

**ІНФОРМАЦІЙНІ ТЕХНОЛОГІЇ ПРИЙНЯТТЯ РІШЕНЬ В ФОРМУВАННІ  
ЄДИНОЇ БАЗИ ДАНИХ ОБ'ЄКТІВ-АНАЛОГІВ ДОСЛІДЖУВАНОВОГО  
ОБ'ЄКТА**

Н. М. Єршова

---

Придніпровська державна академія будівництва та архітектури  
м. Дніпро, 49000, вул. Чернишевського, 24 а E-mail:  
nersoval07@gmail.com

---

Експеримент посідає особливе місце серед способів отримання інформації про внутрішні взаємозв'язки явищ у природі та техніці. Відповідно ускладненню досліджуваних процесів та явищ зростають витрати на апаратуру та проведення експерименту. У ході випробувань збирається велика кількість експериментальних даних, що потребують обробки та аналізу. При цьому тривалість аналізу, осмислення результатів випробувань та їхнього обліку для коригування характеристик нових виробів дуже значна. Під час проведення спостережень чи експерименту дуже важливо відібрати методи та інструментальні засоби обробки даних експерименту. Незважаючи на те, що починаючи з 2001 року в багатьох роботах доводиться ефективність пакету аналізу Excel для обробки даних експерименту, досі в науковій та навчальній літературі використовується при обробці даних спостережень та експерименту обчислення розрахункових значень критеріїв рівності дисперсій та середніх за формулами, а вибір їх критичних значень за таблицями. Іншою проблемою є формування вибірки з вибірок різного обсягу прямим розрахунком за формулами навіть у середовищі Excel. У даній роботі пропонується методика дисперсійного аналізу однорідності вибірок різного обсягу з використанням інструментів пакету аналізу Excel. Інструмент «Описова статистика» виконує статистичну обробку багатовимірних вибірок різного обсягу та видає значення 13 параметрів, у тому числі: середнє значення, мода, медіана, ексцес, асиметричність, дисперсія та обсяг вибірки. За цією інформацією легко визначити розрахункове та критичне значення статистик. Критичне значення критеріїв визначаються за допомогою статистичних функцій майстра функцій. Вихідна інформація інструменту «Однофакторний дисперсійний аналіз» містить розрахункове та критичне значення  $F$  – критерію Фішера, що дозволяє легко перевірити однорідність вибірок рівного обсягу. Методика значно спрощує процедуру формування єдиної бази даних об'єктів-аналогів досліджуваного об'єкта.

**Ключові слова:** експеримент, обробка даних, дисперсійний аналіз, формування вибірки, вибірки різного обсягу, пакет аналізу Excel.

**Вступ**

В багатьох роботах доводиться ефективність пакету аналізу Excel для обробки даних експерименту, але досі в науковій та навчальній літературі використовується при обробці даних спостережень та експерименту обчислення розрахункових значень критеріїв рівності дисперсій та середніх за формулами, а вибір їх критичних значень за таблицями. Іншою проблемою є формування вибірки з вибірок різного обсягу прямим розрахунком за формулами. У даній роботі пропонується методика дисперсійного аналізу однорідності вибірок різного обсягу з використанням інструментів пакету аналізу Excel. Інструмент «Описова статистика» виконує статистичну обробку багатовимірних вибірок різного обсягу та видає значення 13 параметрів, у тому числі: середнє значення, мода, медіана, ексцес, асиметричність, дисперсія та обсяг вибірки. За цією інформацією легко визначити розрахункове та критичне значення статистик. Критичне значення критеріїв визначаються за допомогою статистичних функцій

майстра функцій. Вихідна інформація інструменту «Однофакторний дисперсійний аналіз» містить розрахункове та критичне значення  $F$  – критерію Фішера, що дозволяє легко перевірити однорідність вибірок рівного обсягу. Методика значно спрощує процедуру формування єдиної бази даних об'єктів-аналогів досліджуваного об'єкта.

