

**Дисперсійний аналіз однорічних
результатів спостережень в
однофакторному досліді**

Дисперсійний аналіз - це метод статистичного аналізу, методологічною основою якого є теорія ймовірності. Він є найбільш поширеним методом математичної статистики у науковій агрономії, який дозволяє досліднику визначити достовірність впливу того чи іншого фактора (удобрення, обробіток ґрунту, сорти та ін.) на предмети досліджень, які досліджуються (ріст і розвиток рослин, їхню урожайність, якість урожаю і т. ін.). Статистичною достовірністю явища називають факт наявності істотного впливу фактора на дослідну функцію доведений математичними критеріями.

Сутність дисперсійного аналізу полягає:

1) у розкладанні загальної суми квадратів відхилень окремих спостережених величин від середньої арифметичної (C_y) та загального ступеня свободи на складові, що відповідають структурі досліджу;

2) визначенні середніх квадратів відхилень (дисперсій s^2) та відношення їх до дисперсії помилки ($F_{\phi} - sv^2/sz^2$);

3) порівнянні фактичного та теоретичного відношення дисперсій (F_{ϕ} та F_{05}). Якщо величина $F_{\phi} \geq F_{05}$, нульова гіпотеза відкидається ($H : d \neq 0$), а коли $F_{\phi} \leq F_{05}$ - її приймають, що свідчить про відсутність істотної різниці між варіантами.

Методика аналізу. Послідовність дій з дисперсійного аналізу однофакторного однорічного дослідження покажемо на прикладі завдання.

Завдання. Проведіть дисперсійний аналіз даних однофакторного польового дослідження, дайте йому загальну оцінку та визначте істотність різниць між варіантами.

Дисперсійний аналіз проводять в декілька етапів. Кількість їх залежить від структури експерименту. В однофакторному досліді він об'єднує 7 етапів.

Перший етап.

Складають підсумкову таблицю, де визначають суму величин у варіантах ($\sum V$) та повтореннях ($\sum P$), середню арифметичну величину в цілому досліді ($\bar{x}_0$).

Урожайність коренеплодів гібридів буряків цукрових, <i>т/га</i>						
Варіанти (<i>l</i>)	Величини у повтореннях (<i>n</i>), <i>x</i>				$\sum V$	Середня величина у варіантах
	I	II	III	IV		
1(контроль)	45,2	45,7	46,1	46,2	183,2	45,8
2	49,7	48,4	48,1	48,6	194,8	48,7
3	45,3	45,6	45,8	46,0	182,7	45,7
4	40,1	40,2	41,4	40,8	162,5	40,6
5	47,5	48,1	47,7	48,0	191,3	45,7
$\sum P$	227,8	228,0	229,1	229,6	$\sum X = 914,5$	$\bar{x}_0 = 45,7$

Другий етап

Якщо в табл. 4 1 цифрові показники мають великі дво- тризначні величини, то визначають відхилення всіх величин ознаки (x) від середньої величини в досліді (x_0). $X_1 = x - x_0 = 45,2 - 45,7 = -0,5$ і так далі. Ці відхилення можна визначити і від довільного числа, близького до середньої арифметичної по досліді. В цьому прикладі відхилення визначені від середньої арифметичної.

Відхилення X_1 всіх величин ознаки (x) від середньої величини в досліді ($\bar{x}_0$)					
Варіанти (l)	Відхилення X_1 у повтореннях (n)				ΣV_1
	I	II	III	IV	
1(St)	-0,5	0,0	0,4	0,5	0,4
2	4,0	2,7	2,4	2,9	12,0
3	-0,4	-0,1	0,1	0,3	-0,1
4	-5,6	-5,5	-4,3	-4,9	-20,3
5	1,8	2,4	2,0	2,3	8,5
ΣP_1	-0,7	-0,5	0,6	1,1	$\Sigma V_1 = \Sigma X_1 = 0,5$

Третій етап

за даними таблиці розраховують квадрати відхилень

Квадрати відхилень окремих величин ознаки від її середньої величини					
Варіанти (<i>l</i>)	X_1^2 у повтореннях (<i>n</i>)				$(\sum V_1)^2$
	I	II	III	IV	
1(St)	0,25	0,00	0,16	0,25	0,16
2	16	7,29	5,76	8,41	144,00
3	0,16	0,01	0,01	0,09	0,01
4	31,36	30,25	18,49	24,01	412,09
5	3,24	5,76	4,0	5,29	72,25
$(\sum P_1)^2$	0,49	0,25	0,36	1,21	$(\sum X_1)^2 = 0,25$

Четвертий етап

Визначення видів дисперсій величин дослідженої ознаки.

Якщо аналізують однофакторні статистичні комплекси із певною кількістю варіантів (L) і повторень (n), то загальну дисперсію ознаки, виражену загальною сумою квадратів (C_y) розділяють на три частини: дисперсію обумовлену повтореннями (C_p), варіантами (C_v) і помилками (C_z). В даному випадку загальна дисперсія (C_y) і загальне число ступенів свободи ($N - 1$) можна представити наступними символічними виразами:

$$C_y = C_p + C_v + C_z$$
$$(N - 1) = (n - 1) + (l - 1) + (n - 1)(l - 1)$$

Перелічені види дисперсій ознаки розраховують у такій послідовності:

1. Коригуючий фактор(поправка): $C = (\sum X_1)^2 / n = 0,25 / 5 \times 4 = 0,01$. (4.2)

2. Загальна дисперсія: $C_y = \sum X_1^2 - C = (0,25 + 0,0 + 0,16 + \dots + 4,0 + 5,29) - 0,01 = 160,78$. (4.3)

