Розділ 11

ОСНОВИ ДИСПЕРСІЙНОГО АНАЛІЗУ

У практиці обробки результатів спостережень часто виникає питання про те, наскільки істотний вплив чинить зміну деякого фактора або групи факторів на вимірювану величину. Наприклад, можна досліджувати спільний вплив декількох економічних факторів, які не підлягають кількісному вимірюванню щодо досліджуваного економічного показника, можна оцінити вплив властивостей сировини на показники якості продукції, кількості внесених добрив на врожайність і т. д.

У розділі 10 було розглянуто перевірку гіпотези про рівність математичних сподівань двох сукупностей при невідомих і однакових дисперсіях. Однак на практиці часто виникає необхідність узагальнення задачі, тобто перевірки при заданому рівні значущості за вибірковим середнім нульової гіпотези про рівність математичних сподівань p $(p > 2)$ сукупностей, у яких також дисперсії невідомі та однакові. Для ефективного розв'язку значущості таких задач застосовується новий підхід, який зоснований на порівнянні дисперсій, і тому називається *дисперсійним аналізом* (в основному розроблений англійським математиком-статистом Р. А. Фішером). Коротко сутність цього аналізу зводиться до розчленовування загальної дисперсії ознаки на компоненти, які обумовлені впливом конкретних факторів, і перевірці гіпотез про значущість їхнього впливу. Під *рівнем фактора* розуміють деяку його міру або стан. Моделі дисперсійного аналізу, залежно від числа факторів, класифікують на *одно-*, *двофакторні* й т. д. За метою дослідження виділяють наступні моделі: *детермінована* – тут рівні всіх факторів заздалегідь фіксовані та перевіряють саме їхній вплив, *випадкова* – тут рівні кожного фактора отримані як випадкова вибірка з генеральної сукупності рівнів фактора, і *змішана* – тут рівні одних факторів заздалегідь фіксовані, а рівні інших – випадкова вибірка.

В основі однофакторного дисперсійного аналізу лежить наступна теоретико-імовірнісна модель:

$$x_{ij} = \overline{x} + \alpha_j + \varepsilon_{ij}, \tag{11.1}$$

де x_{ij} – значення ознаки *X*, отримане при *i*-му випробуванні на *j*-му рівні фактора; $\overline{x}$ – загальна середня величина ознаки X; α_j – ефект фактора на j-му рівні; ε_{ij} – випадковий компонент, що впливає на значення ознаки X в i-му спостереженні на j-му рівні фактора. Приймається припущення, що ε_{ij} взаємно незалежні та мають нормальний закон розподілу $N(0;\sigma^2)$.

11.1 Однофакторний дисперсійний аналіз

1. Однакове число випробувань на різних рівнях. Нехай на кількісну нормально розподілену ознаку (випадкову величину) X впливає фактор F, що має p постійних рівнів. Будемо припускати, що число спостережень (випробувань) на кожному рівні однакове і дорівнює q.

Нехай спостерігалося $n = pq$ значень x_{ij} ознаки X, де i – номер випробування $(i = \overline{1, q})$, j – номер рівня фактора $(j = \overline{1, p})$.

Результати спостережень наведено в табл. 11.1.

Таблиця 11.1

Номер випробування	Рівні фактору F_j			
	F_1	F_2	…	F_p
1	x_{11}	x_{12}	…	x_{1p}
2	x_{21}	x_{22}	…	x_{2p}
…	…	…	…	…
q	x_{q1}	x_{q2}	…	x_{qp}
Групова середня	$\overline{x}_1$	$\overline{x}_2$	…	$\overline{x}_p$

Необхідно перевірити, чи чинить істотний вплив деякий якісний фактор F, який має p рівнів $F_1, F_2, ..., F_p$, на досліджувану ознаку X.

Якщо вважати, що елементи стовпців табл. 11.1, позначених F_j $(j = \overline{1, p})$, без елементів групової середньої, є числовими значеннями випадкових величин X_1, $X_2, ..., X_p$, які мають нормальний закон розподілу з математичними сподіваннями відповідно $m_1,\ m_2,...,m_p$ та однаковими дисперсіями, то дана задача зводиться до перевірки нульової гіпотези $H_0 : m_1 = m_2 = ... = m_p$, здійснюваної в дисперсійному аналізі.

Позначимо через $\overline{x}_j$ – *групове середнє j*-ї групи

$$\overline{x}_j = \frac{1}{q}\sum_{i=1}^{q} x_{ij}\ , \tag{11.2}$$

а через $\overline{x}$ – *загальне середнє*

$$\overline{x} = \frac{1}{pq}\sum_{j=1}^{p}\sum_{i=1}^{q} x_{ij} = \frac{1}{p}\sum_{j=1}^{p}\overline{x}_j\ , \tag{11.3}$$

оскільки $\sum_{i=1}^{q} x_{ij} = q\overline{x}_j$.

Розглянемо суму квадратів відхилень спостережуваних значень x_{ij} від загальної середньої $\overline{x}$:

$$\sum_{j=1}^{p}\sum_{i=1}^{q}(x_{ij}-\overline{x})^2 = \sum_{j=1}^{p}\sum_{i=1}^{q}(x_{ij}-\overline{x}_j+\overline{x}_j-\overline{x})^2 = \sum_{j=1}^{p}\sum_{i=1}^{q}(x_{ij}-\overline{x}_j)^2 +$$

$$+2\sum_{j=1}^{p}\sum_{i=1}^{q}(x_{ij}-\overline{x}_j)(\overline{x}_j-\overline{x}) + \sum_{j=1}^{p}\sum_{i=1}^{q}(\overline{x}_j-\overline{x})^2\ , \tag{11.4}$$

причому

$$\sum_{j=1}^{p}\sum_{i=1}^{q}(x_{ij}-\overline{x}_j)(\overline{x}_j-\overline{x}) = \sum_{j=1}^{p}(\overline{x}_j-\overline{x})\sum_{i=1}^{q}(x_{ij}-\overline{x}_j) = 0\ ,$$

оскільки

$$\sum_{i=1}^{q}(x_{ij}-\overline{x}_j) = \sum_{i=1}^{q} x_{ij} - q\overline{x}_j = \sum_{i=1}^{q} x_{ij} - q\frac{1}{q}\sum_{i=1}^{q} x_{ij} = 0\ .$$

