

УДК 519.85:519.233.4

DOI: 10.30838/J.BPSACEA.2312.241023.74.995

КОНТРОЛЬ ІНФОРМАЦІЇ СТАТИСТИЧНИХ ДАНИХ НА ОСНОВІ ДИСПЕРСІЙНОГО АНАЛІЗУ

ЄРШОВА Н. М., *докт. техн. наук, проф.*

Кафедра комп'ютерних наук, інформаційних технологій та прикладної математики, Придніпровська державна академія будівництва та архітектури, вул. Архітектора Олега Петрова, 24-а, 49005, Дніпро, Україна, тел. +38 (095) 918-01-02, e-mail: nersova107@gmail.com, ORCID ID: 0000-0003-0198-0883

Анотація. Постановка проблеми. Точність і достовірність вихідного статистичного матеріалу – найважливіше завдання статистичного спостереження. Навіть за гарної організації спостережень або проведення експерименту з тих чи інших причин можуть виявитися неточності, похибки, тобто помилки реєстрації. Аналіз існуючого способу контролю інформації показує, що він занадто складний, використовує різні критерії залежно від обсягу вибірки, критичні значення критеріїв обираються з таблиць, розрахунок критеріїв виконується за формулами та статистичними функціями майстра функцій Excel. **Мета статті** – розроблення способу контролю інформації на основі дисперсійного аналізу даних спостережень. **Результати.** Створено єдину базу даних об'єктів-аналогів. Для коректного створення об'єднаної бази даних потрібно бути впевненим, що об'єднані дані належать до однієї генеральної сукупності. Складність проведення дисперсійного аналізу залежить від обсягу вибірок. Якщо об'єднуються кілька вибірок одного обсягу, легко перевірити їх однорідність за допомогою інструменту «Однофакторний дисперсійний аналіз», у вихідній інформації якого видаються розрахункове і критичне значення F -критерію Фішера. На прикладах показано, що за допомогою дисперсійного аналізу можна встановити не тільки однорідність вибірок, а і причину її порушення. Створено спосіб дисперсійного аналізу однорідності вибірок різного обсягу за допомогою інструменту «Описова статистика» пакета аналізу. Виконані перевірки адекватності моделі регресії активного експерименту. План експерименту визначає точність моделі регресії. У факторному просторі обирається деяка точка і розглядається безліч точок її околиці. У цій околиці проводиться експеримент, на основі якого будується перша модель. Головна вимога до моделі – здатність передбачати напрямки подальших дослідів із необхідною точністю. І точність цього передбачення в усіх напрямках пошуку повинна бути однаковою. Модель, що задовольняє таку вимогу, називається адекватною. Перевірка здійсненості цієї умови називається аналізом адекватності моделі. У процесі проведення експерименту необхідно переконатися, що вимірювані значення відгуку належать до однієї генеральної сукупності і технологічний процес не вимагає регулювання. Для цього здійснюються паралельні досліді. Після проведення чотирьох дослідів за матрицею планування експерименту необхідно переконатися в однорідності отриманих вибірок відгуку та можливості відтворюваності дослідів. **Висновки.** Виконані розрахунки доказують, що: існуючий спосіб контролю інформації занадто складний та має суттєві недоліки – використовуються різні критерії залежно від обсягу вибірки, критичні значення критеріїв обираються з таблиць, розрахунок критеріїв виконується за формулами та статистичними функціями майстра функцій Excel; спосіб контролю інформації на основі дисперсійного аналізу не має цих недоліків і є універсальним, тому що існує один критерій для малих і великих вибірок, розрахункове і критичне значення F -критерію видаються у вихідній інформації інструменту «Однофакторний дисперсійний аналіз», його можна використовувати для створення єдиної бази даних об'єктів-аналогів та перевірки адекватності моделі регресії активного експерименту.