#### **Аналіз останніх досліджень і публікацій**

Експеримент посідає особливе місце серед способів отримання інформації про внутрішні взаємозв'язки явищ у природі та техніці. Відповідно ускладненню досліджуваних процесів та явищ зростають витрати на апаратуру та проведення експерименту. У ході випробувань збирається велика кількість експериментальних даних, що потребують обробки та аналізу. При цьому тривалість аналізу, осмислення результатів випробувань та їхнього обліку для коригування характеристик нових виробів дуже значна. Під час проведення спостережень чи експерименту дуже важливо відібрати методи та інструментальні засоби обробки даних експерименту. Незважаючи на те, що починаючи з 2001 року в роботах [5-6, 10] доводиться ефективність використання пакету аналізу Excel для обробки даних експерименту, до цих пір у науковій та навчальній літературі при обробці даних спостережень та експерименту використовується обчислення розрахункових значень критеріїв рівності дисперсій та середніх за формулами, а вибір їх критичних значень за таблицями [1, 4, 7, 9].

Іншою проблемою є формування вибірки з вибірок різного обсягу прямим розрахунком за формулами навіть в середовищі Excel. Постає питання, чи не можна спростити цю процедуру? Виявляється, можна, якщо для цього використовувати інструмент «Описова статистика» пакету аналізу.

#### **Мета роботи**

Отже метою даної роботи є розробка методики дисперсійного аналізу однорідності вибірок різного обсягу з використанням інструменту «Описова статистика» пакету аналізу Excel.

#### **Основна частина**

Методи дисперсійного аналізу дозволяють формувати єдину базу даних об'єктів-аналогів та оцінювати величину впливу конкретних факторів на досліджувану результативну ознаку.

Під час експерименту для кожного об'єкту часто можна зміряти (отримати) значення декількох ознак. У результаті виходить багатовимірна вибірка. Смысл обробки багатовимірних вибірок полягає у встановленні зв'язків між ознаками. Для цього їх ділять на ознаки факторні і результативні. Факторна ознака викликає зміну інших, пов'язаних з ним, ознак. Результативна ознака змінюється під дією факторних ознак. Дисперсійний аналіз призначений для кількісного дослідження впливу факторних ознак на результативну ознаку у разі малих вибірок.

Для порівняння впливу факторних ознак на результативну ознаку необхідний певний статистичний матеріал – кожному рівню фактора повинна відповідати певна вибірка значень результативної ознаки. Статистичний матеріал зручно представляти у вигляді таблиці 1. Перш ніж судити про кількісний вплив фактора, необхідно встановити наявність такого впливу. Можливо, розбіжність значень результативної ознаки для різних рівнів фактора пояснюється дією чистої випадковості.

Загальне число спостережень  $n = n_1 + n_2 + \dots + n_p$ .

Таблиця 1

Матриця експериментів для однофакторного аналізу

	Рівні фактора (Номер вибірки)			
	1	2	...	$p$
Значення результативної ознаки	$x_{11}$	$x_{12}$		$x_{1p}$
	$x_{21}$	$x_{22}$		$x_{2p}$
	...	...		...
	$x_{n_11}$	$x_{n_22}$		$x_{n_p p}$
Обсяг вибірки	$n_1$	$n_2$		$n_p$

На статистичній мові це припущення означає перевірку однорідності всіх вибірок таблиці 1, тобто перевірку приналежності всіх значень результативної ознаки однієї генеральної сукупності. Основною процедурою дисперсійного аналізу є перевірка цієї гіпотези за допомогою статистичних критеріїв.

Допустимо фактор  $A$  має  $p$  різних рівнів, на кожному з яких виконано  $n$  спостережень. Отже, спостерігалось  $N = pn$  значень  $x_{ij}$  ознаки (властивості)  $X$ , де  $i$  - номер спостереження ( $i = 1, 2, \dots, n$ ),  $j$  - номер рівня фактора ( $j = 1, 2, \dots, p$ ).

Існують поняття [8]:

- загальна сума квадратів - сума квадратів відхилень всіх можливих значень ознаки від їх загального середнього значення

$$S = \sum_{i=1}^n \sum_{j=1}^p (x_{ij} - \bar{X})^2; \quad (1)$$

- сума квадратів між групами або за факторами - зважена сума квадратів відхилень середніх значень за групами від загального середнього значення

$$S_1 = n \sum_{j=1}^p (\bar{x}_j - \bar{X})^2; \quad (2)$$

- сума квадратів усередині груп - сума квадратів відхилень можливих значень ознаки кожної групи (рівня фактора) від середнього значення цієї групи

$$S_2 = \sum_{i=1}^n \sum_{j=1}^p (x_{ij} - \bar{x}_j)^2, \quad (3)$$

де  $\bar{x}_j, \bar{X}$  - відповідно середнє значення групи і загальне середнє значення результативної ознаки, що визначаються за формулами

$$\bar{x}_j = \frac{\sum_{i=1}^n x_{ij}}{n}; \quad \bar{X} = \frac{\sum_{j=1}^p \bar{x}_j}{p}. \quad (4)$$

