UDC 519.237.4
DOI: 10.20535/SRIT.2308-8893.2024.2.10

STUDY OF THE FACTOR INFLUENCE ON THE UNIFORMITY OF COFFEE GRAIN GRINDING BY METHODS OF STATISTICAL ANALYSIS

I.V. HRYHORENKO, S.I. KONDRASHOV, S.M. HRYHORENKO, O.S. OPRYSHKIN

Abstract. In order to assess the impact of each of the factors that affect the quality and uniformity of grinding coffee beans and to compare the impact of these factors, it is worth establishing a quantitative indicator of this impact. To solve this problem, dispersion analysis was used as a method of organizing sample data according to possible sources of dispersion. The chosen method made it possible to decompose the total dispersion into components caused by the influence of factor levels. Grinding time, geometric dimensions of the grain, moisture content of the grain, speed of rotation of the motor shaft were selected as factors influencing the homogeneity of grinding. The justification and assessment of the reliability of statistical conclusions about the informational significance of indicators affecting the homogeneity of coffee grinding was carried out to ensure the highest possible probability of the obtained result.

Keywords: dispersion analysis, homogeneity of grinding, factor influence, model, indicator of control, coffee bean.

INTRODUCTION

The problems of determining the factor influence on the quality and uniformity of grinding and the creation of systems for controlling the grinding process are of interest to both domestic scientists and the world scientific community. The paper [1] states that one of the most common problems in the preparation of coffee drinks in coffee machines is the unsatisfactory quality of coffee, which is associated with improper grinding of coffee beans, and the taste of the coffee drink depends on the size of the ground particles and the uniformity of coffee grinding. The paper [2] examines the influence of the origin of coffee beans and the temperature during the grinding of roasted coffee. It is noted that the extraction depends on the temperature, the chemical composition of the water, as well as the available surface area of the coffee. The study [3] reported that some physicochemical characteristics such as extraction yield, total dissolved solids, total phenol content, pH, and titrated acidity can strongly depend on the degree of grinding of the coffee bean. The work [4; 5] is aimed at developing an experiment to study the key factors affecting different methods of coffee preparation. The need to obtain different degrees of grinding of coffee beans in order to be able to provide a high flow rate of coffee in the preparation of espresso is discussed in the paper [6].

The structure or scheme of an experiment is described, by the factors involved in it and the ways in which different levels of different factors are combined [7]. The variance is used here as the simplest measure of dispersion, provid-

© *I.V. Hryhorenko, S.I. Kondrashov, S.M. Hryhorenko, O.S. Opryshkin*, 2024

ing an opportunity to compare the influence of the factor under study and the factor of chance [8]. If the dispersion is due to the joint action of random causes and the change in the levels of the factors, then by obtaining an estimate of the total response variance and the estimation of the variances of the factors, one can find an estimate of the residual variance, and then, using statistical variance comparison criteria, rank the factors according to the degree of their effect on the response dispersion.

PRESENTATION OF THE TASK

The quality of grinding coffee beans is influenced by a number of factors that negatively affect the uniformity of grinding, which are difficult to stabilize to reduce the impact on the original value. As previously mentioned, it is necessary to carry out a procedure for randomizing factors in order to make their impact random and to be able to use statistical criteria.

Suppose that H — is the parameter of the control object characterizing the homogeneity of coffee grinding, which needs to be determined; $F_1, \dots F_n$ — factors affecting the quality of the grinding process (e.g. grinding time, speed of rotation of the motor shaft, temperature at the engine stator, distance between mills, grain moisture, geometry of the bean: width, thickness, length). The result that takes into account the action of each of the factors can be written in the form of a mathematical model in which the influencing factors are H and $(n-1)$ factors due to the variability of the remaining control indicators. This statement is due to the fact that the remaining indicators characterize quantitatively $(n-1)$ the physical properties of the control object and can be directly measured.

To assess the homogeneity of coffee grinding, consider the model of the effect on the measurement result of the control index F taking into account the effects of four factors (H and factors whose levels are quantified by the values of the three control indicators (grinding time, geometric grain sizes, grain moisture). Data in cross-classification are denoted by symbols with four indices $\alpha, \beta, \gamma, \delta$. Since the control indicators are not additive, it is necessary to introduce components into the model characterizing the interaction between the indicators. Thus, the mathematical model has the form:

$$F_{\alpha\beta\gamma\delta i} = \overline{F} + f_\alpha + A_\beta + B_\gamma + C_\delta + (fA)_{\alpha\beta} + (fB)_{\alpha\gamma} + (fC)_{\alpha\delta} + (AB)_{\beta\gamma} + \\ + (AC)_{\beta\delta} + (BC)_{\gamma\delta} + (fAB)_{\alpha\beta\gamma} + (fAC)_{\alpha\beta\delta} + (fBC)_{\alpha\gamma\delta} + \\ + (ABC)_{\beta\gamma\delta} + (fABC)_{\alpha\beta\gamma\delta} + \varepsilon_{\alpha\beta\gamma\delta i}. \quad (1)$$

