

3. ДИСПЕРСІЙНИЙ АНАЛІЗ

Дисперсійний аналіз є сукупністю статистичних методів, призначених для перевірки гіпотез про зв'язок між певною ознакою та досліджуваними факторами, які не мають кількісного опису, а також для встановлення ступеня впливу факторів та їх взаємодії. У спеціальній літературі дисперсійний аналіз часто називають ANOVA (від англomовної назви Analysis of Variations). Вперше цей метод було розроблено Р. Фішером в 1925 р.

Факторами називають контрольовані чинники, що впливають на кінцевий результат. **Рівнем фактора**, або **способом обробки**, називають значення, що характеризують конкретний прояв цього фактора. Ці значення зазвичай подають у номінальній або порядковій шкалі вимірювань. Значення вимірюваної ознаки називають **відгуком**.

Часто вихідні значення факторів вимірюють у кількісних або порядкових шкалах. Тоді постає проблема групування вихідних даних у ряди спостережень, що відповідають приблизно однаковим значенням фактора. Якщо кількість груп взяти надмірно великою, то кількість спостережень у них може виявитися недостатньою для отримання надійних результатів. Якщо її взяти надмірно малою, це може призвести до втрати суттєвих особливостей впливу досліджуваного фактора на систему. Загальну методологію групування описано в розділі 1. Вибір конкретного способу групування даних залежить від їх обсягу і характеру варіювання значень фактора.

Кількість і розміри інтервалів при однофакторному аналізі найчастіше визначають за принципом рівних інтервалів або за принципом рівних частот. При багатofакторному аналізі застосовують три типи групування:

- групи з рівною кількістю спостережень;
- групи з різною кількістю спостережень;
- групи, кількості спостережень у яких відповідають певній пропорції.

При цьому існують певні особливості обробки даних, залежно від типу групування, які не розглядаються у цьому посібнику.

3.1. Однофакторний аналіз

Основною метою однофакторного аналізу зазвичай є оцінка величини впливу конкретного фактора на досліджуваний відгук. Іншою метою може бути порівняння двох або декількох факторів один з одним з метою визначення різниці їх впливу на відгук, яку часто називають **контрастом факторів**. Попереднім етапом є перевірка нульової гіпотези про відсутність будь-якого впливу досліджуваного фактора (факторів), тобто гіпотези про те, що зміни значень ознаки в порівнюваних вибірках є випадковими, і всі дані належать до однієї генеральної сукупності.

Якщо нульову гіпотезу відкидають, то наступним етапом є кількісне оцінювання впливу досліджуваного фактора і побудова довірчих інтервалів для отриманих характеристик. У випадку, коли нульова гіпотеза не може бути відкинута, зазвичай її приймають і роблять висновок про відсутність впливу. Але, якщо є підстави вважати, що такий вплив має бути присутнім (наприклад, це може випливати з теоретичних уявлень про об'єкт дослідження), то необхідно перевірити наявність інших факторів, що можуть його маскувати.

При **однофакторному дисперсійному аналізі** вихідні дані подають у вигляді таблиць, у яких кількість стовпчиків дорівнює кількості рівнів фактора, а кількість значень у кожному стовпчику – кількості спостережень при відповідному рівні фактора (табл. 3.1). Для різних рівнів фактора кількість спостережень може бути різною. При цьому виходять з припущення, що результати спостережень для різних рівнів є вибірками з нормально розподілених сукупностей, середні значення та дисперсії яких є однаковими і не залежать від рівнів. Завданням аналізу є перевірка нульової гіпотези про рівність середніх значень сукупностей, що розглядаються.

Таблиця 3.1

**Форма таблиці спостережень при проведенні
однофакторного дисперсійного аналізу**

Результати вимірювань	Рівні фактора			
	1	2	...	k
1	x_{11}	x_{12}	...	x_{1k}
2	x_{21}	x_{22}	...	x_{2k}
...
n_i	x_{n_i1}	x_{n_i2}	...	x_{n_ik}

Метод базується на основній тотожності дисперсійного аналізу, згідно з якою сума квадратів відхилень спостережень від загального середнього (**загальна варіація**) дорівнює:

$$\sum_{j=1}^k \sum_{i=1}^{n_j} (x_{ij} - \bar{x})^2 = \sum_{j=1}^k n_j (\langle x_j \rangle - \bar{x})^2 + \sum_{j=1}^k \sum_{i=1}^{n_j} (x_{ij} - \langle x_j \rangle)^2, \quad (3.1)$$

де $\bar{x} = \frac{1}{N} \sum_{j=1}^k \sum_{i=1}^{n_j} x_{ij}$ – загальне середнє; $\langle x_j \rangle = \frac{1}{n_j} \sum_{i=1}^{n_j} x_{ij}$ $j = 1, \dots, k$;

$N = \sum_{j=1}^k n_j$ – загальна чисельність; k – кількість вибірок; n_j ($j = 1, 2, \dots, k$) –

кількість елементів у j -й вибірці; $\langle x_j \rangle$ – середнє значення j -ї вибірки.

У правій частині (3.1) перший доданок (**факторна, або міжгрупова варіація**) є зваженою сумою квадратів відхилень групових середніх від

загального середнього. Він характеризує коливання значень, зумовлені фактором, на основі якого здійснено групування даних. Другий доданок (**залишкова, або внутрішньогрупова варіація**) є сумою квадратів відхилень спостережень від відповідних групових середніх. Він характеризує коливання значень досліджуваної ознаки, зумовлені неврахованими факторами або випадковими чинниками.

