Statistical metrics for providing the population with drinking water of standard quality

D V Parshukov, Z E Shaporova and I Yu Koloskova

Krasnoyarsk state agrarian university, 90 Mira Avenue, Krasnoyarsk, Russia

E-mail: fub@kgau.ru

Abstract. The problem of providing the population with drinking water is global in nature and is one of the main tasks for civilization in the long term. It is necessary not only to provide the population with water, but also to meet the regulatory requirements for hygienic, microbiological and sanitary-chemical standards. This article presents the results in statistical studies of drinking water quality for rural population. Object of research is water supply systems in rural areas of the Krasnoyarsk territory. Main research methods are descriptive statistics, Bootstrap method, and correlation and regression analysis. Based on the application of these methods statistical metrics of drinking water quality from various sources of water supply are identified, and a statistical relationship between the drinking water quality and the health of the rural population for the first time registered diseases of the digestive and endocrine systems is described. The most acute problem is the unsatisfactory condition of water supply facilities, the quality of water from underground and surface sources of centralized water supply, as well as from distribution networks.

1. Introduction

The provision the population with drinking water is at the heart in the United Nations concept of sustainable development for the twenty-first century. Water supply of territories with water of standard quality is one of the key tasks in the implementation of the state strategy for sustainable development of territories in the Russian Federation.

The development of water supply systems is closely related to the tasks of improving the life quality of the population, improving living conditions, and above all, preserving the health of the nation. Providing the population with clean water has a preventive effect, reducing the risks of various epidemiological diseases. Therefore, an important task is to establish the relationship between the observed parameters of drinking water quality and the health status of the population.

In rural areas there are either no or insufficient centralized water supply systems. It is also worth noting the high wear and tear of water supply systems and the increase in the pollution level of local natural sources. In this regard, there is need to find comprehensive solutions to problems of drinking water supply in compliance with regulatory requirements for the quality of drinking water.

2. Problem statement

The results study of water quality testing and their statistical characteristics determines the need to clarify the following issues that are of an applied and practical nature:
2.1. The problem clarification of providing the population in rural areas with drinking water
Providing the rural population with drinking water of standard quality is a difficult task due to the unsatisfactory state of communications and high wear and tear of water supply systems (Dzyubo, 2007; Poluhkina, 2017). Also, most of the rural population receives water from non-centralized sources; and water quality control is carried out selectively and does not always meet the requirements of reliability and accuracy. At the same time, part of the private sector population uses natural water without proper pre-treatment (Yasankin, 2015).

2.2. Definition of quality standards for drinking water in the Russian Federation
The study about the state of drinking water and factors affecting its compliance with regulatory quality is a control task assigned to the office of the Federal service for supervision in consumer protection. The results are published in the annual state report "About the state of sanitary and epidemiological well-being of the population in the Krasnoyarsk territory" (Report, 2018). Comparison of domestic and international drinking water quality standards was carried out in the work (Bashketova et al., 2018).

2.3. Study of the main research results to identify the statistical relationship between the drinking water quality and the health of the population
Testing the hypothesis that there is a statistical relationship between the drinking water quality and the health of the population was confirmed in studies conducted both in domestic works (Bakumenko, 2011; Borzunova at al, 2018; Konshina and Lezhnin, 2019) and in the works of foreign researchers (Hellard, 2001; Strosnider, 2017; LiandWu, 2019). In this study it is necessary to check the existence of a similar relationship for rural areas of the Krasnoyarsk territory and to determine its quantitative measurement.

3. Research Questions
Highlight the main statistical characteristics in the study results of drinking water quality and the state of water supply facilities for rural areas in the Krasnoyarsk territory.
Determine whether there is a link between the drinking water quality in rural areas and diseases in the digestive and endocrine systems of the rural population.

4. Purpose of the Study
The purpose of the research is to establish a statistical metric of the rural population in the Krasnoyarsk territory's drinking water supply and to assess the impact of water quality on individual health indicators in rural areas.

5. Research Methods
Initial empirical data were collected for the following indicators, initially divided into criterial and independent variables:
   a) Criterial variables Y:
      Y\(_1\) – first identified diseases of the digestive system in all population groups, cases per 1000 people;
      Y\(_2\) – for the first time identified diseases of the endocrine system, eating disorders and metabolic disorders, cases per 1000 people;
   b) Independent variable X:
      X\(_1\) – Specific weight of water samples from sources (underground, surface) of the centralized water supply in the Krasnoyarsk territory that do not meet hygienic standards, %
      X\(_2\) – Percentage of drinking water sources that do not meet sanitary requirements, %
      X\(_3\) – Percentage of water supply facilities that do not meet sanitary requirements, %
Specific weight of water samples from the distribution network of centralized water supply in the Krasnoyarsk territory that do not meet hygienic standards, %;

weight of water samples from sources of non-centralized water supply in rural settlements that do not meet hygienic standards, %

For their systematization, visual representation, and determination of primary statistical metrics descriptive statistics were used: the average value, sample variance, standard deviation, and distribution characteristics. Descriptive statistics were supplemented with the Bootstrap method for investigating the distribution of statistical data based on the Monte - Carlos method with a given 95% reliability level for 1000 samples. The initial data for the research are shown in figure 1 as distribution graphs with a trend line that provides the highest value of the approximation coefficient.

