POSIBILITIES OF STATISTIC IN ANTHROPOGENIC RISKS RESEARCH IN INDUSTRIAL AREA

Ladislav Novák[1], Mária Lusková[2]

Abstract : *There are used several qualitative and quantitative methods for risk assessment of industrial processes. In case we have sufficient historical data from previous industrial accidents we can use quantitative methods based on statistical methods. Within results interpretation we can use especially classification of statistical data, frequency and probability of risk sources parameters occurrence, interdependency of risk sources and their consequences. Statistical methods in combination with other mathematic methods allow us to quantify risk also based on risk sources estimation and their consequences. This article presents methods and examples for their possible using.*
Keywords: *risk, industrial accident, risk assessment, statistical methods.*

INTRODUCTION

Within the risk assessment of industrial processes we can identify risks resulting from the human factor activities, risks connected with failures of machines, tools, production lines, or defects of material. External environment risks that can have direct or after effect on industrial processes, e.g. natural disasters, terrorist attacks or unexpected drop out of energy medium are not also negligible. Industrial processes risks do not act separately. We can expect that risks activation can be followed by accidents chaining in the form of "domino effect" or concurrent occurrence and effect of more industrial accidents in one industrial process, enterprise or branch with fatal impacts on workers, industrial production, enterprise economy and inhabitants in the neighbourhood.

The above mentioned risks can be reduced by preventive measures. The industrial accidents consequences can be reduced by measures applicable for effective response to accidents. Determination of range and flexibility of mentioned measures is questionable and directly dependent upon the enterprise financial resources. In this situation determining the level (value, degree…) of risk of industrial accidents can be significantly helpful for management decision making concerning the risks reduction preventive measures range and preparation of forces, resources and response to anticipated industrial accidents.

At present a lot of methods for industrial processes risk assessment are available. They are based on mathematical qualitative and quantitative methods. Decision making concerning their use is directly dependent upon range of available historical data about relevant accidents and knowledge of

[1] Ladislav Novák, doc,. Ing., PhD., Katedra krízového manažmentu, Fakulta špeciálneho inžinierstva ŽU v Žiline, Ulica 1. mája 32, 010 26 Žilina, Slovakia, e-mail: Ladislav.Novak@fsi.uniza.sk.
[2] Mária Lusková, Ing., PhD., Pracovisko výskumu krízového riadenia, Fakulta špeciálneho inžinierstva ŽU v Žiline, Ulica 1. mája 32, 010 26 Žilina, Slovakia., e-mail: Maria.Luskova@fsi.uniza.sk.

mathematical methods by people carrying out the risk assessment. In this situation knowledge of statistics, relevant statistical methods and ability to interpret the results can play an important role. Exactly formulated statistical project for risk assessment has to be the base for statistical methods application [3, 4].

1. RISK ASSESSMENT OF INDUSTRIAL PROCESSES FROM THE VIEW OF STATISTICS

The use of statistical methods for industrial processes risk assessment is dependent on several facts and limiting conditions:

- risk of which process, failure or accident should be expressed,
- which risk factors of respective industrial process or external environment will be investigated,
- towards which subject or in which territory should be the risk expressed,
- to which date should be the risk expressed,
- how many statistical data we are able to acquire,
- what way or form should be the risk expressed,
- which statistical methods or SW environment are available ?

Within the risk investigation we usually ask a question: "What is the risk….?" that should expressly consider previous facts and conditions. But if the question of industrial accident is, e.g. "What is the risk of industrial accident?" this is so called ambiguous question from the view of statistics that does not include type of accident, time, subject or territory to which the risk should be related.

From this follow these conclusions:

- The question of time and type of accident can be solved by considering all known historical accidents that have existed up till now.
- The question of subject or territory to which the accidents are related cannot be omitted. Probability of accident occurrence would approximate to 1 in too large territory and for too long a period of time.
- Form of risk expression is dependent upon the requirement of order or our decision but always in the form understandable for the risk assessment order.

The use of statistical methods depends on previous mentioned facts. It starts especially from data range we are able to acquire and combination of risk factors and consequences of respective accidents mutual relations of which we want to explore. The single accidents are considered to be statistical units with identified statistical symbols. Statistical symbols can include risk factors that caused the accident and consequences induced by respective accident.

From this view the basic statistical methods appropriate for assessment of industrial processes risks are as follows:

- basic statistical analysis,
- single classification of statistical data of one statistical symbol (in case the statistical data have verbal modifications or a small number of numerical modifications) that enables to express frequencies, shares or probabilities of occurrence of the single symbol modifications,
- group classification of statistical data of one statistical symbol (in case the numerical statistical data have a large number of modifications) that enables to express the same parameters but expressed in so called continuous environment by classes, intervals or groups,
- classification according to statistical data of more statistical symbols in two-dimensional (x, y) or three-dimensional (x, y, z) environment that enables to express frequencies, shares or probabilities of combinations of statistical symbols modifications,

- predicting values of one statistical symbol with use of relative frequency, theory of probability and distribution of random variable,
- estimations of values of one statistical symbol on the basis of expert's estimations of its values (minimum, median/arithmetic average, maximum) and with use of modified theory 6 σ,
- correlation task through which we explore dependence of two numerical statistical symbols,
- regression analysis through which we seek the course of dependence of two numerical statistical symbols and predict the values of dependent symbol,
- contingent task through which we explore dependence of two verbal statistical symbols,
- time series where we explore frequencies, shares or probabilities of modifications of one statistical symbol (on the axis „y“) in dependence on time (on the axis „x“),
- simulation (modelling) of occurrence and course of industrial accidents based on the theory of probability and probability distribution.