Сутність методу полягає в тому, що за умови правильності нульової гіпотези величини

$$\sigma_1^2 = \frac{1}{N-k} \sum_{j=1}^k \sum_{i=1}^{n_j} (x_{ij} - \langle x_j \rangle)^2 \quad (3.2)$$

та

$$\sigma_2^2 = \frac{1}{k-1} \sum_{j=1}^k n_j (\langle x_j \rangle - \bar{x})^2 \quad (3.3)$$

є незміщеними оцінками дисперсії похибок спостережень σ^2 і мають бути приблизно рівними одна одній. Перша з них є мірою варіації всередині вибірок і не пов'язана з припущенням про рівність середніх значень, тому $\sigma^2 \approx \sigma_1^2$ незалежно від справедливості нульової гіпотези. Друга оцінка характеризує варіацію між вибірками. При справедливості нульової гіпотези $\sigma_2^2 \approx \sigma^2$, а при її порушенні неї величина σ_2^2 є тим більшою, чим більше відхилення від неї.

Значення критерію розраховують за формулою:

$$F = \frac{(N-k) \sum_{j=1}^k n_j (\langle x_j \rangle - \bar{x})^2}{(k-1) \sum_{j=1}^k \sum_{i=1}^{n_j} (x_{ij} - \langle x_j \rangle)^2}. \quad (3.4)$$

Ця величина має F -розподіл Фішера з параметрами $k-1$ та $N-k$. Нульову гіпотезу відхиляють, якщо ймовірність $P(F \geq F^*)$, де F^* – значення, розраховане за емпіричними даними за формулою (3.3) є достатньо малою.

Непараметричним аналогом однофакторного дисперсійного аналізу є **ранговий однофакторний аналіз Краскела – Уолліса**. Він розроблений американськими математиком Вільямом Краскелом та економістом Вільсоном Уоллісом в 1952 р. Цей критерій призначено для перевірки нульової гіпотези про рівність ефектів впливу на досліджувані вибірки з невідомими, але рівними середніми. При цьому кількість вибірок має бути більшою ніж дві. Нульова гіпотеза полягає в тому, що k вибірок обсягами n_1, n_2, \dots, n_k отримані з однієї і тієї самої генеральної сукупності. Критерій

Краскела – Уолліса є узагальненням U -критерію Манна – Уїтні на випадок, коли кількість вибірок $k > 2$.

Рангові методи, у тому числі й метод Краскела – Уолліса, не передбачають нормальності розподілу результатів спостережень і можуть застосовуватися як для кількісних даних з невідомим законом розподілу, так і для порядкових ознак.

У табл. 3.1 замість спостережень заносять їх ранги r_{ij} , отримані шляхом впорядкування за зростанням усієї сукупності спостережень x_{ij} . При цьому одержують табл. 3.2.

Таблиця 3.2

Загальний вигляд вихідної таблиці рангового однофакторного аналізу

№ результату	№ вибірки			
	1	2	...	k
1	r_{11}	r_{12}	...	r_{1k}
2	r_{21}	r_{22}	...	r_{2k}
...
n_i	r_{n_1}	r_{n_2}	...	r_{n_k}

Для кожного рівня фактора, тобто для кожного стовпця, розраховують суму рангів $R_j = \sum_{i=1}^{n_j} r_{ij}$ або відповідні середні ранги $\langle R_j \rangle = \frac{1}{n_j} \sum_{i=1}^{n_j} r_{ij}$.

Для контролю можна використовувати тотожність:

$$\sum_{i=1}^k R_i = \frac{N(N+1)}{2}, \quad (3.5)$$

де $N = \sum_{i=1}^k n_i$ – загальна чисельність.

Якщо між стовпцями немає систематичної різниці, то останні будуть близькими до середнього рангу, розрахованого за усією сукупністю, який дорівнює $(N+1)/2$. Тому величини $\langle R_j \rangle - (N+1)/2$ мають бути відносно малими, якщо нульова гіпотеза є правильною.

Обчислення критерію здійснюють за формулами:

$$H = \frac{12}{N(N+1)} \sum_{j=1}^k \frac{R_j^2}{n_j} - 3(N+1), \quad (3.6)$$

або

$$H = \frac{12}{N(N+1)} \sum_{j=1}^k n_j \left(\langle R_j \rangle - \frac{N+1}{2} \right)^2. \quad (3.7)$$

За $n_i \geq 5$ й $k \geq 4$ статистика критерію асимптотично наближається до χ^2 -розподілу з кількістю степенів вільності $k - 1$. У цьому випадку нульову гіпотезу відхиляють на рівні значущості α , якщо $H > \chi_{1-\alpha}^2$, де $\chi_{1-\alpha}^2$ – квантиль рівня $1 - \alpha$ розподілу χ^2 з $k - 1$ степенем вільності. При $k = 2$ статистика Краскела – Уолліса стає еквівалентною статистиці W Уїлкоксона.

Якщо серед спостережень є рівні значення, описану вище схему аналізу можна застосовувати як наближену. Надійність її висновків буде тим нижчою, чим більшою є кількість збігів. Для підвищення надійності можна використовувати середні ранги, при цьому у випадку, коли вони не є цілими числами, їх не округляють. Якщо кількість збігів велика, використовують модифіковану форму статистики Краскела – Уолліса:

$$H' = \frac{H}{1 - \left(\sum_{j=1}^g \frac{T_j}{N^3 - N} \right)}, \quad (3.8)$$

де g – кількість груп спостережень, що збігаються;

$T_j = (t_j^3 - t_j)$; t_j – кількість спостережень, що збігаються в j -ї групі.

Критерій Джонкхієра (Джонкхієра – Терпстра) запропонований незалежно один від одного нідерландським математиком Т.Дж. Терпстрою в 1952 р. й британським психологом Е.Р. Джонкхієром в 1954 р. Його застосовують тоді, коли заздалегідь відомо, що наявні групи результатів упорядковані за зростанням впливу досліджуваного фактора, який вимірюють у порядковій шкалі. Таблиця даних має такий самий вигляд, як і в попередньому випадку. Будемо вважати, що її перший стовпчик відповідає найменшому рівню фактора, другий – наступному за величиною тощо, останній стовпчик відповідає найбільшому рівню. При виконанні таких припущень критерій Джонкхієра є більш потужним, ніж критерій Краскела – Уолліса, стосовно гіпотези про монотонний вплив фактора.

