

ТЕМА: ДИСПЕРСІЙНА АНАЛІЗА

1.1. Статистична рівність двох генеральних дисперсій

1. Нехай маємо два варіаційних ряди для двох випадкових величин X та Y:

$$X: x_1 \leq x_2 \leq x_3 \leq \dots \leq x_i \leq \dots \leq x_{N1}: \bar{x}, S_x^2, S_x, N_x \quad (1.1)$$

$$Y: y_1 \leq y_2 \leq y_3 \leq \dots \leq y_i \leq \dots \leq x_{N2}: \bar{y}, S_y^2, S_y, N_y \quad (1.2)$$

$$2. \text{ Висуваємо нульову гіпотезу: } H_0: \begin{matrix} \sigma_x^2 & = & \sigma_y^2 \\ \uparrow & & \uparrow \text{ оцінки} \\ S_x^2 & \neq & S_y^2 \end{matrix}$$

(дві генеральні дисперсії σ_x^2 і σ_y^2 , яким дана оцінка за відповідними вибірковими дисперсіями S_x^2 і S_y^2 , статистично рівні).

Альтернативною гіпотезою H_1 до нульової H_0 буде:

$$H_1: \begin{matrix} \sigma_x^2 & \neq & \sigma_y^2 \\ \uparrow & & \uparrow \text{ оцінки} \\ S_x^2 & \neq & S_y^2 \end{matrix}$$

(дві генеральні дисперсії статистично нерівні).

$$\text{Нехай: } S_x^2 = \frac{SS_x}{f_x} > S_y^2 = \frac{SS_y}{f_y}, \quad (1.3)$$

де $f_x = N_x - 1$, $f_y = N_y - 1$ – числа відповідних ступенів вільностей.

3. Перевірку H_0 проведемо за критерієм Фішера, розрахувавши статистику F_p :

$$F_p = \frac{S_{\max}^2}{S_{\min}^2} = \frac{S_x^2}{S_y^2} \quad (1.4)$$

а) якщо $F_p \leq F_\alpha = F_T\{\alpha; f_{\max}=f_x=N_x-1; f_{\min}=f_y=N_y-1\}$, то H_0 приймається з рівнем значущості α (з рівнем ймовірності $p=1-\alpha$) при обсягах двох виборок N_x та N_y , тобто дві генеральні дисперсії статистично рівні (дві відповідні вибіркові дисперсії статистично однорідні).

Введемо у науковий обіг означення **ступеня рівності** двох генеральних дисперсій:

$$\xi_1(F) = \frac{F_T(\alpha)}{F_p} \geq 1, \quad (1.5)$$

а так як ми маємо справу зі статистичними оцінками, то при цій статистичній рівності залишковий **ступінь нерівності** двох генеральних дисперсій дорівнює:

$$\xi_2(F) = \frac{F_p}{F_T(\alpha)} < 1. \quad (1.6)$$

б) якщо $F_p > F_\alpha = F_T\{\alpha; f_{\max}=f_x=N_x-1; f_{\min}=f_y=N_y-1\}$, то H_0 відкидається з α ($p=1-\alpha$), приймається альтернативна гіпотеза H_1 з α ($p=1-\alpha$) та обсягами виборок N_x та N_y , тобто генеральні дисперсії статистично нерівні (відповідні вибіркові дисперсії статистично неоднорідні)

Введемо у науковий обіг означення **ступеня нерівності** двох генеральних дисперсій:

$$\xi_2(F) = \frac{F_p}{F_T(\alpha)} > 1. \quad (1.7)$$

При цьому залишковий **ступінь рівності** двох генеральних дисперсій:

$$\xi_1(F) = \frac{F_T(\alpha)}{F_p} \leq 1. \quad (1.8)$$

1.2. Статистична рівність двох математичних сподівань

Для рядів (1.1) і (1.2) висуваємо нульову гіпотезу: $H_0: a_x = a_y$
 $\uparrow \quad \uparrow$ оцінки
 $\bar{x} \neq \bar{y}$

(дві генеральні середні a_x і a_y , яким дана оцінка за відповідними вибірковими середніми $\bar{x}$ і $\bar{y}$, статистично рівні).

Альтернативною відносно H_0 гіпотезою буде гіпотеза H_1 :

$H_1: a_x \neq a_y$
 $\uparrow \quad \uparrow$ оцінки
 $\bar{x} \neq \bar{y}$

(дві генеральні середні a_x і a_y , яким дана оцінка за відповідними вибірковими середніми $\bar{x}$ і $\bar{y}$, статистично нерівні).

Перевірку H_0 проведемо за критерієм Стюдента $t_T = t_\alpha$.

1. Якщо попередньо доведено, що генеральні дисперсії двох сукупностей випадковиз величин рівні $\sigma_x^2 = \sigma_y^2$, то розраховуємо об'єднану вибіркову дисперсію:

$$S_{xy}^2 = \frac{SS_x + SS_y}{f_x + f_y} = \frac{SS_{xy}}{f_{xy}} = \frac{f_x S_x^2 + f_y S_y^2}{N_x + N_y - 2} \quad (1.9)$$

2. Розраховуємо за абсолютною величиною статистику t_p :

$$t_p = \left| \frac{\bar{x} - \bar{y}}{S_{xy} \sqrt{\frac{1}{N_x} + \frac{1}{N_y}}} \right|. \quad (1.10)$$

3. а) якщо при цьому $|t_p| \leq t_\alpha = t_T\{\alpha; f_{xy}=N_x+N_y-2\}$, то H_0 приймаємо з α ($p=1-\alpha$), тобто генеральні середні статистично рівні $a_x = a_y$, а вибіркові середні статистично однорідні.

