2(n+\nu-1).$

It is proved that in order to equally distributed production system competing work processes to be effective, it is necessary and sufficient that the following conditions are satisfied:

$$\alpha \leq \begin{cases} f\left(1+\sqrt{\nu}\right), \ \sqrt{\nu}\text{-integer}, \\ \max\left\{f\left(1+\left[\sqrt{\nu}\right]\right),\ f\left(2+\left[\sqrt{\nu}\right]\right)\right\}, \\ \sqrt{\nu}\text{-non-integer}; \end{cases}$$

$$f(x) = \frac{(\nu-1)t^n(x-1)}{x(x+\nu-1)},$$

where $[x]$ – is the largest integer not exceeding x.

A production system with equally distributed competing workflows will be effective when $p = \nu = 2,$ if $\frac{\alpha}{t^n} \leq \frac{n-1}{n(n+1)}.$

Production modes are classified into synchronous, asynchronous and combined. The synchronous production mode is a mode of sequential execution of all production processes without delay. Asynchronous mode is a mode of execution of production processes in violation of sequential cycles, with delays. For statistical research, this process can be considered as a discrete random vector [4, 5]. Asynchronous production mode leads to a number of negative consequences in production: accumulation of unnecessary products in warehouses and long-term storage; reducing the overall speed of the process and production efficiency; loss of work schedule, customers, working hours, usable area, etc. [6].

Suppose that the interaction of production processes, work processors and blocks is subject to the following conditions: none of the resource blocks can be processed simultaneously by more than one processor, and none of the processors can process more than one block at a time if each of them is processed without interruptions; the procedure for providing processor to processes is cyclical; a block is ready for execution if it is not running on any of the processors. Then when $n = kp$ minimum total time $T(p, n, s)$ of execution of the set of competing heterogeneous production processes determined by the critical path length in the network vertex-weighted graph, given by matrix A^*, in which starting and ending vertices of the graph – are the first and last elements of the matrix.

Let $T(p, n, s)$ be the minimum total execution time of all n concentrated processes $n \geq 2$ from the moment of the beginning of the first and to the moment of completion of the latter within an asynchronous production process. If $k = 1$, then $n = p$. In this case, it is enough to consider the execution of p processes on p processors. According to the matrix $A_{p\times s} = [t_{ij}],\ i = \overline{1, p},\ \ j = \overline{1, s},$ we obtain network vertex-weighted graph with the number of vertices $p\times s$. The vertices and arcs of the graph must match the nodes and arcs of the rectangular grid $p\times s$. Horizontal arcs in a graph indicate connections between production units, and vertical ones – a linear order of production processes. Let the numbering of the vertices and their weight – are the relevant elements a_{ij} of the matrix A^*, $i = \overline{1, p},\ \ j = \overline{1, s},$ where a_{ij} – production units. If $n > p$, then:

$$\begin{cases} n = kp, & \text{if } k > 1; \\ n = kp + r, & \text{if } k \geq 1, 1 \leq r < p. \end{cases}$$

Suppose $n = kp$, then we divide the matrix $A_{p\times s}$ into k block matrices with dimensions $p\times s$ of each, as a result n production processes are divided into k groups of p processes in each. For each of the block matrices, i. e. submatrices $A_{p\times s}^{(m)},\ \ m = \overline{1, k},$ we construct k linear diagrams Gantt. Each of the diagrams displays in time the execution of the next p_k processes with the help of PR_k processors, $k = \overline{1, p}.$