where $\alpha, \beta, \gamma, \delta$ are the number of factor levels; $\overline{F}$ — is the general average; f_α — is the deviation of the measurement result of the control indicator F from its average value $\overline{F}$, which is due to the influence of the parameter H; $A_\beta, B_\gamma, C_\delta$ — deviation of the measurement result $F_{\alpha\beta\gamma\delta}$ from $\overline{F}$, due to three factors; $(fA)_{\alpha\beta}, (fB)_{\alpha\gamma}, (fC)_{\alpha\delta}, (AB)_{\beta\gamma}, (AC)_{\beta\delta}, (BC)_{\gamma\delta}$ — deviations due to pairwise interactions of all influencing factors; $(fAB)_{\alpha\beta\gamma}, (fAC)_{\alpha\beta\delta}, (fBC)_{\alpha\gamma\delta}$, $(ABC)_{\beta\gamma\delta}$ — deviations due to the interaction of three influencing factors; $(fABC)_{\alpha\beta\gamma\delta}$ — deviation, which is due to the interaction of four influencing factors; $\varepsilon_{\alpha\beta\gamma\delta i}$ — random residue; *i* — is the number of multiple measurement at fixed levels $\alpha, \beta, \gamma, \delta$.

The initial conditions for model (1) will be:

1. $\sum_{\alpha} f_{\alpha} = 0;\ \sum_{\beta} A_{\beta} = 0;\ \sum_{\gamma} B_{\gamma} = 0;\ \sum_{\delta} C_{\delta} = 0;$

$\sum_{\alpha}(fA)_{\alpha\beta} = 0;\ \sum_{\beta}(fA)_{\alpha\beta} = 0; \sum_{\alpha}(fB)_{\alpha\gamma} = 0; \sum_{\gamma}(fB)_{\alpha\gamma} = 0; \sum_{\alpha}(fC)_{\alpha\delta} = 0;$

2. $\sum_{\delta}(fC)_{\alpha\delta} = 0; \sum_{\beta}(AB)_{\beta\gamma} = 0; \sum_{\gamma}(AB)_{\beta\gamma} = 0; \sum_{\beta}(AC)_{\beta\delta} = 0; \sum_{\delta}(AC)_{\beta\delta} = 0;$

$\sum_{\gamma}(BC)_{\gamma\delta} = 0; \sum_{\delta}(BC)_{\gamma\delta} = 0;$

3. $\sum_{\alpha}(fAB)_{\alpha\beta\gamma} = \sum_{\beta}(fAB)_{\alpha\beta\gamma} = \sum_{\gamma}(fAB)_{\alpha\beta\gamma} = 0;$

$\sum_{\alpha}(fAC)_{\alpha\beta\delta} = \sum_{\beta}(fAC)_{\alpha\beta\delta} = \sum_{\delta}(fAC)_{\alpha\beta\delta} = 0;$

$\sum_{\alpha}(fBC)_{\alpha\gamma\delta} = \sum_{\gamma}(fBC)_{\alpha\gamma\delta} = \sum_{\delta}(fBC)_{\alpha\gamma\delta} = 0;$

$\sum_{\beta}(ABC)_{\beta\gamma\delta} = \sum_{\gamma}(ABC)_{\beta\gamma\delta} = \sum_{\delta}(ABC)_{\beta\gamma\delta} = 0;$

4. $\sum_{\alpha}(fABC)_{\alpha\beta\gamma\delta} = \sum_{\beta}(fABC)_{\alpha\beta\gamma\delta} = \sum_{\gamma}(fABC)_{\alpha\beta\gamma\delta} = \sum_{\delta}(fABC)_{\alpha\beta\gamma\delta} = 0;$

5. $\sum_{\alpha}\sum_{\beta}\sum_{\gamma}\sum_{\delta}\sum_{i}\varepsilon_{\alpha\beta\gamma\delta i} = 0.$

In addition to these conditions, restrictions are imposed on the random balance:

1) all $\varepsilon_{\alpha\beta\gamma\delta i}$ are mutually independent;

2) $M[d_{abcq}{}^2] = \sigma^2 \ \ M[\varepsilon_{\alpha\beta\gamma\delta i}{}^2] = \sigma^2$;

3) random variables $\varepsilon_{\alpha\beta\gamma\delta i}$ are distributed according to the normal law.

Regarding the type of deviations f_{α}, $A_{\beta}, B_{\gamma}, C_{\delta}$ we note the following:

1) f_{α} — is a random variable because it reflects the effect of a priori uncertain levels of the parameter of the control object *H;*

2) $A_{\beta}, B_{\gamma}, C_{\delta}$ — are parameters by virtue of the metrological values of the control indicators.

Since due to the randomness of the levels of the parameter *H* the resulting model (1) is not exclusively parametric, but can be attributed to mixed models.

TRANSITION TO THE MOST COMMON FACTOR FUEL MODEL

It is a well-known fact that the multifactor model (1) requires for its study a sample size of $k^4 \cdot r$, where k — is the number of levels for each of the factors, and r — is the number of multiple observations for all possible combinations of levels of the influencing factors. In order for the results obtained during the variance analysis to be statistically significant, the value must satisfy the conditions $k > 4$, $r > 1$.