Спочатку для кожної пари вибірок з номерами u, v ($1 \leq u < v \leq k$), де k – кількість рівнів фактора, розраховують статистику Манна – Уїтні:

$$U_{u,v} = \sum_{\substack{i=1, \dots, n_u \\ j=1, \dots, n_v}} \varphi(x_{iu}, y_{jv}), \quad (3.9)$$

де

$$\varphi(x_i, y_j) = \begin{cases} 1 & (x_i < y_j); \\ 1/2 & (x_i = y_j); \\ 0 & (x_i > y_j). \end{cases} \quad (3.10)$$

Потім розраховують статистику Джонкхіера:

$$J = \sum_{1 \leq u \leq v \leq k} U_{u,v}. \quad (3.11)$$

Великі значення J свідчать проти гіпотези про однорідність вибірок.

Для вибірок великого обсягу статистика Джонкхіера апроксимується нормальним розподілом з параметрами:

$$MJ = \frac{1}{4} \left(N^2 - \sum_{j=1}^k n_j^2 \right); \quad DJ = \frac{1}{72} \left[N^2 (2N + 3) - \sum_{j=1}^k n_j^2 (2n_j + 3) \right]. \quad (3.12)$$

Свідченням проти гіпотези однорідності є великі, порівняно з відсотковими точками стандартного нормального розподілу, значення статистики $\frac{J - MJ}{\sqrt{DJ}}$.

М-критерій Бартлетта запропонований британським статистиком Маурісом Стівенсоном Бартлеттом в 1937 р. Його застосовують для перевірки нульової гіпотези про рівність дисперсій кількох нормальних генеральних сукупностей, з яких взяті досліджувані вибірки, що у загальному випадку мають різні обсяги (обсяг кожної вибірки має бути не менше чотирьох). Обчислення критерію здійснюють за формулою:

$$B = V / C, \quad (3.13)$$

де

$$V = k \ln \bar{s}^2 - \sum_{i=1}^{\ell} k_i \ln s_i^2;$$

$$C = 1 + \frac{1}{3(\ell - 1)} \left[\sum_{i=1}^{\ell} \frac{1}{k_i} - \frac{1}{k} \right],$$

$$\bar{s}^2 = \frac{\sum_{i=1}^{\ell} k_i s_i^2}{k} \text{ — зважене за кількістю степенів вільності середнє арифме-}$$

тичне стандартних відхилень вибірок; $k = \sum_{i=1}^{\ell} k_i$ — загальна чисельність; k_i — чисельність i -ї вибірки; ℓ — кількість вибірок; s_i^2 — стандартне відхилення i -ї вибірки. При великих n_j статистика критерію асимптотично наближається до χ^2 -розподілу з кількістю ступенів вільності $\ell - 1$.

Г-критерій Кокрена (Кочрена) запропонований американським статистиком Вільмом Геммелом Кочреном в 1941 р. Його використовують

для перевірки нульової гіпотези про рівність дисперсій k ($k \geq 2$) нормальних генеральних сукупностей за незалежними вибірками рівного обсягу. Значення критерію обчислюють за формулою:

$$G = \frac{\max_{1 \leq j \leq k} \sigma_j^2}{\sum_{j=1}^k \sigma_j^2}, \quad (3.14)$$

де σ_j^2 – дисперсія j -ї вибірки. Для вибірок рівного обсягу він є потужнішим за критерій Бартлетта. Критичні точки критерію Кокрена визначають за спеціальними таблицями.

Непараметричний **критерій Левене**, запропонований американським математиком Ховардом Левене в 1960 р. є альтернативою критерію Бартлетта в умовах, коли немає впевненості у тому, що досліджувані вибірки підпорядковуються нормальному розподілу. Розрахункове значення критерію обчислюють за формулою:

$$W = \frac{(N - k) \sum_{j=1}^k n_j (Z_j^* - \bar{Z})^2}{(k - 1) \sum_{j=1}^k \sum_{i=1}^{n_j} n_j (Z_{ij} - \bar{Z}_j)^2}, \quad (3.15)$$

де $Z_{ij} = |x_{ij} - \bar{x}_j|$; x_{ij} – значення i -го спостереження в j -ої вибірці; \bar{x}_j – середнє арифметичне спостережень, що потрапили до j -ої вибірки; \bar{Z} – загальне середнє арифметичне значень Z за усіма спостереженнями; \bar{Z}_j – середнє арифметичне Z за спостереженнями, що потрапили до j -ої вибірки. Розрахункове значення критерію порівнюють з відповідним квантилем F -розподілу з кількостями степенів вільності $(k - 1)$ та $(N - k)$.

В 1974 р. американські статистики Мортон Б. Браун та Алан Б. Форсайт запропонували більш робастний тест (**критерій Брауна – Форсайта**), який відрізняється від критерію Левене тим, що значення $Z_{ij} = |x_{ij} - \tilde{x}_j|$, де \tilde{x}_j – медіана спостережень, що потрапили до j -ої вибірки.

Розглянуті вище критерії дають змогу встановити різницю дисперсій сукупностей, але не дають можливості дати кількісну оцінку впливу фактора на досліджувану ознаку, а також встановити, для яких саме сукупностей дисперсії є різними.

Для встановлення кількісного впливу досліджуваного фактора часто застосовують **адитивну модель**, яка передбачає, що значення відгуку є сумою впливу фактора і незалежної від нього випадкової величини:

$$x_{ij} = a_j + \varepsilon_{ij} \quad (j = 1, \dots, k; i = 1, \dots, n), \quad (3.16)$$

де a_j – не випадкові невідомі величини, що визначаються значеннями рівнів фактора;

ε_{ij} – незалежні випадкові величини, які мають однаковий розподіл і відображають внутрішню мінливість, що не пов'язана із значеннями рівнів фактора.

