Reliability of the Estimate of Interrelation of Parameters Efficienty Their Work of Objects of Ees

Farhadzadeh E.M., Muradaliyev A.Z., Farzaliyev Y.Z.,
Rafiyeva T.K. Abdullayeva S.A.

Azerbaijan Scientific-Research and Design-Prospecting Institute of Energetic
AZ1012, Ave. H.Zardabi-94, e-mail: elmeht@rambler.ru

Abstract

Given is computational method of statistical distribution function of possible realizations of coefficients of linear correlation of Pearson and grade correlation of Spirmen for dependent random quantities. Dependence provided with ranging of random quantities of selections. Random quantities are modeled and have uniform distribution in the range of [0,1]. These distributions allow to evaluate the error of II type and to compare it with the error of I type. The error of II type established on statistical distribution function of coefficient correlation of independent selections. These distributions are in a basis of criteria for estimate of the importance of coefficients of correlation of technical and economic parameters of objects of electrical power systems.

Keywords: Correlation coefficient, selection, ranging, distribution, technical and economic indicator

I. Introduction

The indispensable condition of objectivity of ranging of objects of electric energy systems (EES) by comparison of integral parameters (IP) of efficiency (reliability, profitability and non-failure operation) their works is reliability of IP. In turn, reliability of the estimate of IP depends on the level of statistical communications of the consisting IP of technical and economic indicators (TEI). Reliability of IP is possible at independence TEI the objects of the same name. If the number of dependent TEI more, then the reliability of IP will be is lower and the risk of the wrong decision will is higher. Need of ranging of objects arises systematically at the organization of their maintenance and repair, load distribution, carrying out tests. And, first of all, - for objects, perioditon recovery of wear is not reglamented and is defined by their technical condition. Main characteristic of technical condition of objects of EES serves the set of formalized TEI [1].

At classification of statistical data passport data and data on service conditions are not less important. A set of indicators characterizing of objects selected expressly consciously, since the personnel of the power supply system should solve operational problems, being guided by technical condition of specific objects. If besides to add that many of these indicators differ with units, the scale and the scale of measurement, then difficulties of ensuring reliability of ranging of objects with calculation methods become obvious. For this reason in practice of the decision, following from ranging of objects, are accepted intuitively, in many respects heuristically. The risk of the wrong decision at decrease in qualification of personnel increases. The multidimensionality noted above and natural difficulty of accounting of all information is not less important reason of risk of the wrong decision.

In these conditions the possibility of the automated methodical and information support of personnel is very tempting. Thus:

- application of modern methods of the analysis and synthesis of statistical data, bulkiness of calculations, knowledge volumity, cause the high probability of "mechanical" mistakes by production of calculations of IP manually;
- methodical support of personnel consists in the recommendations based on objective comparison and ranging of objects;
- information support consists in justification of some decisions in the form of specialized calculation tables [2].

It should be noted that it is inexpedient and to absolutize the importance of methodical support of personnel as IP not always consider all factors on which objectivity of the decision depends.

To such factors:

- availability of the defective nodes of object demanding replacement;
- materials, relevant repair crews, possibility of shutdown of object, etc.

Estimate of the importance of statistical communication TEI in practice it is carried out by comparison of design values of the correlation factors (CF), with the CF critical value for the set error of I Type α. The CF critical values are tabulated for the number of values α and the sample size of n_v. They can be also determined by formulas of calculation of boundary values of the reliability interval provided that distribution function of CF corresponds to the normal law. Among the set of possible CF we will use the most often used by coefficient of linear correlation of Pearson (γ) and the index of rank of Spirmen (r). Let's remind that Pearson's CF (γ) characterizes only statistical interrelation of selections which implementations are set in the quantitative scale, and Spearmen (ρ) CF characterizes statistical interrelation of selections of ranks (ordinal values) of implementations. Critical values γ and ρ are considered equal, i.e. $\gamma_\alpha=\rho_\alpha$ [3].