2. EXAMPLES OF BASIC STATISTICAL METHODS APPLICATION

2.1. Basic statistical analysis

Basic statistical analysis enables to investigate characteristics of level and variability of risk factors. The respective characteristics are usable for all quantitative risk factors. Concerning the qualitative risk factors the characteristics are usable only with limitation.

The level characteristics are especially:
- the most often repeated value of risk factor - modus,
- the values that divide data file in parts with the same number of values - quintiles (the most widely known is median - 50% quintile),
- various types of averages, especially arithmetic average.

The variability characteristics of risk factors are especially:
- the smallest value occurring within the data file – minimum,
- the greatest value occurring within the data file– maximum,
- difference between maximum and minimum values – variation range,
- dispersion of data around arithmetic average,
- data deviations around important values – standard deviation, deviation around arithmetic average, deviation around median, etc.

Separate assessment of risk factors according to their characteristics of level or variability is not respectable since it could lead to the essential information misrepresentation about respective risk factor.

2.2. Classification of statistical data according to one risk factor

Statistical classification is basic method for determination of frequencies, share or probabilities of values occurrence of industrial processes risk factor. Frequency is absolute count; share or probability is relative count of this value occurrence. We use single or interval (group) classification. The way of classification is dependent upon the type of risk factor. We specify if it is quantitative (numerical) or qualitative (verbal or numerical indicating quality) symbol. In case of quantitative symbol, we assess if it is continuous or non-continuous (discrete) and the number of alternatives the values of statistical symbol achieve. In some specific cases we can work also with alternative and plural symbols.

Single classification is used in case the qualitative (verbal) or quantitative symbols achieve a small number of their values alternatives. In case of quantitative symbol values of which achieve more number of alternatives than 15 we use interval (group) classification. Results interpretation achieved in

respective classification is given by distribution of absolute and relative frequencies in relation to their graphical illustration by specific graph (mostly column graph).

2.3. Classification of statistical data according to two risk factors

Investigation of relation of two risk factors is based on combination frequency of all values of these factors and calculation of probability of these combinations occurrence. We distinguish these types of classification of two statistical symbols:

– Hierarchical classification — consists in classification of mutually subordinated statistical symbols when within intervals of one symbol are formed intervals of other (subordinated) symbol. The typical result of classification is hierarchical tree – dendogram (evolution tree). For example, employees stricken by industrial accident are classified first with regard to their age, in age intervals according to their working position and within their working position according to their gender.

– Combination classification — classification according to two symbols mutually unsubordinated. Combination tables are typical result of this type of classification. According to the type of classified statistical symbols we distinguish these types of combination tables:

- correlation table - classification according to two numerical symbols,
- contingency table - classification according to two verbal symbols (one symbol can be also numerical),
- association table - classification according to two alternative verbal symbols.

The results achieved within data classification enable widespread interpretation from the view of the statistical symbols and serve as the basis for investigation of their dependences. Absolute and relative frequencies are starting point for graphical presentation of mentioned frequencies.

2.4. Predicting values of one risk factor

Estimation (calculation) of risk factor values is one of the statistical tasks in industrial processes risk assessment. These values are situated:

– among values obtained by statistical survey or,
– outside variation range of statistical file.

This task is typical for cases when the investigating all industrial accidents would be time-consuming, non-economical and physically impossible.

The simplest way to solve the above mentioned case is to carry out an approximation for values being closest from the left and right. The better solution of this task provides statistical definition of probability that takes relative frequency as probability of occurrence of statistical symbol respective value. Solution of this task consists in finding such functions that are able to describe the course of relative frequency of statistical symbol values with sufficient accuracy. The most usual ways are:

– description by frequency function or so called probability density function. Shape of this function gives a picture about important properties of distribution,
– description by distribution function.

Binomial, Poisson and hyper geometric distributions are the typical probability distributions usable in statistical research of discrete (non-continuous) risk factors.
Normal (Gaus), exponential and Erlang distributions are the typical probability distributions usable in statistical research of continuous risk factors.

2.5. Estimation of values of one risk factor

Estimation of values of risk factor can be carried out in case of quantitative (numerical) risk factors. The typical data are e.g. time data, data presenting intensity of accident and consequences of crisis event, e.g. material losses or casualties. In these cases we can estimate:

– expected risk factor minimum value *xmin ,*

- expected risk factor maximum value *xmax* ,
- expected average value (arithmetic average) *xp*, modus or median of risk factor.

