Berlin Studies Transnational Journal of Science and Humanities ISSN 2749-0866
Vol.1 Issue 1.10 Sociological sciences
http://berlinstudies.de/

Dispersional analysis of the development of higher educational institutions

Alohon Alikarieva

Phd, assistant professor of Sociology, National University of Uzbekistan named Mirzo Ulugbek, Department of Sociology, Tashkent, Uzbekistan, alohon@mail.ru

Abstract Variance analysis is a statistical method for analyzing the results of observations that depend on various simultaneously acting factors, based on comparing estimates of the variances of the corresponding groups of sample data. The purpose of variance analysis is to test the significance of the difference between the averages by comparing the variances. The variance of the measured feature is decomposed into independent terms, each of which characterizes the influence of a particular factor or their interaction. The subsequent comparison of these terms allows us to assess the significance of each factor under study, as well as their combinations. Variance analysis is a powerful modern statistical method for processing and analyzing experimental data in the humanities. It is very closely related to the specific methodology of planning and conducting experimental studies. The article describes the quantitative assessment of factors for improving the efficiency of personnel in higher education institutions. The author offers mathematical models of socio-economic factors that affect the effectiveness of teachers ' work. The analysis of variance also revealed the influence of a number of factors on the main socio-economic indicators of higher education institutions.

Keywords – statistical method, modeling, variance, regression, factors, variance analysis, higher education institutions, teacher labor efficiency.

Introduction

Analysis of variance (from the Latin Dispersio – dispersion) is a statistical method that allows you to analyze the influence of various factors on the variable under study. The method was developed by the biologist R. Fischer in 1925 and was originally used to evaluate experiments in plant growing. Later, the general scientific significance of the analysis of variance for experiments in psychology, pedagogy, medicine, etc.

The purpose of analysis of variance is to test the significance of the difference between means by comparing variances. The dispersion of the measured trait is decomposed into independent terms, each of which characterizes the influence of a particular factor or their interaction. Subsequent comparison of such terms makes it possible to assess the significance of each studied factor, as well as their combination [1].

Most of the phenomena and processes are in constant mutual and all-encompassing objective communication. The study of dependencies and interrelations between objectively existing phenomena and processes plays an important role in various spheres of public life. It provides an opportunity to understand the complex mechanism of cause-effect relations between phenomena better. Correlation-regression analysis is widely used to study the intensity, type and form of dependencies, which is a methodological toolkit for solving problems of forecasting, planning and analyzing the activities of universities.

Economic and statistical models describe and reproduce in a formalized form real socio-economic systems, imitating their behavior in a changing environment. At the same time, the model itself is a system that transforms a set of factors (factor attributes) at the input or output results (performance attributes).

The quality of models and their adequacy to real processes are determined not only by the set of input values but also by the chosen form of connection. It is practically impossible to display the entire variety of conditions, factors and interrelationships of the real phenomenon, so in the process of economic and statistical modelling consider the most significant of them [2].

When studying and analyzing socio-economic factors affecting the effectiveness of teachers' work, one of the most important moments is to identify the significance of the influence of these or those factors in their total population on the effective indicator. In practice, this is very difficult. The task is easier if one can use the method of dispersion analysis, which is one of the sections of mathematical statistics.

Dispersion analysis is the identification and evaluation of individual factorial traits that cause the variability of the effective trait. Each factor trait varies in the total population of units. The accepted way of measuring and analyzing the variation of these features is the basis of the dispersion analysis as a method of studying the significance of factors.

The task of the dispersion analysis is to isolate the sign from the total variability of the sign:

- variability caused by the action of each of the investigated independent variables;
- variability caused by the interaction of the studied characteristics;
- random variability caused by all other unknown variables.

The idea of the dispersion analysis is to decompose the overall dispersion of the efficacious trait into the parts caused by the influence of controlled factors and the residual dispersion caused by uncontrolled influence or random circumstances. Conclusions about the materiality of the influence of controlled factors on the result are made by comparing the parts of the total dispersion when meeting the requirement of normal distribution of the resultant attribute.

Many dispersion analysis models are known. They are classified, on the one hand, by the mathematical nature of the factors (deterministic, random and mixed) and, on the other hand, by the number of controlled factors (single-factor and multi-factor models). Models with more than one factor make it possible to investigate not only the influence of individual controlled factors (main influences) on the result, but also their superposition (interactions). Complete and incomplete t-factor plans, complete and incomplete block plans and randomized block plans [2] stand out by the way of organizing the initial data among the models of dispersion analysis.