Calculation formulas γ and ρ are received for the assumption of compliance of implementations of selections to the normal distribution law. The premises above noted when calculating γ_α and ρ_α for TEI objects of EES, as a rule, are not carried out. And if to be more exact, distribution of implementations TEI it is unknown, changes and depends on the large number of factors. To such, volumes of selection are small.

The preference of strategy of minimization of the error of I Type of α, when effects of these errors are accepted much more dangerously than effects of the error of II Type β is not justified. Explicit preference of one of two assumptions by comparison TEI is absent. However computational methods of risk of the wrong decision for the assumption "statistical communication between TEI it is significant" and the known CF are not developed.

The present article is devoted to development of the computational method and accounting of the error of II Type.

II. Results of the analysis of fiducially distributions of correlation factors of independent random quantities

Fiducial distribution is understood as distribution of possible implementations of complex indicators. As complex it is considered to be indicators which implementations can be received as a result of calculations. Complex indicators the arithmetic average (harmonious, geometrical), dispersion indicators (dispersion, the average quadratic deviation, scope), maintenance coefficient, the availability quotient are among random quantities. CF also belongs to their number. Method and algorithm of modeling of possible implementations of CF, creation of statistical function of fiducial distributions (s.f.f.d.) and the assessment of the CF critical values for the set α are brought in [4].

The carried-out calculations allowed establishing:

2.1. The requirement of compliance of selections of random quantities ξ to the normal law (for the reliable assessment of coefficient of linear correlation of Pearson (γ) at small selections ($n_v\leq 15$) and uniform distribution of random quantities ξ in the range of [0,1] was not confirmed. At number modeled γ_M equal N=10000 discrepancy of critical values $\gamma_{M,\alpha}$, an and tabular values given in reference books γ_α for α equal 0,1; 0,005; 0,02; 0,01; 0,005; and 0,001 did not exceed 1,5%. In conditions when for practical calculations value α is accepted, as a rule, equal 0,05 and occasionally 0,01, such discrepancy of the quantile of distribution can be considered admissible.

2.2. Irrespective of n_v and α equality of critical values γ_α and ρ_α is confirmed. At the same time, for the same modeled (м) couples of selections $\{\xi_1\}_{nv}$ and $\{\xi_2\}_{nv}$ CF $\gamma_м$ and $\rho_м$ can significantly differ. We met this phenomenon also by comparison of criteria of check of the hypothesis of nature of discrepancy of statistical distribution functions (s.f.d.) random quantities [5]. Such discrepancy explained by distinction of estimates of statistical parameters of selections there. By comparison of the CF experimental values γ_e and ρ_e for the same selections, by analogy, influences observed discrepancy distinction of physical essence γ and ρ. But, and this is important, the recommendation [5] remains invariable: the importance of statistical connection has to be established on the basis *of the principle of the addition* considering independence of properties γ and ρ [6]. In other words, the decision is made by results of comparison of experimental and critical value for both CF. Statistical communication is accepted significant if at least one of CF is significant;

2.3. If implementations TEI change in the quantitative scale, then it is recommended to calculate CF γ_{e1}, γ_{e2} both ρ_{e2} on one and oh to the formula for Pearson's CF:

- CF γ_{e1} – on selections $\{\tau_1\}_{nv}$ and $\{\tau_2\}_{nv}$;
- CF γ_{e2} – on selections of points $\{b_1\}_{nv}$ and $\{b_2\}_{nv}$;
- CF ρ_{e2} - on selections of ranks (ordinal values) $\{r_1\}_{nv}$ and $\{r_2\}_{nv}$

For γ_2 and ρ_2 as mean value is accepted the median. If implementations TEI are measured in the serial scale, then it is recommended to calculate CF γ_{e2} and ρ_{e2}

2.4. S.f.f.d. $F^*(\rho)$ have discrete character. Maximum step of discretization at n_v=5. In the absence of identical implementations of selections the step is equal 0,1 and not linearly decreases with growth of n_v. Discrete character of $F^*(\rho)$ excludes possibilities of the reliable assessment of critical ρ_α values. Use of the tabular ρ_α values recommended in reference books is connected with essential risk of the wrong decision which increases at reduction of n_v. The new criterion of identification of statistical communications offered. The difference with the existing criterion is that not compared ρ_e and ρ_α, but are compared $R^*(\rho_e)=1-F^*(\rho_e)$ and α.