Ключові слова: експеримент; інформація; спосіб контролю; критерій; дисперсійний аналіз; адекватність моделі регресії

CONTROL INFORMATION OF STATISTICAL DATA BASED ON DISPERSION ANALYSIS

YERSHOVA N.M., *Dr. Sc. (Tech.), Prof.*

Department of Computer Science, Information Technologies and Applied Mathematics, Prydniprovsk State Academy of Civil Engineering and Architecture, 24-a, Architect Oleh Petrov St., Dnipro, 49005, Ukraine, tel. +38 (095) 918-01-02, e-mail: nersova107@gmail.com, ORCID ID: 0000-0003-0198-0883

Abstract. Problem statement. The accuracy and reliability of the source statistical material is the most important

task of statistical observation. Even with a good organization of observations or conducting an experiment, for one reason or another, inaccuracies, errors, that is, registration errors, may appear. Analysis of the existing method of information control shows that it is too complicated, uses different criteria depending on the sample size, critical criteria values are selected from tables, calculation is performed using formulas and statistical functions of the Excel function master. **The purpose of the article is** to develop a method of information control based on dispersion analysis of observation data. **Results.** A single database of analogue objects is created. To correctly create a combined database, you need to be sure that the combined data belong to the same general population. The complexity of dispersion analysis depends on the size of the samples. If several samples of the same volume are combined, then it is easy to check their homogeneity with the help of the “One-factor dispersion analysis” tool, in the output of which the calculated and critical value F -Fisher's test - are given. The examples show that with the help of dispersion analysis it is possible not only to establish the homogeneity of samples, but also the reason for its violation. A method of dispersion analysis of the homogeneity of samples of different volumes was created using the “Descriptive statistics” tool of the analysis package. The adequacy check of the regression model of the active experiment was performed. The experiment plan determines the accuracy of the regression model. Some point is selected in the factor space and many points in its neighborhood are considered. An experiment is being conducted in this neighborhood, on the basis of which the first model is being built. The main requirement for the model is the ability to predict the direction of further experiments with the required accuracy. And the accuracy of this prediction in all search directions should be the same. A model that satisfies this requirement is called adequate. Checking the feasibility of this condition is called model adequacy analysis. In the process of conducting the experiment, it is necessary to make sure that the measured response values belong to the same general population and the technological process does not require regulation. For this purpose, is carried parallel experiments. After conducting 4 experiments according to the experiment planning matrix, it is necessary to make sure of the received response samples' homogeneity and the reproducibility of the experiments. **Conclusions.** The performed calculations prove that: the existing method of information control is too complex and has significant disadvantages – different criteria are used depending on the sample size, critical values of the criteria are selected from tables, the calculation is performed using formulas and statistical functions of the Excel function master; the method of information control based on dispersion analysis does not have these disadvantages and is universal, because there is one criterion for small and large samples, the calculated and critical value of F -criterion are given in the initial information of the tool “One-factor dispersion analysis”, it can be used to create a single database analogue objects and to check the adequacy of the regression model at the active experiment.

Keywords: *experiment; information; method of control; criterion; dispersion analysis; regression model adequacy*

Постановка проблеми. Перш ніж проводити кореляційно-регресійний аналіз даних статистичних спостережень, необхідно перевірити вибірку на наявність викидів (занадто великі і занадто малі значення ознаки, що помірно відрізняються в ряді вимірювань). Точність і достовірність вихідного статистичного матеріалу – найважливіше завдання статистичного спостереження. Навіть за гарної організації спостережень або проведення експерименту з тих чи інших причин можуть виявитися неточності, похибки, тобто помилки реєстрації. У існуючому способі контролю інформації для перевірки вихідної інформації на наявність викидів використовують залежно від обсягу вибірки критерії [1]:

- S -критерій – для вибірки, що містить понад 25 елементів;

- M -критерій – для вибірки малого обсягу.

Процедуру перевірки екстремальних значень щодо їх виключення з вибірки як викидів за M -критерієм можна здійснити так:

- переглянути багатовимірну вибірку та скопіювати ознаку з екстремальним значенням у вільний стовпець поруч із вибіркою;

- значення скопійованої ознаки вибірки впорядкувати по мірі зростання чи спадання, щоб екстремальне значення виявилось першим елементом масиву;

- позначити перший елемент масиву – x_1 ;

- за формулою обчислити розрахункове значення критерію M (M_p) та порівняти його з критичним значенням M , що наведене у таблиці 1.