The execution of the next group of p_k processes should begin after the completion of the previous group. The total sum of the execution time of all P_i ($i = \overline{1, n}$) the processes in this case is defined as the sum of the lengths of critical paths in each of successive misalignment diagrams Gantt defined by a straight sum of matrices $A_{p\times s}^{(m)}$ ($m=\overline{1,k}$). To reduce time, you must use the technique of combining k – sequential diagrams along the time axis from right to left. It should be taken into consideration that the combination, starting from the second diagram, is carried out at the maximum possible value so that the technological conditions for the execution of many processes and production units are not violated. Block resultant matrix A^* of execution times of the blocks of resource production with taking into the account the maximum consecutive overlapping diagrams, has blocks of submatrix $A_{p\times s}^{(m)}$ ($m=\overline{1,k}$). In this case, the submatrix $A_{p\times s}^{(m)}$ in the resulting matrix A^* is positioned so that the nature of the interaction of the production resource blocks not violated, performed both using the same processor – an intelligent machine (as indicated by the horizontal connections between the blocks), so and by other processors (it's vertical communication). The first row of the matrix A^* consists of blocks of submatrices $A_{p\times s}^{(m)}$, which reflects the nature of the interaction of resource blocks and processes, using the same processors, and this, as a rule, is the intellectual resources of production. The step of combining diagrams is determined by the corresponding displacement of blocks (submatrices) $A_{p\times s}^{(m)}$ matrix A^*. An offset for vertical communications occurs starting from the first row, so that the next row, whose elements are blocks (submatrices) of the form $A_{p\times s}^{(m)}$ is shifted from right to left by the maximum value that does not violate the linear order of the execution of the same blocks for different production processes. Since all blocks (submatrices) $A_{p\times s}^{(m)}$ have the same dimensions $p\times s$, the displacement at each step is equal to the value s. Instead of the rightmost submatrix (block) shifted at each step, zeros are set. Having thus completed the $(k-1)$ step of shifting, we obtain the matrix structure A^* corresponding to the combined Gantt diagram. The matrix A^* takes into account both all horizontal connections between blocks and all vertical ones, as well as connections between blocks from Gantt diagrams. Matrix A^* as a block is symmetrical about a second upper triangular diagonal, type Hankel order k, is:

$$A^*=\begin{pmatrix} A^{(1)} & A^{(2)} & A^{(3)} & \dots & A^{(k-1)} & A^{(k)} \\ A^{(2)} & A^{(3)} & A^{(4)} & \dots & A^{(k)} & 0 \\ A^{(3)} & A^{(4)} & A^{(5)} & \dots & 0 & 0 \\ \dots & \dots & \dots & \dots & \dots & \dots \\ A^{(k-1)} & A^{(k)} & 0 & \dots & 0 & 0 \\ A^{(k)} & 0 & 0 & \dots & 0 & 0 \end{pmatrix},$$

where $A_{p\times s}^{(m)}$ ($m=\overline{1,k}$) – matrix of the form:

$$A_{p\times s}^{(m)}=\begin{pmatrix} a_{11}^{(m)} & a_{12}^{(m)} & \dots & a_{1s}^{(m)} \\ a_{21}^{(m)} & a_{22}^{(m)} & \dots & a_{2s}^{(m)} \\ \dots & \dots & \dots & \dots \\ a_{p1}^{(m)} & a_{p2}^{(m)} & \dots & a_{ps}^{(m)} \end{pmatrix}.$$

In the general case, the description of production processes at $n=kp$ will be given by the matrix $A_{kp\times ks}^*$. In this case, as with $n=p$, a vertex-weighted network graph is constructed with the weights given by the matrix $A_{kp\times ks}^*$. The vertices of the graph are located at the nodes of the rectangular grid ($kp\times ks$). Obviously, the weights of the vertices corresponding to zero values of the matrix $A_{kp\times ks}^*$, including the final vertex with a number (kp, ks), are equal to zero. Graph defined by the matrix $A_{kp\times ks}^*$, fully displays in time the execution of $n=kp$, $k>1$ processes and thus account for all possible connections between the blocks, the production processes and processors, given by the conditions of their interaction [7, 8]. If the number of production processes is not a multiple of the number of processors, i. e. $n=kp+r$, $1\le r\le(p-1)$, then the matrix of production execution times of the resource block $A_{(k+1)p\times(k+1)s}$ is constructed similarly as in the case $n=kp$. The block matrix (submatrix) $A_{(k+1)p\times(k+1)s}$ will contain $(p-r)$ zero rows. The minimum total runtime $T(p, kp+r, s)$ of a variety of heterogeneous competing processes in this case is also determined by the length of the critical path in the corresponding network graph.