Тому, узявши до уваги, що $\sum_{j=1}^{p}\sum_{i=1}^{q}(\overline{x}_j-\overline{x})^2 = q\sum_{j=1}^{p}(\overline{x}_j-\overline{x})^2,$ ми можемо основну тотожність (11.4) записати у наступному вигляді:

$$\sum_{j=1}^{p}\sum_{i=1}^{q}(x_{ij}-\overline{x})^2 = q\sum_{j=1}^{p}(\overline{x}_j-\overline{x})^2 + \sum_{j=1}^{p}\sum_{i=1}^{q}(x_{ij}-\overline{x}_j)^2\ ,$$

або в скороченому вигляді

$$S_{\text{заг}} = S_{\text{факт}} + S_{\text{зал}}\ . \tag{11.5}$$

У (11.5) $S_{\text{заг}}$ – *загальна сума* квадратів відхилень спостережуваних значень від загальної середньої, $S_{\text{факт}}$ – *факторна сума* квадратів відхилень групових середніх від загальної середньої, котра характеризує розсіювання між групами, $S_{\text{зал}}$ – *залишкова сума* квадратів відхилень спостережуваних значень групи від своєї групової середньої, котра характеризує розсіювання всередині групи.

Практично залишкову суму знаходять за рівністю

$$S_{\text{зал}} = S_{\text{заг}} - S_{\text{факт}}, \tag{11.6}$$

яке випливає з (11.5).

Одержимо більш зручні формули для розрахунків $S_{\text{заг}}$ і $S_{\text{факт}}$:

$$S_{\text{заг}} = \sum_{j=1}^{p}\sum_{i=1}^{q}(x_{ij}-\overline{x})^2 = \sum_{j=1}^{p}\sum_{i=1}^{q}(x_{ij}^2 - 2x_{ij}\overline{x} + \overline{x}^2) = \sum_{j=1}^{p}\sum_{i=1}^{q}x_{ij}^2 -$$

$$-2\overline{x}\sum_{j=1}^{p}\sum_{i=1}^{q}x_{ij} + \sum_{j=1}^{p}\sum_{i=1}^{q}\overline{x}^2 = |11.3| = \sum_{j=1}^{p}\sum_{i=1}^{q}x_{ij}^2 - 2\frac{1}{pq}\left(\sum_{j=1}^{p}\sum_{i=1}^{q}x_{ij}\right)^2 +$$

$$+\frac{1}{pq}\left(\sum_{j=1}^{p}\sum_{i=1}^{q}x_{ij}\right)^2 = \sum_{j=1}^{p}P_j - \frac{1}{n}\left(\sum_{j=1}^{p}R_j\right)^2; \tag{11.7}$$

$$S_{\text{факт}} = q\sum_{j=1}^{p}(\overline{x}_j-\overline{x})^2 = q\sum_{j=1}^{p}(\overline{x}_j^2 - 2\overline{x}_j\overline{x} + \overline{x}^2) = q\sum_{j=1}^{p}\overline{x}_j^2 - 2q\overline{x}\sum_{j=1}^{p}\overline{x}_j +$$

$$+qp\overline{x}^2 = |11.2;\ 11.3| = \frac{1}{q}\sum_{j=1}^{p}\left(\sum_{i=1}^{q}x_{ij}\right)^2 - 2q\frac{1}{p}\sum_{j=1}^{p}\overline{x}_j\sum_{j=1}^{p}\overline{x}_j +$$

$$+\frac{q}{p}\left(\sum_{j=1}^{p}\overline{x}_j\right)^2 = \frac{1}{q}\sum_{j=1}^{p}\left(\sum_{i=1}^{q}x_{ij}\right)^2 - \frac{q}{p}\left(\sum_{j=1}^{p}\overline{x}_j\right)^2 = \frac{1}{q}\sum_{j=1}^{p}\left(\sum_{i=1}^{q}x_{ij}\right)^2 -$$

$$-\frac{1}{pq}\left(\sum_{j=1}^{p}\sum_{i=1}^{q}x_{ij}\right)^2 = \frac{1}{q}\sum_{j=1}^{p}R_j^2 - \frac{1}{pq}\left(\sum_{j=1}^{p}R_j\right)^2, \tag{11.8}$$

де $P_j = \sum_{i=1}^{q}x_{ij}^2$ – сума квадратів значень ознаки на рівні F_j; $R_j = \sum_{i=1}^{q}x_{ij}$ – сума значень ознаки на рівні F_j.

Зауваження 1. Обчислення за формулами (11.7) і (11.8) можна спростити, якщо виконати заміну змінних $y_{ij} = x_{ij} - C$, де C наближено дорівнює загальній середній.

Підставляючи до цих формул $x_{ij} = y_{ij} + C$ і $R_j = \sum_{i=1}^{q} x_{ij} =$ $= \sum_{i=1}^{q}(y_{ij} + C) = \sum_{i=1}^{q} y_{ij} + qC = T_j + qC$, де $T_j = \sum_{i=1}^{q} y_{ij}$, з урахуванням позначення $Q_j = \sum_{i=1}^{q} y_{ij}^2$, одержимо

$$S_{\text{заг}} = \sum_{j=1}^{p}\sum_{i=1}^{q}(y_{ij} + C)^2 - \frac{1}{pq}\left(\sum_{j=1}^{p}(T_j + qC)\right)^2 = \sum_{j=1}^{p}\sum_{i=1}^{q}(y_{ij}^2 + 2y_{ij}C + C^2) -$$

$$-\frac{1}{pq}\left(\sum_{j=1}^{p} T_j + pqC\right)^2 = \sum_{j=1}^{p}(Q_j + 2CT_j + qC^2) - \frac{1}{pq}\left(\left(\sum_{j=1}^{p} T_j\right)^2 +\right.$$

$$\left. +2pqC\sum_{j=1}^{p} T_j + p^2q^2C^2\right) = \sum_{j=1}^{p} Q_j + 2C\sum_{j=1}^{p} T_j + pqC^2 - \frac{1}{pq}\left(\sum_{j=1}^{p} T_j\right)^2 -$$

$$-2C\sum_{j=1}^{p} T_j - pqC^2 = \sum_{j=1}^{p} Q_j - \frac{1}{n}\left(\sum_{j=1}^{p} T_j\right)^2 \tag{11.9}$$

$$S_{\text{факт}} = \frac{1}{q}\sum_{j=1}^{p} T_j^2 - \frac{1}{pq}\left(\sum_{j=1}^{p} T_j\right)^2. \tag{11.10}$$

Зауваження 2. Якщо спостережувані значення x_{ij} – десяткові дроби з k знаками після коми, то доцільно перейти до чисел $y_{ij} = 10^k x_{ij} - C$, де C – наближене середнє значення чисел $10^k x_{ij}$. Хоча при цьому факторна і залишкова дисперсії збільшуються в 10^{2k} разів, їхнє відношення не змінюється.