2.5. The new way of decrease in influence of discrete nature of distribution $F^*(\rho_1)$ is offered.
This way is based on transformation of continuous random quantities to discrete (points) and taking note of identical implementations of points on their ranks. And the invariance of critical $\rho_\alpha=\gamma_\alpha$ values was the most surprising at these calculations.

2.6. Statistical analysis of CF $\gamma_м$ testifies to full symmetry of distribution of positive and negative values. This feature allows passing to s.f.f.d. absolute values of CF.

III. Method and algorithm of calculation of fiducial distributions of correlation factors of dependent random quantities

Transition from boundary values of the confidential interval to boundary values of the fiducial interval completely changes physical essence of the KK critical values, makes understanding of intervals more available and physically clear. The matter is that all implementations of the variational series of the range $\gamma_м$ and $\rho_м$ are real, and, in fact, as critical value we have to choose the greatest importance of implementations. Each value of this row both when modeling and according to operation perhaps, but differs in probability of emergence on the set of implementations. And only comparison of these probabilities of risk of the wrong decision allows preferring one or the other assumptions (H). Namely: compared TEI are independent (H_1) or statistical communication between TEI it is significant (H_2). And before passing to such type of criterion it is necessary to be able to build fiducial distributions of CF of statistically dependent selections of random quantities. Let's notice that the set of possible fiducial communications is infinite and significant statistical communication - only one. It or is (for example, $\gamma_e>\gamma_\alpha$), or it is absent ($\gamma_e<\gamma_\alpha$).

Modeling method essence s.f.f.d. CF of statistically dependent selections of random quantities is reduced to the way of forming of statistically dependent selections. All other calculations are similar to algorithm of modeling s.f.f.d. CF of independent random quantities [5].

To receive two statistically dependent selections of statistically independent selections of random quantities with positive CF it is offered to range these independent selections in ascending order of random quantities. At this CF will change in the range of [0,1]. We will receive two dependent selections with negative CF if we execute ranking of the first selection in ascending order, and the second - in decreasing order. Or on the contrary. The enlarged algorithm of calculation s.f.f.d. CF of dependent selections is given in fig. 1.