The object of the dispersion analysis study is stochastic relations between the response (reaction) and factors when the latter are not quantitative but qualitative in nature [3]. The main idea of the dispersion analysis is to compare the "factor dispersion" generated by the influence of the factor and the "residual dispersion" caused by random causes [4]. If the difference between these dispersions is significant, the factor has a significant impact on X; in this case, the average values of observed values at each level (group averages) also differ significantly. If it has already been established that the factor significantly affects X, and it is necessary to find out which of the levels has the greatest impact, the comparison of averages is additionally made in pairs. There are two models of variance analysis:

- ✓ with fixed levels of factors,
- ✓ with random factors.

Depending on the number of factors determining the variation of the effective feature, the dispersion analysis is subdivided into one-factor and multifactor.

The main schemes of the organization of the initial data with two or more factors are:

- cross-classification, which is characteristic of models with fixed levels of factors.
- hierarchical (nested) classification characteristic of random factor models.

A dispersion analysis is based on the division of the dispersion into parts or components. Intra-group dispersion explains the influence of factors not taken into account in grouping, and inter-group dispersion explains the influence of grouping factors on the group average.

Single-factor dispersion analysis can be used to identify the most significant links between variables in the qualitative study of objects of different nature. Single-factor dispersion analysis is used to compare average values for three or more samples.

As a disadvantage, it may be impossible to identify samples that are different from others. For this purpose, it is necessary to use the Sheffe method or to conduct paired comparisons of samples. In addition to the functions of single-factor dispersion analysis, the multifactor analysis evaluates the inter factor interaction [5].

In practice, it is often necessary to check the materiality of the difference between the sample averages of m samples ($m > 2$). For example, it is necessary to assess the influence of various factors on the efficiency of teachers' work, the increase of students' cognitive activity on the indicators of education quality, the improvement of mechanisms of the social management system on the studied indicators, etc. To solve this problem effectively, a new approach is needed, which is implemented in the dispersion analysis.

As it was noted, dispersion analysis is a statistical method of analysis of test results, the purpose of which is to assess the impact of one or more qualitative factors on the considered value of X, as well as for subsequent planning. That is, it is the analysis of the variability of the attribute under the influence of controlled variables. In foreign literature, variance analysis is often referred to as ANOVA, which is translated as Analysis of Variance. The author of the method is R.A. Fisher (Fisher R.A., 1918, 1938).

By the number of factors, the influence of which is studied, distinguishes between single-factor and multifactor dispersion analyses. The essence of the dispersion analysis consists in the dismemberment of the general dispersion of the studied feature on the separate components caused by the influence of concrete factors, and the testing of hypotheses about the significance of the influence of these factors on the investigated feature.

Table 1. Basic concepts and formulas

<table>
<tr><td colspan="3">Dispersion Types</td></tr>
<tr><td>• group and intra group
• inter group and general</td><td colspan="2">D_{jgr} and $D_{\boldsymbol{ingr}}$
$D_{\boldsymbol{ingr}}$and D_{gen}</td></tr>
<tr><td colspan="3">Dispersion analyses</td></tr>
<tr><td>X – research factor</td><td colspan="2">μ – overall average feedback Y;</td></tr>
<tr><td>Y – feedback (experiment result) Model: $Y_{ij} = \mu + F_i + \varepsilon_{ij}$</td><td colspan="2">$F_i$– factor influence X_i on Y
ε_{ij} – random balance</td></tr>
<tr><td>Factor (intergroup) dispersion
S^2_{fact} (X)</td><td>←comparison→
⇓
Factor influence degree</td><td>Residual (intergroup) dispersion
S^2_{in}</td></tr>
</table>

The two-factor dispersion analysis is used when the simultaneous effect of two factors on different samples of objects is studied, i.e. when different samples are under the influence of different combinations of two factors. It can happen that one variable has a significant effect on the studied feature only at certain values of another variable. The essence of the method remains the same as in the case of a single-factor model, but more hypotheses can be tested in the two-factor dispersion analysis [6].

The solution of the two-factor dispersion analysis problem depends on the number of observations made at each combination of factor levels, if, in other words, in each cell of the two-factor complex.

The dispersion analysis is designed to evaluate the influence of different but controlled factors on the result of the experiment. Let the result of the experiment be some random value of Y, also called a response. The values of the random value of Y are influenced by the factor X, consisting of n-levels. Depending on the number of factors included in the analysis, a distinction is made between single-factor, two-factor and multi-factor dispersion analysis.