У розкладанні (11.5) полягає основна ідея дисперсійного аналізу. Однак у ньому аналізуються не власне суми квадратів відхилень, а так звані середні квадрати, які є незміщеними оцінками відповідних дисперсій, які отримують за допомогою ділення сум квадратів відхилень на відповідне число ступенів вільності.

Нагадаємо, що *число ступенів вільності* визначається як загальне число ступенів спостережень мінус число єднальних їхніх рівнянь. Тому число ступенів вільності загальної дисперсії дорівнює $pq - 1 = n - 1$, оскільки один ступінь вільності губиться при визначенні середньої. Аналогічне число ступенів вільності факторної дисперсії дорівнює $p - 1$, тому що групові середні варіюють навколо однієї загальної середньої. Нарешті, число ступенів вільності залишкової дисперсії дорівнює $n-p$, тому що використовуються p співвідношень при обчисленні p групових середніх $\overline{x}_j$.

Використовуючи отримані значення сум квадратів і чисел ступенів вільності, можна обчислити незміщені оцінки трьох дисперсій:

$$s^2_{\text{заг}} = \frac{S_{\text{заг}}}{n-1}; \qquad s^2_{\text{факт}} = \frac{S_{\text{факт}}}{p-1}; \qquad s^2_{\text{зал}} = \frac{S_{\text{зал}}}{n-p}. \qquad (11.11)$$

Зауваження 3. Число ступенів вільності залишкової дисперсії $pq-p$ дорівнює різниці між числами ступенів вільності загальної та факторної дисперсій. Дійсно, $pq-1-(p-1) = pq-p = n-p$.

Перевірка гіпотези H_0 про рівність групових математичних сподівань ґрунтується на порівнянні дисперсій $s^2_{\text{факт}}$ і $s^2_{\text{зал}}$. Виявляється, якщо гіпотеза H_0 правильна, то правильна і гіпотеза про рівність факторної та залишкової дисперсій, яка перевіряється за критерієм Фішера-Снедекора (див. **10.3**)

$$F = s^2_{\text{факт}} / s^2_{\text{зал}}, \qquad (11.12)$$

який має ступеня вільності $k_1 = p-1$ і $k_2 = n-p$.

Якщо нульова гіпотеза про рівність групових середніх помилкова, то помилкова і гіпотеза про рівність $s^2_{\text{факт}}$ і $s^2_{\text{зал}}$. Правильні також і обернені твердження.

Наведемо схему перевірки нульової гіпотези у випадку правобічної критичної області:

$$\left.\begin{array}{c} p; \quad q; \quad n = pq \\ k_1 = p-1;\ k_2 = n-p \\ \alpha \end{array}\right\} \to H_0 : m_1 = m_2 = \ldots = m_p \to F_{\text{спос}} =$$

$$= s^2_{\text{факт}} \big/ s^2_{\text{зал}} \to H_1 : (\text{правобіч.крит.обл.}) \to F_{\text{кр}}(\alpha; k_1; k_2) \xrightarrow{Д\,8} x^{\text{пр}}_{\text{кр}} \to$$

$$\to \begin{cases} F_{\text{спос}} < x^{\text{пр}}_{\text{кр}} \to \mathbf{П}, \\ F_{\text{спос}} > x^{\text{пр}}_{\text{кр}} \to \mathbf{В}. \end{cases} \qquad (11.13)$$

Зауваження 4. Якщо $s^2_{\text{факт}} < s^2_{\text{зал}}$, то звідси випливає правильність нульової гіпотези H_0 і, значить, немає необхідності використовувати критерій F.

Зауваження 5. Якщо немає впевненості у правильності припущення про рівність дисперсій розглянутих p сукупностей, то це припущення варто перевірити попередньо, наприклад за критерієм Кочрена.

Приклад 1. Зроблено по чотири випробування на кожному із чотирьох рівнів. Результати випробувань наведено в табл. 11.2. Методом дисперсійного аналізу при рівні значущості $\alpha = 0,05$ перевірити нульову гіпотезу про рівність групових середніх. Передбачається, що вибірки витягнуті з нормальних сукупностей.

○ Знайдемо за формулами

$$\overline{x}_j = \frac{1}{4}\sum_{i=1}^{4} x_{ij}, \qquad \overline{\sigma}_j^2 = \frac{1}{4}\sum_{i=1}^{4}(x_{ij} - \overline{x}_j)^2, \qquad j = \overline{1,4}$$

групові вибіркові середні та дисперсії, результати обчислень яких занесемо до табл. 11.2.

Таблиця 11.2

Номер випробування	Рівні фактора F_j			
i	F_1	F_2	F_3	F_4
1	140	150	148	150
2	144	149	149	155
3	142	152	146	154
4	145	150	147	152
Групове середнє, $\overline{x}_j$	142,75	150,25	147,5	152,75
Вибіркова дисперсія, $\overline{\sigma}_j^2$	3,6875	1,1875	1,25	3,6875

Перш ніж проводити дисперсійний аналіз, переконаємося в тому, що нульова гіпотеза $H_0 : \sigma_1^2 = \sigma_2^2 = \sigma_3^2 = \sigma_4^2$ про рівність групових дисперсій для вибірок, які витягнуті з нормальних генеральних сукупностей, не суперечить результатам спостережень. Оскільки вибірки містять однаковий об'єм $q = 4$, то при перевірці нульової гіпотези доцільно скористатися критерієм Кочрена.

Для цього:

1. За формулою (8.21) знайдемо виправлені вибіркові дисперсії:

$$S_1^2 = \frac{\bar{\sigma}_1^2 q}{q-1} = \frac{3,6875 \cdot 4}{3} = 4,917; \qquad S_2^2 = 1,583;$$

$$S_3^2 = 1,667; \qquad S_4^2 = 4,917\,.$$

2. Знайдемо спостережуване значення критерію Кочрена за формулою (10.37). Маємо

$$G_{\text{спос}} = S_{\max}^2 \big/ (S_1^2 + S_2^2 + S_3^2 + S_4^2) = 4,917/13,084 \approx 0,376\,.$$

3. Знайдемо з таблиці (Д. 9), за рівнем значущості $\alpha = 0,05$, числом ступенів вільності $k = 4 - 1 = 3$ і числом вибірок $l = 4$ критичну точку $G_{\text{кр}}(0,05;3;4) = g_{\text{кр}} = 0,6841$.