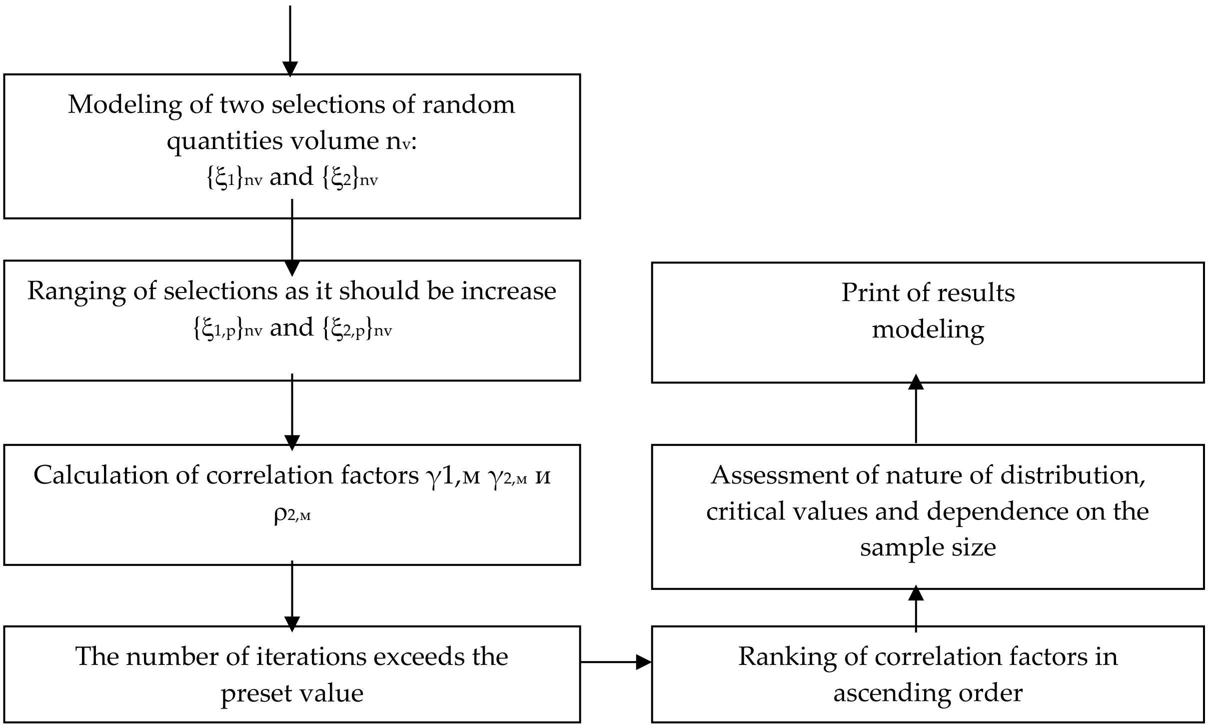


Fig.1. Enlarged flowchart of algorithm of modeling s.f.f.d. CF

Let's notice that CF $\rho_{1,м}$, to m for the ranged implementations of selections $\{\xi_1\}_{nv}$ and $\{\xi_2\}_{nv}$ will be equal to unit and therefore it is excluded from consideration. Results of modeling $\gamma_{1,м}$ $\gamma_{2,м}$ and $\rho_{2,м}$, for dependent selections $\{\xi_1\}_{nv}$ and $\{\xi_2\}_{nv}$ allowed to establish:

3.1. By analogy with independent selections for any couple of dependent selections of implementation $\gamma_{1,м}$ $\gamma_{2,м}$ and $\rho_{2,м}$ are also different. Some implementations of these CF for n_v=10 are given in the illustrative purposes in table 1.

3.2. Unlike the CF critical values of independent selections the CF critical values for dependent selections and small n_v significantly differ. With increase in n_v it is a divergence decreases. For confirmation the CF critical values, respectively, $\gamma_{1,м,\beta}$, $\gamma_{2,м,\beta}$ and $\rho_{2,м,\beta}$ are given in tables 2, 3 and 4

3.3. Comparison of critical values of coefficients of linear correlation of Pearson $\gamma_{1,м}$ for errors of I Type and II Type is given in table 5. This table confirms a possibility of consent with both assumptions (H_1 and H_2). In other words a number of realization $\gamma_{1,м}$, belongs to set of values $\{\gamma_{1,n,\alpha}\}_N$ and to a set of values $\{\gamma_{1,n,\beta}\}_N$