Figure 1. The original data with the trend line based on criterion variable.

Figure 2. The original data with the trend line based on the independent variable.
To assess the tightness of the relationship the correlation analysis was performed and pairwise correlation coefficients between the criterion and independent variables were calculated. Statistical dependence between variables was determined using one-factor linear regression analysis.

The information base of the research were the proceedings of the annual state report "About the state of sanitary and epidemiological well-being of the population in the Krasnoyarsk territory in 2018," the data of Federal state statistics service in the Krasnoyarsk territory.

6. Findings

The results of calculating statistical indicators based on the initial data are shown in table 1.

Table 1. Descriptive metrics of drinking water status in water supply systems in rural areas of the Krasnoyarsk territory.

| Variables | Statistical indicators | Value | Offset | The Results Of Bootstrap | 95% confidence interval |
|-----------|------------------------|-------|--------|--------------------------|------------------------|
|           |                        |       |        | Average                  | Standard error         | Lower     | Top       |
| X_1       | Average                | 19,571| 0,039  | 0,579                    | 18,400                 | 20,571    |
|           | Standard deviation     | 1,710 | -0,181 | 0,387                    | 0,763                  | 2,082     |
|           | Dispersion             | 2,926 | -0,435 | 1,036                    | 0,583                  | 4,334     |
|           | Asymmetry              | -0,882| 0,052  | 0,918                    | -2,641                 | 0,821     |
|           | Excess                 | -0,831| 0,712  | 2,838                    | -2,799                 | 6,980     |
|           | Average                | 48,200| 0,000  | 0,264                    | 47,644                 | 48,700    |
| X_2       | Standard deviation     | 0,755 | -0,074 | 0,156                    | 0,350                  | 0,971     |
|           | Dispersion             | 0,570 | -0,082 | 0,210                    | 0,122                  | 0,943     |
|           | Asymmetry              | -0,303| 0,132  | 0,707                    | -1,661                 | 1,209     |
|           | Excess                 | -0,509| 0,052  | 1,551                    | -2,612                 | 3,363     |
|           | Average                | 30,457| -0,013 | 0,644                    | 29,130                 | 31,629    |
| X_3       | Standard deviation     | 1,867 | -0,154 | 0,295                    | 1,096                  | 2,069     |
|           | Dispersion             | 3,486 | -0,463 | 0,823                    | 1,202                  | 4,280     |
|           | Asymmetry              | -0,312| 0,005  | 1,022                    | -2,341                 | 1,326     |
|           | Excess                 | -2,515| 1,521  | 2,584                    | -2,797                 | 6,093     |
|           | Average                | 17,043| 0,048  | 1,595                    | 14,400                 | 20,670    |
| X_4       | Standard deviation     | 4,644 | -0,390 | 0,738                    | 3,139                  | 4,792     |
|           | Dispersion             | 21,566| -2,929 | 4,453                    | 9,852                  | 22,967    |
|           | Asymmetry              | 0,366 | -0,004 | 1,116                    | -1,240                 | 2,618     |
|           | Excess                 | -2,779| 1,791  | 2,996                    | -2,798                 | 6,952     |
|           | Average                | 29,471| -0,049 | 3,569                    | 22,149                 | 36,027    |
| X_5       | Standard deviation     | 10,365| -0,891 | 1,703                    | 5,476                  | 12,291    |
|           | Dispersion             | 107,429| -14,785| 30,869                   | 29,986                 | 151,062   |
|           | Asymmetry              | -0,114| 0,001  | 0,777                    | -1,702                 | 1,302     |
|           | Excess                 | -1,811| 0,951  | 1,612                    | -2,696                 | 4,039     |

According to the table, the asymmetry for all variables except X_1 can be considered small. This indicates slight shifts in the variation series to the left for the variables X_2, X_3 and X_5, and to the right for the variable X_4. The excess for all variables is less than 0, which means that the curves of the variables are flat-topped. The largest variance is observed for variables X_4 and X_5, which indicates an increased spread of values for individual shares of non-compliant drinking water samples from centralized and non-centralized sources over the measurement periods. The values of the limits in confidence intervals of the average values in variables obtained by the Bootstrap method indicate that the most reliable measurements are as follows: percentage of drinking water sources that do not meet sanitary requirements; specific weight of water samples from sources (underground, surface) of the centralized water supply in the Krasnoyarsk territory that do not meet hygienic standards.
The results of correlation analysis for criteria and independent variables are presented in the form of a correlation matrix (table 2).