A dispersion analysis is possible if the measurement results are independent random variables subject to the normal distribution law with the same dispersions. Single-factor dispersion analysis reveals the degree of influence of one factor *X* on the mathematical expectation of *M(Y)* response. The factor can be quantitative or qualitative. In the course of the experiment, the factor *X* is supported at *n*-levels. At each level of the factor *m* duplicate experiments are carried out. Value *m* can be the same or different for each level. The results of all measurements are presented in the form of a table called an observation matrix [7].

Table 2. Surveillance Matrix

Factor level number	**Factor level**	**Observation**	**Number of duplicate experiments**
1	X_1	$Y_{11}, Y_{12}, \dots, Y_{1j}, \dots, Y_{1m_1}$	m_1
…	…	…	…
i	X_i	$Y_{i1}, Y_{i2}, \dots, Y_{ij}, \dots, Y_{im_i}$	m_i
…	…	…	…
n	X_n	$Y_{n1}, Y_{n2}, \dots, Y_{nj}, \dots, Y_{nm_n}$	m_n
			$N = \sum_{i=1}^{n} m_i$

First, for each series of duplicate experiments, the arithmetic mean μ_i is calculated, which are the estimations of $M(Yi)$ and S^2_{bi} reproducibility dispersion (Tab.3).

Then the homogeneity of a series of dispersions S^2_{bi} is checked either in pairs using the Fisher criterion (if m_i are different) or using the Cochrane criterion (if $m_i$are constant). For this purpose, we formulate a null hypothesis $H_0: D(X_1) = D(X_2) = \dots = D(X_L)$. The observed value of the Kohren criterion is determined by samples of one volume:

$G_{\text{obs}} = S^2_{max} / \sum_{i=1}^{L} S_i^2$. The observed value of the criterion is compared with the critical point of the right critical region $Gcr\ (\alpha;\ k;\ L)$, where $k = m - 1$(Appendix 8) and the homogeneity of dispersions is concluded. If the dispersions are not homogeneous, no further analysis is carried out.

After confirming the hypothesis about the homogeneity of dispersions, you can proceed to the analysis. It is believed that the result of any measurement Y_{ij} can be represented by a model:

$$Y_{ij} = \mu + F_i + \varepsilon_{ij},$$

where Y_{ij} is the value of the studied variable obtained at the i-th level of the factor with the j-th serial number;

μ is the overall average of the response Y;

F_i - the effect of the influence of factor X_i on Y: deviation of the mean values of μ_i at the i-th level (group means) from the general average μ (i.e F_i = μ_i-μ);

ε_{ij} is a random remainder reflecting the influence of all other uncontrolled (unaccounted) factors on the value of Y_{ij}.

The main assumptions of ANOVA are as follows:

- the remainders ε_{ij} are mutually independent for any i and j;
- the ε_{ij} values are subject to the normal law [8].

The task of analysis of variance is to assess the significance of the effect of changes in the level of a factor. The dispersion of the response values caused by the controlled factor is estimated by the factor variance (the sum of the squares of the deviations of the group means from the total mean) – S^2_{fact} (X).

The influence of uncontrollable factors (contribution ε_{ij}) can be estimated by the average variance of reproducibility (residual variance) – S^2_{in} .

The total dispersion of the response values caused by both controlled and uncontrolled factors is estimated by the total (or total) variance (total sum of squares of deviations) – S^2_{total}.

Table 3. Formulas for calculating dispersions in dispersion analysis

Average arithmetic (group)	**Dispersion reproducibility (group)**	**Residual (intra-group) dispersion**
$\mu_i = \frac{1}{m_i}\sum_{j=1}^{m_i} Y_{ij}$	$S^2_{in_i} = \frac{1}{m_i}\sum_{j=1}^{m_i}(Y_{ij} - \mu_i)^2$	$S^2_{in} = \frac{1}{n}\sum_{i=1}^{n} S^2_{in_i}$
Total average	Factor (intergroup) dispersion	Total (full) dispersion
$\mu = \frac{1}{n}\sum_{i=1}^{n} \mu_i$	$S^2_{fact}(X) = \frac{1}{n-1}\sum_{i=1}^{n} m_i(\mu_i - \mu)^2$	$S^2_{total} = \frac{1}{N-1}\sum_{i=1}^{n}\sum_{j=1}^{m_i}(Y_{ij} - \mu)^2$

To identify the degree of influence of factor X and compare it with the spread (caused by random, uncontrolled reasons), the homogeneity of the variances of factorial and reproducibility (residual) is checked according to Fisher's criterion: $F_{obser} = (S_{fact}{}^2) / (S_{in}{}^2)$. The observed value of the criterion is compared with the critical F_{cr} (α; k_1; k_2), which is found from the tables of the F-distribution for the significance level α, the number of degrees of freedom k_1 = n-1 and k_2 = N – n.