Thus, at $k=5$, and $r=2$ the minimum volume of the number of four-dimensional observations of the control index *F* should be $5^4 \cdot 2 = 1250$ values. It turns out that it is very difficult, and sometimes almost impossible to provide such a large number of non-standard samples of coffee beans while maintaining the complete uniformity of the measurement experiment.

In order to avoid this complexity, we simplify model (1), leaving only the main deviations f_α, $A_\beta, B_\gamma, C_\delta$ and the deviations caused by pairwise interactions $(fA)_{\alpha\beta}, (fB)_{\alpha\gamma}, (fC)_{\alpha\delta}$. The resulting model will contain an increased residual $\psi_{\alpha\beta\gamma\delta\nu}$, which also includes a random residual $\varepsilon_{\alpha\beta\gamma\delta i}$, and deviations caused by the action of three: $(fAB)_{\alpha\beta\gamma}, (fAC)_{\alpha\beta\delta}, (fBC)_{\alpha\gamma\delta}, (ABC)_{\beta\gamma\delta}$ and four — $(fABC)_{\alpha\beta\gamma\delta}$, influencing factors, as well as pairwise deviations $(AB)_{\beta\gamma}$, $(AC)_{\beta\delta}$, $(BC)_{\gamma\delta}$:

$$F_{\alpha\beta\gamma\delta\nu} = \overline{F} + f_\alpha + A_\beta + B_\gamma + C_\delta + (fA)_{\alpha\beta}, (fB)_{\alpha\gamma}, (fC)_{\alpha\delta} + \psi_{\alpha\beta\gamma\delta\nu} \quad (2)$$

In order to reduce the complexity of the model (1) it can be reduced to three two-factor simplified cross-classification models:

$$F_{\alpha\beta\nu} = \overline{F} + f_\alpha + A_\beta + (fA)_{\alpha\beta} + \psi(A)_{\alpha\beta\nu}; \;\; \gamma = \delta = i = \nu; \quad (3)$$

$$F_{\alpha\gamma\nu} = \overline{F} + f_\alpha + B_\gamma + (fB)_{\alpha\gamma} + \psi(B)_{\alpha\gamma\nu}; \;\; \alpha = \delta = i = \nu; \quad (4)$$

$$F_{\alpha\delta\nu} = \overline{F} + f_\alpha + C_\delta + (fC)_{\alpha\delta} + \psi_{(C)\alpha\delta\nu}; \;\; \beta = \gamma = i = \nu. \quad (5)$$

where ν — is the number of multiple measurements of the *F* indicator in the table cell with the original model data (3), (4) i (5); $\psi_{(A)\alpha\beta\nu}$, $\psi_{(B)\alpha\gamma\nu}$, $\psi_{(C)\alpha\delta\nu}$ — random residues due to three factors: grinding time, geometric grain size, grain moisture, respectively.

A comparison of the residuals of the presented models with the residuals of models (1) and (2) shows that simplifying the model obviously reduces its accuracy, because these residuals are more than $\psi_{\alpha\beta\gamma\delta\nu}$, a $\psi_{\alpha\beta\gamma\delta\nu} > \varepsilon_{\alpha\beta\gamma\delta i}$.

Let's analyze the condition that allows us to synthesize model (2) on the basis of models (3), (4), and (5). For each of these models, there is the same main deviation s, additional deviations $A_\beta, B_\gamma, C_\delta$, and deviations due to the effects of pairwise interaction.

The first and most important condition is to ensure that the standard deviations in models (3), (4), and (5) are equal to each other. To do this, the number of groups of observations of the control indicator *F* must be the same for all models, which corresponds to the same number of rows in the original data table. At the same time, the number of values of the control indicator *F* in each group should be the same *b*, and the value of $F_{\alpha\beta\gamma\delta}$ in the middle of each group should remain unchanged (where *N* — is the number of measurements) for any of the models (3), (4), and (5). The grouping of $F_{\alpha\beta\gamma\delta i}$ values should be carried out according to the specified groups of values of the parameter *H*.

It should be borne in mind that the method of forming columns (subgroups) in the table of the established source data should be determined by the established additional influencing factor, and provide a single procedure for selecting $F_{\alpha\beta\gamma\delta}$ values for each of the subgroups of the group with a fixed number $\alpha, \alpha = \overline{1, b}$. The number of subgroups in each of the groups must also be the same *m*, for each of the models (3), (4) and (5). Each of the subgroups of any of the groups (cells of the source data table) will have the same $g = N/mb$ — the number of observations.

In order to model additional factor injection (based on all vortex data) on display *F*, you need to carry out the following operations:

– rank the intragroup values of the control indicator F_p, you need to carry out the following operations;

– break the ranked (for all *b* groups) series of values of the F_p indicator into *m* subgroups;

– in each of the subgroups, select the *g* values of the information index *F* corresponding to the *g* values of the F_p indicator and enter them into the source data cell.

The resulting $b \times m$ table of observation results of control measure *F* values with *g* multiple observations in each of $b \times m$ cells can now be used for variance analysis of any of the models (3), (4) and (5) cross classifications corresponding to a given additional factor influencing F_p, $p = \overline{1,3}$.