4. Оскільки $G_{\text{спос}} < g_{\text{кр}}$, то гіпотезу H_0 приймаємо.

Гіпотеза, яка перевіряється в дисперсійному аналізі, має вигляд: $H_0 : m_1 = m_2 = m_3 = m_4$, де m_i – математичне сподівання i-ї генеральної сукупності.

Для спрощення розрахунку віднімемо $C = 148$ з кожного спостережуваного значення: $y_{ij} = x_{ij} - 148$.

Складемо розрахункову табл. 11.3.

Таблиця 11.3

Номер випробування	Рівні фактора F_j								Підсумковий стовпець
i	F_1		F_2		F_3		F_4		
	y_{i1}	y_{i1}^2	y_{i2}	y_{i2}^2	y_{i3}	y_{i3}^2	y_{i4}	y_{i4}^2	
1	–8	64	2	4	0	0	2	4	–
2	–4	16	1	1	1	1	7	49	
3	–6	36	4	16	–2	4	6	36	
4	–3	9	2	4	–1	1	4	16	
$Q_j = \sum_{i=1}^{4} y_{ij}^2$	–	125	–	25	–	6	–	105	$\sum_{j=1}^{4} Q_j = 261$
$T_j = \sum_{i=1}^{4} y_{ij}$	–21	–	9	–	–2	–	19	–	$\sum_{j=1}^{4} T_j = 5$
T_j^2	441	–	81	–	4	–	361	–	$\sum_{j=1}^{4} T_j^2 = 887$

Користуючись табл. 11.3 і враховуючи, що число рівнів фактора $p = 4$, число випробувань на кожному рівні $q = 4$, знайдемо за формулами (11.9) і (11.10) загальну і факторну суми квадратів відхилень:

$$S_{\text{заг}} = 261 - 25/16 = 259,4375;$$

$$S_{\text{факт}} = 887/4 - 25/16 = 220,1875 .$$

Знайдемо залишкову суму квадратів відхилень:

$$S_{\text{зал}} = S_{\text{заг}} - S_{\text{факт}} = 39,25 .$$

Факторну і залишкову дисперсії знайдемо за формулами (11.11):

$$s^2_{\text{факт}} = 220,1875/3 \approx 73,4;$$

$$s^2_{\text{зал}} = 39,25/(16-4) \approx 3,27 .$$

Перевірку гіпотези H_0 про рівність групових математичних сподівань наведемо за схемою (11.13):

$$\left.\begin{array}{r} p=4; \quad q=4; \\ n=16; \quad k_1=3 \\ k_2=12; \alpha=0,05 \end{array}\right\} \to H_0 : m_1 = m_2 = m_3 = m_4 \to F_{\text{спос}} =$$

$$= s^2_{\text{факт}} \Big/ s^2_{\text{зал}} = 22,45 \to H_1 : (\text{правобіч. крит. обл.}) \to$$

$$\to F_{\text{кр}}(0,05;3;12) \xrightarrow{Д8} x^{\text{пр}}_{\text{кр}} = 3,49 \to F_{\text{спос}} > x^{\text{пр}}_{\text{кр}} \to \mathbf{В}$$

Таким чином, групові середні істотно розрізняються. Якщо потрібно зрівняти середні попарно, то варто скористатися критерієм Стьюдента. ●

2. Неоднакове число випробувань на різних рівнях. Нехай число випробувань на різних рівнях різне, а саме: зроблено q_1 випробувань на рівні F_1, q_2 випробувань – на рівні F_2, …, q_p випробувань – на рівні F_p. У цьому випадку загальну суму квадратів відхилень також знаходять за формулою (11.7), де

$P_1=\sum_{i=1}^{q_1}x_{i1}^2$, $P_2=\sum_{i=1}^{q_2}x_{i2}^2$, …, $P_p=\sum_{i=1}^{q_p}x_{ip}^2$ – суми квадратів спостережуваних значень ознаки відповідно на рівнях $F_1,\ F_2,...,F_p$;

$R_1=\sum_{i=1}^{q_1}x_{i1}$; $R_2=\sum_{i=1}^{q_2}x_{i2}$, …, $R_p=\sum_{i=1}^{q_p}x_{ip}$ – суми спостережуваних значень ознаки відповідно на рівнях $F_1,\ F_2,...,F_p$;

$n=q_1+q_2+...+q_p$ – об'єм вибірки, а факторну суму квадратів відхилень знаходять за формулою

$$S_{\text{факт}}=\sum_{j=1}^{p}(R_j^2/q_j)-\left(\sum_{j=1}^{p}R_j\right)^2\Big/n. \qquad (11.14)$$

Якщо для спрощення обчислень вводиться заміна змінних $y_{ij}=x_{ij}-C$, де C наближено дорівнює загальної середньої, то формула для загальної суми квадратів відхилень має вигляд (11.9),
де

$$Q_1=\sum_{i=1}^{q_1}y_{i1}^2,\quad Q_2=\sum_{i=1}^{q_2}y_{i2}^2,\ \ ...,\ Q_p=\sum_{i=1}^{q_p}y_{ip}^2,$$

$$T_1=\sum_{i=1}^{q_1}y_{i1},\quad T_2=\sum_{i=1}^{q_2}y_{i2},\ \ ...,\ T_p=\sum_{i=1}^{q_p}y_{ip},$$

а формула для факторної суми відхилень має вигляд

$$S_{\text{факт}}=\sum_{j=1}^{p}(T_j^2/q_j)-\left(\sum_{j=1}^{p}T_j\right)^2\Big/n. \qquad (11.15)$$

Інші обчислення роблять, як і у випадку однакового числа випробувань за формулами (11.6) і (11.11).

Приклад 2. Зроблено 22 випробування, з яких 7 на першому рівні фактора, 6 – на другому, 5 – на третьому і 4 – на четвертому. Результати випробувань наведено в табл. 11.4. Методом дисперсійного аналізу при рівні значущості $\alpha=0,05$ перевірити нульову гіпотезу про рівність групових середніх. Передбачається, що вибірки витягнуті з нормальних сукупностей з однаковими дисперсіями.