Таблиця 1

Критичні значення критерію M

Обсяг вибірки	Рівень значущості			M_p
	$\alpha = 0,1$	$\alpha = 0,05$	$\alpha = 0,01$	
3	0,886	0,941	0,988	$\frac{ x_1 - x_2 }{ x_1 - x_n }$
4	0,679	0,765	0,889	
5	0,557	0,642	0,780	
6	0,482	0,560	0,698	
7	0,434	0,507	0,637	
8	0,479	0,554	0,683	$\frac{ x_1 - x_2 }{ x_1 - x_{n-1} }$
9	0,441	0,512	0,635	
10	0,409	0,477	0,597	- ' -
11	0,517	0,576	0,679	$\frac{ x_1 - x_3 }{ x_1 - x_{n-1} }$
12	0,490	0,546	0,642	
13	0,467	0,521	0,615	
14	0,492	0,546	0,641	$\frac{ x_1 - x_3 }{ x_1 - x_{n-2} }$
15	0,472	0,525	0,616	
16	0,454	0,507	0,595	$\frac{ x_1 - x_3 }{ x_1 - x_{n-2} }$
17	0,438	0,490	0,577	
18	0,424	0,475	0,561	
19	0,412	0,462	0,547	
20	0,401	0,450	0,535	
21	0,391	0,440	0,524	
22	0,382	0,430	0,514	
23	0,374	0,421	0,505	
24	0,367	0,413	0,497	
25	0,360	0,406	0,489	

Примітка. Індекси у формулах M_p відповідають розташуванню елемента в упорядкованому масиві ознаки.

Якщо $M_p < M$, то екстремальне значення x_1 з ймовірністю $1 - \alpha$ не є викидом, тобто його не можна виключати з вибірки. І тут екстремальне значення пояснюється проявом мінливості, що притаманно даній генеральній сукупності. Якщо $M_p \geq M$, то виключається з багатовимірної вибірки рядок з екстремальним значенням ознаки.

У разі обсягу вибірки $n > 25$ екстремальні значення можуть бути перевірені за S -критерієм:

$$S_p = (x_1 - \bar{X}) / \sigma, \tag{1}$$

де \bar{X} – вибіркове середнє та σ – стандартне відхилення всієї вибірки ознаки; x_1 – передбачуваний викид. Критичне значення критерію визначається за таблицею 2.

Якщо $|S_p| < s$, екстремальне значення x_1 не є викидом. При $|S_p| \geq s$ із багатовимірної вибірки виключається рядок з

екстремальним значенням.

Таблиця 2

Критичні значення критерію S

Обсяг вибірки	Рівень значущості	
	$\alpha = 0,05$	$\alpha = 0,01$
30	2,929	3,402
50	3,082	3,539
100	3,283	3,718
1000	3,884	4,264

Приклад 1. Серед 15 значень ознаки X (прибуток банку) зустрічається негативне значення. Перевірити, чи воно є викидом?

Оскільки обсяг вибірки малий ($n = 15$), використовуємо M -критерій, розрахункове значення якого визначається за формулою:

$$M_p = \frac{|x_1 - x_3|}{|x_1 - x_{n-2}|},$$

де x_1, x_3, x_{n-2} – відповідно перший, третій та 13-й елементи впорядкованого масиву X .

Допустимо $x_1 = -0,8; x_3 = 1; x_{13} = 6,2$.

Тоді $M_p = \left| \frac{-0,8-1}{-0,8-6,2} \right| = 0,257$. Критичне

значення критерію M для обсягу вибірки $n=15$ при рівні значимості $\alpha = 0,05$ дорівнює 0,525 (табл.1). Отже, розрахункове значення критерію менше критичного і екстремальне значення не є викидом.

Приклад 2. Перевірити на викид значення 526,6 результативної ознаки Y при вибірці обсягом $n=50$. Вибіркове середнє $\bar{X}=113,1945$; стандартне відхилення $\sigma=116,937$. Розрахункове значення критерію дорівнює $S_p = (526,6 - 113,1945) / 116,937 = 3,5353$.

Критичне значення критерію $S = 3,082$ для обсягу вибірки $n=50$ за рівнем значимості $\alpha = 0,05$. Отже, значення 526,6 є викидом і з багатовимірної вибірки слід виключити рядок із цим значенням.

Приклад 3. Грубих помилок у вихідній інформації багатовимірної вибірки обсягом $n=53$ немає, але серед елементів масиву ознаки X_8 (премії та винагороди на одного працівника, %) є два елементи, значення яких різко виділяються на загальному фоні: 0,03 та 4,44. Перевіримо, чи ці значення є викидами. Результати розрахунку наведені у таблиці 3.