We will introduce the designation of these three factors through F_s, F_t, F_u. In general, we will denote any of these factors through F_p. Therefore, any of the models (3), (4), and (5) can be represented in the form:

$$F_{\alpha z \nu} = \overline{F} + f_\alpha + p_z + (fp)_{\alpha z} + \psi_{\alpha z \nu}, \qquad (6)$$

where $\psi_{\alpha z \nu}$ — is a random residue.

The complete decomposition of the sum of the squares of deviations of the values $F_{\alpha z \nu}$ from $\overline{F}$, under the initial conditions and constraints of the model (1), will have the following form:

$$W = W_f + W_p + W_{fp} + W_{\psi p}. \qquad (7)$$

The results of the variance analysis of the model (6) are presented in Table 1, where $F_{\alpha z \nu} = \overline{F}_\alpha, \overline{F}_z, \overline{F}_{\alpha z}$ — are the average values in rows, columns and in cells.

Table 1. Results of dispersion analysis

Source of variability	**Number of degrees of freedom *k***	**Sum of squares**
The main factor *H*	$b-1$	$W_f = gp\sum_{\alpha=1}^{b}(\overline{F}_\alpha - \overline{F})^2$
Additional factor F_p	$m-1$	$W_p = gb\sum_{z=1}^{m}(\overline{F}_z - \overline{F})^2$
Interaction between *H* and F_p	$(b-1)(m-1)$	$W_{fp} = g\sum_{\alpha=1}^{b}\sum_{z=1}^{m}(\overline{F}_{\alpha z} - \overline{F}_\alpha - \overline{F}_z + \overline{F})^2$
Remainder (in the middle of the cell)	$bm\,(g-1)$	$W_{\psi p} = \sum_{\alpha=1}^{b}\sum_{z=1}^{m}\sum_{\nu=1}^{g}(F_{\alpha z \nu} - \overline{F}_{\alpha z})^2$
General	$N-1$	$W = \sum_{\alpha=1}^{b}\sum_{z=1}^{m}\sum_{\nu=1}^{g}(F_{\alpha z \nu} - \overline{F})^2$

Unfortunately, for models (3), (4) that (5), the same $\overline{F}$ and the sums of W and W_f is represented by the sum of squares from $F_{\alpha\beta\gamma\delta}$ from F model (2) as the union of the sum of (7), $p = \overline{1,3}$ with the fictitious residual sum W_ψ^*, which characterizes the influence of factors not taken into account in the model:

$$W = W_f + W_s + W_t + W_u + W_{fs} + W_{ft} + W_{fu} + W_\psi^* \ . \quad (8)$$

The resulting ratio (8) makes it possible to simplify the previous model (2) and represent it in the following form:

$$F_{\alpha\beta\gamma\delta\nu} = \overline{F} + f_\alpha + A_\beta + B_\gamma + C_\delta + (fA)_{\alpha\beta} + (fB)_{\alpha\gamma} + (fC)_{\alpha\delta} + \psi^*_{\alpha\beta\gamma\delta\nu}, \quad (9)$$

where $\psi^*_{\alpha\beta\gamma\delta\nu}$ — is a random residue due to the action of three factors: grinding time, geometric grain size, grain moisture. The values of the sums of the right part of the expression (8), in addition to W_ψ^* are calculated according to the equations of the sums of the squares of the tables of the results of the variance analysis of the models (3), (4) and (5), similar to Table 2, replacing the F_p factor with a specific additional factor — F_s, F_t, F_u. The sum of W_ψ^* can be calculated from any of the equations:

$$\begin{cases} W_\psi^* = W_{\psi s} - W_t - W_u - W_{f\,t} - W_{f\,u}\ ; \\ W_\psi^* = W_{\psi t} - W_s - W_u - W_{f\,s} - W_{f\,u}\ ; \\ W_\psi^* = W_{\psi u} - W_t - W_s - W_{f\,t} - W_{f\,s}\ ; \end{cases} \quad (10)$$

$$W_\psi^* = W_{\psi s} + W_{\psi t} + W_{\psi u} + 2W_f + 2W\ . \quad (11)$$

Table 2 presents the results of the variance analysis of the simplified model (9), the random residue of which will be fictitious (determines the sum of W_ψ^* in equation (8)).

Table 2. Results of the variance analysis with the model (8)

Source of variability	**Source of variability**	**Sum of the squares of deviations**
The main factor H	$k_f = b - 1$	$\overline{W}_f = W_f / k_f$
Factor F_s	$k_s = m - 1$	$\overline{W}_s = W_s / k_s$
Factor F_t	$k_t = m - 1$	$\overline{W}_t = W_t / k_t$
Factor F_u	$k_u = m - 1$	$\overline{W}_u = W_u / k_u$
Interaction HF_s	$k_{fs} = (b-1)\cdot(m-1)$	$\overline{W}_{fs} = W_{fs} / k_{fs}$
Interaction HF_t	$k_{ft} = (b-1)\cdot(m-1)$	$\overline{W}_{ft} = W_{ft} / k_{ft}$
Interaction HF_u	$k_{fu} = (b-1)\cdot(m-1)$	$\overline{W}_{fu} = W_{fu} / k_{fu}$
Remainder	$k_\psi = N - b((3m-2)$	$\overline{W}_\psi = W_\psi^* / k_\psi$
General	$k = N$-1	$\overline{W} = W / k$