Модель (3.16) можна записати у вигляді:

$$x_{ij} = \mu + \tau_j + \varepsilon_{ij} \quad (j = 1, \dots, k; i = 1, \dots, n), \quad (3.17)$$

де $\mu = \frac{1}{k} \sum_{j=1}^k a_j$ – середній рівень;

$\tau_j = a_j - \mu$ – відхилення від середнього рівня при j -му значенні рівня фактора.

У такій формі модель має на один невідомий параметр більше (середній рівень і k значень відхилень від нього), але кількість незалежних невідомих параметрів залишилася рівною k , оскільки відхилення пов'язані співвідношенням $\sum_{j=1}^k \tau_j = 0$.

Розглянемо різницю відгуків для двох значень рівня фактора, яку часто називають **зсувом**. Як оцінку зсуву можна взяти **медіану Ходжеса – Лемана**:

$$z_{ij} = \text{med} \{x_{ui} - x_{vj}\} \quad (u = 1, \dots, n_i; v = 1, \dots, n_j). \quad (3.18)$$

Вона має властивість: $z_{ij} = -z_{ji}$. Статистика z_{ij} може застосовуватися для оцінювання величини $\tau_i - \tau_j$. Її суттєвим недоліком є невиконання рівності: $z_{ij} = z_{ik} + z_{kj}$. Тому частіше використовують зважені скореговані оцінки зсуву (**оцінки Спетволя**):

$$W_{ij} = \bar{\Delta}_i - \bar{\Delta}_j, \quad (3.19)$$

де величини

$$\bar{\Delta}_i = \frac{\sum_{u=1}^k n_u z_{iu}}{N} \quad (i = 1, \dots, k) \quad (3.20)$$

відображають зсув i -ї вибірки відносно всіх інших, усереднений з ваговими коефіцієнтами n_1, \dots, n_k . Оцінки Спетволя задовольняють співвідношення $W_{ij} = W_{ik} + W_{kj}$. Але вони мають інший недолік: оцінка зсуву двох вибірок одна відносно одної залежить від усіх інших вибірок.

Якщо гіпотезу про рівність середніх відхиляють, то наступним кроком може бути визначення вибірок, для яких ця різниця є суттєвою. Для цього використовують метод лінійних контрастів.

Лінійним контрастом у моделі адитивного впливу фактора на відгук називають лінійну функцію середніх значень k незалежних нормальних вибірок з невідомими рівними дисперсіями:

$$L = \sum_{j=1}^k c_j m_j, \quad (3.21)$$

де c_i – відомі сталі, які задовольняють вимогу $\sum_{i=1}^k c_i = 0$; m_j – математичне сподівання для j -ої вибірки, яке для оцінювання лінійного контрасту заміняють відповідним середнім арифметичним.

Оцінку дисперсії лінійного контрасту розраховують за формулою:

$$s_L^2 = \sigma_1^2 \sum_{j=1}^k \frac{c_j^2}{n_j}, \quad (3.22)$$

де σ_1^2 визначається формулою (3.2).

Довірчим інтервалом для лінійного контрасту є:

$$L \pm s_L \sqrt{(k-1) F_{1-\alpha}(k-1; N-k)}. \quad (3.23)$$

Найпростішим прикладом лінійних контрастів є різниці $m_i - m_j$, яким відповідають значення: $c_i = 1$, $c_j = -1$, $c_u = 0$ при всіх $u \neq i, j$. Нульові гіпотези, що перевіряються, полягають у тому, що $m_i = m_j$ для всіх можливих пар вибірок. Їх приймають, якщо нульове значення потрапляє до відповідного довірчого інтервалу.

Для встановлення вибірок, що належать певній множині даних, дисперсії яких є різними, найчастіше застосовують **метод множинних порівнянь (Шеффе)**, запропонований американським статистиком Генрі Шеффе. Обчислення критерію при перевірці нульової гіпотези $L = L_0$ здійснюють за формулою:

$$t = \frac{\sum_{i=1}^k c_i \bar{x}_i - L_0}{\sqrt{M \sum_{i=1}^k \frac{c_i^2}{n_i}}}, \quad (3.24)$$

де N – загальна кількість; n_i – кількість елементів в i -ї вибірці;
 \bar{x}_i – середнє значення для i -ї вибірки;

$$M = \frac{1}{N - k} \sum_{i=1}^k \sum_{j=1}^{n_i} (x_{ij} - \bar{x}_i)^2$$

– середній квадратичний залишок. Розраховане значення критерію для порівняння з критичним необхідно брати за модулем.

3.2. Двофакторний аналіз

Двофакторний дисперсійний аналіз застосовують для пов'язаних нормально розподілених вибірок. Дані подають у вигляді табл. 3.3, у стовпчиках якої наводять дані, що відповідають певному рівню першого фактора, а в рядках – дані, що відповідають рівням другого. Таблиця даних має розмірність $n \times k$, де n і k – кількість рівнів першого та другого факторів, відповідно.

Таблиця 3.3

Таблиця даних двофакторного дисперсійного аналізу

Рівні фактора В	Рівні фактора А			
	1	2	...	k
1	x_{11}	x_{12}	...	x_{1k}
2	x_{21}	x_{22}	...	x_{2k}
...
n	x_{n1}	x_{n2}	...	x_{nk}

Основною відмінністю від таблиці однофакторного дисперсійного аналізу є можлива неоднорідність даних у стовпцях, якщо вплив другого фактора є суттєвим. На практиці часто використовують і складніші таблиці двофакторного дисперсійного аналізу, зокрема такі, у яких кожна комірка містить набір даних (повторні вимірювання), що відповідають фіксованим значенням рівнів обох факторів.

Для опису даних табл. 3.3 часто можна застосовувати адитивну модель, яка передбачає, що значення відгуку є сумою внесків окремо кожного із факторів b_i і t_j , а також незалежної від факторів випадкової компоненти ε_{ij} :

$$x_{ij} = b_i + t_j + \varepsilon_{ij}. \tag{3.25}$$

На практиці модель (3.24) часто подають в еквівалентному вигляді:

$$x_{ij} = \mu + \beta_i + \tau_j + \varepsilon_{ij}, \tag{3.26}$$

де $\mu = \frac{1}{kn} \sum_{j=1}^k x_{ij}$ – загальне середнє за всіма спостереженнями;

β_i і τ_j – відхилення від середнього, зумовлені факторами А і В, відповідно.