Таблиця 3

Контроль інформації про наявність викидів

Контроль інформації	
x1	середнє
0,03	1,0718868
стандартне відхилення	
	0,6759838
Sp	S
-1,54129	3,082
x1=0,03 -	не викид
x1	
4,44	викид
Sp	S
4,9825356	3,082

Визначаємо вибіркове середнє, стандартне відхилення та розрахункове значення S -критерію. В окрему осередку заносимо критичне значення цього

критерію. Отже, перше екстремальне значення не є викидом, тобто у такому розмірі премії та винагороди на одного працівника для аналізованої сукупності підприємств можливі. За розрахунками друге екстремальне значення є викидом.

Аналіз існуючого способу контролю інформації показує, що він занадто складний: використовує різні критерії залежно від обсягу вибірки, критичні значення критеріїв обираються з таблиць, розрахунок критеріїв виконується за формулами та статистичними функціями майстра функцій Excel.

Мета статті – розроблення способу контролю інформації на основі дисперсійного аналізу даних спостережень.

Результати досліджень. Створено єдину базу даних об'єктів-аналогів. Для коректного створення об'єднаної бази даних потрібно бути впевненим, що об'єднані дані належать до однієї генеральної сукупності.

Складність проведення дисперсійного аналізу залежить від обсягу вибірок. Якщо об'єднуються кілька вибірок одного обсягу, тоді легко перевірити їх однорідність за допомогою інструменту «Однофакторний дисперсійний аналіз», у вихідній інформації якого видаються розрахункове і критичне значення F -критерію Фішера. Розрахункове значення критерію обчислюється дисперсійним відношенням

$$F = s_1^2 / s_0^2, \quad s_1^2 > s_0^2, \quad (2)$$

де s_1^2 – факторна дисперсія; s_0^2 – залишкова дисперсія.

Якщо $F \leq F_{kp}$, то на прийнятому рівні значимості робиться висновок про однорідність вибірок.

Приклад 4 [2]. Маємо вибірку трудомісткості одиниці виробленої продукції обсягом 40 елементів. Потрібно перевірити їх приналежність до однієї генеральної сукупності.

Із вибірки будемо 8 малих вибірок обсягом 5 елементів (табл. 5) і перевіряємо їх однорідність за допомогою інструменту *Однофакторний дисперсійний аналіз* (табл. 4).

Таблиця 4

Дисперсійний аналіз однорідності вибірок трудомісткості одиниці виробленої продукції

Однофакторний дисперсійний аналіз						
ВИСНОВОК						
Групи	Счет	Сума	Середнє	Дисперсія		
1	5	1,72	0,344	0,01248		
2	5	1,89	0,378	0,00097		
3	5	1,6	0,32	0,00385		
4	5	1,49	0,298	0,00327		
5	5	1,91	0,382	0,00872		
6	5	1,41	0,282	0,00237		
7	5	1,49	0,298	0,00317		
8	5	1,79	0,358	0,01527		
Дисперсійний аналіз						
Джерело варіації	SS	df	MS	F	P-Значення	F критичне
Між групами	0,052	7	0,00742	1,185058	0,338923	2,3127412
Усередині груп	0,2	32	0,00626			
Разом	0,252	39				

Таблиця 5

Вибірki малого обсягу

1	2	3	4	5	6	7	8
0,23	0,36	0,42	0,26	0,29	0,36	0,25	0,42
0,43	0,37	0,3	0,37	0,41	0,23	0,31	0,51
0,31	0,43	0,32	0,29	0,41	0,26	0,38	0,31
0,26	0,35	0,25	0,34	0,29	0,27	0,24	0,37
0,49	0,38	0,31	0,23	0,51	0,29	0,31	0,18

Дисперсійний аналіз показує: вибірки однорідні, тому що $F < F_{кр}$, середні значення всіх вибірок практично однакові, тобто елементи вибірки обсягом 40 елементів належать до однієї генеральної сукупності.

Приклад 5. Маємо вибірку премії та винагороди на одного працівника (%) обсягом 44 елементи. Потрібно перевірити їх приналежність до однієї генеральної сукупності.

Із чотирьох малих вибірок одна вибірка має максимальне число ознаки (табл. 6).

Перевіряємо максимальне число на викид за допомогою інструменту *Однофакторний дисперсійний аналіз*. Вихідна інформація наведена в таблиці 7.