Таблиця 11.4

Номер випробування	Рівні фактора F_j			
i	F_1	F_2	F_3	F_4
1	1,3	1,4	1,44	1,27
2	1,27	1,3	1,4	1,05
3	1,21	1,28	1,28	1,24
4	1,09	1,27	1,28	1,22
5	1,03	1,24	1,06	–
6	1,01	1,08	–	–
7	1,09	–	–	–

○ Для спрощення розрахунку помножимо кожне спостережуване значення x_{ij} на 100 і віднімемо $C = 122$ (зауваження 2). Складемо розрахункову таблицю 11.5.

Таблиця 11.5

Номер випробування	Рівні фактору F_j								Підсумковий стовпець
i	F_1		F_2		F_3		F_4		
	y_{i1}	y_{i1}^2	y_{i2}	y_{i2}^2	y_{i3}	y_{i3}^2	y_{i4}	y_{i4}^2	
1	8	64	18	324	22	484	5	25	–
2	5	25	8	64	18	324	–17	289	
3	–1	1	6	36	6	36	2	4	
4	–13	169	5	25	6	36	0	0	
5	–19	361	2	4	–16	256	–	–	
6	–21	441	–14	196	–	–	–	–	
7	–13	169	–	–	–	–	–	–	
$Q_j = \sum_{i=1}^{q_j} y_{ij}^2$	–	1230	–	649	–	1136	–	318	$\sum_{j=1}^{4} Q_j = 3333$
$T_j = \sum_{i=1}^{q_j} y_{ij}$	–54	–	25	–	36	–	–10	–	$\sum_{j=1}^{4} T_j = -3$
T_j^2	2916	–	625	–	1296	–	100	–	–

Використовуючи табл. 11.5, знайдемо загальну і факторну суми квадратів відхилень:

$$S_{\text{заг}} = \sum_{j=1}^{4} Q_j - \left(\sum_{j=1}^{4} T_j\right)^2 \Big/ n = 3333 - 9/22 \approx 3332,6\,;$$

$$S_{\text{факт}} = \sum_{j=1}^{4} \left(T_j^2 / q_j\right) - \left(\sum_{j=1}^{4} T_j\right)^2 \Big/ n = 416,571 + 104,167 +$$

$$+259,2 + 25 - 9/22 = 804,938 - 9/22 \approx 804,529\,.$$

Знайдемо залишкову суму квадратів відхилень:

$$S_{\text{зал}} = S_{\text{заг}} - S_{\text{факт}} = 2528,071\,.$$

Факторну і залишкову дисперсії знайдемо за формулами (11.11):

$$s_{\text{факт}}^2 = S_{\text{факт}} / (p-1) = 804,529/3 \approx 268,176\,;$$

$$s_{\text{зал}}^2 = S_{\text{зал}} / (n-p) = 2528,071/18 \approx 140,448\,.$$

Зрівняємо $s_{\text{факт}}^2$ і $s_{\text{зал}}^2$ за критерієм Фішера-Снедекора, використовуючи схему (11.13)

$$\left.\begin{array}{l} p = 4;\ n = 22; \\ k_1 = 3;\ k_2 = 18 \\ \quad \alpha = 0,05 \end{array}\right\} \to H_0 : m_1 = m_2 = m_3 = m_4 \to F_{\text{спос}} = s_{\text{факт}}^2 \Big/ s_{\text{зал}}^2 =$$

$$= 268,176/140,448 \approx 1,9094 \to H_1 : (\text{правобіч. крит. обл.}) \to$$

$$\to F_{\text{кр}}(0,05;3;18) \xrightarrow{Д8} x_{\text{кр}}^{\text{пр}} = 3,16 \to F_{\text{спос}} < x_{\text{кр}}^{\text{пр}} \to \mathbf{П}\,. \bullet$$

11.2 Поняття про двофакторний дисперсійний аналіз

В основі двофакторного дисперсійного аналізу лежить наступна теоретико-імовірнісна модель:

$$x_{ijk} = \overline{x} + \alpha_i + \beta_j + \gamma_{ij} + \varepsilon_{ijk}\,, \qquad (11.16)$$

де x_{ijk} – значення ознаки *X* у *k*-му спостереженні на *i*-му рівні фактора *A* і на *j*-му рівні фактора *B*;

$\overline{x}$ – загальна середня величина ознаки *X*;

α_i – ефект впливу фактора A на i-му рівні;

β_j – ефект впливу фактора B на j-му рівні;

γ_{ij} – ефект спільного впливу факторів (позначимо $A\times B$);

ε_{ijk} – незалежні нормально розподілені випадкові компоненти $N(0;\sigma^2)$, які являють собою відхилення ознаки від відповідних середніх.

Нехай необхідно виявити вплив двох факторів (A і B) і їхніх взаємодій на деяку ознаку X. Спостереження проводяться при фіксованих рівнях факторів A і B. Оскільки для кожного поєднання факторів спостереження повторюється n разів, то відповідно одержимо n значень ознаки.

Дані спостережень зручно подати у вигляді таблиці 11.6, у якій значення ознаки X позначене через x_{ijk},

де

$i=\overline{1,q}$ (q – число спостережень фактора A);

$j=\overline{1,p}$ (p – число спостережень фактора B);

$k=\overline{1,n}$ (k – порядковий номер спостереження для кожного сполучення рівнів).

Таблиця 11.6

Рівень фактора A	Рівень фактора B				Сума P_i
	B_1	B_2	...	B_p	
A_1	x_{111}, x_{112} x_{113}, ..., x_{11n}	x_{121}, x_{122} x_{123}, ..., x_{12n}	...	x_{1p1}, x_{1p2} x_{1p3}, ..., x_{1pn}	$\sum_{j=1}^{p}\sum_{k=1}^{n}x_{1jk}$
A_2	x_{211}, x_{212} x_{213}, ..., x_{21n}	x_{221}, x_{222} x_{223}, ..., x_{22n}	...	x_{2p1}, x_{2p2} x_{2p3}, ..., x_{2pn}	$\sum_{j=1}^{p}\sum_{k=1}^{n}x_{2jk}$
...	...	...	...	...	...
A_q	x_{q11}, x_{q12} x_{q13}, ..., x_{q1n}	x_{q21}, x_{q22} x_{q23}, ..., x_{q2n}	...	x_{qp1}, x_{qp2} x_{qp3}, ..., x_{qpn}	$\sum_{j=1}^{p}\sum_{k=1}^{n}x_{qjk}$
Сума R_j	$\sum_{i=1}^{q}\sum_{k=1}^{n}x_{i1k}$	$\sum_{i=1}^{q}\sum_{k=1}^{n}x_{i2k}$	...	$\sum_{i=1}^{q}\sum_{k=1}^{n}x_{ipk}$	$\sum_{i=1}^{q}\sum_{j=1}^{p}\sum_{k=1}^{n}x_{ijk}$