Введемо у науковий обіг означення ступеня рівності двох генеральних середніх:

$$\xi_1(t) = \frac{t_T(\alpha)}{|t_p|} \geq 1, \quad (1.11)$$

а так як ми маємо справу зі статистичними оцінками, то при статистичній рівності залишковий ступінь нерівності двох генеральних середніх дорівнює:

$$\xi_2(t) = \frac{|t_p|}{t_T(\alpha)} < 1; \quad (1.12)$$

б) якщо при цьому $|t_p| \geq t_\alpha = t_T\{\alpha; f_{xy}=N_x+N_y-2\}$, то H_0 відкидаємо з α ($p=1-\alpha$), приймаємо H_1 з α ($p=1-\alpha$) – генеральні середні статистично нерівні $a_x \neq a_y$, а вибіркові середні статистично неоднорідні.

Введемо у науковий обіг означення ступеня нерівності двох генеральних середніх:

$$\xi_2(t) = \frac{|t_p|}{t_T(\alpha)} > 1, \quad (1.13)$$

При цьому за нерівністю залишковий ступінь рівності двох генеральних середніх дорівнює:

$$\xi_1(t) = \frac{t_T(\alpha)}{|t_p|} \leq 1. \quad (1.14)$$

4. Якщо попередньо доведено, що генеральні дисперсії двох сукупностей випадкових величин статистично нерівні $\sigma_x^2 \neq \sigma_y^2$, то розраховуємо за абсолютною величиною статистику t_p :

$$t_p = \frac{\left| \bar{x} - \bar{y} \right|}{\sqrt{\frac{S_x^2}{N_x} + \frac{S_y^2}{N_y}}}. \quad (1.15)$$

а) якщо при цьому $|t_p| \leq t_\alpha = t_T\{\alpha; f\}$, то H_0 приймаємо з α ($p=1-\alpha$), тобто генеральні середні статистично рівні $a_x = a_y$, а вибіркові середні статистично однорідні зі ступенем рівності двох генеральних середніх:

$$\xi_1(t) = \frac{t_T(\alpha)}{|t_p|} \geq 1, \quad (1.16)$$

а залишковий ступінь нерівності становить:

$$\xi_2(t) = \frac{|t_p|}{t_T(\alpha)} < 1. \quad (1.17)$$

б) якщо при цьому $|t_p| \geq t_\alpha = t_T\{\alpha; f\}$, то H_0 відкидаємо з α ($p=1-\alpha$), тобто приймаємо H_1 з α ($p=1-\alpha$) – генеральні середні статистично нерівні $a_x \neq a_y$, а вибіркові середні статистично неоднорідні зі ступенем нерівності:

$$\xi_2(t) = \frac{|t_p|}{t_T(\alpha)} > 1, \quad (1.18)$$

а залишковий ступінь рівності становить:

$$\xi_1(t) = \frac{t_T(\alpha)}{|t_p|} \leq 1. \quad (1.19)$$

Теоретичні значення критерію Стюдента вибирають із таблиць для числа ступенів вільностей:

$$f = \frac{f_x \cdot f_y}{f_y c^2 + (1 - c)^2 f_x}, \quad (1.20)$$

$$\text{де } c = \frac{\frac{S_x^2}{N_x}}{\frac{S_x^2}{N_x} + \frac{S_y^2}{N_y}} \quad (\text{якщо } S_x^2 > S_y^2). \quad (1.21)$$

1.3. Однофакторний експеримент

1.3.1. Статистична рівність ряду генеральних дисперсій

1. Нехай виконується реальний експеримент під час дослідження $y=f(x)$.

2. Експеримент сплановано таким чином, що фактор X варіюється на N рівнях (i); на кожному рівні плануємо повторні досліді n_i -раз.

3. Модель експерименту:

$$a_y = a_0 + \Phi_i + \varepsilon_{ij}, \quad (1.22)$$

де a_0 – середній результат;

Φ_i – ефект від x -фактора на i -рівні;

ε_{ij} – помилка досліді на i -рівні при j -повторі.

План експерименту подано в табл. 1.1.

4. Всім дослідідам, включаючи і повторні, присвоєно номери, згідно цих чисел досліді рандомізовано у часі за таблицею випадкових чисел.

5. Результати експерименту занесені в табл. 1.1.

6. Розрахуємо середню кожного ряду:

$$\bar{y}_i = \frac{\sum_{j=1}^{n_i} y_{ij}}{n_i}. \quad (1.23)$$

7. Розраховуємо загальну середню:

$$\bar{y} = \frac{\sum_{i=1}^N \sum_{j=1}^{n_i} y_{ij}}{\sum_{i=1}^N n_i} = \frac{\sum_{i=1}^N n_i \bar{y}_i}{\sum_{i=1}^N n_i}. \quad (1.24)$$

Таблиця 1.1

План та результати експерименту

Рівень фактора X (i)	Повторні досліди						n _i	$\bar{y}_i$	S _i ²	S _i
	1	2	...	j	...	n _i				
1	y ₁₁	y ₁₂	...	y _{1j}	...	y _{1n1}	n ₁	$\bar{y}_1$	S ₁ ²	S ₁
2	y ₂₁	y ₂₂	...	y _{2j}	...	y _{2n2}	n ₂	$\bar{y}_2$	S ₂ ²	S ₂
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
i	y _{i1}	y _{i2}	...	y _{ij}	...	y _{ini}	n _i	$\bar{y}_i$	S _i ²	S _i
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
N	y _{N1}	y _{N2}	...	y _{Nj}	...	y _{NnN}	n _N	$\bar{y}_N$	S _N ²	S _N