Table 1. Realization of coefficients of correlation of dependent selections

N	coefficients of correlation			Note
	$\gamma_{1,м}$	$\gamma_{2,м}$	$\rho_{2,м}$	
1	0,988	0,946	0,939	$n_v=10$
2	0,756	0,747	0,890	$\gamma_{1,м}=\frac{\sum_{j=1}^{n_v}[\Delta\xi_{1,j}\cdot\Delta\xi_{2,j}]}{\sqrt{\sum_{j=1}^{n_v}\Delta\xi_{1,j}^2}\cdot\sqrt{\sum_{j=1}^{n_v}\Delta\xi_{2,j}^2}}$
3	0,923	0,879	0,928	
4	0,956	0,970	0,990	$\gamma_{2,м}=\frac{\sum_{j=1}^{n_v}[\Delta b_{1,j}\cdot\Delta b_{2,j}]}{\sqrt{\sum_{j=1}^{n_v}\Delta b_{1,j}^2}\cdot\sqrt{\sum_{j=1}^{n_v}\Delta b_{2,j}^2}}$
5	0,931	0,937	0,980	
6	0,879	0,830	0,895	$\rho_{2,м}=\frac{\sum_{j=1}^{n_v}[\Delta r_{1,j}\cdot\Delta r_{2,j}]}{\sqrt{\sum_{j=1}^{n_v}\Delta r_{1,j}^2}\cdot\sqrt{\sum_{j=1}^{n_v}\Delta r_{2,j}^2}}$
7	0,354	0,894	0,916	
				$\Delta\xi_j=\xi_j-\bar{\xi}$; $\Delta b_j=b_j-\bar{b}$; $\Delta r_j=r_j-\bar{r}$ $\bar{r}$ - median

Table 2. Critical values of coefficients of linear correlation of Pearson $\gamma_{1,м,\beta}$ for selections of dependent random variables.

Selection volume	Error II Type β					
	0,5	0,1	0,05	0,01	0,005	0,001
5	0,919	0,787	0,743	0,646	0,602	0,538
8	0,935	0,848	0,815	0,738	0,694	0,625
10	0,945	0,874	0,845	0,776	0,746	0,679
15	0,967	0,913	0,892	0,851	0,832	0,787
30	0,979	0,954	0,943	0,919	0,908	0,883

Table 3. Critical values of coefficients of linear correlation of Pearson $\gamma_{2,м,\beta}$ for discrete selections of random variables.

Selection volume	Error II Type β					
	0,5	0,1	0,05	0,01	0,005	0,001
5	0,877	0,690	0,612	0,408	0,375	0
8	0,897	0,778	0,733	0,643	0,612	0,520
10	0,906	0,809	0,768	0,673	0,638	0,575
15	0,923	0,858	0,833	0,777	0,750	0,663
30	0,941	0,902	0,890	0,856	0,823	0,783

Table 4. Critical values of coefficients of rank correlation of Spirmen $\rho_{2,м,\beta}$ for selections of dependent random variables.

Selection volume	Error II Type β					
	0,5	0,1	0,05	0,01	0,005	0,001
5	0,894	0,740	0,646	0,408	0,335	0
8	0,918	0,833	0,803	0,713	0,679	0,589
10	0,926	0,864	0,835	0,759	0,729	0,648
15	0,938	0,899	0,883	0,849	0,817	0,764
30	0,950	0,929	0,921	0,898	0,895	0,868

Table 5. Comparison $\gamma_{1,м,\alpha}$ and $\gamma_{1,м,\beta}$ at $\alpha=\beta$ and $n_v=5$

Errors Types	Wrong decision risk					
	0,5	0,1	0,05	0,01	0,005	0,001
α	0,403	0,812	0,885	0,962	0,976	0,989
β	0,913	0,787	0,749	0,646	0,602	0,538

3.4. Some characteristic combinations of realization $\gamma_{1,м}$ $\gamma_{2,м}$ и $\rho_{2,м}$ and confirmation of statistical communication of selections for β=0,05 are given in table 6. As appears from these data:

- selections $\{\xi_1\}$ and $\{\xi_2\}$for which statistical communication is confirmed all three CF, two of three CF ($\gamma_{2,м}$ and $\rho_{2,м}$) and only one CF are observed ($\gamma_{1,м}$).
- *CF $\gamma_{1,м}$ $\gamma_{2,м}$ и $\rho_{2,м}$, supplement each other*