Для оцінки впливу фактора слід розкласти загальну суму квадратів на складові: суму квадратів між групами (за факторами) і суму квадратів усередині груп. Отже

$$S = S_1 + S_2. \quad (5)$$

Сума  $S_1$  відображає вплив на результативну ознаку рівнів фактора, а сума  $S_2$  - вплив погрішностей вимірювань. Оскільки  $S_2 = S - S_1$ , то суму  $S_2$  називають ще залишковою сумою квадратів.

Суми квадратів  $S$ ,  $S_1$ ,  $S_2$  ділені на відповідні числа ступенів свободи, дають три незміщені оцінки дисперсії  $\sigma^2$  генеральної сукупності:

$$s^2 = \frac{S}{N-1}; \quad (6)$$

$$s_1^2 = \frac{S_1}{p-1}; \quad (7)$$

$$s_0^2 = \frac{S_2}{p(n-1)} = \frac{S_2}{N-p}. \quad (8)$$

Перша оцінка називається загальною оцінкою дисперсії (або вибірковою дисперсією), друга – оцінкою дисперсії за факторами (оцінкою дисперсії між групами або факторної дисперсією) і третя – залишковою оцінкою дисперсії (оцінкою дисперсії усередині груп або залишковою дисперсією).

Для порівняння дисперсій двох вибірок використовують  $F$  – критерій Фішера. Визначають розрахункове значення  $F$  – критерію у вигляді відношення більшої дисперсії до меншої

$$F = \frac{s_1^2}{s_2^2}. \quad (9)$$

Критичне значення - критерію ( $F_{kp}$ ) обчислюємо за допомогою статистичної функції  $F$ . ОБР.ПХ( $\alpha; m_1; m_2$ ). Число ступенів свободи приймають відповідно  $m_1 = n_1 - 1; m_2 = n_2 - 1$ , де  $n$  – обсяг вибірки. Гіпотеза про рівність дисперсій підтверджується, якщо  $F \leq F_{kp}$ .

Для порівняння двох вибірових середніх використовують  $t$  – статистику. Після перевірки гіпотези про рівність двох вибірових дисперсій, обчислюють загальну дисперсію двох вибірок та розрахункове значення  $t$  – статистики за формулами:

$$s^2 = \frac{m_1 s_1^2 + m_2 s_2^2}{m}; \quad (10)$$

$$t = \frac{(\bar{X}_1 - \bar{X}_2) \sqrt{n_1 n_2 / (n_1 + n_2)}}{s}. \quad (11)$$

Критичне значення  $t$  – статистики ( $t_{kp}$ ) визначаємо за допомогою статистичної функції СТЬЮДЕНТ.ОБР.2Х( $\alpha; m$ ). Число ступенів свободи  $m = m_1 + m_2$ . Гіпотеза про рівність середніх значень підтверджується, якщо  $|t| \leq t_{kp}$ .

На основі дисперсійного аналізу можна приймати рішення у багатьох галузях науки та практики. Особливо це важливо при створенні нових матеріалів, виробів, технологічних процесів та ін., коли є мало інформації про властивості об'єкта, що досліджується.

В даній статті розглядаються інформаційні технології прийняття рішень при формуванні єдиної бази даних об'єктів-аналогів досліджуваного об'єкта.

Мала вибірка містить мало інформації про цікаву властивість. Для отримання більш надійних висновків потрібно об'єднати малі вибірки в одну, але при цьому необхідно встановити їх однорідність.

Коли фактор приймає тільки два значення, тоді в розпорядженні дослідника є дві вибірки, що відповідно характеризують зміну результативної ознаки в залежності від зміни рівня фактора. В цьому випадку процедуру факторного аналізу називають перевіркою однорідності двох вибірок. При цьому в залежності від типу даних розрізняють дві ситуації:

– вибірки незалежні, коли вимір значень ознаки проводиться на

різних, досить однорідних об'єктах;

– вибірки представляють собою парні спостереження, коли безліч об'єктів зафіксовано, а спостереження проводяться в різні моменти часу.