We will conduct a study of the simplified model (9). This model, occupies an intermediate position between model (1) and models (3), (4) and (5). From (10) it follows that the residual sum of the simplified model (11) is less than the residual sums of the models (3), (4) and (5). This indicates the increased accuracy of this model compared to the cross-classification models (3), (4), and (5). Based on the results of Table. 2, we conclude that the number of degrees of freedom of the residual sum W_{ψ}^{*} decreases with an increase in the number of *b* groups (by the level of parameter *H*) and the number of *m* subgroups (by the level of additional factors F_s, F_t, F_u).

Let's write k_{ψ} from Table 2 in the following form:

$$k_{\psi} = N - b \cdot m\left(3 - \frac{b}{m}\right). \qquad (12)$$

From the resulting expression (12) it follows that the number of degrees of freedom is the greater, at $b \cdot m = \text{const}$, the greater the ratio b/m.

This finding makes it possible to further plot the number of groups and subgroups in the tables of the original data of the models (3), (4), and (5). In this case, the number of *b* groups (subranges of measurement of the control parameter *H* (homogeneity) is desirable to increase, and the number of subgroups should be reduced, reducing *m* to a minimum. For example, take *m* = 2. This will increase the number of degrees of freedom of the residual sum W_{ψ}^{*}. Thus, it is obvious that the main advantage of the simplified model is the possibility of simultaneous testing of the H_0, hypothesis, that is, when the influence of factors *H*, F_s, F_t, F_u, on the information measure of control *F* is absent. In this way you can write H_0: $f_p = \ldots = f_b = 0$.

This is the main hypothesis and its components have the following form:

$$H_0^s : s_p = \ldots = s_m = 0; \quad H_0^{fs} : fs_p = \ldots = fs_{bm} = 0;$$

$$H_0^t : t_p = \ldots = t_m = 0; \quad H_0^{ft} : ft_p = \ldots = ft_{bm} = 0;$$

$$H_0^u : u_p = \ldots = u_m = 0; \quad H_0^{fu} : fu_p = \ldots = fu_{bm} = 0.$$

The test of the hypotheses presented is done in relation to the corresponding mean squares $(\overline{W}_f, \overline{W}_s, \overline{W}_t, \overline{W}_u, \ldots, \overline{W}_{fu})$ to the mean residual square $\overline{W}_{\psi}$ followed by comparison of the obtained *F* — statistics with the corresponding percentage points for *F* — distributions.

This advantage of the simplified model (9) makes it possible to estimate the amount of expected information about the levels of parameter *H* for the information measure *F* when considering the levels of both influencing factors and their interactions:

$$I = \log\sqrt{1 + \left(\frac{\sigma_F}{\sigma_{\Delta F}}\right)^2},$$

where $\sigma_F^2 = \overline{W}_f$;

$$\sigma_F^2 = \overline{W}_f = \frac{1}{(n-1)} \sum_{i=1}^{n} (F_i - \overline{F})^2 ,$$

where $\overline{F} = \frac{1}{n} \sum_{i=1}^{n} F_i$; $\sigma_{\Delta F}^2$ — is a function of the sum of the squares of deviations $(\overline{W}_f, \overline{W}_s, \overline{W}_t, \overline{W}_u, ..., \overline{W}_{fu})$ in Table 1 and Table. 2.

Table 3 presents the equation for calculating $\sigma_{\Delta F}^2$ for the simplified model (9) with different combinations of factors affecting the information indicator *F*.

Table 3. Calculated Ratios for Parameter $\sigma_{\Delta F}^2$

Factorial influences taken into account	$\sigma_{\Delta F}^2$
Additional factor F_s	$(W_t + W_u + W_{fs} + W_{ft} + W_{fu} + W_\psi^*) / (k_t + k_u + k_{fs} + k_{ft} + k_{fu} + k_\psi)$
Additional factor F_t	$(W_s + W_u + W_{fs} + W_{ft} + W_{fu} + W_\psi^*) / (k_s + k_u + k_{fs} + k_{ft} + k_{fu} + k_\psi)$
Additional factor F_u	$(W_s + W_t + W_{fs} + W_{ft} + W_{fu} + W_\psi^*) / (k_s + k_t + k_{fs} + k_{ft} + k_{fu} + k_\psi)$
F_s , F_t	$(W_u + W_{fs} + W_{ft} + W_{fu} + W_\psi^*) / (k_u + k_{fs} + k_{ft} + k_{fu} + k_\psi)$
F_s , F_u	$(W_t + W_{fs} + W_{ft} + W_{fu} + W_\psi^*) / (k_t + k_{fs} + k_{ft} + k_{fu} + k_\psi)$
F_t , F_u	$(W_s + W_{fs} + W_{ft} + W_{fu} + W_\psi^*) / (k_s + k_{fs} + k_{ft} + k_{fu} + k_\psi)$
F_s , F_t , F_u	$(W_{fs} + W_{ft} + W_{fu} + W_\psi^*) / (k_{fs} + k_{ft} + k_{fu} + k_\psi)$
F_s , F_t , F_u , HF_s	$(W_{ft} + W_{fu} + W_\psi^*) / (k_{ft} + k_{fu} + k_\psi)$
F_s , F_t , F_u , HF_t	$(W_{fs} + W_{fu} + W_\psi^*) / (k_{fs} + k_{fu} + k_\psi)$
F_s , F_t , F_u , HF_u ,	$(W_{fs} + W_{ft} + W_\psi^*) / (k_{fs} + k_{ft} + k_\psi)$
F_s , F_t , F_u , HF_s , HF_t , HF_u	$\overline{W}_\psi$