Table 2. Correlation matrix.

|       | Y₁    | Y₂    | X₁    | X₂    | X₃    | X₄    | X₅    |
|-------|-------|-------|-------|-------|-------|-------|-------|
| Y₁    | 1     |       |       |       |       |       |       |
| Y₂    | 0.7   | 1     |       |       |       |       |       |
| X₁    | -0.57 | -0.71 | 1     |       |       |       |       |
| X₂    | -0.27 | -0.48 | 0.123 | 1     |       |       |       |
| X₃    | 0.83  | 0.54  | -0.62 | 0.005 | 1     |       |       |
| X₄    | -0.9  | -0.62 | 0.732 | 0.014 | -0.98 | 1     |       |
| X₅    | -0.63 | -0.1  | 0.228 | -0.04 | -0.42 | 0.503 | 1     |

The variable Y₁ has the closest relationship with the variables X₃ and X₄, while the correlation coefficient between these independent variables also indicates a very close relationship. Only the relationship between Y₁ and X₂ should be considered weak. The variable Y₂ is most closely related to the variable X₁. The relationship between Y₂ and X₄ is less pronounced. To describe the statistical dependence it is advisable to use the multiple regression model. To avoid the influence of multicollinearity the model should include variables X₂ and X₃, or a combination of two variables, one of which is X₅.

The results of one-factor linear regression analysis are shown in table 3.

Table 3. Results of one-factor regression analysis.

| Kind of model | The value of the correlation coefficient | The value of determination coefficient | Value of the Fisher criterion | Student's t-value for the regression coefficient |
|---------------|----------------------------------------|--------------------------------------|-------------------------------|-----------------------------------------------|
| Y₁ = 14.019 + 0.765 × X₁ | 0.83 | 0.697 | 11.51 | 3.39 |
| Y₁ = 42.67 − 0.31 × X₄ | 0.85 | 0.725 | 13.23 | -3.64 |
| Y₂ = 25.08 − 0.6256 × X₃ | 0.71 | 0.51 | 5.2 | -2.28 |

Thus, there is a close relationship between digestive diseases diagnosed for the first time and the state of water supply facilities in rural areas (the percentage of the described variance of the variable Y₁, the variance of the variable X₃ is 69.7%). The occurrence of the endocrine system diseases, eating disorders and metabolic disorders is influenced by the water quality from underground and surface sources of centralized water supply. Both criteria variables are indirectly affected by the quality of drinking water from centralized distribution systems.

7. Conclusion

The noted indicators for the study period tend to improve the drinking water quality according to hygienic standards, sanitary-chemical and microbiological indicators. The obtained statistical metrics show that estimates of drinking water quality and characteristics of rural water supply systems are generally reliable. The problem of unsatisfactory state in water supply systems is relevant. Natural pollution of drinking water from surface and underground sources creates serious risks, which is aggravated by the lack of a system for decontamination and disinfection of waste water in rural areas.

References

[1] Bakumenko L P and Korotkov P A 2011 Statistical analysis of the impact of drinking water quality on the health of the region's population Applied econometrics 2(22)

[2] Bashketova N S, Vyucheyskaya D S, Sladkova Yu N, Eremin G B and Fridman K B 2018 Regulation of drinking water Quality Comparison of national and international standards
Health - the basis of human potential: problems and ways of their solution 13(3) 1136-48

[3] Borzunova E A, Kuzmin S V et al 2007 Impact assessment of drinking water quality on public health Hygiene and sanitation 3 32-4

[4] State report (2018) About the state of sanitary and epidemiological well-being of the population in the Krasnoyarsk territory in 2018 Department of the Federal service for supervision of consumer protection and human welfare in the Krasnoyarsk territory p 325

[5] Dzyubo V V 2007 Preparation of underground water for drinking water supply in small settlements in the West Siberian region (St. Petersburg)

[6] Konshina L G and Lezhnin V L 2014 Assessment of drinking water quality and public health risks Hygiene and sanitation 93 3

[7] Polianskikh A A and Borisenko M S 2019 Problems of rural water supply and ways to solve them Patterns and trends of innovative development in society pp 99-101

[8] Li P and Wu J 2019 Drinking water quality and public health Exposure and Health 11(2) 73-9

[9] Polukhina M G 2017 Economic analysis of housing and engineering infrastructure development in rural areas Russian Journal of Agricultural and Socio-Economic Sciences 63(3)

[10] Stroshnider H, Kennedy C, Monti M and Yip F 2017 Rural and urban differences in air quality MMWR Surveillance Summaries 66(13) 1

[11] Hellard M E, Sinclair M I, Forbes A B and Fairley C K 2001 A randomized, blinded, controlled trial investigating the gastrointestinal health effects of drinking water quality Environmental Health Perspectives 109(8) 773-8

[12] Yasakin B B 2015 High-quality drinking water in rural settlements of the Volgograd region-the basis for the development of agro-tourism Agroecological tourism as a tool for sustainable development of rural areas in the regions of Russia and abroad pp 250-4