За даними таблиці одержимо наступні середні:
загальна середня

$$\overline{x} = \frac{1}{qpn}\sum_{i=1}^{q}\sum_{j=1}^{p}\sum_{k=1}^{n} x_{ijk}\ ; \tag{11.17}$$

середні за рядками

$$\overline{P}_i = \frac{1}{pn} P_i = \frac{1}{pn}\sum_{j=1}^{p}\sum_{k=1}^{n} x_{ijk},\ i = \overline{1,q}\ ; \tag{11.18}$$

середні за стовпцями

$$\overline{R}_j = \frac{1}{qn} R_j = \frac{1}{qn}\sum_{i=1}^{q}\sum_{k=1}^{n} x_{ijk},\ j = \overline{1,p}\ ; \tag{11.19}$$

середні для кожного окремого блоку таблиці

$$\overline{Q}_{ij} = \frac{1}{n} Q_{ij} = \frac{1}{n}\sum_{k=1}^{n} x_{ijk}\ . \tag{11.20}$$

Суми квадратів відхилень від загальної середньої $\sum_{i=1}^{q}\sum_{j=1}^{p}\sum_{k=1}^{n}(x_{ijk} - \overline{x})^2$ розкладемо на складові. Для цього подамо $x_{ijk} - \overline{x}$ в еквівалентному вигляді

$$x_{ijk} - \overline{x} = (\overline{R}_j - \overline{x}) + (\overline{P}_i - \overline{x}) + (\overline{Q}_{ij} - \overline{R}_j - \overline{P}_i + \overline{x}) + (x_{ijk} - \overline{Q}_{ij})\,.$$

Оскільки всі перехресні добутки, при піднесенні правої частини останнього виразу до квадрата, дорівнюють нулю, то в кінцевому рахунку одержимо

$$\sum_{i=1}^{q}\sum_{j=1}^{p}\sum_{k=1}^{n}(x_{ijk} - \overline{x})^2 = \sum_{i=1}^{q}\sum_{j=1}^{p}\sum_{k=1}^{n}(\overline{R}_j - \overline{x})^2 + \sum_{i=1}^{q}\sum_{j=1}^{p}\sum_{k=1}^{n}(\overline{P}_i - \overline{x})^2 +$$

$$+\sum_{i=1}^{q}\sum_{j=1}^{p}\sum_{k=1}^{n}(\overline{Q}_{ij} - \overline{R}_j - \overline{P}_i + \overline{x})^2 + \sum_{i=1}^{q}\sum_{j=1}^{p}\sum_{k=1}^{n}(x_{ijk} - \overline{Q}_{ij})^2\,,$$

або в скороченому вигляді

$$S = S_1 + S_2 + S_3 + S_4\,. \tag{11.21}$$

У (11.21) чотири складові: сума квадратів, пов'язана із впливом фактора B, фактора A, їхньої взаємодії, та складова, яка характеризує залишкову суму квадратів (суму квадратів усередині кожного блоку таблиці).

Загальне число ступенів вільності, очевидно, дорівнює $N-1$, де N – загальна кількість спостережень $(N = qpn)$.

Число ступенів вільності між стовпцями дорівнює $p-1$, між рядками $q-1$, для взаємодії $(p-1)(q-1)$, усередині осередків $pq(n-1) = N - pq$. Для перевірки знайдемо суму

$$p-1+q-1+ (p-1)(q-1) + +N - pq = N-1 .$$

Схему дисперсійного аналізу, яка зоснована на отриманих вище даних, наведена в табл. 11.7.

Таблиця 11.7

Фактор	Сума квадратів	Число ступенів вільності	Оцінка дисперсії
B	$S_1 = qn\sum_{j=1}^{p}(\overline{R}_j - \overline{x})^2$	$p-1 = k_1$	$s_1^2 = \frac{S_1}{p-1}$
A	$S_2 = pn\sum_{i=1}^{q}(\overline{P}_i - \overline{x})^2$	$q-1 = k_2$	$s_2^2 = \frac{S_2}{q-1}$
$A \times B$	$S_3 = n\sum_{i=1}^{q}\sum_{j=1}^{p}(\overline{Q}_{ij} - \overline{R}_j - \overline{P}_i + \overline{x})^2$	$(p-1)\times$ $\times(q-1) = k_3$	$s_3^2 = \frac{S_3}{(p-1)(q-1)}$
Залишковий	$S_4 = \sum_{i=1}^{q}\sum_{j=1}^{p}\sum_{k=1}^{n}(x_{ijk} - \overline{Q}_{ij})^2$	$N - pq = k_4$	$s_4^2 = \frac{S_4}{N-pq}$
Сума	$S = \sum_{i=1}^{q}\sum_{j=1}^{p}\sum_{k=1}^{n}(x_{ijk} - \overline{x})^2$	$N-1$	$s^2 = \frac{S}{N-1}$

Оскільки при проведенні аналізу нас цікавить вплив кожного фактора порізно і вплив їхньої взаємодії, то знаходимо відповідно три значення:

$$F_B = s_1^2 \big/ s_4^2 ; \quad F_A = s_2^2 \big/ s_4^2 ; \quad F_{AB} = s_3^2 \big/ s_4^2 . \qquad (11.22)$$

Необхідні для аналізу суми квадратів відхилень (11.21) можемо одержати також за наступними формулами:

$$S_1 = \frac{1}{nq}\sum_{j=1}^{p} R_j^2 - \frac{G^2}{N}; \quad (11.23)$$

$$S_2 = \frac{1}{np}\sum_{i=1}^{q} P_i^2 - \frac{G^2}{N}; \quad (11.24)$$

$$S_3 = \frac{1}{n}\sum_{i=1}^{q}\sum_{j=1}^{p} Q_{ij}^2 - \frac{1}{nq}\sum_{j=1}^{p} R_j^2 - \frac{1}{np}\sum_{i=1}^{q} P_i^2 + \frac{G^2}{N}; \quad (11.25)$$

$$S_4 = \sum_{i=1}^{q}\sum_{j=1}^{p}\left(\sum_{k=1}^{n} x_{ijk}^2 - \frac{Q_{ij}^2}{n}\right), \quad (11.26)$$

де

$$G = \sum_{i=1}^{q}\sum_{j=1}^{p}\sum_{k=1}^{n} x_{ijk}, \qquad Q_{ij} = \sum_{k=1}^{n} x_{ijk}, \qquad F_{ij} = \sum_{k=1}^{n} x_{ijk}^2 - \frac{Q_{ij}^2}{n}. \quad (11.27)$$

При рівні значущості α визначаємо критичні точки:

$$F_{\text{кр}}(\alpha; k_1; k_4) = x_{\text{кр}}^{B}; \qquad F_{кр}(\alpha; k_2; k_4) = x_{\text{кр}}^{A}; \qquad F_{\text{кр}}(\alpha; k_3; k_4) = x_{\text{кр}}^{AB}.$$

Якщо:

1) $F_B > x_{\text{кр}}^{B}$, то нульова гіпотеза про відсутність впливу фактора B відхиляється;

2) $F_A > x_{\text{кр}}^{A}$, то нульова гіпотеза про відсутність впливу фактора A відхиляється;

3) $F_{AB} > x_{\text{кр}}^{AB}$, то нульова гіпотеза про відсутність спільного впливу факторів A і B відхиляється.