Дисперсійний аналіз показує $F > F_{кр}$, тобто вибірки не однорідні, середнє значення вибірки 1 суттєво більше середніх значень інших вибірок. Отже, це є

причиною порушення однорідності вибірок. Вибірka 1 має елемент 4,44 більший величини порівнянно з іншими елементами, тобто 4,44 є викид.

Таблиця 6

Вибірki малого обсягу

1	2	3	4
0,67	1,23	0,84	0,68
0,98	1,04	0,67	0,86
1,16	1,8	1,04	0,96
0,54	0,43	0,66	0,33
1,23	0,88	0,86	0,45
0,78	0,57	0,79	0,74
1,16	0,99	0,34	0,03
4,44	0,24	0,82	0,99
1,06	0,84	0,84	0,24
2,13	0,6	0,67	0,57
1,21	0,82	0,58	1,22

Таблиця 7

Дисперсійний аналіз однорідності вибірок премії та винагороди на одного працівника

Однофакторний дисперсійний аналіз						
ВИСНОВОК						
Групи	Счет	Сума	Середнє	Дисперсія		
1	11	15,36	1,3963636	1,19034545		
2	11	9,44	0,8581818	0,17931636		
3	11	8,11	0,7372727	0,03350182		
4	11	7,07	0,6427273	0,12904182		
Дисперсійний аналіз						
Джерело варіації	SS	df	MS	F	P-Значення	F критичне
Між групами	3,745464	3	1,2484879	3,25932237	0,03129705	2,838745398
Усередині груп	15,32205	40	0,3830514			
Раом	19,06752	43				

Приклад 6. Сформувані вибірку премії та винагороди на одного працівника (%) із вибірок малого обсягу. Розглянуто можливі варіанти.

Варіант 1

№ п/п	Вибірки		
	1	2	3
1	1,23	1,72	0,001
2	1,04	1,7	0,99
3	1,8	0,84	0,24
4	0,43	0,6	0,57
5	0,88	0,82	0,01

Варіант 2

№ п/п	Вибірки		
	1	2	3
1	1,22	4,44	0,67
2	0,68	1,06	0,98
3	1	2,13	1,16
4	0,81	1,21	0,54
5	0,27	2,2	1,23

Варіант 1. Результати дисперсійного аналізу наведені в таблиці 8.

Таблиця 8

Дисперсійний аналіз однорідності вибірок премії та винагороди на одного працівника

Однофакторний дисперсійний аналіз						
ВИСНОВОК						
Групи	Счет	Сума	Середнє	Дисперсія		
1	5	5,38	1,076	0,25123		
2	5	5,68	1,136	0,28348		
3	5	1,811	0,3622	0,17669		
Дисперсійний аналіз						
Джерело варіації	SS	df	MS	F	P-Значення	F критичне
Між групами	1,853	2	0,9266	3,90736	0,0493356	3,885294
Усередині груп	2,846	12	0,2371			
Разом	4,699	14				

Дисперсійний аналіз показує: $F > F_{кр}$, середнє значення вибірки 3 суттєво відрізняється від середніх значень вибірок 1 та 2, тобто вибірки неоднорідні. Аналізуючи

елементи вибірок, відмічаємо, що вибірка 3 має два елементи дуже малої величини. Замінімо значення елемента 0,01 на 1,22 (табл. 9).

Таблиця 9

Дисперсійний аналіз однорідності вибірок

Дисперсійний аналіз						
Джерело варіації	SS	df	MS	F	P-Значення	F критичне
Між групами	0,848344	2	0,42417	1,6084626	0,240506071	3,885293835
Усередині груп	3,164553	12	0,26371			
Разом	4,012897	14				

Дисперсійний аналіз показує: вибірки однорідні, тому що $F < F_{кр}$. Отже, можна отримати вибірку обсягом 15 елементів, коли у вибірці 3 буде тільки один елемент дуже малої величини. Тобто елемент

фактора премії і винагороди 0,03 не є викидом.

Варіант 2. Результати дисперсійного аналізу наведені в таблиці 10. Вони показують, що середнє значення вибірки 2 суттєво відрізняється від середніх значень вибірок 1 та 3.