У випадку, коли випадкова компонента ε_{ij} підпорядковується нормальному розподілу з нульовим середнім і рівними для всіх i, j дисперсіями σ^2 застосовують **двофакторний дисперсійний аналіз (дисперсійний аналіз за двома ознаками)**.

Нульова гіпотеза може полягати в рівності ефектів стовпчиків між собою $H_{01} : \tau_1 = \tau_2 = \dots = \tau_k = 0$, або рівності ефектів рядків між собою $H_{02} : \beta_1 = \beta_2 = \dots = \beta_n = 0$, тобто в першому випадку припускають відсутність впливу фактора А, а у другому – фактора В.

Як і у випадку однофакторного дисперсійного аналізу в цьому разі розраховують дві оцінки дисперсії. При перевірці гіпотези H_{01} першу з них розраховують за формулою:

$$\sigma_1^2 = \frac{1}{(n-1)(k-1)} \sum_{i=1}^n \sum_{j=1}^k (x_{ij} - \langle x_i \rangle - \langle x_j \rangle + \langle x \rangle)^2, \quad (3.27)$$

де $\langle x \rangle = \frac{1}{nk} \sum_{i=1}^n \sum_{j=1}^k x_{ij}$ – загальне середнє за всіма спостереженнями;

$\langle x_i \rangle = \frac{1}{k} \sum_{j=1}^k x_{ij}$ – середнє за i -м рядком;

$\langle x_j \rangle = \frac{1}{n} \sum_{i=1}^n x_{ij}$ – середнє за j -м стовпчиком. Оцінка дисперсії σ_1^2 є незміщеною і не залежить від справедливості нульової гіпотези.

Другу оцінку розраховують за формулою:

$$\sigma_2^2 = \frac{n}{k-1} \sum_{j=1}^k (\langle x_j \rangle - \langle x \rangle)^2. \quad (3.28)$$

Вона є незміщеною лише за умови справедливості нульової гіпотези. Що більшою є різниця між результатами дії фактора А, то більшим є значення, розраховане за формулою (3.28).

Для перевірки справедливості гіпотези H_{01} необхідно розрахувати відношення дисперсій:

$$F = \frac{\sigma_2^2}{\sigma_1^2} = \frac{n(n-1)(k-1) \sum_{j=1}^k (\langle x_j \rangle - \langle x \rangle)^2}{(k-1) \sum_{i=1}^n \sum_{j=1}^k (x_{ij} - \langle x_i \rangle - \langle x_j \rangle + \langle x \rangle)^2}. \quad (3.29)$$

Воно має F -розподіл Фішера з кількостями степенів вільності $(k-1)$ і $(n-1)(k-1)$. Нульову гіпотезу приймають на рівні значущості α , якщо $F < F_{1-\alpha}$, де $F_{1-\alpha}$ – α -квантиль F -розподілу з відповідними кількостями степенів вільності.

Для перевірки нульової гіпотези H_{02} можна використовувати формулу (3.28), в якій необхідно попарно поміняти місцями величини n і k , а також i та j .

Якщо припущення, необхідні для застосування двофакторного дисперсійного аналізу, не виконуються, то використовують непараметричний **ранговий критерій Фрідмана (Фрідмана, Кендалла та Сміта)**, розроблений американським економістом Мілтоном Фрідманом наприкінці 1930 р. Цей критерій не залежить від типу розподілу. Передбачається лише, що розподіл величин ε_{ij} є однаковим і неперервним, а самі вони незалежні одна від одної.

При перевірці нульової гіпотези $H_{01} : \tau_1 = \tau_2 = \dots = \tau_k = 0$ вихідні дані подають у формі прямокутної матриці, у якій n рядків відповідають рівням фактора В, а k стовпців – рівням фактора А. Кожна комірка таблиці (блок) може бути результатом вимірювань параметрів на одному об'єкті або на групі об'єктів при сталих значеннях рівнів обох факторів. У цьому випадку відповідні дані подають як середні значення певного параметра за всіма вимірюваннями або об'єктами досліджуваної вибірки. Для застосування критерію в таблиці вихідних даних необхідно перейти від безпосередніх результатів вимірювань до їх рангів. Ранжирування здійснюють за кожним рядком окремо, тобто величини x_{ij} впорядковують для кожного фіксованого значення i , отримуючи при цьому k значень відповідних рангів r_{ij} . Це дає можливість усунути вплив фактора В, значення якого для кожного рядка є однаковим.

Обчислення критерію здійснюють за формулою:

$$S = \left[\frac{12}{nk(k+1)} \sum_{j=1}^k \left(\sum_{i=1}^n r_{ij} \right)^2 \right] - 3n(k+1). \quad (3.30)$$

Якщо необхідно перевірити нульову гіпотезу $H_{02} : \beta_1 = \beta_2 = \dots = \beta_n = 0$, то вихідні дані необхідно ранжувати за стовпчиками і повторити описану вище процедуру із заміною n на k і навпаки.

При справедливості нульової гіпотези і $n \rightarrow \infty$ S -статистика Фрідмана асимптотично наближається до статистики χ^2 з $k-1$ степенем вільності, тому нульову гіпотезу можна прийняти на рівні значущості α , якщо $S < \chi_{1-\alpha}^2(k-1)$.