Приклад 3. При рівні значущості $\alpha = 0,05$ перевірити, чи існує вплив факторів A, B, а також їхнього спільного впливу на ознаку X, для результатів випробувань, наведених у табл. 11.8.

Таблиця 11.8

Рівень фактора A	Рівень фактора B		
	B_1	B_2	B_3
A_1	10; 8; 7; 10	8; 12; 14; 12	15; 8; 10; 10
A_2	12; 8; 8; 7	12; 13; 11; 14	13; 15; 12; 10

○ Використовуючи табл. 11.8 і формули (11.23) ÷ (11.27), складемо табл. 11.9 і обчислимо суми квадратів відхилень $S_i \ (i=\overline{1,4})$.

Таблиця 11.9

$A \backslash B$	B_1		B_2		B_3		$\sum_{j=1}^{3} Q_{ij}^2$	P_i	P_i^2
	x_{i1k}	x_{i1k}^2	x_{i2k}	x_{i2k}^2	x_{i3k}	x_{i3k}^2			
A_1	10	100	8	64	15	225	–	–	–
	8	64	12	144	8	64			
	7	49	14	196	10	100			
	10	100	12	144	10	100			
Q_{1j}	35	–	46	–	43	–	–	124	15376
Q_{1j}^2	1225	–	2116	–	1849	–	5190	–	–
$\sum_{k=1}^{4} x_{1jk}^2$	–	313	–	548	–	489	–	–	–
F_{1j}	6,75		19		26,75		$\sum_{j=1}^{3} F_{1j} = 52,5$		
A_2	12	144	12	144	13	169	–	–	–
	8	64	13	169	15	225			
	8	64	11	121	12	144			
	7	49	14	196	10	100			
Q_{2j}	35	–	50	–	50	–	–	135	18225
Q_{2j}^2	1225	–	2500	–	2500	–	6225	–	–
$\sum_{k=1}^{4} x_{2jk}^2$	–	321	–	630	–	638	–	–	–
F_{2j}	14,75		5		13		$\sum_{j=1}^{3} F_{2j} = 32,75$		
R_j	70	–	96	–	93	–	–	$G = 259$	$\sum_{i=1}^{2} P_i^2 = 33601$
R_j^2	4900	–	9216	–	8649	–	$\sum_{j=1}^{3} R_j^2 = 22765$		

З табл. 11.9 маємо:

$$S_1 = \frac{1}{8}\sum_{j=1}^{3} R_j^2 - \frac{G^2}{24} = \frac{1}{8}\cdot 22765 - \frac{529}{24} \approx 50,58\,;$$

$$S_2 = \frac{1}{12}\sum_{i=1}^{2} P_i^2 - \frac{G^2}{24} = \frac{1}{12}\cdot 33601 - \frac{529}{24} \approx 5,04\,;$$

$$S_3 = \frac{1}{4}\sum_{i=1}^{2}\sum_{j=1}^{3} Q_{ij}^2 - \frac{1}{8}\sum_{j=1}^{3} R_j^2 - \frac{1}{12}\sum_{i=1}^{2} P_i^2 + \frac{G^2}{24} =$$

$$= \frac{1}{4}(5190 + 6225) - 2845,625 - 2800,083 + 2795,0417 \approx 3,083\,;$$

$$S_4 = \sum_{i=1}^{2}\sum_{j=1}^{3}\left(\sum_{k=1}^{4} x_{ijk}^2 - Q_{ij}^2/4\right) = 52,5 + 32,75 = 85,25\,.$$

Дані дисперсійного аналізу наведемо в табл. 11.10.

Таблиця 11.10

Джерело варіації	S_i	Число ступенів вільності	Оцінка дисперсії
Фактор *B*	50,58	2	25,29
Фактор *A*	5,04	1	5,04
$A \times B$	3,083	2	1,54
Залишкова варіація	85,25	18	4,74
Σ	143,95	23	–

Визначимо спостережувані значення критерію:

$$F_{\text{спос}}^{B} = 25,29/4,74 \approx 5,34; \qquad F_{\text{спос}}^{A} = 5,04/4,74 \approx 1,06;$$

$$F_{\text{спос}}^{AB} = 1,54/4,74 \approx 0,32$$

і порівняємо їх з відповідними критичними значеннями, знайденими з таблиці (Д. 8):

$$F_{\text{кр}}(0,05;2;18) = x_{\text{кр}}^{B} = 3,55;$$

$$F_{\text{кр}}(0,05;1;18) = x_{\text{кр}}^{A} = 4,41; \qquad F_{\text{кр}}(0,05;2;18) = x_{\text{кр}}^{AB} = 3,55\,.$$

Оскільки $F_{\text{спос}}^{B} > x_{\text{кр}}^{B}$, то нульова гіпотеза про відсутність впливу фактора *B* відхиляється, і варто зробити висновок про значущість впливу цього фактора.