If we do not take into account all the influencing factors, and this corresponds to obtaining information under multifactorial influence, then

$$\sigma_{\Delta F}^2 = \frac{W_s + W_t + W_u + W_{fs} + W_{ft} + W_{fu} + W_\psi^*}{N - b - 2} .$$

In order to perform the verification of statistical findings, we present model (9) as a one-factor model of one-sided classification, i.e., when the influence of additional factors is determined solely by the magnitude of the random residue $\psi_{\alpha i}$:

$$F_{\alpha i} = \overline{F} + f_i + \psi_{\alpha i} , , \quad (13)$$

where $\alpha = \overline{1, b}$; $i = \overline{1, n}$; $n = N/b$.

In this case, the magnitude of *n* multifactorial observations in each of the *b* groups is the same.

Model (13), due to the uncertainty of the levels of the control parameter *H*, refers to the variance component models.

We will use the formal analysis of the model (13) and write the expression of the full sum of the squares of deviations *W* through the sum of the two terms

$$W_f = n\sum_{i=1}^{b}(\overline{F}_\alpha - \overline{F})^2 ,$$

where $\overline{F}_\alpha$ — group mean values of the performance trait; $\overline{F}$ — is the overall average; *n* — is the number of units of the population in each group.

Residual variance (random) is the sum of the group sums of the squares of deviations of all variant of the resultant trait in groups from the mean values of the trait in them:

$$W_\psi = \sum_{\alpha=1}^{b}\sum_{i=1}^{g}(F_{\alpha i} - \overline{F}_\alpha)^2 .$$

Consider now the ratio of the mean squares for the recorded sums W_f and W_ψ:

$$F = \frac{W_f/(b-1)}{W_\psi/(N-b)} .$$

F-statistics can be used to test one of two hypotheses:

$$H_0 : M[\overline{F}_1] = ... = M[\overline{F}_b] \quad H_1 : M[\overline{F}_1] \neq ... \neq M[\overline{F}_b] .$$

A rule follows from the theory: if statistical conclusions indicate the validity of the main hypothesis H_0, then the parameter *H* does not affect the change in the control indicator *F*. That is, the indicator *F* does not carry information about the change in the levels of the control parameter *H*, and if the hypothesis H_1, is valid, then the indicator *F* is informative in relation to the control parameter *H*. Decisions μ_0 (valid hypothesis H_0) and μ_1 (valid hypothesis H_1) are accepted, comparing the *F* statistic with the critical value F_K.

The conditional densities of the probability distribution of *F*-statistics are a linearly transformed random variable with central $F_{b-1,N-b}$, the distribution:

$$f(F/H_0) \approx F_{b-1,N-b} ,$$

$$f(F/H_1) \approx \left(1 + g\frac{\sigma_f^2}{\sigma^2}\right) \cdot F_{b-1,N-b} , \qquad (14)$$

where the variances σ_f^2 and σ^2 refer, respectively, to the random deviations of f_α i $\psi_{\alpha i}$ in the model (13).

The variance of σ_f^2 can be represented as the sum:

$$\sigma_f^2 = \sigma_H^2 + \sigma_r^2 ,$$

where σ_H^2 — the variance of the control parameter *H*, which is due to the uncertainty of its values in the measurement range D_H; σ_r^2 — the variance of the measurement result of the values of the control parameter *H*, which is due to the uncertainty of the reproduction of the values of *H* by the technical means of control (in fact, this is the variance of the measurement result).

Testing hypotheses about the significance of the effects of factors and their interactions is carried out using the Fisher criterion. To do this, calculate the ratio of the corresponding mean squares to the remaining middle square. The obtained values are compared with the F_K values found in the *F*-distribution tables for the accepted significance level $\alpha = 0{,}05$, or higher (0,01 – 0,001) and the number of degrees of freedom.

Hence, if the control measure *F* is sensitive to a change in the level of the control parameter *H* over the entire range of its measurements, then the *F*-statistic will be characterized by the distribution density (14) for all measured levels of the indicator *H*.