8. Розрахуємо дисперсію в ряду:

$$S_i^2 = \frac{\sum_{j=1}^{n_i} (y_{ij} - \bar{y}_i)^2}{n_i - 1} = \frac{SS_i}{f_i}, \quad (1.25)$$

де число ступенів вільностей для дисперсії результатів i-ряду:

$$f_i = n_i - 1. \quad (1.26)$$

9. Розраховуємо загальну дисперсію всього експерименту:

$$S^2 = \frac{\sum_{i=1}^N f_i \cdot S_i^2}{\left(\sum_{i=1}^N n_i \right) - N} = \frac{SS}{f_2}, \quad (1.27)$$

де $f_2 = \left(\sum_{i=1}^N n_i \right) - N$ - число ступенів вільностей для дисперсії всіх

результатів табл. 1.1. (1.28)

У разі, коли $n_i = \text{const} = n$:

$$f_2 = N(n - 1). \quad (1.29)$$

10. Процедурою доведемо або приймаємо, якщо для цього є підстави, що результати у кожному ряду підпорядковані нормальному закону розподілу (н.з.р.).

11. Висуваємо нульову гіпотезу:

$$H_0: \sigma_1^2 = \sigma_2^2 = \dots = \sigma_i^2 = \dots = \sigma_N^2$$

$\uparrow \quad \uparrow \quad \quad \uparrow \quad \quad \uparrow$ оцінки
 $S_1^2 \neq S_2^2 \neq \dots \neq S_i^2 \neq \dots \neq S_N^2$

(ряд генеральних дисперсій статистично рівний).

12. Альтернативною до нульової H_0 буде гіпотеза H_1 :

$$H_1: \sigma_1^2 \neq \sigma_2^2 \neq \dots \neq \sigma_i^2 \neq \dots \neq \sigma_N^2$$

$\uparrow \quad \uparrow \quad \quad \uparrow \quad \quad \uparrow$ оцінки
 $S_1^2 \neq S_2^2 \neq \dots \neq S_i^2 \neq \dots \neq S_N^2$

(ряд генеральних дисперсій статистично нерівний).

Перевірка нульової гіпотези H_0 :

1) якщо $n_i = \text{const} = n$, то H_0 перевіряють за **критерієм Кохрана (Кохрена, Кокрана, Кокрена)**, розраховуючи статистику:

$$G_p = \frac{S_{i \max}^2}{\sum_{i=1}^N S_i^2} \quad (1.30)$$

де $S_{i \max}^2$ – максимальна дисперсія у ряду дисперсій;

$\sum_{i=1}^N S_i^2$ – сума N дисперсій всього ряду, включаючи $S_{i \max}^2$.

а) якщо при цьому $G_p \leq G_\alpha = G_T \{ \alpha; N; f_i = n_i - 1 = n - 1 \}$, де $G_\alpha = G_T$ – табличне значення критерія Кохрана, то H_0 приймаємо з α ($p = 1 - \alpha$) – ряд генеральних дисперсій статистично рівний (відповідний ряд вибірових дисперсій статистично однорідний).

Введемо в науковий обіг означення ступеня рівності ряду дисперсій за критерієм Кохрана:

$$\xi_1(G) = \frac{G_T(\alpha)}{G_p} \geq 1, \quad (1.31)$$

при цьому через статистичні оцінки залишковий ступінь нерівності ряду дисперсій за критерієм Кохрана становить:

$$\xi_2(G) = \frac{G_p}{G_T(\alpha)} < 1. \quad (1.32)$$

б) якщо при цьому $G_p > G_T \{ \alpha; N; f_i = n_i - 1 = n - 1 \}$, то H_0 відкидаємо з α ($p = 1 - \alpha$), тобто приймаємо H_1 з α ($p = 1 - \alpha$) – ряд генеральних дисперсій

статистично нерівний (відповідний ряд вибірових дисперсій статистично неоднорідний).

Введемо в науковий обіг означення ступеня нерівності ряду дисперсій за критерієм Кохрана:

$$\xi_2(G) = \frac{G_p}{G_T(\alpha)} > 1. \quad (1.33)$$

при цьому через статистичні оцінки залишковий ступінь рівності ряду дисперсій за критерієм Кохрана становить:

$$\xi_1(G) = \frac{G_T(\alpha)}{G_p} \leq 1. \quad (1.34)$$

2) якщо $n_i = \text{const}$ або $n_i \neq \text{const}$, то H_0 перевіряють за критерієм Фішера, розраховуючи статистику F_p :

$$F_p = \frac{S_{\max}^2}{S_{\min}^2}, \quad (1.35)$$

де $S_{\max}^2$, $S_{\min}^2$ – максимальна та мінімальна дисперсія відповідно, які вибрані із ряду дисперсій;

а) якщо при цьому $F_p \leq F_\alpha = F_T\{\alpha; f_{\max}=N_{i\max}-1; f_{\min}=N_{i\min}-1\}$, де $F_\alpha = F_T$ – табличне значення критерія Фішера, то H_0 приймається з α ($p=1-\alpha$) – ряд генеральних дисперсій статистично рівний (відповідний ряд вибірових дисперсій статистично однорідний).