Оскільки $F_{\text{спос}}^{A} > x_{\text{кр}}^{A}$ і $F_{\text{спос}}^{AB} > x_{\text{кр}}^{AB}$, то немає підстави для відхилення відповідних нульових гіпотез. ●

Питання для самоперевірки

1. У чому сутність дисперсійного аналізу?
2. Наведіть класифікацію моделей дисперсійного аналізу за числами факторів і за метою дослідження.
3. Запишіть теоретико-імовірнісну модель для однофакторного дисперсійного аналізу.
4. Що називається груповою і загальною середніми? Наведіть формули для їхнього обчислення.
5. Вивести основну тотожність для суми квадратів відхилень спостережуваних значень x_{ij} від загальної середньої $\overline{x}$.
6. Вивести формули для розрахунку $S_{заг}$ і $S_{факт}$ через P_j і R_j.
7. Вивести формули для розрахунку $S_{заг}$ і $S_{факт}$ через Q_j і T_j.
8. Наведіть формули для обчислення незміщених оцінок трьох дисперсій: $s^2_{заг}$; $s^2_{факт}$; $s^2_{зал}$.
9. Наведіть схему для перевірки нульової гіпотези про рівність групових середніх.
10. Наведіть основні та спрощені формули для розрахунку $S_{заг}$ і $S_{факт}$ при неоднаковому числі випробувань на різних рівнях.
11. Запишіть теоретико-імовірнісну модель для двофакторного дисперсійного аналізу.
12. Наведіть формули для обчислення середніх при двофакторному дисперсійному аналізі.
13. Дайте характеристику чотирьох складових розкладання суми квадратів відхилень від загальної середньої.
14. Наведіть формули для розрахунку S_1, S_2, S_3 і S_4.

Індивідуальні завдання 11. За відповідною вибіркою *B* (табл. 8.5) скласти таблицю результатів 22-х випробувань, з яких п’ять значень для першого рівня фактора F_1 вибираються з номерами 1 ÷ 5, шість значень для другого рівня фактора F_2 вибираються з номерами 6 ÷ 11, сім значень для третього рівня фактора F_3 вибираються з номерами 12 ÷ 18 і чотири значення для четвертого рівня фактора F_4 вибираються з номерами 19 ÷ 22.

У цій же таблиці обчислити групові середні $\overline{x}_j \ (j=\overline{1,4})$. Методом дисперсійного аналізу при рівні значущості

$$\alpha=\begin{cases}0,05, & V-\text{парне},\\ 0,01, & V-\text{непарне}\end{cases}$$

перевірити нульову гіпотезу про рівність групових середніх. Передбачається, що вибірки витягнуті з нормальних сукупностей з однаковими дисперсіями.

Розв'язок варіанта ІЗ

○ Складемо табл. 11.11 за вибіркою B розв'язаного варіанта індивідуального завдання 8 і знайдемо групові середні за формулою

$$\overline{x}_j=\left(\sum_{i=1}^{q_j}x_{ij}\right)\Big/q_j\ ,$$

де

$$q_1=5;\ q_2=6;\ q_3=7;\ q_4=4\,.$$

Для спрощення обчислень віднімемо з кожного спостережуваного значення те саме число $C=70$, яке наближено дорівнює загальному середньому.

Складемо розрахункову таблицю 11.12.

Таблица 11.11

	Рівні фактора F_j			
	F_1	F_2	F_3	F_4
1	67	70	70	80
2	73	74	69	69
3	68	79	76	69
4	72	65	71	60
5	67	72	63	–
6	–	71	65	–
7	–	–	73	–
$\overline{x}_j$	69,4	71,83	69,57	69,5

Використовуючи табл. 11.12, знайдемо загальну і факторну суми квадратів відхилень:

$$S_{\text{заг}} = 485 - 9/22 \approx 484,6; \qquad S_{\text{факт}} = 24,26 - 0,41 = 23,85 .$$

Залишкову суму квадратів відхилень знайдемо за формулою (11.6)

$$S_{\text{зал}} = S_{\text{заг}} - S_{\text{факт}} = 460,756 .$$

Таблиця 11.12

<table>
<tr><td>Номер випроб-ня</td><td colspan="8">Рівні фактора F_j</td><td rowspan="3">Підсумковий стовпець</td></tr>
<tr><td rowspan="2">i</td><td colspan="2">F_1</td><td colspan="2">F_2</td><td colspan="2">F_3</td><td colspan="2">F_4</td></tr>
<tr><td>y_{i1}</td><td>y_{i1}^2</td><td>y_{i2}</td><td>y_{i2}^2</td><td>y_{i3}</td><td>y_{i3}^2</td><td>y_{i4}</td><td>y_{i4}^2</td></tr>
<tr><td>1</td><td>–3</td><td>9</td><td>0</td><td>0</td><td>0</td><td>0</td><td>10</td><td>100</td><td rowspan="7">–</td></tr>
<tr><td>2</td><td>3</td><td>9</td><td>4</td><td>16</td><td>–1</td><td>1</td><td>–1</td><td>1</td></tr>
<tr><td>3</td><td>–2</td><td>4</td><td>9</td><td>81</td><td>6</td><td>36</td><td>–1</td><td>1</td></tr>
<tr><td>4</td><td>2</td><td>4</td><td>– 5</td><td>25</td><td>1</td><td>1</td><td>–10</td><td>100</td></tr>
<tr><td>5</td><td>–3</td><td>9</td><td>2</td><td>4</td><td>–7</td><td>49</td><td>–</td><td>–</td></tr>
<tr><td>6</td><td>–</td><td>–</td><td>1</td><td>1</td><td>–5</td><td>25</td><td>–</td><td>–</td></tr>
<tr><td>7</td><td>–</td><td>–</td><td>–</td><td>–</td><td>3</td><td>9</td><td>–</td><td>–</td></tr>
<tr><td>$Q_j = \sum_{i=1}^{q_j} y_{ij}^2$</td><td>–</td><td>35</td><td>–</td><td>127</td><td>–</td><td>121</td><td>–</td><td>202</td><td>$\sum_{j=1}^{4} Q_j = 485$</td></tr>
<tr><td>$T_j = \sum_{i=1}^{q_j} y_{ij}$</td><td>–3</td><td>–</td><td>11</td><td>–</td><td>–3</td><td>–</td><td>–2</td><td>–</td><td>$\sum_{j=1}^{4} T_j = 3$</td></tr>
<tr><td>T_j^2/q_j</td><td>1,8</td><td>–</td><td>20,17</td><td>–</td><td>1,29</td><td>–</td><td>1</td><td>–</td><td>$\sum_{j=1}^{4} T_j^2/q_j = 24,26$</td></tr>
</table>

За формулами (11.11) знайдемо факторну і залишкову дисперсії: $s_{\text{факт}}^2 = 23,844/3 \approx 7,948$; $\quad s_{\text{зал}}^2 = 460,756/18 \approx 25,6$.

Оскільки $s_{\text{факт}}^2 < s_{\text{зал}}^2$ (зауваження 4), то вже звідси випливає правильність гіпотези про рівність групових середніх і, випливає, що немає потреби звертатися до критерію Фішера-Снедекора. ●