

Дисперсійний аналіз впливу певного фактора на досліджуваний біологічний показник.

Теоретичні відомості

Біологічні показники залежать від великої кількості зовнішніх та внутрішніх факторів. Розроблені спеціальні методи за допомогою яких можна було б виділити вплив окремих факторів та оцінити їх відносну роль. Цей метод отримав назву дисперсійного аналізу, який допомагає біологу вирішити задачі по вивченню і прогнозу впливу факторів на біологічні об'єкти. Методологію дисперсійного аналізу розробив англійський математик та біолог Р.Фішер. Дисперсійний аналіз дозволяє визначити, по-перше, вплив окремих факторів та їх взаємодію, по-друге, вплив неврахованих факторів, по-третє, оцінити як якісні, так і кількісні впливи.

Метод одержав таку назву тому, що в його основі лежить аналіз різноманітності ознак, показником якої є дисперсія:

Дисперсія = Сума квадратів відхилень/кількості ступенів вільності.

Суть методу полягає у вивченні статистичних впливів одного або декількох факторів на результативну ознаку. Статистичний вплив слід відрізнити від фізіологічного. Якщо фактор узагалі змінює ознаку об'єкта, вважають, що він чинить фізіологічну дію. Про статистичний вплив говорять у тих випадках, коли виявляється, що різні градації неоднаково впливають на ознаку. Про відсутність статистичного впливу фактора свідчить однакова дія різних градацій на ознаку. Статистичний вплив проявляється у відмінностях між середніми арифметичними, варіантами різних груп. Загальним методичним прийомом дисперсійного аналізу є порівняння декількох груп за різноманітністю об'єктів.

Причиною мінливості ознаки є певний фактор або декілька факторів. Фактор – це будь-який вплив (температура, час, хімічна речовина, екологічна ситуація та ін.) чи стан організму (стать, вік та ін.), який може вплинути на мінливість ознаки. Кожен фактор може мати одне або декілька значень, які називають градаціями фактора (різна температура, різний час дослідження, різні дози речовини, різні генотипи, пори року і т.п.). Наявність декількох градацій фактора – необхідна умова щодо проведення дисперсійного аналізу. Вивчаючи вплив антибіотиків на організм, розглядаються її різні концентрації. Досліджуючи вплив радіації на об'єкти, діють різними дозами випромінювання.

Дослідження впливу одного фактора здійснюють за допомогою *однофакторного дисперсійного аналізу*, а дослідження впливу двох факторів – за допомогою *двофакторного аналізу*, дослідження впливу трьох та більше факторів здійснюють за допомогою *трифакторного та багатфакторного дисперсійного аналізу*. У дисперсійному аналізі вивчають три види статистичних впливів: факторіальний, випадковий, загальний. Факторіальний вплив – це вплив досліджуваного фактора. Однак, на ознаку діють не тільки фактори, що організовані дослідником, а безліч інших, які невраховані експериментатором. Ці фактори називають випадковими, а їх вплив – випадковим впливом. Чим більше факторіальний вплив відрізняється від випадкового впливу, тим

статистично більш достовірним вважається дія фактора. Сумарний вплив усіх організованих і неорганізованих факторів називають загальним впливом.

Окремі значення заносять у таблицю, в якій окремі варіанти розподілені за групами відповідно до градацій фактора. Таку таблицю називають *дисперсійним комплексом*. Якщо в усіх градаціях фактора міститься однакова кількість варіант, комплекс називають рівномірним, якщо неоднакова – нерівномірним. Рівномірний комплекс на практиці складно організувати, тому що в процесі роботи окремі особини внаслідок випадкових причин (хвороба, загибель) вилучаються з аналізу, і комплекс, що спочатку задуманий як рівномірний перетворюється на нерівномірний.

Таблиця 4.1 Загальний вид вихідних даних для однофакторного дисперсійного аналізу

Фактор А	1	y_{11}	y_{12}	y_{13}	y_j
	2	y_{21}	y_{22}	y_{23}	y_{2j}
	3	y_{31}	y_{32}	y_{33}	y_{3j}
	4	y_{41}	y_{42}	y_{43}	y_{4j}

При формуванні дисперсійних комплексів необхідно дотримуватись наступних умов: фактори, що діють на ознаку мають бути незалежними, а вибірки повинні формуватись у випадковому порядку. Для складання дисперсійного комплексу потрібно розподілити варіанти за градаціями. Дисперсійний аналіз дозволяє вивчати як кількісні, так і якісні ознаки.

Типи варіювання варіант і його характеристика. У мінливості можна виділити три типи варіювання:

- 1) загальне варіювання всіх варіант незалежно від того, в якій групі вони знаходяться навколо загального середнього;
- 2) варіювання групових середніх арифметичних кожного рівня для даного фактору навколо загального середнього арифметичного ;
- 3) варіювання варіант всередині кожної групи навколо кожного групового середнього арифметичного.

Для усіх трьох типів варіювання розраховують суми квадратів відхилення та дисперсії.