Складність проведення дисперсійного аналізу залежить від обсягу вибірок. Якщо об'єднуються кілька вибірок одного обсягу, то легко перевірити їхню однорідність за допомогою інструмента «Однофакторний дисперсійний аналіз» пакета аналізу. У разі вибірок різного обсягу виникає проблема. До цього часу доводять однорідність вибірок різного обсягу безпосереднім розрахунком за формулами (1)-(11). Критичні значення критеріїв визначають за таблицями.

Доведемо шляхом моделювання можливість поєднання вибірок різного розміру за допомогою інструмента «Описова статистика» пакета аналізу.

**Моделювання однорідності незалежних вибірок**

В якості прикладу дисперсійного аналізу однорідності вибірок у випадку нерівного числа спостережень за факторами розглянемо дані спостережень терміну служби електричних ламп [8].

**Приклад 1.** Для виготовлення кожної партії ламп взято дріт різних сортів, інші умови виробництва були однакові. Потрібно встановити однорідність партій ламп між собою за терміном служби.

Вихідна інформація та результати розрахунку за формулами (1)..(11) представлені у табл.2.

**Таблиця 2**

Дисперсійний аналіз однорідності партій електричних ламп

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P
1	формування вибірки з вибірок різного обсягу															
2	термін служби електричних ламп															
3	номер партії	1	2	3	4											
4	X1j	1.6	1.58	1.46	1.51			Xij-Xcp				(Xij-Xcp)^2				
5	X2j	1.61	1.64	1.55	1.52			-0	-0.06	-0.181	-0.13081	0.002	0.004	0.033	0.017	
6	X3j	1.55	1.64	1.6	1.53			-0	-0	-0.091	-0.12081	9E-04	7E-07	0.008	0.015	
7	X4j	1.68	1.7	1.62	1.67			0.01	-0	-0.041	-0.11081	8E-05	7E-07	0.002	0.012	
8	X5j	1.7	1.75	1.64	1.6			0.04	0.06	-0.021	0.029188	0.002	0.004	4E-04	9E-04	
9	X6j	1.72		1.66	1.68			0.06	0.11	-8E-04	-0.04081	0.004	0.012	7E-07	0.002	
10	X7j	1.8		1.74				0.08		0.019	0.039188	0.006		4E-04	0.002	
11	X8j			1.82		SSj		0.16		0.099		0.025		0.01		
12	Sj	11.8	8.31	13.1	9.51	42.67				0.179				0.032		S
13	nj	7	5	8	6	Xcp						0.039	0.019	0.085	0.048	0.192
14	Xjcp	1.68	1.66	1.64	1.585	1.641										
15	Xjcp-Xcp	0.04	0.02	-0	-0.06											
16	(Xjcp-Xcp)^2	0	0	0	0.003	S1	S2									
17	nj*(Xjcp-Xcp)^2	0.01	0	0	0.019	0.032	0.16									
18						s1^2	s0^2									
19						0.011	0.007									
20	n	p	m1	m2	m											
21		26	4	3	22	25										
22	alfa	F	Fkp													
23		0,05	1,46	3,05												

В результаті розрахунку за формулами отримали  $F < F_{kp}$ , тобто немає причин відкидати гіпотезу щодо однорідності терміну служби електролам.

Вихідна інформація інструменту «Описова статистика» для даного прикладу наведена в таблиці 3.

## Статистична обробка вибірок

1		2		3		4	
Среднее	1,68	Среднее	1,662	Среднее	1,61	Среднее	1,585
Стандартная	0,026095	Стандартная	0,02905	Стандартная	0,03331	Стандартная	0,031278
Медиана	1,68	Медиана	1,64	Медиана	1,62	Медиана	1,565
Мода	#Н/Д	Мода	1,64	Мода	#Н/Д	Мода	#Н/Д
Стандартное	0,069041	Стандартное	0,06496	Стандартное	0,088129	Стандартное	0,076616
Дисперсия в	0,004767	Дисперсия в	0,00422	Дисперсия в	0,007767	Дисперсия в	0,00587
Эксцесс	0,265861	Эксцесс	-0,4379	Эксцесс	0,881162	Эксцесс	-2,34249
Асимметрич	0,650875	Асимметрич	0,24878	Асимметрич	-0,4234	Асимметрич	0,40824
Интервал	0,2	Интервал	0,17	Интервал	0,28	Интервал	0,17
Минимум	1,6	Минимум	1,58	Минимум	1,46	Минимум	1,51
Максимум	1,8	Максимум	1,75	Максимум	1,74	Максимум	1,68
Сумма	11,76	Сумма	8,31	Сумма	11,27	Сумма	9,51
Счет	7	Счет	5	Счет	7	Счет	6