Введемо в науковий обіг означення ступеня рівності ряду дисперсій за критерієм Фішера:

$$\xi_1(F) = \frac{F_T(\alpha)}{F_p} \geq 1, \quad (1.36)$$

при цьому через статистичні оцінки залишковий ступінь нерівності ряду дисперсій при їх статистичній рівності за критерієм Фішера становить:

$$\xi_2(F) = \frac{F_p}{F_T(\alpha)} < 1. \quad (1.37)$$

б) якщо при цьому $F_p > F_\alpha = F_T\{\alpha; f_{\max}; f_{\min}\}$, то H_0 відкидається з α ($p=1-\alpha$), приймається альтернативна гіпотеза H_1 з α ($p=1-\alpha$) – ряд генеральних

дисперсій статистично нерівний (відповідний ряд вибірових дисперсій статистично неоднорідний).

Введемо в науковий обіг означення ступеня нерівності ряду дисперсій за критерієм Фішера:

$$\xi_2(F) = \frac{F_p}{F_T(\alpha)} > 1, \quad (1.38)$$

при цьому через статистичні оцінки залишковий ступінь рівності ряду дисперсій при їх статистичній нерівності за критерієм Фішера становить:

$$\xi_1(F) = \frac{F_T(\alpha)}{F_p} \leq 1. \quad (1.39)$$

3) якщо $\mathbf{n}_i = \mathbf{var}$, то для перевірки H_0 застосовують більш точний **критерій Бартлета** (точніше критерій Пірсона χ^2 у формі Бартлета), розраховуючи статистику χ_p^2 :

$$\chi_p^2 = \frac{1}{c} \left[(f_2 \ln S_2^2) - \sum_{i=1}^N (f_i \ln S_i^2) \right], \quad (1.40)$$

$$\text{де } c = \frac{1}{\ln 10} \left[1 + \frac{1}{3f_1} \left(\sum_{i=1}^N \frac{1}{f_i} - \frac{1}{f_2} \right) \right]. \quad (1.41)$$

$$f_i = n_i - 1; f_1 = N - 1; f_2 = \left(\sum_{i=1}^N n_i \right) - N;$$

$$S_i^2 = \frac{\sum_{j=1}^{n_i} (y_{ij} - \bar{y}_i)^2}{f_i};$$

$$S_2^2 = \frac{\sum_{i=1}^N (f_i \cdot S_i^2)}{f_2};$$

де N – число дисперсій (табл. 1.1).

Якщо при цьому $\chi_p^2 \leq \chi_{\alpha}^2 = \chi_{\tau}^2 \{ \alpha; f_1 \}$, де $\chi_{\alpha}^2 = \chi_{\tau}^2$ – табличне значення критерія Пірсона, то H_0 приймається з α ($p=1-\alpha$) – ряд генеральних дисперсій статистично рівний (відповідний ряд вибірових дисперсій статистично однорідний).

Введемо в науковий обіг означення ступеня рівності ряду дисперсій за критерієм Пірсона у формі Бартлета:

$$\xi_1(\chi^2) = \frac{\chi_T^2(\alpha)}{\chi_p^2} \geq 1. \quad (1.42)$$

при цьому через статистичні оцінки залишковий ступінь нерівності ряду дисперсій при їх статистичній рівності за критерієм Пірсона у формі Бартлета становить:

$$\xi_2(\chi^2) = \frac{\chi_p^2}{\chi_T^2(\alpha)} < 1. \quad (1.43)$$

Якщо при цьому $\chi_p^2 > \chi_T^2 \{ \alpha; f_1 \}$, то H_0 відкидається з α ($p=1-\alpha$). Приймаємо H_1 – ряд генеральних дисперсій статистично нерівний (відповідний ряд вибірових дисперсій статистично неоднорідний).

Введемо в науковий обіг означення ступеня нерівності ряду дисперсій за критерієм Пірсона у формі Бартлета:

$$\xi_2(\chi^2) = \frac{\chi_p^2}{\chi_T^2(\alpha)} > 1. \quad (1.44)$$

при цьому через статистичні оцінки залишковий ступінь рівності ряду дисперсій при їх статистичній нерівності за критерієм Пірсона у формі Бартлета становить:

$$\xi_1(\chi^2) = \frac{\chi_T^2(\alpha)}{\chi_p^2} \leq 1. \quad (1.45)$$

1.3.2. Порівняння двох виборок за їх ознаками

1. У табл. 1.2 приведений приклад статистичних даних за ознаками для двох об'єктів порівняння.

У табл. 1.2: n_{1i} , n_{2i} – абсолютні частоти зустрічі ознак у відповідній виборці;

ω_{1i} , ω_{2i} – відносні частоти зустрічі ознак у відповідній виборці;

k – кількість ознак;

N_1 , N_2 – загальна сума абсолютних ознак у відповідній виборці.

Таблиця 1.2

i	Назва ознаки	Об'єкт 1		Об'єкт 2	
		n_{1i}	$\omega_{1i} = \frac{n_{1i}}{N_1}$	n_{2i}	$\omega_{2i} = \frac{n_{2i}}{N_2}$
1	LiCl	n_{11}	ω_{11}	n_{21}	ω_{21}
2	NaCl	n_{12}	ω_{12}	n_{22}	ω_{22}
$\vdots$	NaF, NaBr...	$\vdots$	$\vdots$	$\vdots$	$\vdots$
i	KCl	n_{1i}	ω_{1i}	n_{2i}	ω_{2i}
$\vdots$	KF, KBr...	$\vdots$	$\vdots$	$\vdots$	$\vdots$
k	CsCl	n_{1k}	ω_{1k}	n_{2k}	ω_{2k}
		$N_1 = \sum_{i=1}^k n_{1i}$	1,0	$N_2 = \sum_{i=1}^k n_{2i}$	1,0

2. Висуваємо нульову гіпотезу:

H_0 : немає статистичної різниці (відмінності) між об'єктами 1 і 2 за їх структурами.

