

Статистичний аналіз варіаційного ряду показників за кількісної мінливості ознаки

Мета роботи: оволодіння методикою статистичного аналізу варіаційного ряду з кількісною мінливістю спостереженої ознаки.

Приклад.

На полі пшениці з густрою посівів 4,5 млн рослин на гектар, потрібно визначити середню кількість зерен в колосі. Оскільки зробити це для всіх 4,5 млн рослин практично неможливо, то розглядають частину цієї кількості, вибірку із 500 рослин. По середній кількості зерен в колосі для цієї вибірки роблять висновок про генеральну сукупність всіх 4,5 млн рослин пшениці. Величини показників дослідженої ознаки в генеральній сукупності об'єктів називають параметрами, а у вибірці – статистиками.

Генеральна сукупність являє собою точно обмежену і однозначно визначену сукупність всіх об'єктів (рослини, їхні морфологічні частини і т.д.), що підлягають дослідженню для отримання певної інформації. Оцінювання генеральної сукупності обмежується розглядом її репрезентативної частини, а саме, **вибірки**. Важливо, щоб вибірка була виділена випадковим чином.

Варіювання (мінливість) – генетично і екологічно обумовлена властивість піддослідних об'єктів відрізнятися між собою, навіть в однорідних зведених сукупностях (групах).

Є два типи мінливості ознак: **кількісна та якісна**. *Ознаками кількісної мінливості є ті, що можна виміряти, зважити, порахувати.* Відрізняють неперервну та перервну (дискретну) кількісну мінливість. *Неперервна мінливість* властива для ознак, показники яких визначають у вигляді дробових чисел: довжина стебла, висота рослин, площа листя, маса врожаю тощо. *Перервна мінливість* характеризує поштучні об'єкти: кількість шкідників, листків, коренеплодів, рослин чи їхніх органів тощо.

Для повного аналізу кількісної мінливості ознаки розраховують ряд похідних статистик. Основними із них є **середня арифметична, стандартне відхилення**, також розраховують дисперсію (s^2); помилку середньої арифметичної (s_x); коефіцієнт варіації (V); коефіцієнт вирівняності (st); відносну помилку (точність досліду, $s_x\%$); довірчий інтервал на певному рівні значущості (скажімо, 5 %-му $x \pm t_{05}s_x$).

В агрономічній практиці найчастіше використовують середню арифметичну, як просту і зважену. Середня арифметична проста використовується для характеристики малих вибірок, а зважена — великих. Середня арифметична є основною статистикою кожного варіаційного ряду. Всі інші лиш всебічно характеризують середню арифметичну величину.

Методика аналізу варіаційного ряду за кількісної мінливості ознаки.

Малими вважають вибірки з об'ємом менше 30. Їх характеризують статистики: середня арифметична, дисперсія, стандартне відхилення, коефіцієнт варіації, коефіцієнт вирівняності, абсолютна помилка вибірки і відносна помилка вибірки (точність). Для прикладу розглянемо алгоритм аналізу на прикладі величини урожайності пшениці озимої, визначеної у п'яти повтореннях.

В результаті обліку урожайності пшениці озимої були отримані наступні дані: кількість повторень $n = 5$; урожайність по ділянках $m/га$: 5,21; 4,93; 5,55; 4,88; 5,00. Розрахунок статистик цього варіаційного ряду здійснимо в такій послідовності:

Середня арифметична проста

Середня арифметична проста ($\bar{x}$) – відношення суми показників всіх варіантів ($\sum X$) до їхнього числа (n).

$$\bar{x} = \frac{\sum X}{n} = \frac{5,21+4,93+5,55+4,88+5,00}{5} = 5,11 \text{ m/za.}$$

Показники неперервної ознаки, вираженої дробом, округлюють з точністю до тисячної частки – для чисел без цілих величин, до соті – для однозначних цілих, до десятої – двозначних і до одиниці – тризначних. Основною властивістю середньої арифметичної є рівність суми всіх додатних і всіх від’ємних відхилень від неї, тобто сума відхилень всіх окремих варіант від $\bar{x}$ дорівнює 0 ($\sum(X - \bar{x}) = 0$).