Вихідна інформація щодо дисперсійного аналізу вибирається з таблиці 3 і зводиться в таблицю 4. 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 ламп (таблиця 5).

Таблиця 4

Зведення вихідних даних

	V	W	X	Y
17	партії	дисперсія	середня	обсяг
18	1	0,00476667	1,68	7
19	2	0,00422	1,662	5
20	3	0,00776667	1,61	8
21	4	0,00587	1,585	6

Таблиця 5

Дисперсійний аналіз однорідності вибірок

	V	W	X	Y	Z
23	alfa=	0,05			
24	партії 1-2	F	F <sub>кр</sub>	дисперсія	t <sub>кр</sub>
25		1,129541864	6,16313228	0,004538889	2,22814
26		m1		t	
27			6	0,456290187	
28		m2		m	
29			4		10
30	партії 1-3	F	F <sub>кр</sub>	дисперсія	t <sub>кр</sub>
31		1,629370629	4,28386571	0,006366667	2,17881
32		m1		t	
33			6	-1,695080656	
34		m2		m	
35			6		12
36	партії 1-4	F	F <sub>кр</sub>	дисперсія	t <sub>кр</sub>
37		1,231468531	4,38737419	0,005275897	2,20099
38		m1		t	
39			5	-2,350868854	
40		m2		m	
41			6		11

Порівняння розрахункових значень критеріїв з відповідними критичними значеннями показує, що для всіх поєднань партій ламп  $F < F_{kp}$  і  $t \leq t_{kp}$ .

Для доказу достатньо використати  $F$  – критерій. Отже, вибірки однорідні і можуть бути об'єднані в одну вибірку обсягом 26 елементів.

### Моделювання однорідності парних спостережень

**Приклад 2 [10].** Є вибірки, що містять вартість 1 м<sup>2</sup> внутрішньої площі об'єктів нерухомості у місті  $N$  за 1998-2000 р. р. Необхідно виконати дисперсійний аналіз із метою перевірки однорідності вибірок.

Вихідні дані та результати дисперсійного аналізу в середовищі ЕТ за формулами наведені в таблиці 6.

Отримано  $F < F_{kp}$ , тобто аналізовані вибірки однорідні.

Для підтвердження цього розглянемо додатково можливість об'єднання вибірок, отриманих в результаті парних спостережень. Виконаємо дисперсійний аналіз вибірок за 1998 та 1999 роки. Вони мають однаковий обсяг – 26 спостережень.

Зведення дисперсійного аналізу, що отримано за допомогою інструменту «Однофакторний дисперсійний аналіз», представлено у таблиці 7.

В результаті дисперсійного аналізу отримано  $F < F_{kp}$ , тобто вибірки за 1998 та 1999 роки однорідні.

Покажемо можливість об'єднання вибірок різного обсягу – вибірки за 1999 та 2000 роки. Вихідна інформація інструмента «Описова статистика» наведена в табл. 8.