Критерій Пейджа (L -критерій), запропонований американським статистиком Є.Б. Пейджем в 1963 р., призначений для перевірки нульової гіпотези $H_{01} : \tau_1 = \tau_2 = \dots = \tau_k = 0$ або $H_{02} : \beta_1 = \beta_2 = \dots = \beta_n = 0$, проти альтернатив $\tau_1 \leq \tau_2 \leq \dots \leq \tau_k$ або, відповідно, $\beta_1 \leq \beta_2 \leq \dots \leq \beta_n$, у яких принаймні одна

із нерівностей є строгою. Для впорядкованих альтернатив він є потужнішим за критерій Фрідмана. Значення критерію обчислюють за формулами:

$$L_1 = \sum_{j=1}^k jr_j; \quad L_2 = \sum_{i=1}^n ir_i, \quad (3.31)$$

де $r_j = \sum_{i=1}^n r_{ij}$, $r_i = \sum_{j=1}^k r_{ij}$.

Для великих вибірок застосовують апроксимацію статистики Пейджа:

$$L_1^* = \frac{L - nk(k+1)^2/4}{\sqrt{n(k^3 - k)^2/144(k-1)}}; \quad L_2^* = \frac{L - nk(n+1)^2/4}{\sqrt{k(n^3 - n)^2/144(n-1)}}, \quad (3.32)$$

які за умови справедливості відповідних нульових гіпотез підпорядковуються стандартному нормальному розподілу.

У разі, коли в рядках вихідної таблиці є однакові значення, необхідно використовувати середні ранги. При цьому точність висновків буде тим гіршою, чим більшою є кількість таких збігів.

Q-критерій Кокрена запропонований В. Кочреном в 1937 р. Його використовують у випадках, коли групи однорідних суб'єктів піддаються впливам, кількість яких перевищує два, і для яких можливі два варіанти відгуків – умовно-негативний (0) та умовно-позитивний (1). Нульова гіпотеза полягає в рівності ефектів впливу.

Значення критерію розраховують за формулою:

$$Q = \frac{(c-1) \left(c \sum_{j=1}^c T_j^2 - \left(\sum_{j=1}^c T_j \right)^2 \right)}{c \sum_{i=1}^r T_i - \sum_{i=1}^r T_i^2}, \quad (3.33)$$

де $T_j = \sum_{i=1}^r x_{ij}$ ($j = 1, 2, \dots, c$) – суми стовпців; $T_i = \sum_{j=1}^c x_{ij}$ ($i = 1, 2, \dots, r$) – суми рядків; c – кількість стовпців (вибірок); r – кількість рядків (параметрів). Довірчий рівень визначається функцією розподілу χ^2 з кількістю степенів вільності, яка дорівнює $c - 1$.

Двофакторний дисперсійний аналіз дає можливість визначити існування ефектів обробки, проте не дає змоги встановити, для яких саме стовпців існує цей ефект.

При вирішенні цієї проблеми застосовують метод множинних порівнянь Шеффе для пов'язаних вибірок. Значення критерію розраховують за формулою:

$$t = \frac{\left(\sum_{i=1}^r c_i \bar{x}_i \right)^2}{(r-1)S \sum_{i=1}^r c_i^2}, \quad (3.34)$$

де c_i ($i = 1, 2, \dots, r$) – константи;

$$S = \sum_{i=1}^r \sum_{j=1}^c x_{ij}^2 - \frac{T^2}{rc} - \text{залишковий середній квадрат};$$

$$T = \sum_{i=1}^r \sum_{j=1}^c x_{ij} - \text{загальна сума};$$

c – кількість стовпців (вбірок);

r – кількість рядків (параметрів). Довірчий рівень визначається функцією розподілу Фішера з параметрами $(r-1)$ та $(r-1)(c-1)$ при дослідженні ефекту рядків і $(c-1)$ та $(r-1)(c-1)$ при дослідженні ефекту стовпців.

3.3. Приклад виконання дисперсійного аналізу

За допомогою вбудованого генератора випадкових чисел електронних таблиць MS Excel сформуємо чотири нормально розподілені вибірки обсягом по 200 елементів з параметрами: $\bar{x}_1 = 50$; $s_1 = 5$; $\bar{x}_2 = 49$; $s_2 = 5$; $\bar{x}_3 = 51$; $s_3 = 6$; $\bar{x}_4 = 49,5$; $s_4 = 5,5$.

На першому етапі перевіряємо відповідність вибірок нормальному закону розподілу. Оскільки приклади такої перевірки було розглянуто в попередньому розділі й ураховуючи, що ми будували саме нормально розподілені послідовності, будемо вважати, що вони задовольняють цю вимогу. Тому можемо використовувати параметричні методи аналізу.

Спочатку перевіримо нульову гіпотезу про рівність середніх значень досліджуваних вибірок. При розрахунку безпосередньо за формулою (3.4) маємо: $n_1 = n_2 = n_3 = n_4 = 200$; $N = 800$; $k = 4$; $F = 1,544$. Кількості степенів вільності $k - 1 = 3$; $N - k = 796$. За допомогою функції =FРАСПОБР(0,05;3;796), що знаходиться у бібліотеці статистичних функцій електронних таблиць MS Excel, визначаємо критичне значення для рівня значущості 0,05 $F_{кр} = 2,62$. Бачимо, що розрахункове значення критерію є меншим, ніж критичне. За допомогою статистичної функції = FРАСП(F7;3;796) ми також можемо визначити рівень значущості, якому відповідає розрахункове значення критерію $F^* = 0,202$. Бачимо, що імовірність

припущення помилки першого роду при відхиленні нульової гіпотези є досить високою. Тому приймаємо нульову гіпотезу про рівність середніх значень.

В пакеті аналізу електронних таблиць MS Excel є вбудована процедура “Однофакторний дисперсійний аналіз”. На рис. 3.1 показано результати її застосування, для згенерованих вибірок.