Альтернативною до H_0 буде гіпотеза:

H_1 : між об'єктами 1 і 2 існує статистична різниця (відмінність) за їх структурами.

3. Розраховуємо статистику χ^2_p :

$$\chi^2_p = N_1 N_2 \sum_{i=1}^k \left[\frac{\left(\frac{n_{1i}}{N_1} - \frac{n_{2i}}{N_2} \right)^2}{\frac{n_{1i} + n_{2i}}{N_1 + N_2}} \right] = N_1 N_2 \sum_{i=1}^k \left[\frac{(\omega_{1i} - \omega_{2i})^2}{\omega_{1i} + \omega_{2i}} \right], \quad (1.46)$$

а) якщо $\chi^2_p \leq \chi^2_{\alpha} = \chi^2_{\tau} \{ \alpha; f \}$, де $\chi^2_{\alpha} = \chi^2_{\tau}$ – табличне значення критерія Пірсона; $f=k-1$, то H_0 приймається з α ($p=1-\alpha$) – між двома вибірками відсутня статистична різниця (відмінність).

Ступінь статистичної однаковості:

$$\xi_1(\chi^2) = \frac{\chi^2_T(\alpha)}{\chi^2_p} \geq 1. \quad (1.47)$$

Ступінь статистичної відмінності:

$$\xi_2(\chi^2) = \frac{\chi^2_p}{\chi^2_T(\alpha)} < 1. \quad (1.48)$$

б) якщо $\chi^2_{розр} > \chi^2_T \{ \alpha; f \}$, то H_0 відкидається з α ($p=1-\alpha$). Приймаємо H_1 .

Ступінь статистичної відмінності:

$$\xi_2(\chi^2) = \frac{\chi^2_p}{\chi^2_T(\alpha)} > 1. \quad (1.49)$$

Ступінь статистичної однаковості:

$$\xi_1(\chi^2) = \frac{\chi^2_T(\alpha)}{\chi^2_p} \leq 1. \quad (1.50)$$

Ступінь сумарної статистичної однаковості та відмінності:

$$\xi_{12} = \xi_1 + \xi_2. \quad (1.51)$$

1.3.3. Рівність ряду математичних сподівань (генеральних середніх)

1. За результатами табл. 1.1 перевіримо рівність ряду генеральних середніх за відповідними вибірковими середніми.

2. Висуваємо нульову гіпотезу $H_0: a_{y1} = a_{y2} = \dots = a_{yi} = \dots = a_{yN}$
 $\uparrow \quad \uparrow \quad \quad \uparrow \quad \quad \uparrow$ оцінка
 $\bar{y}_1 \neq \bar{y}_2 \neq \dots \neq \bar{y}_i \neq \dots \neq \bar{y}_N.$

(генеральні середні статистично рівні).

3. Доводимо однорідність вибіркових дисперсій рядків (рівність відповідних генеральних дисперсій).

4. Доводимо або приймаємо, що результати у кожному ряду підпорядковані нормальному закону розподілу.

5. Розрахунок дисперсій:

а) Дисперсія за рівнями факторів:

$$s_1^2 = \frac{SS_1}{f_1} = \frac{\sum_{i=1}^N (\bar{y}_i - \bar{y})^2}{N-1}, f_1 = N-1. \quad (1.52)$$

б) Дисперсія внутрішня (залишкова) [сумарна в середині повторних дослідів]:

$$S_2^2 = \frac{SS_2}{f_2} = \frac{\sum_{i=1}^N \sum_{j=1}^{n_i} (y_{ij} - \bar{y}_i)^2}{\left(\sum_{i=1}^N n_i \right) - N}, \quad f_2 = \left(\sum_{i=1}^N n_i \right) - N. \quad (1.53)$$

$S_2^2 = S_\varepsilon^2$ застосовують для оцінки дисперсії помилки всього експерименту.

в) Загальна (повна) дисперсія:

$$S_3^2 = \frac{SS_3}{f_3} = \frac{\sum_{i=1}^N \sum_{j=1}^{n_i} (y_{ij} - \bar{y})^2}{\left(\sum_{i=1}^N n_i \right) - 1}, \quad f_3 = \left(\sum_{i=1}^N n_i \right) - 1. \quad (1.54)$$

6. Аналіз:

1) Вплив фактора:

Якщо $F_p = \frac{S_1^2}{S_2^2} = \frac{S_1^2}{S_\varepsilon^2} \leq F_T\{\alpha; f_1; f_2\}$, H_0 приймаємо з α ($p=1 - \alpha$) (ряд генеральних середніх за рівнями факторів статистично рівний).

Ступінь рівності:

$$\xi_1(F) = \frac{F_T(\alpha)}{F_p} \geq 1. \quad (1.55)$$

2) тоді всі результати експерименту належать до одного фрагменту генеральної сукупності з нормальним законом розподілу обсягом $\sum_{i=1}^N n_i$ з параметрами:

$$a_y \leftarrow \bar{y}$$

$$\sigma_y^2 \leftarrow S_3^2$$

3) Довірчі інтервали та довірча ймовірність для a_y та σ_y^2 :

$$P \left\{ \bar{y} - \frac{S_3 t_T\{\alpha; f_3\}}{\sqrt{\sum_{i=1}^N n_i}} < a_y < \bar{y} + \frac{S_3 t_T\{\alpha; f_3\}}{\sqrt{\sum_{i=1}^N n_i}} \right\} = 1 - \alpha; \quad (1.56)$$

$$p\left(z_1^2 S_3^2 \frac{f_3}{f_3 + 1} < \sigma_y^2 < z_2^2 S_3^2 \frac{f_3}{f_3 + 1}\right) = 1 - \alpha ; \quad (1.57)$$

або

$$p\left(\frac{f_3 S_3^2}{\chi_T^2 \left\{\frac{\alpha}{2}; f_3\right\}} < \sigma_y^2 < \frac{f_3 S_3^2}{\chi_T^2 \left\{1 - \frac{\alpha}{2}; f_3\right\}}\right) = 1 - \alpha . \quad (1.58)$$

4) Якщо $F_p > F_T$, то H_0 відкидаємо.