**Таблиця 6**

Дисперсійний аналіз однорідності вибірок у середовищі ЕТ

	A	B	C	D	E	F	G	H	I	J	K
1	формування вибірки з вибірок різного розміру										
2	оцінка об'єктів нерухомості										
3	рік оцінки	1998	1999	2000		Xij-Xcp			(Xij-Xcp)^2		
4		75,92	84,352	91,8		-22,6261	-14,1941	-6,74608	511,939	201,472	45,5096
5		117,318	77,844	90,561		18,7719	-20,7021	-7,98508	352,385	428,576	63,7615
6		90,254	84,352	70,053		-8,29208	-14,1941	-28,4931	68,7585	201,472	811,855
7	вартість 1м^2	88,958	149,152	98,368		-9,58808	50,6059	-0,17808	91,9312	2560,96	0,03171
8	внутрішньої	100,743	55,031	109,603		2,19692	-43,5151	11,0569	4,82647	1893,56	122,256
9	площі	86,522	119,997	111,352		-12,0241	21,4509	12,8059	144,578	460,142	163,992
10		76,98	149,152	133,252		-21,5661	50,6059	34,7059	465,096	2560,96	1204,5
11		147,115	87,98	95,366		48,5689	-10,5661	-3,18008	2358,94	111,642	10,1129
12		71,005	135,026	128,222		-27,5411	36,4799	29,6759	758,511	1330,78	880,66
13		114,415	73,594	66,016		15,8689	-24,9521	-32,5301	251,823	622,606	1058,21
14		114,423	88,822	126,799		15,8769	-9,72408	28,2529	252,077	94,5577	798,228
15		95,031	63,649	101,016		-3,51508	-34,8971	2,46992	12,3558	1217,81	6,10052
16		37,7	77,874	89,627		-60,8461	-20,6721	-8,91908	3702,25	427,335	79,5499
17		110,341	72,935	90,033		11,7949	-25,6111	-8,51308	139,12	655,927	72,4725
18		72,379	68,051	98,267		-26,1671	-30,4951	-0,27908	684,716	929,95	0,07788
19		124,033	103,055	165,528		25,4869	4,50892	66,9819	649,583	20,3304	4486,58
20		92,27	109,923	95,301		-6,27608	11,3769	-3,24508	39,3891	129,434	10,5305
21		136,626	110,65	65,948		38,0799	12,1039	-32,5981	1450,08	146,505	1062,63
22		187,005	71,069	115,067		88,4589	-27,4771	16,5209	7824,98	754,99	272,941
23		152,979	79,207	101,137		54,4329	-19,3391	2,59092	2962,94	374	6,71288
24		156,401	92,418	98,454		57,8549	-6,12808	-0,09208	3347,19	37,5533	0,00848
25		43,208	137,198	63,15		-55,3381	38,6519	-35,3961	3062,3	1493,97	1252,88
26		29,336	99,033	80,464		-69,2101	0,48692	-18,0821	4790,03	0,23709	326,962
27		35,483	99,148	154,443		-63,0631	0,60192	55,8969	3976,95	0,36231	3124,47
28		52,771	141,003	82,74		-45,7751	42,4569	-15,8061	2095,36	1802,59	249,832
29		49,176	113,967	73,919		-49,3701	15,4209	-24,6271	2437,4	237,805	606,493
30				107,458				8,91192			79,4224
31				83,558				-14,9881			224,642
32				126,977				28,4309			808,317
33				88,272				-10,2741			105,557

34				115,554				17,0079			289,269	
35				107,558				9,01192			81,2148	
36				110,997				12,4509			155,025	
37				79,999				-18,5471			343,994	
38				111,731				13,1849			173,842	
39				70,165				-28,3811			805,486	
40				104,15				5,60392			31,404	
41				187,315				88,7689			7879,92	
42				135,36	SSj			36,8139			1355,26	S
43	Sj	2458,39	2544,48	4025,58	9028,45				42435,5	18695,5	29050,7	90181,8
44	nj	26	26	39	Xc						s^2=	1002,02
45	Xcpj	94,5535	97,8647	103,22	98,5461							
46	Xcpj-Xcp	-3,99254	-0,68138	4,67392								
47	(Xcpj-Xcp)^2	15,9404	0,46428	21,8456	S1							
48	nj*(Xcpj-Xcp)^2	414,449	12,0714	851,977	1278,5							
49				s1^2	639,249							
50				S2	s0^2							
51				88903,3	1010,26							
52	n	p	m1	m2	m							
53		91	3	2	88	90						
54	alfa	F	Fkp									
55		0.05	0.63275	3.10007								

Результати статистичної обробки вибірок показують, що вартість 1 м<sup>2</sup> внутрішньої площі об'єктів нерухомості підпорядковується нормальному закону розподілу, оскільки середнє значення, мода та медіана мають один порядок, а значення ексцесу та асиметричності близькі до нуля.