If $F_{obser} \le F_{cr}$ (α; k_1; k_2), then the influence of factor X is insignificant. Consequently, all the obtained measurement results belong to one general population, distributed normally with parameters m and $S_{total}{}^2$. When $F_{obser} > F_{cr}$ (α; k_1; k_2), the influence of the factor is taken to be significant. It is believed that in this case there are n normally distributed populations, each of which has a corresponding mathematical expectation μ_i and the same variance $S_{in}{}^2$. The assessment of the effect of the influence of the i-th level of the factor is equal to the difference between the total and group means ($F_i = \mu_i - \mu$). Suppose that factor X affects the response Y. To measure the degree of this influence, a sample coefficient of determination is used, which is equal to the ratio of variances:

$$\overline{d} = \frac{S_{fact}^2(n-1)}{S_{total}^2(N-n)} = \frac{S_{fact}^2(n-1)}{S_{fact}^2(n-1) + S_{in}^2 n}$$

The sample determination factor shows what proportion of the sample total dispersion is the factor dispersion (group averages), i.e. what proportion of the total dispersion is explained by the dependence of the response *Y* on factor *X* [9].

There are conditions of application of dispersion analysis:

1. The aim of the study is to determine the strength of the influence of one (up to 3) factor on the result or to determine the strength of the combined influence of various factors (sex and age, physical activity and nutrition, etc.).
2. The factors under study should be independent (unrelated). For example, it is not possible to study the joint effect of work experience and age, height and weight of children, etc. on the morbidity of the population.
3. The selection of groups for the study is done randomly (random selection). Organization of the dispersion complex with the implementation of the principle of random selection of variants is called randomization, i.e. the chosen at random.
4. It is possible to apply both quantitative and qualitative (attributive) characteristics.

Classic dispersion analysis is carried out in the following stages:

1. Construction of the dispersion complex.
2. Calculation of mean squares of deviations.
3. Calculation of dispersion.
4. Comparison of factor and residual variance.
5. Evaluation of the results using theoretical values of the Fischer-Snedecor distribution.
6. Modern applications of dispersion analysis cover a wide range of tasks of

economics, biology and technology and are usually interpreted in terms of statistical theory of revealing systematic differences between the results of direct measurements made under different changing conditions.

7. Thanks to the automation of dispersion analysis, the researcher can conduct various statistical researches with the use of a computer, while spending less time and effort on data calculations. Currently, there are many application packages, in which the apparatus of dispersion analysis is implemented. The most common are such software products as MS Excel, Statistica; Stadia; SPSS [10].

The majority of statistical methods are implemented in modern statistical software products. With the development of algorithmic programming languages, it became possible to create additional blocks for processing statistical data. Dispersion analysis is a powerful modern statistical method of processing and analysis of experimental data in humanities. It is very closely related to the specific methodology of planning and conducting experimental research.

If we deal with a multifactor process, then with the help of dispersion analysis it is possible to determine the dispersions caused by the action of each factor separately, and to estimate the statistical significance of these values. With the help of the dispersion analysis there are values obtained as a result of special calculations of factorial dispersion relations ($\boldsymbol{S^2}_{\boldsymbol{A}}$) to random ($\boldsymbol{S^2}_{\boldsymbol{R}}$). This latter is compared with the theoretical value of $\boldsymbol{F}$ (Fisher's criterion): tables of such values are given in the corresponding manuals on mathematical statistics. If the calculated value of $\boldsymbol{F}$ is less than the tabular value, there is no reason to take into account the influence of the factor under consideration; if the calculated value of $\boldsymbol{F}$ is greater than the tabular value, then it should be assumed that the factors under consideration influence the phenomenon under study.

Let us briefly describe the calculation technique based on analysis of variance. Let the number of factors belonging to one object of variability be equal to p and under the influence of the factor $\boldsymbol{A_g}$ ($\boldsymbol{q} = \overline{\boldsymbol{1}, \boldsymbol{p}}$) ng values of the quantity x_g were observed. Let us denote by x_{gh} the value of the h_{th} number in this group. Then the average value for each group will be:

$$\overline{\boldsymbol{x_g}} = \frac{\sum_{\boldsymbol{g=1}}^{\boldsymbol{n_g}} \boldsymbol{x_{gh}}}{\boldsymbol{n_g}}, \qquad (1)$$

and the overall average over the entire set of observations is –

$$\boldsymbol{x} = \frac{\sum_{\boldsymbol{g=1}}^{\boldsymbol{p}} \sum_{\boldsymbol{n=1}}^{\boldsymbol{n_g}} \boldsymbol{x_{gh}}}{\boldsymbol{n}}. \qquad (2)$$

For dispersion analysis, the sum of the squares $\boldsymbol{n}$ deviations of $\boldsymbol{xgh}$ values from the total mean value should be broken down into constituent parts, one of which corresponds to the object of variability and the other to the influence of random causes.