Однофакторный дисперсионный анализ						
ИТОГИ						
Группы	Счет	Сумма	Среднее	Дисперсия		
Столбец 1	200	10068,6	50,343	26,802066		
Столбец 2	200	9874,115	49,37058	23,848429		
Столбец 3	200	9974,982	49,87491	32,535911		
Столбец 4	200	9877,329	49,38664	27,908546		
Дисперсионный анализ						
Источник вариации	SS	df	MS	F	P-Значение	F критическое
Между группами	128,6168	3	42,87225	1,5436256	0,20180279	2,616088979
Внутри групп	22107,9	796	27,77374			
Итого	22236,51	799				

Рис. 3.1. Вікно виведення результатів однофакторного дисперсійного аналізу MS Excel

Крім визначених раніше параметрів, таблиця результатів застосування процедури однофакторного дисперсійного аналізу містить також середні арифметичні й дисперсії вибірок, значення міжгрупової та внутрішньогрупової дисперсії (MS) й відповідних варіацій (SS), які збігаються з величинами, обчислюваними за формулами (3.1–3.3), та деякі інші дані.

Для визначення однорідності вибірок нам необхідно перевірити також нульову гіпотезу про рівність їх дисперсій. Оскільки у нашому випадку обсяги вибірок є рівними, цю гіпотезу перевірятимемо за критерієм Кокрена. В електронних таблицях MS Excel немає вбудованих засобів для розрахунку критеріїв Бартлетта й Кокрена. Тому розрахункове значення обчислюємо за формулою (3.14). Воно дорівнює $G = 0,293$.

Критичне значення критерію Кокрена для рівня значущості 0,05 можна визначити з відповідних таблиць. Для випадку, що розглядається $G_c < 0,05$ (при тих самих умовах $G_c \approx 0,0495$, якщо обсяги вибірок дорівнюють 120), тобто є значно меншим, ніж розрахункове значення. Тому ми маємо відхилити нульову гіпотезу про рівність дисперсій і зробити висновок, що досліджувані вибірки не є однорідними.

У пакеті SPSS також передбачено можливість здійснення однофакторного дисперсійного аналізу. Для цього необхідно використовувати вікно Analyze/Compare means/One way ANOVA. Вихідні вибірки розміщуємо по-

слідовно одна за одною в одному й тому самому стовпчику, як значення змінної VAR00001. Як значення змінної VAR00002 беремо номери вибірок, до яких належить відповідне значення VAR00001. Додатково ми можемо отримати основні показники описової статистики для кожної з вибірок та усієї сукупності в цілому; перевірити однорідність дисперсій (тест Левена); визначити, які саме вибірки істотно відрізняються від інших (тест Дункана), й отримати іншу корисну інформацію. Якщо для вибірок, що розглядаються у прикладі, замовити у пункті меню “Options” розрахунок параметрів описової статистики й перевірку однорідності дисперсій, а пункті “Post Hoc” – виконання тесту Дункана, одержимо результати, показані на рис. 3.2.

Descriptives

VAR00001

	N	Mean	Std. Deviation	Std. Error	95% Confidence Interval for Mean		Minimum	Maximum	Between-Component Variance
					Lower Bound	Upper Bound			
1,00	200	50,3430	5,17707	,36607	49,6211	51,0649	36,15	64,18	
2,00	200	49,3706	4,88349	,34531	48,6896	50,0515	36,24	60,87	
3,00	200	49,8749	5,70403	,40334	49,0796	50,6703	34,38	62,45	
4,00	200	49,3866	5,28285	,37355	48,6500	50,1233	35,89	64,92	
Total	800	49,7438	5,27546	,18652	49,3777	50,1099	34,38	64,92	
Model									
Fixed Effects			5,27008	,18633	49,3780	50,1095			
Random Effects				23150	49,0071	50,4805			,07549

Test of Homogeneity of Variances

VAR00001

Levene Statistic	df1	df2	Sig.
1,132	3	796	,335

ANOVA

VAR00001

	Sum of Squares	df	Mean Square	F	Sig.
Between Groups	128,617	3	42,872	1,544	,202
Within Groups	22107,895	796	27,774		
Total	22236,512	799			

VAR00001

	VAR00002	N	Subset for alpha = .05
			1
Duncan ^a	2,00	200	49,3706
	4,00	200	49,3866
	3,00	200	49,8749
	1,00	200	50,3430
	Sig.		,093

Means for groups in homogeneous subsets are displayed.
a. Uses Harmonic Mean Sample Size = 200,000.

Рис. 3.2. Результати, отримані у пакеті SPSS

Бачимо, що результати однофакторного дисперсійного аналізу збігаються з даними, отриманими за допомогою MS Excel, а результати перевірки однорідності дисперсій є точнішими й не потребують додаткового застосування таблиць. Крім того, додатково визначено, немає поділу вибірок на групи, що суттєво відрізняються одна від іншої за середнім значенням.

3.4. Приклад виконання рангового однофакторного аналізу

Нехай ми маємо чотири вибірки, сформовані за допомогою пакета аналізу MS Excel як суміш 100 елементів нормально розподіленої вибірки з параметрами $\bar{x} = 20$, $s = 3$ та рівномірно розподілених вибірок обсягом по 100 елементів кожна заданих, відповідно, на відрізках: [17; 22], [18; 22], [18; 22] та [17; 23]. В електронних таблицях MS Excel немає вбудованих засобів для реалізації рангового однофакторного аналізу Краскела – Уолліса. Але його неважко здійснити за допомогою наявних функцій. Спочатку необхідно перетворити таблицю вихідних даних у таблицю значень рангів. Для цього можна використати функцію РАНГ (). Вікно задання її параметрів показано на рис. 3.3.