Таблиця 10

Дисперсійний аналіз однорідності вибірок

Однофакторний дисперсійний аналіз						
ВИСНОВОК						
Групи	Счет	Сума	Середнє	Дисперсія		
1	5	3,98	0,796	0,12793		
2	5	11,04	2,208	1,82547		
3	5	4,58	0,916	0,09103		
Дисперсійний аналіз						
Джерело варіації	SS	df	MS	F	P-Значення	F критичне
Між групами	6,129	2	3,0645	4,49686	0,0348779	3,885294
Усередині груп	8,178	12	0,6815			
Разом	14,31	14				

Це причина того, що $F > F_{кр}$, тобто вибірки неоднорідні і їх не можна об'єднати в одну. Аналізуючи елементи вибірок, відмічаємо, що вибірка 2 має елемент 4,44 більший величини порівнянно з іншими

елементами. Замінімо 4,44 на 0 і перевіримо однорідність вибірок.

Результати дисперсійного аналізу наведені в таблиці 11.

Таблиця 11

Дисперсійний аналіз однорідності вибірок

Дисперсійний аналіз						
Джерело варіації	SS	df	MS	F	P-Значення	F критичне
Між групами	0,753653	2	0,37683	1,0953096	0,365661635	3,88529383
Усередині груп	4,12844	12	0,34404			
Разом	4,882093	14				

Дисперсійний аналіз показує $F < F_{кр}$, тобто вибірки однорідні і їх можна об'єднати в одну вибірку обсягом 14 елементів. Отже, число 4,44 є викид.

Таким чином, за допомогою дисперсійного аналізу можна встановити не тільки однорідність вибірок, а і причину її порушення.

Дисперсійний аналіз однорідності вибірок різного обсягу. Як приклад дисперсійного аналізу однорідності вибірок у випадку нерівного числа спостережень за факторами розглянемо дані спостережень терміну служби електричних ламп.

Приклад 7 [3]. Для виготовлення кожної партії ламп взято дрiт різних сортів, інші умови виробництва були однакові. Потрібно встановити однорідність партій ламп між собою за терміном служби. Вихідна інформація наведена у таблиці 12.

Покажемо можливість поєднання вибірок різного обсягу за допомогою інструменту «Описова статистика» пакета аналізу. Вихідна інформація наведена в таблиці 13.

Таблиця 12

Дисперсійний аналіз однорідності терміну служби партій електричних ламп

Номер партії	Термін служби електричних ламп			
	1	2	3	4
	1,6	1,58	1,46	1,51
Термін служби	1,61	1,64	1,55	1,52
	1,65	1,64	1,6	1,53
	1,68	1,7	1,62	1,67
	1,7	1,75	1,64	1,6
	1,72		1,66	1,68
	1,8		1,74	

Таблиця 13

Статистична обробка вибірок

1	2	3	4
Середнє	1,68	Середнє	1,662
Стандартна помилка	0,02609506	Стандартна помилка	0,029
Медіана	1,68	Медіана	1,64
Мода	#Н/Д	Мода	1,64
Стандартне відхилення	0,06904105	Стандартне відхилення	0,065
Дисперсія вибірки	0,00476667	Дисперсія вибірки	0,004
Експес	0,26586141	Експес	-0,438
Асиметричність	0,65087455	Асиметричність	0,249
Інтервал	0,2	Інтервал	0,17
Мінімум	1,6	Мінімум	1,58
Максимум	1,8	Максимум	1,75
Сума	11,76	Сума	8,31
Рахунок	7	Рахунок	5

Таблиця 14

Зведення вихідних даних та дисперсійний аналіз

партії	дисперсія	середнє	обсяг
1	0,004766667	1,68	7
2	0,00422	1,662	5
3	0,012169643	1,63625	8
4	0,00587	1,585	6
Дисперсійний аналіз			
alfa=	0,05		
партії 1-2	m1	m2	F
		6	4
			1,12954
			6,163132
партії 1-3		6	6
			2,55307
			4,283866
партії 1-4		5	6
			1,23147
			4,387374

Аналіз статистичної обробки вибірок показує, що термін служби електричних ламп підпорядковується нормальному закону розподілу, оскільки середнє значення, мода і медіана мають один порядок, а значення ексцесу та асиметричності близькі до нуля.

Вихідна інформація щодо дисперсійного аналізу вибирається з таблиці 13, зводиться в таблицю 14 та виконується дисперсійний аналіз. Дисперсійний аналіз однорідності вибірок різного обсягу виконується шляхом перевірки однорідності двох вибірок, одна з яких – перша партія ламп. Другою вибіркою послідовно є: друга, третя та четверта партії.