Вплив фактора за рівнями статистично суттєвий, порівняно з помилкою експерименту.

Ступінь нерівності:

$$\xi_2(F) = \frac{F_p}{F_T(\alpha)} > 1. \quad (1.59)$$

5) Тоді ми маємо ситуацію, коли результати експерименту кожного рядка (системи спостережень) як виборки, що взяті з відповідних генеральних сукупностей з нормальним законом розподілу обсягом n_i кожна з параметрами:

$$a_{yi} \leftarrow \bar{y}_i$$

$$\sigma_y^2 \leftarrow S_2^2.$$

6) Довірчі інтервали та довірча ймовірність для a_y та σ_y^2 :

$$p\left(\bar{y}_i - \frac{S_2 t_T\{\alpha; f_2\}}{\sqrt{n_i}} < a_y < \bar{y}_i + \frac{S_2 t_T\{\alpha; f_2\}}{\sqrt{n_i}}\right) = 1 - \alpha ; \quad (1.60)$$

$$p\left(z_1^2 S_2^2 \frac{f_2}{f_2 + 1} < \sigma_y^2 < z_2^2 S_2^2 \frac{f_2}{f_2 + 1}\right) = 1 - \alpha ; \quad (1.61)$$

або

$$p\left(\frac{f_2 S_2^2}{\chi_T^2 \left\{\frac{\alpha}{2}; f_2\right\}} < \sigma_y^2 < \frac{f_2 S_2^2}{\chi_T^2 \left\{1 - \frac{\alpha}{2}; f_2\right\}}\right) = 1 - \alpha . \quad (1.62)$$

7) Оцінка дисперсії середніх:

$$S_{yi}^2 = S_{\bar{y}_i}^2 = \frac{N-1}{\sum_{i=1}^N n_i} (S_1^2 - S_2^2) = \frac{f_1}{f_3 + 1} (S_1^2 - S_2^2). \quad (1.63)$$

1.3.4. Обґрунтування нижніх і верхніх гарантованих толерантних меж

Нижню толерантну межу знаходять для заданого рівня довірчої ймовірності γ та рівня значущості $\alpha = 1 - \rho$, виходячи з умови $X \geq x_H$ (1). При нормальному законі розподілу результат дослідження характеризується $x_H = \bar{x} - k\sigma$ (2), яке з довірчою ймовірністю γ гарантує в $\rho \cdot 100\% = (1 - \alpha) \cdot 100\%$ випадках дотримання умов (1). Рівняння (2) є оцінкою рівняння $x_H = a_x - k\sigma$.

Для верхньої толерантної межі $X \leq x_B$:

$$x_B = \bar{x} + k\sigma \xrightarrow{\text{оцінка}} x_B = a_x + k\sigma,$$

$$\text{де } k = z_p \left(1 + \frac{z_\gamma}{\sqrt{2N}} + \frac{5Z_\gamma^2 + 10}{12N} \right), \quad (1.64)$$

z_p, z_γ - квантилі нормального закону розподілу для ймовірностей ρ і γ .

1.4. Двофакторний експеримент

1) $y = f(x_1, x_2)$ на рівнях $N_1(i)$ та $N_2(j)$ відповідно. Кількість повторних дослідів $n = \text{const}$.

2) Модель експерименту:

$$a_y = a_0 + \Phi_i + \Phi_j + \Phi_{ij} + \varepsilon_{ijv}, \quad (1.65)$$

де a_0 - загальний ефект;

Φ_i - ефект від x_1 -фактора на i -рівні;

Φ_j - ефект від x_2 -фактора на j -рівні;

Φ_{ij} - ефект взаємодії двох факторів x_1 на i -рівні та x_2 на j -рівні;

ε_{ijv} - помилка дослідження на i -рівні x_1 -фактора та на j -рівні x_2 -фактора при v -повторних дослідів.

3) План експерименту і результати занесені в табл. 1.3 (попередньо поведена нумерація дослідів, в т.ч. і повторних, і їх рандомізація).

Таблиця 1.3

План та результати двофакторного експерименту

j		1	2	...	j	...	N ₂
i	x _{2j}	x ₂₁	x ₂₁	...	x ₂₁	...	x ₂₁
	x _{1i}	1 2 ... v ... n	1 2 ... v ... n	...	1 2 ... v ... n	...	1 2 ... v ... n
1	x ₁₁	y ₁₁₁ y ₁₁₂ ...y _{11v} ...y _{11n}	y ₁₂₁ y ₁₂₂ ...y _{12v} ...y _{12n}	...	y _{1j1} y _{1j2} ...y _{1jv} ...y _{1jn}	...	y _{1N21} y _{1N22} ... y _{1N2v} ...y _{1N2n}
2	x ₁₂	y ₂₁₁ y ₂₁₂ ...y _{21v} ...y _{21n}	y ₂₂₁ y ₂₂₂ ...y _{22v} ...y _{22n}	...	y _{2j1} y _{2j2} ...y _{2jv} ...y _{2jn}	...	y _{2N21} y _{2N22} ... y _{2N2v} ...y _{2N2n}
⋮	⋮	⋮	⋮	...	⋮	...	⋮
i	x _{1i}	y _{i11} y _{i12} ...y _{i1v} ...y _{i1n}	y _{i21} y _{i22} ...y _{i2v} ...y _{i2n}	...	y _{ij1} y _{ij2} ...y _{ijv} ...y _{ijn}	...	y _{iN21} y _{iN22} ... y _{iN2v} ...y _{iN2n}
⋮	⋮	⋮	⋮	...	⋮	...	⋮
N ₁	x _{1N1}	y _{N111} y _{N112} ... y _{N11v} ...y _{N11n}	y _{N121} y _{N122} ... y _{N12v} ...y _{N12n}	...	y _{N1j1} y _{N1j2} ... y _{N1jv} ...y _{N1jn}	...	y _{N1N21} y _{N1N22} ... y _{N1N2v} ...y _{N1N2n}