The total sum of the squares of deviations is

$$\boldsymbol{Q} = \boldsymbol{Q_R} + \boldsymbol{Q_A}, \qquad (3)$$

where $\boldsymbol{Q_R}$,$\boldsymbol{Q_A}$are sums of squares of deviations a) group averages from the general average, b) within groups, i.e. under the influence of random unexplored factors [11].

The value of $\boldsymbol{Q}$ can be calculated by the formula:

$$\boldsymbol{Q} = \sum_{g=1}^{p} \sum_{n=1}^{\boldsymbol{n_g}} \left(x_{gh} - \bar{x}\right)^2, \qquad (4)$$

Q_A value - according to the formula:

$$\boldsymbol{Q_A} = \sum_{g=1}^{p} \sum_{n=1}^{\boldsymbol{n_g}} \left(x_{gh} - \overline{x_g}\right)^2, \qquad (5)$$

Q_R value - according to the formula:

$$\boldsymbol{Q_R} = \sum_{g=1}^{p} \sum_{n=1}^{\boldsymbol{n_g}} \left(x_{gh} - x\right)^2, \qquad (6)$$

The sum of squares of deviations within groups is the difference between the total sum of squares of deviations and the sum of squares of deviations between groups:

$$\boldsymbol{Q_R} = \boldsymbol{Q} - \boldsymbol{Q_A}. \qquad (7)$$

The most convenient way to calculate $\boldsymbol{Q}$, $\boldsymbol{Q_R}$ and $\boldsymbol{Q_A}$ is to use formulas based on the decomposition of expressions (4) and (6):

$$\boldsymbol{Q} = \sum_{g=1}^{p} \sum_{n=1}^{\boldsymbol{n_g}} x_{gh}^2 - \boldsymbol{n}\bar{x}^2, \qquad (8)$$

$$\boldsymbol{Q_A} = \sum_{g=1}^{p} \boldsymbol{n_g}\bar{x}_g^2 - \boldsymbol{n}\bar{x}^2, \qquad (9)$$

$$Q_R = \sum_{g=1}^{p} \sum_{n=1}^{n_g} x_{gh}^2 - \sum_{g=1}^{p} n_g \bar{x}_{g\cdot}^2. \qquad (10)$$

Division by the corresponding number of degrees of freedom $n–1$, $p–1$ and $n–p$ of the sum of squares of deviations Q, Q_A and $Q_R$will give estimates of dispersion S^2, $S^2{}_A$ and $S^2{}_R$ (S^2 - general dispersion, , $S^2{}_A$ – factorial, $S^2{}_R$ – residual.

$$S^2 = \frac{Q}{n-1}, \qquad S_A^2 = \frac{Q_A}{p-1}, \qquad S_R^2 = \frac{Q_R}{n-p}. \qquad (11)$$

The comparison is made between the variance $S_A{}^2$, caused by the factors under consideration, and the residual $S_R{}^2$, which arises after eliminating the influence of the factors.

If in the case under consideration $S^2{}_A > S^2{}_R$, then the dispersion relation is taken in the form

$$F = \frac{S_A^2}{S_R^2}, \qquad (12)$$

If $S_A^2 < S_R^2$, then

$$F = \frac{S_R^2}{S_A^2}, \qquad (13)$$

In accordance with the number of degrees of freedom, tabular values of the dispersion relations for the probability are selected.

Thus, the general scheme of the dispersion analysis at a single-factor complex can be represented in the form of consecutive operations:

1) grouping;
2) determination of averages by groups and the general average;
3) calculation of the sum of squares of deviations of group averages from the total average;
4) the same - all observed values from the total average;
5) the same - within groups as the difference between the total sum of squares and the sum of squares between groups;
6) finding the number of degrees of freedom of variation in groups and within groups;
7) determination of inter- and intragroup dispersion (taking into account the number of degrees of freedom) and the ratio of the larger value of dispersion to the

smaller one;

8) selection of F-values from tables with a given probability;

9) comparison of the calculated value with the tabular one and conclusion about the reliability or unreliability of the influence of the studied factors [12].