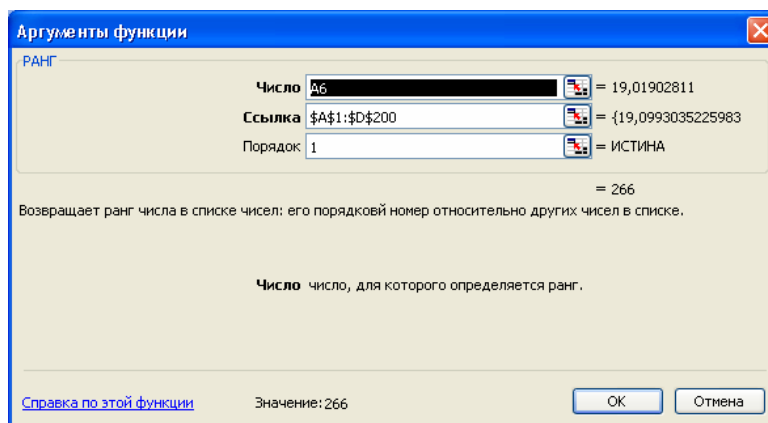


Рис. 3.3. Вікно задання параметрів функції РАНГ ()

У комірці “Число” вказуємо для якого саме значення таблиці вихідних даних необхідно обчислити ранг. У комірці “Ссылка” даємо посилання на весь діапазон, що містить вихідні значення (це посилання має бути абсолютним). У комірці “Порядок” зазначаємо порядок ранжирування: 0 – за убутанням, інше число – за зростанням. Після цього за формулами (3.6) або (3.7) розраховуємо значення критерію. У нашому випадку воно дорівнює 7,22. Як критичне візьмемо значення оберненої функції розподілу χ^2 , яке можна визначити за допомогою функції $= \text{ХИ2ОБР} (0,05;3) = 7,82$. Її аргументами є рівень значущості (0,05) та кількість степенів вільності (у нашому випадку 3). Бачимо, що розрахункове значення критерію дещо менше, ніж критичне. Тому немає підстав для відхилення нульової гіпотези про однорідність досліджуваних вибірок.

У пакеті SPSS реалізовано можливість перевірки однорідності вибірок числових даних, розподіл яких відрізняється від нормального закону, а також порядкових даних за допомогою рангового однофакторного аналізу Краскела – Уолліса.

Для реалізації цієї процедури необхідно увійти до діалогового вікна Analyze/Nonparametric Tests/K Independent Samples й обрати в ньому тест Краскела – Уолліса. Додатково ми можемо здійснити перевірку за медіанним тестом і критерієм Джонкхієра – Терпстри. Результати наведено на рис. 3.4–3.6. Бачимо, що за результатами всіх тестів рівень значущості, що відповідає розрахунковому значенню критерію, є вищим, ніж 0,05. Тому ми можемо прийняти нульову гіпотезу про однорідність досліджуваних вибірок на рівні значущості 0,05.

Ranks

	VAR00002	N	Mean Rank
VAR00001	1,00	200	370,98
	2,00	200	394,42
	3,00	200	432,20
	4,00	200	404,40
	Total	800	

Test Statistics^{a,b}

	VAR00001
Chi-Square	7,221
df	3
Asymp. Sig.	,065

a. Kruskal Wallis Test

b. Grouping Variable: VAR00002

Рис. 3.4. Результати перевірки однорідності вибірок за критерієм Краскела – Уолліса

Frequencies

		VAR00002			
		1,00	2,00	3,00	4,00
VAR00001	> Median	89	95	110	106
	<= Median	111	105	90	94

Test Statistics^b

	VAR00001
N	800
Median	20,0603
Chi-Square	5,640 ^a
df	3
Asymp. Sig.	,131

a. 0 cells (.0%) have expected frequencies less than 5. The minimum expected cell frequency is 100,0.

b. Grouping Variable: VAR00002

Рис. 3.5. Результати перевірки однорідності вибірок за медіанним тестом

Jonckheere-Terpstra Test

	VAR00001
Number of Levels in VAR00002	4
N	800
Observed J-T Statistic	127124,00
Mean J-T Statistic	120000,00
Std. Deviation of J-T Statistic	3654,221
Std. J-T Statistic	1,950
Asymp. Sig. (2-tailed)	,051

a. Grouping Variable: VAR00002

Рис. 3.6. Результати перевірки однорідності вибірок за критерієм Джонкхієра – Терпстрі

Контрольні питання

1. Які завдання вирішують за допомогою дисперсійного аналізу?
2. Що називають факторами й відгуками у дисперсійному аналізі? Наведіть приклади.
3. Що називають рівнем фактора? Наведіть приклади.
4. Як визначають кількість і розміри інтервалів в однофакторному дисперсійному аналізі?
5. Які типи групування використовують у багатфакторному дисперсійному аналізі?
6. Якими є основні умови застосування однофакторного дисперсійного аналізу?
7. Доведіть основну тотожність дисперсійного аналізу.
8. Що являє собою факторна варіація та яку властивість даних вона характеризує?
9. Що являє собою залишкова варіація та яку властивість даних вона характеризує?
10. Які властивості даних характеризують оцінки дисперсії похибок, що використовуються у дисперсійному аналізі?
11. Які завдання вирішують за допомогою рангового однофакторного аналізу Краскела – Уолліса?
12. За яких умов можна використовувати ранговий однофакторний аналіз Краскела – Уолліса?
13. Який критерій є аналогом рангового однофакторного аналізу Краскела – Уолліса при порівнянні двох вибірок?
14. Які завдання вирішують за допомогою критерію Джонкхієра? За яких умов його доцільно використовувати?

15. Які завдання вирішують за допомогою критерію Бартлетта? За яких умов його можна використовувати?
16. Які завдання вирішують за допомогою G -критерію Кокрена? За яких умов його можна використовувати?
17. Яким є загальний вигляд адитивної моделі кількісного впливу досліджуваного фактора на відгук?
18. Що називають зсувом у дисперсійному аналізі? Як можна оцінити зсув кількісно?
19. Що називають лінійним контрастом в адитивній моделі? Наведіть приклади лінійних контрастів.
20. Які фактори визначають дисперсію лінійного контрасту?
21. Які завдання вирішують за допомогою методу множинних порівнянь Шеффе?
22. Які завдання вирішують за допомогою двофакторного дисперсійного аналізу? За яких умов можна використовувати цей метод?
23. Які властивості даних перевіряють за допомогою рангового критерію Фрідмана? За яких умов можна використовувати цей критерій?
24. Які властивості даних перевіряють за допомогою критерію Пейджа? За яких умов можна використовувати цей критерій?
25. Які властивості даних перевіряють за допомогою Q -критерію Кокрена? За яких умов можна використовувати цей критерій?