4) Доводимо за кожною системою спостережень, що результати підпорядковані нормальному закону розподілу.

5) Доводимо однорідність ряду дисперсій для системи спостережень.

6) Середня системи спостережень:

$$\bar{y}_{ij} = \frac{\sum_{v=1}^n y_{ijv}}{n} \quad (1.66)$$

7) Дисперсія в системі спостережень

$$S_{ij}^2 = \frac{1}{n-1} \sum_{v=1}^n (y_{ijv} - \bar{y}_{ij})^2, \quad f_{ij} = n-1 = \text{const} \quad (1.67)$$

8) Зведемо середні в табл. 1.4.

9) Середня ряду:

$$\bar{y}_{i.} = \frac{\sum_{j=1}^{N_2} \bar{y}_{ij}}{N_2} = \frac{\sum_{j=1}^{N_2} \sum_{v=1}^n y_{ijv}}{N_2 n} \quad (1.68)$$

10) Середня графа:

$$\bar{y}_{.j} = \frac{\sum_{i=1}^{N_1} \bar{y}_{ij}}{N_1} = \frac{\sum_{i=1}^{N_1} \sum_{v=1}^n y_{ijv}}{N_1 n} \quad (1.69)$$

Таблиця 1.4

Матриця середніх результатів для двофакторного експерименту
за системами спостережень

j		1	2	...	j	...	N ₂		
i	$\begin{matrix} x_{2j} \\ x_{1i} \end{matrix}$	x ₂₁	x ₂₁	...	x ₂₁	...	x ₂₁	$\bar{y}_{i.}$	S _{i.} ²
1	x ₁₁	$\bar{y}_{11}$	$\bar{y}_{12}$	...	$\bar{y}_{1j}$	...	$\bar{y}_{1N_2}$	$\bar{y}_{1.}$	S _{1.} ²
2	x ₁₂	$\bar{y}_{21}$	$\bar{y}_{22}$	...	$\bar{y}_{2j}$	...	$\bar{y}_{2N_2}$	$\bar{y}_{2.}$	S _{2.} ²
⋮	⋮	⋮	⋮	...	⋮	...	⋮	⋮	⋮
i	x _{1i}	$\bar{y}_{i1}$	$\bar{y}_{i2}$	...	$\bar{y}_{ij}$	...	$\bar{y}_{iN_2}$	$\bar{y}_{i.}$	S _{i.} ²
⋮	⋮	⋮	⋮	...	⋮	...	⋮	⋮	⋮
N ₁	x _{1N₁}	$\bar{y}_{N_1 1}$	$\bar{y}_{N_1 2}$	...	$\bar{y}_{N_1 j}$	...	$\bar{y}_{N_1 N_2}$	$\bar{y}_{N_1 .}$	S _{N₁ .} ²
	$\bar{y}_{.j}$	$\bar{y}_{.1}$	$\bar{y}_{.2}$		$\bar{y}_{.j}$		$\bar{y}_{.N_2}$	$\bar{y}_{..}$	
	S _{j.} ²	S _{1.} ²	S _{2.} ²		S _{j.} ²		S _{N₂ .} ²	S _{..} ²	

11) Загальна середня:

$$\bar{y}_{..} = \frac{\sum_{i=1}^{N_1} \bar{y}_{i.}}{N_1} = \frac{\sum_{j=1}^{N_2} \bar{y}_{.j}}{N_2} = \frac{\sum_{i=1}^{N_1} \sum_{j=1}^{N_2} \bar{y}_{ij}}{N_1 N_2} = \frac{\sum_{i=1}^{N_1} \sum_{j=1}^{N_2} \sum_{v=1}^n y_{ijv}}{N_1 N_2 n} \quad (1.70)$$

12) Дисперсія між середніми за рядами:

$$S_1^2 = \frac{1}{N_1 - 1} \sum_{i=1}^{N_1} N_2 n (\bar{y}_{i.} - \bar{y}_{..})^2 = \frac{SS_1}{f_1}, \text{ де } f_1 = N_1 - 1. \quad (1.71)$$

13) Дисперсія між середніми за графами:

$$S_2^2 = \frac{1}{N_2 - 1} \sum_{j=1}^{N_2} N_1 n (\bar{y}_{.j} - \bar{y}_{..})^2 = \frac{SS_2}{f_2}, \text{ де } f_2 = N_2 - 1. \quad (1.72)$$

14) Дисперсія взаємодії двох факторів:

$$S_3^2 = \frac{1}{(N_1 - 1)(N_2 - 1)} \sum_{i=1}^{N_1} \sum_{j=1}^{N_2} n [\bar{y}_{ij} - (\bar{y}_{i.} + \bar{y}_{.j} - \bar{y}_{..})]^2 = \frac{SS_3}{f_3}, \text{ де } f_3 = (N_1 - 1)(N_2 - 1). \quad (1.73)$$