As an example of calculation, let's consider the dispersion analysis of the influence of production of one teacher (in sums) on the movement of personnel. The initial information is based on the conditional data of different groups. Production per teacher, average age of the group and other average indicators are calculated by the author.

The grouping method divides the values of the factor under consideration into four groups. The number of observations for each group was as follows:

$$\boldsymbol{n = n_{\mathrm{I}} + n_{\mathrm{II}} + n_{\mathrm{III}} + n_{\mathrm{IV}} = 8 + 14 + 11 + 8 = 41}$$

Let's find the sum of the factors' values and average values of one teacher's output for each group (Table 4). It is necessary to determine whether the difference between the average values is significant or it can be explained by a random composition of private populations.

Table 4

Group	**One teacher's output, soum.**								$\sum x_i$	n_g	x
I	308;	257;	355;	336;	441;	338;	365;	316;	2716	8	339,50
II	307;	301;	349;	292;	331;	318;	350;	368;	4636	14	331,14
	359;	331;	339;	372;	305;	312;					
III	311;	381;	316;	208;	358;	175;					
	351;	357;	411;	454;	426;				3748	11	340,72
IV	116;	135;	149;	178;	164;	385;	331;	370;	1828	8	228,50
				Total					12 928	41	310,00

To facilitate calculations, we will consider their deviations from the general average value instead of the x values. The deviations of the values of the features, their sums, the average values, the sums of the squares of the average values, and the corresponding number of observations are shown in Table 5.

To facilitate the calculations, instead of x values, we will consider their deviations x_{gh} from the total average value. The deviations of the values of the features, their sum $\sum \boldsymbol{x_{gh}}$ mean values $\overline{\boldsymbol{x_g}}$, the sum of the squares of the mean values $\sum \boldsymbol{x_{gh}^2}$, the corresponding number of observations n_g are given in Table. 5.

Table 5

Group	**Deviations from the average** x_{gh}	$\sum x_{gh}$	n_g	$\overline{x_g}$	$\sum x^2{}_{gh}$	$n_g\overline{x_g^2}$
I	–2; –53; 45; 26; 131; 28; 55; 6	136,0	8	17,0	26 520	2312
II	–3; –9; 39; –18; 21; 8; 30; 58; 49; 21; 29; 62; – 4; 2	285	14	20,35	14 251	5797,68
III	1; 71; 6; 102; 48; –135; 41; 47; 101; 144; 116	238	11	21,63	84 294	5146,35
IV	–206; 175; –161; –132; –146; 75; 21; 60	–664	8	–83	14 7383	55112
	Total	–5	41	–24,02	272 448	68368,03

Using Table 5 and formulas (8), (9) and (10), let's determine the dispersion of characters:

1. $Q = \sum_{g=1}^{p}\sum_{n=1}^{n_g} x_{gh}^2 - n\overline{x}^2 = 272448 - 41\,(-24,03)^2 = 272448 - 23675,04 = 248\,772,96$;

2. $Q_A = \sum_{g=1}^{p} n_g\overline{x}_g^2 - n\overline{x}^2 = 68\,368,03 - 23\,675,04 = 44\,692,99$;

3. $Q_R = \sum_{g=1}^{p}\sum_{n=1}^{n_g} x_{gh}^2 - \sum_{g=1}^{p} n_g\overline{x}_g^2 = 272448 - 68368,03 = 204079,97$.

Based on these data, we calculate the variance estimates:

$$S^2 = \frac{Q}{n-1} = \frac{248772,96}{40} = 6219,32;$$

$$S_A^2 = \frac{Q_A}{p-1} = \frac{44692,99}{3} = 14897,66;$$

To assess the impact of the development of one teacher on the movement of personnel, one can compare the variance by the factors S^2_A and S^2_R. Since in the considered example $S^2_A > S^2_R$, the variance ratio will be

$$F = \frac{S_A^2}{S_R^2} = \frac{14897,66}{5515,67} = 2,70.$$

In this case, the number of degrees of freedom corresponding to the factorial variance v is 3, the smaller variance (v_2) is 37, and the total variance is – 40.

From the table values, select the corresponding number of degrees of freedom.

For the probability p = 0.05, the table value F_τ will be 2.86, that is, $F_\tau < F_p$ .This proves that the development of one teacher at the studied objects is a significant factor that significantly affects the movement of personnel.