Таблиця 7

## Дисперсійний аналіз однорідності вибірок

Однофакторный дисперсионный анализ						
ИТОГИ						
Группы	Счет	Сумма	Среднее	Дисперсия		
1998	26	2458,392	94,5535385	1680,84294		
1999	26	2544,482	97,8646923	747,3383157		
Дисперсионный анализ						
Источник вариации	SS	df	MS	F	P-Значение	F критическое
Между группами	142,529	1	142,528617	0,117395369	0,733313031	4,034309546
Внутри групп	60704,5	50	1214,09063			
Итого	60847,1	51				



Таблиця 8

## Статистична обробка вибірок

	M	N	O	P	Q	R
21	1998		1999		2000	
22	Среднее	94,5535	Среднее	97,8646923	Среднее	103,22
23	Стандартная ошибка	8,04039	Стандартная ошибка	5,36132273	Стандартная ошибка	4,362050996
24	Медиана	91,262	Медиана	90,62	Медиана	98,454
25	Мода	#Н/Д	Мода	84,352	Мода	#Н/Д
26	Стандартное отклонение	40,9981	Стандартное отклонение	27,3374892	Стандартное отклонение	27,24099974
27	Дисперсия выборки	1680,84	Дисперсия выборки	747,338316	Дисперсия выборки	742,0720669
28	Эксцесс	-0,38574	Эксцесс	-0,7318968	Эксцесс	1,535417751
29	Асимметричность	0,31284	Асимметричность	0,54675199	Асимметричность	1,043068369
30	Интервал	157,669	Интервал	94,121	Интервал	124,165
31	Минимум	29,336	Минимум	55,031	Минимум	63,15
32	Максимум	187,005	Максимум	149,152	Максимум	187,315
33	Сумма	2458,39	Сумма	2544,482	Сумма	4025,58
34	Счет	26	Счет	26	Счет	39

У таблиці 9 виконано дисперсійний аналіз вибірок різного обсягу. Обчислено розрахункові та критичні значення  $F$ -критерію Фішера та  $t$ -статистики Стьюдента за 1999 та 2000 р.р.

Таблиця 9

## Дисперсійний аналіз вибірок різного обсягу

	M	N	O	P	Q	R
40	дисперсія		alfa		F <sub>кр</sub>	F
41	744,1785664		0,05		1,798312276	1,007096681
42	t	t <sub>кр</sub>	m1	m2	m	
43	0,775368492	1,99834	25	38	63	

У результаті дисперсійного аналізу однорідності вибірок за 1999 та 2000 роки встановлено, що  $F < F_{кр}$ . Отже, ці вибірки однорідні.

Аналіз дисперсійного аналізу, що наведено у таблиці 6, показує однорідність трьох вибірок різного обсягу. Тобто можна спростити процедуру проведення дисперсійного аналізу даних спостережень, якщо використовувати інструмент «Описова статистика» пакету аналізу.

#### Алгоритм методики дисперсійного аналізу з використанням інструментів пакету аналізу Excel

Допустимо є кілька вибірок різного обсягу, серед яких є дві вибірки рівного обсягу.

1. Зробити розміщення інформації на робочому аркуші ЕТ в такий спосіб, щоб вибірки рівного обсягу перебували на початку таблиці (табл. 6).

2. Перевірити однорідність цих вибірок за допомогою інструмента «Однофакторний дисперсійний аналіз».

3. Виконати статистичну обробку всіх вибірок, використовуючи інструмент «Описова статистика» (табл. 8).

4. За підсумками аналізу числових характеристик досліджуваної ознаки перевірити її підпорядкування нормальному закону розподілу.

5. Скласти таблицю вихідних даних щодо дисперсійного аналізу вибірок різного обсягу, у якій першої вибіркою буде остання з вибірок рівного обсягу. Вона вважається базовою і з нею порівнюються інші вибірки різного обсягу (табл. 4).

6. Виконати дисперсійний аналіз однорідності вибірок різного обсягу (табл. 5). Для цього достатньо визначити розрахункове та критичне значення  $F$  – критерію Фішера та переконатися в тому, що всі вибірки належать єдиній генеральній сукупності.

#### **Висновки**

1. Вихідна інформація інструменту «Однофакторний дисперсійний аналіз» пакету аналізу містить розрахункове та критичне значення  $F$  – критерію Фішера, що дозволяє легко перевірити однорідність вибірок рівного обсягу.