CONCLUSIONS

1. The results can be used to study reliability systems, which is very important at the stage of designing transport vehicles and for solving problems of optimal use of a multi-structure production resource. The results can also be used to determine the profitability of machine production.

2. To analyze the production process synchronization used statistical approach to the correctness of the selection parameters. Simulation approach allows you to test how much exactly parameters are selected and whether they ensure the smooth operation of production. Statistical analysis of the parameters gives information about failures and about violations of the synchronization of production processes. Statistical estimates are obtained, allowing to evaluate production as a whole [9–11].

3. The cluster analysis was carried out using method of the K-average production processes with minimization of the total error probability. The influence of the principal factors on the work of production is determined. A statistical analysis of the complex data of the production process was carried out in the search for optimal solutions in case of uncertainty and under risk conditions using the following methods: Bayes, Laplace, Germeyer using the simplex-method [12–14].

4. Used network decision-making methods. Statistical methods used with the method of mathematical modeling for further substantiating the optimal size of both individual parts and the volume of local warehouses so that there are no delays in the work of transferring production processes, disruptions in work, and downtime of working equipment. Production processes are considered at different intervals: continuous and discrete. When modeling the production process in this case, impulse functions were used.

5. Concentrated and distributed competing processes, a method of structuring the resources of engineering production are considered. Probabilistic characteristics of the output process are got by solving the auxiliary deterministic task with standard actions on random functions.

REFERENCES

1. Dashchinskaya N. P. (2008) *Statistics of the Enterprise.* Minsk, Belarusian State University. 301 (in Russian).
2. Rao S. R. (1968) *Linear Statistical Methods and their Application.* Moscow, Nauka Publ., 548 (in Russian).
3. Borovikov V. P. (2003) *Statistica. The art of Analyzing Data on a Computer.* St. Petersburg, Piter Publ., 688 (in Russian).
4. Miller B. N. (2002) *Theory of Random Processes.* Moscow, Fizmatlit Publ., 320 (in Russian).
5. Chepeleva T. I. (2011) Stability of the Production Process of Transport Machines. *Vestnik BNTU* [Bulletin of the Belarusian National Technical University], (1), 64–67 (in Russian).
6. Dombrovsky V. V., Bunny V. M. (1998) *Asynchronous Machines: Theory, Calculation, Design Elements.* Leningrad, Energoatomizdat Publ., Leningrad Department. 368 (in Russian).
7. Diestel R. (2017) *Graph Theory.* Springer. 410. https://doi.org/10.1007/978-3-662-53622-3.
8. Swamy M. N. S., Thulasiraman K. (1992*) Graphs: Theory and Algorithms.* John Wiley & Sons, Inc. https://doi.org/10.1002/9781118033104.
9. Day M. V. (2006) Boundary-Influenced Robust Controls: Two Network Examples. *ESAIM: Control, Optimization and Calculus of Variations,* 12 (3), 662–698. https://doi.org/10.1051/cocv:2006016.
10. Chen S. P., Tsai M. J. (2011) Time-Cost Trade-Off Analysis of Project Networks in Fuzzy Environments. *European Journal Operational Research*, 212, 386–397. https://doi.org/10.1016/j.ejor.2011.02.002.
11. Khadzaragova Ye. A. (2015) On Methods of Scalarization of Vector Estimates in Multicriteria Optimization Problems. *Fundamental'nyye i Prikladnyye Issledovaniya = Fundamental and Applied Research*, (18), 123–127 (in Russian).
12. Himmelblau D. (1972) *Applied Nonlinear Programming*. New York, McGraw-Hill. 498.
13. Khoroshev A. N. (1999) *Introduction to the Design Management of Mechanical Systems. Study Guide.* Belgorod. 372 (in Russian).
14. Andrews G. R. (2000) *Foundations of Multithreaded, Parallel, and Distributed Programming.* Reading, Mass: Addison-Wesley. 664.

Received: 17.06.2021
Accepted: 07.09.2021
Published online: 31.03.2022