Аналіз даних таблиці 14 показує, що для всіх поєднань партій середні дуже близькі за значеннями, а розрахункові значення критерію Фішера менші з відповідними критичними значеннями, тобто $F < F_{кр}$. Отже, вибірки однорідні і можуть бути об'єднані в одну вибірку обсягом 26 елементів.

Перевірка адекватності моделі регресії активного експерименту. План експерименту передбачає умови і кількість проведення дослідів, а головне, – визначає точність отриманої в результаті експерименту моделі регресії. У факторному просторі вибирається деяка точка і розглядається безліч точок її околиці. У цій околиці проводиться експеримент, на основі якого будується перша модель. Вона використовується для передбачення результатів у точках, які не увійшли до плану експерименту.

Головна вимога до моделі – здатність передбачати напрямок подальших дослідів із необхідною точністю. І точність цього передбачення в усіх напрямках пошуку повинна бути однаковою, тобто передбачене значення відгуку може відрізнятись від фактичного значення тільки в межах деякої заздалегідь заданої величини. Модель, що задовольняє таку вимогу, називається адекватною. Перевірка здійсненності цієї умови називається аналізом адекватності моделі.

У процесі проведення експерименту необхідно переконатися, що вимірювані значення відгуку належать до однієї генеральної сукупності і технологічний процес не вимагає регулювання.

Для цього кожний дослід матриці планування експерименту проводять кілька разів (паралельні дослідів). Після проведення чотирьох дослідів за матрицею планування експерименту необхідно переконатися в однорідності отриманих вибірок відгуку та можливості відтворюваності дослідів.

Для оцінення адекватності рівняння регресії експериментальним даним залишкова дисперсія відгуку порівнюється з дисперсією фактичних значень відгуку, тобто оцінка адекватності відповідає оцінці однорідності вибірок відгуку.

Перевірка адекватності моделі регресії проводиться за F -критерієм Фішера, розрахункове значення якого визначається відношенням дисперсії фактичних значень відгуку до залишкової дисперсії $F = s_y^2 / s_0^2$.

У разі негативних результатів, тобто $F > F_{кр}$, слід відрегулювати прилади вимірювання, повторити дослідів та їх обробку. Після підтвердження однорідності вибірок відгуків продовжити дослідів за матрицею планування.

Приклад 8 [4]. У кожній точці плану проводили випробування на міцність трьох зразків бетону. В таблиці 15 наведено умови планування експерименту. Досліджується залежність між ознаками: y – міцність на стиснення бетону у віці 28 діб, МПа; x_1 – цементно/водне (Ц/В) відношення бетону М200–М400; x_2 – активність цементу R_c , МПа; x_3 – модуль крупності $M_{кр}$; x_4 – вміст домішок Q_o , що відмучуються.

Перевіримо адекватність моделі регресії за першими чотирма дослідями. Матриця планування експерименту наведена в таблиці 16, де замість y виступає середнє значення міцності бетону на стиск Y_{cp} , отримане за результатами експерименту.

Таблиця 15

Умови планування експерименту

Фактор		Рівень фактора			Інтервал варіювання
натуральний	кодований	-1	0	1	
Ц/В	x_1	1,4	2	2,6	0,6
R_u , МПа	x_2	38,8	45,3	51,8	6,5
$M_{кр}$	x_3	1,4	2,2	3	0,8
Q_o	x_4	1	3	5	2

Таблиця 16

Матриця планування

№	x_1	x_2	x_3	x_4	y_1	y_2	y_3	y_{cp}
1	2,6	51,8	3	5	44,2	43	43,6	43,6
2	2,6	51,8	3	1	49	49,6	47,5	48,7
3	2,6	51,8	1,4	5	42	39,6	41,1	40,9
4	2,6	51,8	1,4	1	45	44	44,2	44,4

Для цього достатньо перевірити однорідність вибірок за допомогою інструменту *Однофакторний дисперсійний* *аналіз*. Вихідна інформація наведена в таблиці 17.