15) Залишкова (внутрішня) дисперсія:

$$S_4^2 = \frac{1}{N_1 N_2 (n-1)} \sum_{i=1}^{N_1} \sum_{j=1}^{N_2} \sum_{v=1}^n (y_{ijv} - \bar{y}_{ij})^2 = \frac{SS_4}{f_4}, \text{ де } f_4 = N_1 N_2 (n-1). \quad (1.74)$$

16) Повна (сумарна, загальна) дисперсія:

$$S_5^2 = \frac{1}{N_1 N_2 n - 1} \sum_{i=1}^{N_1} \sum_{j=1}^{N_2} \sum_{v=1}^n (y_{ijv} - \bar{y}_{..})^2 = \frac{SS_5}{f_5}, \text{ де } f_5 = N_1 N_2 n - 1. \quad (1.75)$$

17) Аналіз:

1. Висуваємо нульову гіпотезу відносно взаємодії двох факторів:

$H_0: \sigma_3^2 = \sigma_4^2 = \sigma_\varepsilon^2$ – відсутність взаємодії між факторами x_1 і x_2 .

↑ ↑ оцінка
 $S_3^2 \neq S_4^2$

(фактори x_1 і x_2 можуть або підсилювати один одного (синергізм), або ослаблювати один одного за впливом на y (антагонізм), або діяти незалежно).

2. Перевірка H_0 :

а) Якщо $F_{0p} = \frac{S_3^2}{S_4^2} = \frac{S_3^2}{S_\varepsilon^2} > F_{0T}\{\alpha; f_3; f_4\}$, то H_0 відкидаємо з α ($p=1-\alpha$) (між

факторами x_1 і x_2 існує взаємодія).

Якщо ступінь взаємодії:

$$\xi_{02}(F) = \frac{F_{0p}}{F_{0T}(\alpha)} \geq 1,0-1,15, \quad (1.76)$$

то взаємодією можна знехтувати.

Якщо $\xi_{02} > 1,15-1,3$, то подальші оцінки та висліди будуть приблизними. Якщо $\xi_{02} > 1,3$, то необхідні спеціальні методи аналізу.

б) Якщо $F_{0p} \leq F_{0T}\{\alpha; f_3; f_4\}$, то H_0 приймаємо з α ($p=1-\alpha$) (взаємодія статистично мала, порівняно з помилкою експерименту)

Ступінь малости взаємодії:

$$\xi_{01}(F) = \frac{F_{0T}(\alpha)}{F_{0p}} \geq 1, \quad (1.77)$$

А ступінь ваговости взаємодії при цьому

$$\xi_{02}(F) = \frac{F_{0p}}{F_{0T}(\alpha)} < 1, \quad (1.78)$$

в) Тоді дисперсію помилки $S_4^2 = S_\varepsilon^2$ об'єднують з дисперсією взаємодії S_3^2 , утворюючи нову помилку експерименту $S_6^2 = S_{34}^2$:

$$S_6^2 = S_{34}^2 = \frac{SS_3 + SS_4}{f_3 + f_4} = \frac{f_3 S_3^2 + f_4 S_4^2}{f_3 + f_4} = \frac{SS_6}{f_6} = \frac{SS_{34}}{f_{34}}, \quad (1.79)$$

$$\text{де } f_6 = f_{34} = f_3 + f_4 = (N_1 - 1)(N_2 - 1) + N_1 N_2 (n - 1) = N_1 N_2 n - N_1 - N_2 + 1. \quad (1.80)$$

18) Далі досліджують значущість впливу на Y фактора x_1 , порівняно з новою помилкою експерименту, перевіряючи нульову гіпотезу:

$$H_0: \sigma_1^2 = \sigma_{34}^2.$$

$\uparrow \quad \uparrow$ оцінка
 $S_1^2 \neq S_{34}^2$

Якщо $F_{1p} = \frac{S_1^2}{S_{34}^2} > F_{1T}\{\alpha; f_1; f_{34}\}$, H_0 відкидаємо з α ($p=1 - \alpha$) (перший

фактор суттєво впливає на Y).

Ступінь впливу:

$$\xi_{2(1)} = \frac{F_{1p}}{F_{1T}(\alpha)} > 1,0 \quad (1.81)$$

19) Значущість впливу другого фактора x_2 :

$F_{2p} = \frac{S_2^2}{S_{34}^2} > F_{2T}\{\alpha; f_2; f_{34}\}$, H_0 відкидаємо з α ($p=1 - \alpha$) (другий фактор

суттєво впливає на Y).

Ступінь впливу:

$$\xi_{2(2)} = \frac{F_{2p}}{F_{2T}(\alpha)} \geq 1,0 \quad (1.82)$$

20) Тоді маємо фрагменти ($N_1 N_2$) генеральної сукупності, яка має н.з.р. кожна обсягом n з параметрами

$$a_{yij} \leftarrow \bar{y}_{ij}$$

$$\sigma_y^2 \leftarrow S_{34}^2$$

21) Довірчі інтервали та довірча ймовірність для a_{yij} та σ_y^2 :

$$P\left(\bar{y}_{ij} - \frac{S_{34} t_T\{\alpha; f_{34}\}}{\sqrt{n}} < a_{yij} < \bar{y}_{ij} + \frac{S_{34} t_T\{\alpha; f_{34}\}}{\sqrt{n}}\right) = 1 - \alpha; \quad (1.83)$$