3. Дисперсія повторень: $C_p = \sum (\sum P_1)^2 / l - C = (0,49 + 0,25 + 0,36 + 1,21) / 5 - 0,01 = 0,45$. (4.4)

4. Дисперсія варіантів: $C_v = \sum (\sum V_1)^2 / n - C = (0,16 + 144,0 + 0,01 + 412,09 + 72,25) / 4 - 0,01 = 157,12$. (4.5)

5. Дисперсія, обумовлена помилками: $C_z = C_y - C_p - C_v = 160,78 - 0,45 - 157,12 = 3,21$ (4.6)

Дві останні суми квадратів C_v і C_z ділять на відповідні їм ступені свободи і отримують дві середні дисперсії – варіантів S_v^2 і помилок S_z^2

$$\text{варіантів } S_v^2 = \frac{C_v}{l-1} \text{ і помилок } S_z^2 = \frac{C_z}{(n-1)(l-1)}$$

$$S_v^2 = \frac{157,12}{5-1} = 39,28$$

$$S_z^2 = \frac{3,21}{(4-1)(5-1)} = 0,27$$

Результати дисперсійного аналізу

Дисперсія	Сума квадратів відхилень	Ступінь свободи, $n - 1$	Середня дисперсія s^2	Критерій достовірності результатів дослідів	
				F_{ϕ}	F_{05}
Загальна C_y	160,78	19			
Повторень C_p	0,45	3			
Варіантів C_v	157,12	4	$S_v^2 = 39,28$	146,82	3,26
Помилки (залишку) C_z	3,21	12	$S_z^2 = 0,27$		

П'ятий етап

підведення підсумків дисперсійного аналізу.

Якщо позначити відношення сум квадратів (C_v/C_y) через показник η_v^2 (η) - грецька літера “ета”), то за результатами дисперсійного аналізу можна встановити частку впливу окремих факторів:

- фактора, що вивчається:

$$\eta_v^2 = C_v/C_y = 157,12/160,78 = 0,977 \text{ або } 97,7 \%;$$

- строкатості родючості ґрунту дослідної ділянки:

$$\eta_p^2 = C_p/C_y = 0,45/160,78 = 0,003 \text{ або } 0,3 \%;$$

- інших, випадкових факторів (помилки):

$$\eta_z^2 = C_z/C_y = 3,21/160,78 = 0,02 \text{ або } 2 \%.$$

Шостий етап

розрахунки точності дослідів та найменшої істотної різниці на 5% рівні значущості

Точність дослідів:

$$s_{\bar{x}\%} = (100\sqrt{S_z^2/n})/\bar{x}_0 = (100\sqrt{0,27^2/4})/45,7 = 0,6\%$$

Якщо точність польового дослідів коливається в межах до 2 %, її вважають відмінною, 2-5 % - доброю, 5-8 % - задовільною і вище 8 % — незадовільною.

Найменша істотна різниця (HIP_{05}):

$$HIP_{05} = t_{05}\sqrt{2S_z^2/n} = 2,18 \times \sqrt{2 * 0,27/4} = 0,80 \text{ т/га},$$

де t_{05} - критерій Ст'юдента на 5 %-му рівні значущості, який знаходять у таблиці I додатків для ступеня свободи помилок $(n-1)(1-1) = 12$, $t_{05} = 2,18$; S_z^2 - дисперсія помилок (залишку) = 0,27.

Поряд з абсолютним показником HIP_{05} доцільно використовувати також і відносний, що дозволить порівнювати за точністю спостереження з різномірними ознаками, користуючись наведеною вище шкалою.

$$HIP_{05\%} = (t_{05}\sqrt{2s^2/n}) \times 100/\bar{x}_0 = 0,82 \times 100/45,7 = 1,8\%,$$

де $\bar{x}_0$ – середня врожайність коренеплодів у досліді.

Сьомий етап

порівняння істотності абсолютних і відносних різниць між варіантами

Відхилення між контрольним (стандартом) і варіантами, що вивчаються, визначають таким чином (див. табл. 4.5):

- а) абсолютне: $48,7 - 45,8 = +2,9$;
- б) відносне: $2,9 \times 100 / 45,8 = 6,3 \%$.

Урожайність коренеплодів гібридів буряків цукрових, т/га				
Варіанти	Середня	Відхилення від контролю (стандарту)		Група відносно різниці до HP_{05}
		абсолютне	відносне, %	
I (контроль)	45,8	-	-	Контроль
2	48,7	+2,9	+6,3	I
3	45,7	-0,1	-0,2	II
4	40,6	-5,2	-11,3	III
5	47,8	+2,0	+4,4	I
HP_{05}	$\bar{x}_0 = 45,7$	0,8	1,8	

Оцінку різниць між варіантами проводять на підставі порівняння їх з HR_{05} .

Якщо різниця між двома варіантами дорівнює або перевищує HR_{05} , то вона істотна; якщо ж менше HR_{05} , то не істотна.

За різницями між варіантами, що вивчаються, і контролем визначають їхні групи:

- I - сорти або гібриди, варіанти, що істотно переважають контроль, маючи між собою різницю, що дорівнює або більша HR_{05} ;
- II - різниця між ними менша від HR_{05} - істотно не відрізняються від контролю;
- III - сорти, варіанти за врожайністю істотно поступаються перед контролем, маючи урожайність менше від контролю на величину, модуль якої дорівнює або більше HR_{05} .

Отже, за врожайністю коренеплодів істотно перевищують контроль варіанти 2 і 5 (I група); у варіанті 3 урожайність істотно не відрізняється від контролю (II група), а у варіанті 4 - істотно поступається перед контролем (III група).