Таблиця 17

Дисперсійний аналіз адекватності моделі регресії

Однофакторний дисперсійний аналіз						
ВИСНОВОК						
Групи	Счет	Сума	Середнє	Дисперсія		
y1	4	180,2	45,05	8,543333		
y2	4	176,2	44,05	17,23667		
y3	4	176,4	44,1	6,94		
Дисперсійний аналіз						
Джерело варіації	SS	df	MS	F	P-Значення	F критичне
Між групами	2,54	2	1,27	0,116443	0,89140046	4,25649473
Усередині груп	98,16	9	10,9067			
Разом	100,7	11				

Таблиця 18

Оскільки $F < F_{кр}$, вибірки міцності бетону на стиск однорідні і лінійна модель регресії адекватна експериментальним даним. Отже, можна продовжувати виконувати досліди за матрицею планування експерименту.

Приклад 9 [5]. Здійснити дисперсійний аналіз впливу на міцність бетону режимів роботи технологічної лінії з виготовлення залізобетонних виробів.

Результати вимірювань міцності бетону

Тиждень	1	2	3	4
Міцність бетону, МПа	22	18,4	21	20,4
	18,5	19,8	22	20,5
	20,1	20,2	18,3	22
	20,4	20,5	19,1	18
	19,7	21	20,3	19

Дисперсійний аналіз впливу технологічного процесу на міцність бетону

Дисперсійний аналіз						
Джерело варіації	SS	df	MS	F	P-Значення	F критичне
Між групами	0,128	3	0,04267	0,02404	0,99474864	3,23887152
Усередині груп	28,4	16	1,775			
Разом	28,528	19				

Результати вимірювань міцності таблиці 18. У таблиці 19 наведено зведення протягом місяця по тижнях показані в дисперсійного аналізу.

СПИСОК ВИКОРИСТАНИХ ДЖЕРЕЛ

1. Ершова Н. М., Деревянко В. Н., Тимченко Р. А., Шаповалова О. В. Обработка данных средствами Excel при планировании эксперимента : учеб. пособ. для вузов. Днепропетровск : ПГАСА, 2012. 350 с.
2. Дубров А. М., Мхитарян В. С., Трошин Л. И. Многомерные статистические методы для экономистов и менеджеров : учеб. Москва : Финансы и статистика, 2000. 352 с.
3. Дворкін Л. Й. Експериментально-статистичне моделювання при проектуванні складів бетонів : навч. посіб. Київ : Видавничий дім «Кондор», 2020. 205 с.
4. Митропольский А. К. Техника статистических вычислений. Москва : Главная редакция физико-математической литературы издательства «Наука», 1971. 576 с.
5. Красовский П. С. Исследование и оптимизация свойств строительных материалов с применением элементов математической статистики : учеб. пособ. Хабаровск : ДВГУПС, 2004. 128 с.

REFERENCES

1. Yershova N.M., Derevyanko V.N., Timchenko R.A. and Shapovalova O.V. *Obrabotka dannykh sredstvami Excel pri planirovani eksperimenta : ucheb. posobiye dlya vuzov* [Data processing using Excel when planning an experiment: textbook manual for universities]. Dnipropetrovsk : PSACEA Publ., 2012, 350 p. (in Russian).
2. Dubrov A.M., Mkhitaryan V.S. and Troshin L.I. *Mnogomernyye statisticheskiye metody dlya ekonomistov i menedzherov : ucheb.* [Multidimensional statistical methods for economists and managers : textbook]. Moscow : Finance and Statistics Publ., 2000, 352 p. (in Russian).
3. Dvorkin L.Yo. *Eksperymental'no-statystychne modelyuvannya pry proektuvanni skladiv betoniv : navch. posib.* [Experimental-statistical modeling in the design of concrete compositions : training manual]. Kyiv : "Condor" Publishing House, 2020, 205 p. (in Ukrainian).
4. Mitropol'skiy A.K. *Tekhnika statisticheskikh vychisleniy* [Statistical computing technique]. Moscow : Main editorial office of physical and mathematical literature of the "Science" Publishing House, 1971, 576 p. (in Russian).
5. Krasovskiy P.S. *Issledovaniye i optimizatsiya svoystv stroitel'nykh materialov s primeneniye elementov matematicheskoy statistiki : ucheb. posob.* [Research and optimization of the properties of building materials using elements of mathematical statistics : textbook allowance]. Khabarovsk : DVGUPS Publ., 2004, 128 p. (in Russian).

Надійшла до редакції: 15.09.2023.