From these input data we can calculate parameters of random variable distributions whereby we proceed from the 6σ theory indicating that certain number of all values of random variable with normal distribution is situated between minimum and maximum value. The theory guarantees that 99, 73% of all values are situated just in the range $\pm 3\sigma$ (σ - standard deviation) from the mean value.

Gauss-Laplace or Erlang distributions are usually used for presentation. Gauss-Laplace distribution is used in case the mean value *xp* is situated exactly in the middle (in median position) between minimum and maximum value.

2.6. Investigation of dependence of two risk factors

Combination classification is the base for investigation of dependence of two risk factors or impacts of industrial accident. Regarding the risk factors types (statistical symbols), the dependences can be divided as follows:

- correlation dependence – dependence among quantitative risk factors,
- association dependence – dependence among qualitative alternative risk factors,
- contingency dependence – dependence among qualitative plural risk factors.

2.7. Time series in investigation of industrial processes

The fact that time is always one of the statistical symbols is the particularity of the time series investigation. In case of the industrial accidents risk assessment it is e.g. time of accident occurrence, year, month or season of accident, duration of accident, etc. Investigation of respective risk factor includes especially these tasks:

- monitoring and evaluation of changes in investigated risk factor development in dependence on time,
- analysis of components affected by investigated risk factor (trend, seasonal component, cyclic component. random component),
- predictions of risk factor expected values.

Solving the tasks is based on the same methods mentioned in previous chapters.

2.8. Simulation of occurrence and course of the industrial processes accidents

Simulation of occurrence and course of the industrial processes accidents is based on computer generation of the risk factors changes and accidents consequences according to random variable distribution that can be used for mathematical description of the risk factors and consequences. Obtaining the large number of representative statistical data, that enable with use of relative frequency of the statistical symbol changes to look for the most reliable random variable distribution of which we use in simulation, is the base. The topic of simulation and modelling of similar processes is individual mathematical discipline and is not subject of this article.

CONCLUSION

Statistical methods and their use in industrial processes risk assessment open broad possibility range of their investigation. They provide diverse view on risks, risk factors and accidents impacts as the qualitative methods. Conceptual apparatus that works with frequencies and probabilities of risk factors values occurrence and impacts of activated emergency events or crisis events in industrial processes is also different.

In case of statistical data lack, the statistical methods in combination with other mathematical theories and methods (e.g. THO or $\boldsymbol{6\sigma}$) enable to use also estimated parameters of risk factors and impacts carried out by representative expert group.

The above mentioned statistical methods and examples only indicate possibilities and directions of their use in industrial processes risk assessment. Especially investigation of risk factors mutual relations and impacts in mutual combinations are open. The important information and conclusions can be obtained in looking for risk factors dependences, e.g. in three-dimensional environment and simulation of accidents and their impacts.

LITERATURE

[1] NOVÁK, L.: *Východiská na kvantifikáciu rizík priemyselných procesov*. In: Zborník (CD) z medzinárodnej vedecko-odbornej konferencie „Manažment, teória, výučba a prax 2011. AOS MRŠ Liptovský Mikuláš 2011. str. 265 – 270. ISBN 978-80-8040-404-8.

[2] NOVÁK, L.: *Štatistický projekt v analýze rizík*. In: Zborník z medzinárodnej vedeckej konferencie "International Conference of Crisis management in Public and Private Sector" (CD). ČR, Uherské Hradiště, FLKR UTB Zlín 2011. str. 43 + text na CD. ISBN 978-80-7454-027-1.

[3] NOVÁK, L., MILATA, I.: *Application of „3 sigma theory“ to Erlangs distribution of random variable*. In: Zborník z X. International Scientific Conference TEMPT´97. Bulharsko, Sofia, Higher Military School of Transport 1996, s. 137 – 142. (90 %). ISBN 954-12-0049-4.

[4] NOVÁK, L.: *Prednášky z predmetu manažérska štatistika*. http://fsi.uniza.sk/kkm/stranka/zamestnanci/ladislav-novak.

[5] SEIDL, M., ŠIMÁK, L., ZAMIAR, Z.: *Aktualne zagadnienia zarządzania kryzysowego (monografia)*. WROCLAW, MWSLiT, Consulting i logistika Spółka z o.o., 2009. 197 s., ISBN 978-83-89908-95-7

[6] VORLÍČEK, M.: Vybrané kapitoly matematické štatistiky. FMNO, Praha 1968.

[7] ZÁNICKÁ HOLLÁ, K., RISTVEJ, J., ŠIMÁK, L.:. *Posudzovanie rizík priemyselných procesov*. Bratislava: Iura Edition, spol. s. r. o., 2010. 155 s. ISBN 978-80-8078-344-0.

This paper was supported by the Slovak Research and Development Agency on the basis of agreement No. APVV-0043-10 for project solution

Complex model of industrial processes risk assessment