Table 6. Illustration of results of verification of the assumption of statistical communication of dependent selections

i	$\gamma_{1,м,i}$	H_2	$\gamma_{2,м,i}$	H_2	$\rho_{2,м,i}$	H_2	Note
1	0.7790	+	0.4082	-	0.4082	-	$\gamma_{1,м,0,05}$=0,6456
2	0.8068	+	0.5929	+	0.6250	+	$\gamma_{2,м,0,05}$0,4082
3	0.5799	-	0.4841	+	0.6250	+	$\rho_{2,м,0,05}$=0,4082
4	0.7676	+	0.4082	-	0.4082	-	
5	0.6065	-	0.3750	-	0.3953	-	
6	0.9618	+	0.5625	+	0.5441	+	
7	0.7427	+	0.5601	+	0.6455	+	
8	0.7439	+	0.6022	+	0.6250	+	
9	0.9437	+	0.6667	+	0.6455	+	

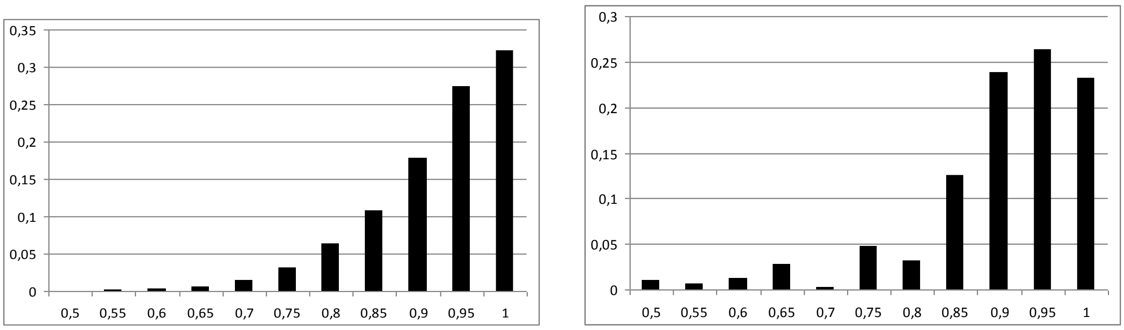
3.5 Criteria of decision-making for selections with a quantitative scale of measurement have an appearance:

$$\left.\begin{array}{l} \text{if } \gamma_{1,м} \le \gamma_{1,м,\beta}\text{, then } H \Rightarrow H_1 \to \text{exit} \\ \text{if } \gamma_{2,м} \le \gamma_{2,м,\beta}\text{, then } H \Rightarrow H_1 \to \text{exit} \\ \text{if } \rho_{2,м} \le \rho_{2,м,\beta}\text{, then } H \Rightarrow H_1 \to \text{exit} \\ \text{otherwise } \quad H \Rightarrow H_2 \end{array}\right\} \quad (1)$$

At a serial scale of measurement the criterion of decision-making (H) has an appearance:

$$\left.\begin{array}{l} \text{if } \gamma_{2,м} \le \gamma_{2,м,\beta}\text{, then } H \Rightarrow H_1 \to \text{exit} \\ \text{if } \rho_{2,м} \le \rho_{2,м,\beta}\text{, then } H \Rightarrow H_1 \to \text{exit} \\ \text{otherwise } \quad H \Rightarrow H_2 \end{array}\right\} \quad (2)$$