Однофакторний дисперсійний аналіз. У найпростішому випадку таблиця вихідних даних має вигляд, представлений у табл.4.1.

Для однофакторних рівномірних комплексів можна використати наступну схему алгоритму:

1. Групування даних у таблиці таким чином, щоб градації фактора розміщались по вертикалі, а варіанти по відповідних градаціях фактору. Далі обчислюють
 - а) вибіркові середні значення досліджуваного біологічного показника для кожного рівня:

$$M_j = \frac{\sum x_j}{n_j}, \text{ де } n_j \text{ – обсяг відповідної вибірки (j=1, 2, \dots, k);}$$

- б) загальне середнє значення показника для всього комплексу даних:

$$M = \frac{\sum x}{n}, \text{ де } n \text{ – сумарний обсяг усіх вибірок.}$$

2. Обчислення сум квадратів відхилень:

а) даних від загального середнього – загальна сума квадратів відхилень

$$SS = \Sigma(x - M)^2;$$

б) вибірових середніх для всіх рівнів фактора від загального середнього – факторіальна (міжгрупова) сума квадратів відхилень.

$$SS_x = n_j \Sigma(M_j - M)^2;$$

в) даних від вибірових середніх (залишок) – залишкова (внутрішньогрупова) сума квадратів відхилень

$$SS_R = SS - SS_x.$$

3. Визначення кількості ступенів вільності:

$\gamma = n - 1$ – для загальної мінливості;

$\gamma_x = k - 1$ – для факторіальної (міжгрупової) варіації;

$\gamma_R = n - k$ або $\gamma_R = (n - 1) - (k - 1)$ – для залишкової внутрішньогрупової варіації, де k – число градацій фактора.

4. Визначення дисперсій як відношення суми квадратів відхилень до відповідного числа ступенів вільності:

$$\sigma_x^2 = \frac{SS_x}{k-1} \text{ факторіальна або між групова дисперсія};$$

$$\sigma_R^2 = \frac{SS_R}{n-k} \text{ залишкова або внутрішньогрупова дисперсія};$$

$$\sigma^2 = \frac{SS}{n-1} \text{ – загальна дисперсія для всього комплексу.}$$

5. Обчислення частки впливу досліджуваного фактора:

$$\eta_x^2 = \frac{SS_x}{SS},$$

Інших факторів (неврахованих в експерименті) –

$$\eta_R^2 = \frac{SS_R}{SS}$$

Частки впливу факторів можуть бути подані у процентах.

6. Встановлення достовірності впливу певного фактора на досліджуваний біологічний показник.

$$F = \frac{\sigma_x^2}{\sigma_R^2}$$

Значення коефіцієнта Фішера порівнюють з $F_{\text{табл}}$ для $p \geq 0.95$, яке знаходять за величинами кількості ступенів вільності $\gamma_1 = k - 1$ $\gamma_2 = n - k$. Якщо $F > F_{\text{табл}}$, то вплив фактора є статистично істотним (з вказаним рівнем імовірності).

Результати дисперсійного аналізу заносять у таблицю.

Таблиця 3. Схема однофакторного дисперсійного аналізу

Джерело мінливості	Сума квадратів відхилень	Кількість ступенів вільності	Дисперсія (σ^2)	Частка впливу (η^2)	Коефіцієнт Фішера
--------------------	--------------------------	------------------------------	--------------------------	----------------------------	-------------------

Фактор А	SS_x	$k-1$	$SS_x/ k-1$	SS_x/ SS	σ^2_x/ σ^2_r
Невраховані фактори	SS_r	$n-k$	$SS_r/ n-k$	SS_r/ SS	—
Загальна мінливість	SS	$n-1$	$SS/ n-1$	100 %	—

Наведена схема використовується як при однофакторному, так і при багатофакторному аналізі. Однак зі збільшенням кількості ознак ускладнюється процес розчленування варіації за джерелами утворення.

Схема розкладання в залежності від числа факторів:

– однофакторний дисперсійний аналіз:

$$D = D_A + D_r,$$

де D – загальна дисперсія, D_A – факторіальна дисперсія, D_r – залишкова дисперсія.

– двофакторний дисперсійний аналіз – факторіальна дисперсія розкладається на складові окремих факторів (А і В) та їх взаємодії (АВ), тому загальну дисперсію визначають за формулою

$$D = D_A + D_B + D_{AB} + D_r .$$

– трифакторний дисперсійний аналіз – міжгрупова дисперсія розкладається на дисперсії трьох факторів та взаємодії між всіма ними. Звідси загальна дисперсія визначається як сума окремих дисперсій, що виникають під дією трьох різних факторів:

$$D = D_A + D_B + D_C + D_{AB} + D_{BC} + D_{AC} + D_{ABC} + D_r$$

Багатофакторні дослідження без повторностей можна опрацювати методом дисперсійного аналізу. Однак, слід пам'ятати, що в двофакторних дослідженнях без повторностей втрачається цінна інформація про парні взаємодії, а менш цінну – взаємодіями більш високих порядків зміщуються з випадковими похибками.