Now consider the problem of one-sided testing of the parameter *H* within the implementation of the alternative hypothesis H_1. This hypothesis must be represented as a complex one, which will check the correspondence between the real value *f* of the control parameter *H* and the control value f_0:

$$H_1^0 : \; f \le f_0 \text{ (parameter } H \text{ is normal);}$$

$$H_1^1 : \; f > f_0 \text{ (parameter } H \text{ is not normal).}$$

Such situations must correspond to the choice of one of two solutions:

$$\mu_0 : \; F \le F_K ,$$

where F — calculated value of the *F*-criterion; F_K — tabular value of the *F*-criterion. Then the influence of the factor on the control parameter is not proven, but the absence of influence of the factor is not proven.

If:

$$\mu_1 : \; F > F_K ,$$

then statistical observation proves with a given probability the influence of the factor on the control parameter.

Enter the following designations: Ω_1 and Ω_2 — probabilities of errors of the first and second kind:

$$\Omega_1 = P(\mu_1 / f \le f_0) ; \quad \Omega_2 = P(\mu_0 / f > f_0) .$$

Then the probabilities of choosing solutions μ_0, μ_1 are respectively determined by the expressions:

$$P[F > F_K] \ge 1 - \Omega_2 , \tag{15}$$

$$P[F > F_K] \le \Omega_1 . \tag{16}$$

The fragment *F* has distribution (14), then from the expressions (15) and (16) it flows

$$\begin{cases} P[F_{b-1,N-b} > F_K (1 + n\nu_1^2)^{-1}] \ge 1 - \Omega_2 ; \\ P[F_{b-1,N-b} > F_K (1 + n\nu_0^2)^{-1}] > \Omega_1 , \end{cases} \tag{17}$$

where

$$\begin{cases} \nu_1^2 = \dfrac{\sigma_H^2 + \sigma_r^2}{\sigma^2}; \\ \nu_0^2 = \dfrac{\sigma_r^2}{\sigma^2}. \end{cases} \quad (18)$$

Considering the system of equations (17), one can find the value of F_K that satisfies both of these expressions. To do this, it is necessary to ensure that the condition is fulfilled

$$\frac{F_{b-1,N-b,1-\Omega_1}}{F_{b-1,N-b,\Omega_2}} \le \frac{1+n\nu_1^2}{1+n\nu_0^2}. \quad (19)$$

The numerator and denominator of the left side of the inequality (19) are $(1-\Omega_1)$ and Ω_2 — are the percentage points of the central F — the distribution with $(b-1)$ and $(N-b)$ degrees of freedom.

It can be shown now that the expression that (19) corresponds to the expression:

$$\frac{\chi_{b-1,1-\Omega_1}^2}{\chi_{b-1,\Omega_2}^2} \le \frac{\nu_1^2}{\nu_0^2}, \quad (20)$$

where $\chi_{b-1,1-\Omega_1}^2$, χ_{b-1,Ω_2}^2 — percentage points of the central χ^2 distribution with $(b-1)$ degrees of freedom.

From formula (20) it follows that given number of levels b of control parameter H (groups of results of observations of parameter F of model (13)) and given ratio $\nu_1^2/\nu_0^2 = \mathrm{A}$ it is possible to estimate the reliability of decision-making μ_0 and μ_1

$$\mathrm{A} = 1 - \frac{\Omega_1 + \Omega_2}{2},$$

fixed, for example, the value of Ω_1 (Naiman – Pearson criterion) and calculate Ω_2, as the interval:

$$\Omega_2 = \frac{1}{2^{\frac{b-1}{2}} \cdot \Gamma\left(\dfrac{b-1}{2}\right)} \int_0^{z_{\min}} \chi^{\left(\frac{b-1}{2}-1\right)} \cdot e^{-\frac{\chi}{2}} \partial\chi,$$

where $z_{\min} = \chi_{b-1,1-\Omega_1}^2 / \mathrm{A}$; $\Gamma\left(\dfrac{b-1}{2}\right)$ — gamma function.

The relation A is determined from the formula (18):

$$\mathrm{A} = 1 + \frac{\sigma_r^2}{\sigma^2}.$$

The results obtained already make it possible to move on to the practical use of the simplified model. At the same time, it must be remembered that in the practice of variance analysis, there may be cases when the number of replicates obtained for each combination of levels of the factors under study is different and equal m_{ij} ; m_{ij} numbers can have any value, including zero, but each row and column must have at least one, and some have two non-zero values. In this case, there are possible situations in which it is impossible to obtain unbiased estimates for all parameters of the model. The need to introduce a system of weight coefficients into the calculated ratios complicates the calculations, the analysis of such plans involves, as a rule, the use of a computer.

SUMMARIES

When executing the partition, the following results are obtained:

– to assess the homogeneity of coffee grinding, a mathematical model of the influence of four factors on the result of measuring the control indicator has been developed;

– a simplified model of cross-classifications was proposed for further use and investigated, which took into account the effects of simultaneous interaction of four factors (grinding time, geometric grain size, grain moisture, speed of rotation of the motor shaft) on the result of measurement of the unit control indicator (uniformity of coffee grinding);

– obtained equations that allow to evaluate the reliability of statistical conclusions in relation to the informational significance of control indicators for the proposed simplified model of cross-classification;

– analytical ratios are obtained, which allow to estimate the amount of information on each of the control indicators under the factor influence on the proposed linear function of the transformation of these indicators.