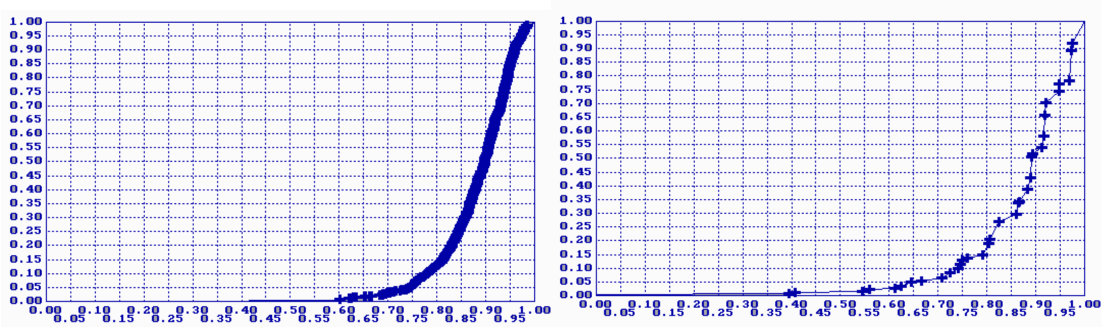
3.6. In fig. 2. Histograms of fiducial distribution of CF $\gamma_{1,м}$ and $\rho_{2,м}$ of dependent selections $\{\xi_1\}_{nv}$ and $\{\xi_2\}_{nv}$ with n_v=5 are provided



a) b)

Fig. 2. The histogram of distribution of possible realization of coefficients of correlation of Pearson $\gamma_{1,м}$ (a) and Spearmen $\rho_{2,м}$, (b) of dependent selections with n_v=5.

And in fig. 3. Statistical functions of fiducial distribution of CF $\gamma_{1,м}$ и $\rho_{2,м}$ for selections of random variables ξ with n_v=5 are given. These drawings demonstrate difference of distribution from normal, a real possibility of an operational assessment of critical values and distinction of distributions of these CF. The dispersion of realization of $\rho_{2,м}$ of fiducial distribution of $F^*(\rho_{2,м}/H_2)$ is caused by discrete nature of change of $\rho_{2,м}$ at small and the discretization step changing on value.



a) b)

Fig. 3. Statistical functions of fiducial distribution $\gamma_{1,м}$ (a) and $\rho_{2,м}$ (b) of dependent selections with n_v=5

Conclusion

1. The method of an assessment of critical values of coefficients of correlation is developed provided that selections are statistically connected;
2. The method based on fiducial approach. The assessment of statistical function of distribution is carried out on possible realization of coefficients of correlation;
3. Realization of coefficients of correlation is calculated on the modeled selections of random variables;
4. The interrelation of these selections is reached by ranging of random variables as their increase with uniform distribution in the range of [0,1];
5. Unlike coefficients of correlation of the independent selections changing within [-1,1] coefficients of correlation of dependent selections change in the range of [0,1] at the ranging type noted above. If ranging of realization of selections is opposite, then coefficients of correlation change in the range of [-1,0]

References

[1] RD 34.08.554 Summary qualifier of electric part of power plants, electric and thermal networks. Service of the best practices of ORGRES. M.: 1991, 104 p.
[2] Farhadzadeh E.M., Muradaliyev A.Z., Farzaliyev Y.Z., Rafiyeva T.K. Increase of reliability and profitability of boiler installations of block power plants. Mathematical questions of research of reliability of big systems of power: Collection of scientific articles of the Issue 66. Editor-in chief N. I. Voropay, M. A. Korotkevich, A. A. Mikhalevich. - Minsk: BNTU, 2015,-454 p.
[3] Orlov A.I. Applied statistics of M.: edition "Examination", 2004, 656 pages.
[4] Kobzar A.I. Applied mathematical statistics of M., Phi math, 2006, 628 p.
[5] Farhadzadeh E.M., Muradaliyev A.Z., Farzaliyev Y.Z., Rafiyeva T.K., Abdullayeva S. A. Assessment of interrelation of technical and economic indicators of objects of EES Kiev, Electronic modeling No. 4, 2017,
[6] Farhadzadeh E.M., Farzaliyev Y.Z., Muradaliyev of A.Z. Efficiency of criteria of expediency of classification of statistical data on refusals of electric equipment of M.: Electricity, No. 3, 2014 19-25 p.