2. Інструмент «Описова статистика» пакету аналізу виконує статистичну обробку багатовимірних вибірок різного обсягу та видає значення 13 параметрів, у тому числі: середнє значення, мода, медіана, ексцес, асиметричність, дисперсія та обсяг вибірки. За цією інформацією легко визначити розрахункове та критичне значення статистик. Критичні значення критеріїв визначаються за допомогою статистичних функцій майстра функцій.

3. Розроблено методику проведення дисперсійного аналізу однорідності вибірок різного обсягу за допомогою інструменту «Описова статистика» пакету аналізу, що значно спрощує процедуру формування вибірки.

#### **Список літератури**

1. Вознесенский В. А., Ляшенко Т. В., Огарков Б. Л. Численные методы решения строительно-технологических задач на ЭВМ. К.: Выща школа, 1989. 328 с.
2. Гарькина И. А., Данилов А. М., Прошин А. П., Бормотов А. Н. Применение математических методов в строительном материаловедении. Пенза: ПГАСА, 1999. 204 с.
3. Гарькина И. А., Данилов А. М., Прошин А. П. Математические методы синтеза строительных материалов. Пенза: ПГАСА, 2001. 106 с.
4. Дворкин Л. И., Шамбан И. Б. Проектирование составов бетона с применением математического моделирования. К.: УМК ВО, 1992. 44 с.
5. Єршова Н. М., Деревянко В. Н., Тимченко Р. А., Шаповалова О. В. Обработка данных средствами Excel при планировании эксперимента: учеб. пособие для вузов Д.: ПГАСА, 2012. 350 с.
6. Єршова Н. М. Дисперсионный анализ данных наблюдений. Днепропетровск: ПГАСА, 2010. 80 с.
7. Красовский П. С. Исследование и оптимизация свойств строительных материалов с применением элементов математической статистики: Учебное пособие. Хабаровск: ДВГУПС, 2004. 128 с.
8. Митропольский А. К. Техника статистических вычислений. М.: Главная редакция физико-математической литературы издательства «Наука», 1971. 576 с.
9. Пінчук С. Й. Організація експерименту при моделюванні та оптимізації технічних систем: навч. посібник для студ. ВНЗ. Д.: Дніпро-VAL, 2009. 289 с.
10. Сивец С. А. Статистические методы в оценке недвижимости и бизнеса. Учебно-практическое пособие по статистике для оценщиков. Запорожье, 2001. 320 с.

## INFORMATION TECHNOLOGIES FOR DECISION-MAKING IN FORMATION UNIFORM DATABASE OF ANALOGUE OBJECTS OF THE RESEARCHED OBJECT

N. M. Yershova

Prydniprovsk State Academy of Civil Engineering and Architecture, 24-a, Chernyshevskiy St., Dnipro,  
49600, Ukraine, E-mail:nersova107@gmail.com

The experiment occupies a special place among the ways of obtaining information about the internal interconnections of phenomena in nature and technology. According to the increasing complexity of the processes and phenomena under research, the expense of equipment and the experiment increase. A large amount of experimental data is collected during the tests and requires processing and analysis. At the same time, the duration of analysis, comprehension of the test results and their consideration in order to adjust the new products' characteristics is significant. It is very important to select methods and means of processing experimental data during conducting an observation or experiment. Since 2001, many publications proved the effectiveness of the Excel analysis tool package for the experimental data processing. Despite this, in the observational and experimental data processing, the calculation of values for the equality of variances criteria and averages under formulas is still used in scientific and educational literature, and to select their critical values in the tables. Another problem is the sample formation of different sample sizes under direct calculation using formulas, even in Excel. This article proposes a variance analysis methodology of homogeneity for samples of different sizes, using the "Descriptive statistics" tool of the Excel analysis tool package. The "Descriptive Statistics" tool performs statistical processing of multivariate samples of various sizes and outputs the values of 13 parameters, including: mean, mode, median, kurtosis, asymmetry, variance, and sample size. Based on this information, it is easy to determine the estimated and critical value of statistics. The critical value of the criteria is determined using the statistical functions of the function wizard. The initial information of the "Univariate analysis of variance" tool contains the calculated and critical value - Fisher's test, which allows you to easily check the homogeneity of samples of equal volume. The technique significantly simplifies the procedure of forming a single database of objects-analogues of the object under study.

**Keywords:** experiment, data processing, variance analysis, sample formation, samples of different sizes, Excel analysis package.