The importance of other factors is determined by a similar calculation method. With a 5% confidence level, the following are significant: average age, profitability, utilization rate, output of one teacher and average salary per person. Other factors turned out to be insignificant. Obviously, each of these factors, taken separately, influences the efficiency of teachers' work indirectly, that is, through other factors.

Thus, the dispersion analysis allows us to determine the impact of a number of factors on the main socio- economic indicators of universities.

Summing up, we can say that the purpose of dispersion analysis is to check the statistical significance of the difference between the averages (for groups or variables). This check is carried out by dividing the sum of squares into components, i.e. by dividing the total variance (variation) into parts, one of which is caused by a random error (i.e., intragroup variability), and the other is related to the difference in mean values. The last dispersion component is then used to analyse the statistical significance of the difference between the mean values. If this difference is significant, the null hypothesis is rejected and an alternative hypothesis about the existence of a difference between averages is accepted.

The formation of a system of quantitative indicators of the effectiveness of teachers and the development of scientifically sound methods of evaluation, taking into account the real labor input of all types of educational and scientific work of the teacher in order to optimize the use of his teaching and research experience and knowledge, taking into account the individual characteristics of each teacher, will increase the productivity of teaching labor and provide a better quality of training of students [13,14,15].

The motivation function provides for actions aimed at encouraging teachers to work effectively to achieve the goals of the university. In general, the effectiveness of teachers' work depends on many factors, among which the main role is played by a clear understanding of the goals of their work, the likelihood of achieving these goals, the resources necessary to achieve these goals, the system of material and moral incentives to ensure the interest of teachers in improving their work efficiency [16].

The study of different approaches to consideration of labor activity efficiency allowed to systematize the classification of factors of growth of labor efficiency of the faculty of higher educational institutions taking into account the specifics of labor in the field of education [17].

And so, the following factors influence economic and social results of labor in the sphere of education [18]:

- organizational-economic, consisting of quantitative (number of lecture hours by discipline, number of practical hours by discipline, number of hours of practice at the enterprise, etc.) and qualitative (quality of lectures, quality of practical and laboratory lessons, quality of practice organization at the enterprise, organization of the workplace, providing the use of progressive methods and organizational forms of work, improvement of management structure, etc.);

- material and technical factors (number of names of laboratory equipment, number of equipment units of one name, useful area of buildings of higher education institution per student, technical level of laboratory equipment, provision of educational, methodical, periodical and scientific literature, etc.);
- scientific and innovative factors (number and significance of scientific articles published by teachers and students in the reporting period;
- the amount of financial resources allocated for research work by external customers;
- the intensity of defending doctoral dissertations (Phd and DSc), obtaining scientific titles;
- the share of doctoral candidates (applicants) who defended their dissertations and remained to work at the university;
- the share of students who have passed final qualifying tests and entered the doctoral studies at the university, etc.);
- social factors (degree of dependence of material reward of students on quality of their knowledge, degree of dependence of material reward of teachers on quality of students' training, level of development of social security system of students and teachers at the university);
- psychophysiological factors, conditioned by individual peculiarities of a person, influencing the time, spent by a lecturer on transfer of educational information and students on its perception, as well as on other activities of lecturers (degree of responsibility, level of professionalism, attitude to one's profession, ability to study, health condition, age, etc.).

Conclusion

So, in this article the following is done:

1. The quantitative assessment of the factors of increasing the efficiency of personnel at universities is characterized.
2. Mathematical models of socio-economic factors influencing the efficiency of teachers' work are proposed.
3. The method of calculation determines the importance of various factors: average age, profitability, utilization factor, output of one employee and average salary per person.
4. The method of dispersion analysis established the influence of a number of factors on the main socio- economic indicators of universities.

References:

1. Kremer N.SH. Teoriya veroyatnosti i matematicheskaya statistika. – M.: Yuniti – Dana, 2002. – 343s.

2. Berezhnaya Ye.V., Berezhnoy V.I. Matematicheskiye metodi modelirovaniya ekonomicheskikh sistem (Mathematical methods of modeling of economic systems): Ucheb. Posobiye (Textbook). – 2-ye izd., pererab. i dop (2 nd edition, revised and supplemented). – Moscow: Finance and statistics, 2011. – 432 s.

3. Matematicheskaya statistika (Mathematical statistics) Ucheb. dlya vuzov (Textbook for universities) / V.B.Goryainov, I.V.Pavlov, G.M.Tsvetkova [i dr.] [and others]; pod red.(by ed.) V.S.Zarubina, A.P.Krishchenko. – M.: Publishing House of Moscow State Technical University. N.E.Bauman, 2001. – 424 s.