– the advantages of the chosen approach are that it makes it possible to assess the validity of multiparameter control results of three or more levels of the control indicator and to select the most informative subsets of these values.

REFERENCES

1. V.P. Misyats, M.M. Rubanka, and S.A. Demishonkova, “System of adaptive control of the drive of automatic coffee machines,” *Bulletin of the Khmelnytskyi National University*, no. 1, pp. 151–159, 2021 (293). doi: 10.31891/2307-5732-2021-293-1-151-159.
2. E. Uman et al., “The effect of bean origin and temperature on grinding roasted coffee,” *Sci. Rep.*, vol. 6, 24483, 2016. doi: https://doi.org/10.1038/srep24483.
3. Nancy Cordoba, Laura Pataquiva, Coralia Osorio, Fabian Leonardo Moreno Moreno, and Ruth Yolanda Ruiz, “Effect of grinding, extraction time and type of coffee on the physicochemical and flavor characteristics of cold brew coffee,” *Sci. Rep.*, 9, 8440, 2019. doi: https://doi.org/10.1038/s41598-019-44886-w.
4. Jonathan D. Walston, Daniel L. Short, and M. Affan Badar, “An Experimental Design on Coffee Extraction Factors Impacting the Measurable Percent of Total Dissolved Solids in Solution,” *Asia-Pacific Journal of Management Research and Innovation*, pp. 1–11, 2023. doi: 10.1177/2319510X221136690.
5. Anderson G. Costa, Eudócio R.O. da Silva, Murilo M. de Barros, and Jonatthan A. Fagundes, “Estimation of percentage of impurities in coffee using a computer vision system,” *Brazilian Journal of Agricultural and Environmental Engineering*, vol. 26, no. 2, pp. 142–148, 2022. doi: http://dx.doi.org/10.1590/1807-1929/agriambi.v26n2p142-148.
6. G. Angeloni et al., “Test of an innovative method to prepare coffee powder puck, improving espresso extraction reliability,” *Eur. Food Res. Technol.*, 248, pp. 163–170, 2022. doi: https://doi.org/10.1007/s00217-021-03868-x.

7. Ihor Hryhorenko, Elena Tverytnykova, Svitlana Hryhorenko, and Viktoria Krylova, "Temperature sensor research as a part of a microprocessor system by statistical analysis method," *2022 IEEE 3rd KhPI Week on Advanced Technology (KhPIWeek), 2022, Kharkiv, Ukraine*, pp. 102–107. doi: 10.1109/ХПИНеделя57572.2022.9916478.
8. I. Hryhorenko, S. Kondrashov, and O. Opryshkin, "Formation of test impacts for the first level of the information and measurement system," *Bulletin of the National Technical University «KhPI». Series: New solutions in modern technology*, no. 1 (15), pp. 19–26, 2023. doi: https://doi.org/10.20998/2413-4295.2023.01.03.

Received 20.07.2023

INFORMATION ON THE ARTICLE

Ihor V. Hryhorenko, ORCID: 0000-0002-4905-3053, National Technical University "Kharkiv Polytechnic Institute", Ukraine, e-mail: grigmaestro@gmail.com

Serhii I. Kondrashov, ORCID: 0000-0002-5191-8562, National Technical University "Kharkiv Polytechnic Institute", Ukraine, e-mail: Serhii.Kondrashov@khpi.edu.ua

Svitlana M. Hryhorenko, ORCID: 0000-0003-0150-4844, National Technical University "Kharkiv Polytechnic Institute", Ukraine, e-mail: sngloba@gmail.com

Oleksandr S. Opryshkin, ORCID: 0009-0008-7094-5129, National Technical University "Kharkiv Polytechnic Institute", Ukraine, e-mail: Aleksandr.Opryshkin@cit.khpi.edu.ua

ДОСЛІДЖЕННЯ ФАКТОРНОГО ВПЛИВУ НА ОДНОРІДНІСТЬ ПОМЕЛУ ЗЕРНА КАВИ МЕТОДАМИ СТАТИСТИЧНОГО АНАЛІЗУ / І.В. Григоренко, С.І. Кондрашов, С.М. Григоренко, О.С. Опришкін

Анотація. Для оцінювання впливу кожного із факторів, які впливають на якість та однорідність помелу зерна кави, і порівняння впливу цих факторів варто встановити кількісний показник цього впливу. Для цього використано дисперсійний аналіз як метод організації вибіркових даних відповідно до можливих джерел розсіювання. Обраний метод дозволив розкласти загальне розсіювання на складові, зумовлені впливом рівнів факторів. Факторами, що впливають на однорідність помелу, обрано: час помелу, геометричні розміри зерна, вологість зерна, швидкість обертання валу двигуна. Проведено обґрунтування та оцінювання достовірності статистичних висновків про інформаційну значущість показників, що впливають на однорідність помелу кави для забезпечення максимально високої вірогідності отриманого результату.

Ключові слова: дисперсійний аналіз, однорідність помелу, факторний вплив, модель, показник контролю, зерно кави.