UDC 517. 22.161.1

APPLICATION OF ONE-FACTOR DISPERSION ANALYSIS IN STATISTICAL PROBLEM OF THE YIELD OF GRAIN CROPS DEPENDING ON THE TYPE OF AGRICULTURAL HOLDINGS

Danil Tsapodoi
National Aviation University, Kyiv

Supervisor Yana Liashenko, PhD in Physics and Mathematics, Associate Professor

Keywords: one-factor dispersion analysis, statistical methods, level of significance, unbiased variance estimates.

One-factor dispersion analysis is a statistical method used to determine the existence of statistically significant differences between the average values of several groups. This method is based on comparing variations (variances) of data within groups and between groups to understand whether one or more factors (independent variables) affect a dependent variable.

One-factor dispersion analysis is an important tool in various fields of science and technology, in particular in biology, psychology, medicine, social sciences, and engineering, where there is a need to compare the average values of several groups. It allows researchers to establish whether the difference between groups is random or if it is the result of the influence of the factor under study.

To perform one-factor variance analysis, research data must be divided into several groups based on one classification criterion (independent variable), and a certain quantitative characteristic (dependent variable) is measured for each group [1].

We will take the yield of enterprises and Households from 1 hectare in 10 regions of Ukraine (statistics are taken from the official website of the State Statistics Service of Ukraine for 2021 [2]) and determine by the one-factor dispersion analysis whether the yield depends on the type of economic activity (enterprise or Households).

Let's check the hypothesis H_0, that the factor (type of Agricultural holdings) does not affect the yield. According to the sample given in Table 1, we find the sum of the squares of deviations among groups (dispersion by factors) $Q_1 = \sum_{j=1}^{m} n_i (\bar{x}_i - \bar{x})^2 = 74.11$, the sum of the squares of deviations of observations from the total average $Q = \sum_{j=1}^{m} \sum_{i=1}^{n_i} (x_{ij} - \bar{x})^2 = 361.07$, the sum of the squares of deviations within the group (residual scattering) $Q_2 = \sum_{j=1}^{m} \sum_{i=1}^{n_i} (x_{ij} - \bar{x}_i)^2 = 286.95$.

Table.1

Regions	Agricultural holdings (yield, centner per ha of the harvested area)		Regions	Agricultural holdings (yield, centner per ha of the harvested area)	
	Enterprises	Households		Enterprises	Households
Vinnytsya	35,1	27,3	Poltava	17,4	22,6
Dnipropetrovsk	23,4	34,9	Cherkasy	21,4	23,1
Donetsk	32,1	24,5	Kherson	27,1	27,5
Zaporizhzhya	22,4	25,8	Mykolaiv	23,1	25,4
Kyiv	23,4	25,2	Odesa	23,4	24,6

Unbiased variance estimates s^2 are $S_1^2 = \frac{Q_1}{m-1} = 74.11$, $S_2^2 = \frac{Q_2}{n-m} = 15.94$, $S^2 = \frac{Q}{n-1} = 19.00$.

The value of the criterion [3] is $F^* = \frac{S_1^2}{S_2^2} = 4.64$.

We compare the obtained value with the table value $F_{cr}(0.05;1;18) = 4,41$ according to the given level of significance $a = 0.05$ [3] and we find out if the type of agricultural holdings is important for growing products, particularly millet.

Since $F^* > F_{cr}$, that is, the observed value occurs in the critical area, and therefore the hypothesis H_0 deviates. Therefore, we can assert with the level of significance $a = 0.05$ that the yield of millet in observed regions depends on the type of agricultural holdings.

So, by the of one-factor dispersion analysis, it was possible to determine that the yield of grain crops depends on who grows the products, and the data in the table, which shows that yields in enterprises higher than in households, are not accidental.

Thus, one-factor dispersion analysis is a powerful tool for analyzing the influence of one factor on the results of the study. Using this method allows researchers to draw objective conclusions about the statistical significance of factor influence and use this knowledge for further research and practical applications.

References:

1. https://www.investopedia.com/terms/a/anova.asp
2. https://www.ukrstat.gov.ua/
3. Руденко В. М. Математична статистика. Навч. посіб. – К.: Центр учбової літератури, 2012. – 304 с.