4. Gmurman V.Ye. Teoriya veroyatnostey i matematicheskaya statistika (Probability theory and Mathematical Statistics): Ucheb. posobiye dlya vuzov (Textbook for universities) / V.Ye.Gmurman. – 9 th ed. – M.: Higher. Shk., 2003. – 479 p.

5. Sheffe G. Dispersionniy analiz (Dispersion analysis). 2-ye izd., pererab. i dop (2 nd edition, revised and supplemented). – M.: Science, 1980. – 512 p.

6. Bogdanova M.G., Starozhilova O.V. Teoriya veroyatnostey i matematicheskaya statistika (Probability theory and mathematical statistics) // Uchebnoye posobiye (Chast' 2. Regressionnyy analiz, dispersionniy analiz) (Tutorial (Part 2. Regression analysis, analysis of variance)). – Samara: INUTPGUTI, 2015. –144 p.

7. Khodarev O.N. Planirovaniye i organizatsiya eksperimenta: uchebn. posobiye (Planning and organization of the experiment) / Yuzh. -Ros.gos.tekhn.un-t (NPI) - Novocherkassk: YURGTU (NPI), 2010- 146 s.

8. Bogdanova M.G., Starozhilova O.V. Teoriya veroyatnostey i matematicheskaya statistika (Probability theory and mathematical statistics) // Uchebnoye posobiye (Chast' 2. Regressionniy analiz, dispersionniy analiz) (Tutorial (Part 2. Regression analysis, analysis of variance)). – Samara: INUTPGUTI, 2015. –144 p.

9. Maksimenko N.V. Teoriya veroyatnostey i matematicheskaya statistika (Probability theory and mathematical statistics) [Elektronniy resurs]: metod. ukazaniya/Ye.N.Smirnova, Orenburgskiy gos. un-t, N.V.Maksimenko. – Orenburg : OGU, 2014 . – 131 s.

10. Blokhin A.V. Teoriya eksperimenta (The theory of the experiment) [Elektronnyy resur]: Kurs lektsiy v dvukh chastyakh: Chast' 1. – Mn.: Nauchno-metodicheskiy tsentr “Elektronnaya kniga BGU”, 2003.

11. Yudenkov V.A. Dispersionniy analiz (Analysis of variance). – Minsk:Biznesovset, 2013. – 76s.

12. Baldin K.V. Osnovi teorii veroyatnestey i matematicheskoy statistiki (Fundamentals of probability theory and mathematical statistics). – M.:Flinta, 2010. – 487 s.

13. Gmurman V.Ye. Teoriya veroyatnostey i matematicheskaya statistika (Probability theory and mathematical statistics): Ucheb.posobiye dlya vuzov / V.Ye.Gmurman. – 9-ye izd., ster. – M.:Vissh.shk., 2007. – 478 s.

14. Alikarieva Alokhon, Alikariev Nuriddin, Aliqoriev Olimkhon, Sabirova Umida, Rashidova Shahnoza. Issues of improving the quality of training of Highly qualified specialists // International Journal of Advanced Science and Technology, 29(05), Retrieved from http: // sersc. Org / journals / index. Php /IJAST/article/view/10090. Vol. 29, No. 5, (2020), pp. 1600-1611.

15. Sabirova U. Innovation in the processes of reforming the higher education system in Uzbekistan (2020) International Journal of Psychosocial Rehabilitation, 24 (Special Issue 1), pp. 551-560.Innovation in the processes of reforming the higher education system in Uzbekistan.

16. Litvinova O.I. Effektivnost truda prepodavateley visshix uchebnix zavedeniy kak faktor rosta konkurentosposobnosti vuza (Efficiency of work of teachers of higher educational institutions as a factor of growth of competitiveness of higher education institution): dissertatsiya ... kandidata ekonomicheskix nauk : 08.00.05.- Omsk, 2012.- 216 s.

17. Kochetkova N.N., Iglina N.A., Ryabova T.V. Upravleniye effektivnostyu truda professorsko-prepodavatelskogo sostava vuza (Management of labor efficiency of the teaching staff of the University) / Astrakhanskiy gosudarstvennyy tekhnicheskiy universitet, 2018. №2. – S.42.

18. Cherkasov V.Ye. Sreda obucheniya i prepodavaniya, effektivnost truda prepodavateley kak faktor rosta konkurentosposobnosti vuza (The environment of training and teaching, the effectiveness of teachers as a factor of increasing the competitiveness of the University). – M.:Vissh.shk., 2012.