І справді $(5,21-5,11) + (4,93-5,11) + (5,55-5,11) + (4,88-5,11) + (5,00-5,11) = 0$.

Дисперсія

Дисперсія (S^2) і стандартне відхилення (S) є основними показниками варіації, розсіювання досліджуваної ознаки.

Дисперсія являє собою частку від ділення суми квадратів відхилень ($\sum(X - \bar{x})^2$) до числа ступенів свободи ($n-1$)

$$S^2 = \frac{\sum(X - \bar{x})^2}{n - 1}$$

Число ступенів свободи – це частина варіант, що вільно змінюються за відомої суми їх показників і середнього показника, визначена числом всіх варіант без однієї, залежної від решти.

Стандартне відхилення

Стандартне відхилення, або середнє квадратичне відхилення – найбільш використовуваний індикатор мінливості об'єкта, що показує, на скільки в середньому відхиляються індивідуальні показники ознаки x від їхньої середньої величини $\bar{x}$. Цю статистику виражають в одиницях виміру самої спостереженої ознаки і розраховують, як квадратний корінь із дисперсії її величини:

$$S = \sqrt{\frac{\sum (X - \bar{x})^2}{n - 1}} = \quad (2.2)$$
$$= \sqrt{\frac{(5.21 - 5.11) + (4.93 - 5.11) + (5.55 - 5.11) + (4.88 - 5.11) + (5.00 - 5.11)}{5 - 1}} = 0.27 \text{ m/ga}$$

Коефіцієнт варіації

Коефіцієнт варіації (V) – відносна величина, що слугує для оцінювання величини коливання (мінливості) ознаки. Використання коефіцієнта варіації є суттєвим лише за вимірювання показників ознаки, яка має позитивні величини. Модуль коефіцієнта варіації визначають відношенням стандартного відхилення до середньої арифметичної, вираженим у відсотках:

$$V = \frac{s}{\bar{x}} \times 100 = \frac{0,27 \times 100}{5,11} = 5,4 \%$$

Мінливість ознаки оцінюють за шкалою: якщо $V =$ до 10 %, показники у вибірці вирівняні, $V = 10$ –20 % – їх строкатість середня і за $V > 20$ % – дуже строкаті.

У нашому прикладі показники урожайності пшениці виявились мало мінливими у повтореннях, вирівняними.

Коефіцієнт варіації застосовується тоді, коли необхідно оцінити мінливість ознак об'єкта, які виражені в різних одиницях вимірювання.

Для більш детальної характеристики вирівняності показників кількісної мінливості використовують коефіцієнт вирівняності (стабільності st).

$$s_t, \% = 100 - V$$

Помилка вибіркової середньої

Помилка вибіркової середньої (S_x) – кількісна різниця між середніми величинами ознаки вибіркової і генеральної сукупності. Ці помилки властиві лише вибіркового методу досліджень, їх чисельна величина залежить від ступеня мінливості досліджуваних ознак (S - стандартне відхилення) і об'єму вибірки (n). Її обчислюють за формулою:

$$S_{\bar{x}} = \frac{s}{\sqrt{n}} = \frac{0,27}{\sqrt{5}} = 0,12 \text{ m/za}$$

Помилка вибіркової середньої зменшується із збільшенням кількості вимірювань, з яких вираховують середню арифметичну, і зменшенням мінливості досліджуваної ознаки.

Відносна помилка вибіркової середньої або точність дослід

Відносна помилка вибіркової середньої або точність дослід ($sx\%$) – це відношення абсолютної величини помилки (sx) вибіркової середньої до величини останньої x , виражене у відсотках:

$$s_{\bar{x}\%} = \frac{s_x}{\bar{x}} \times 100 = 2,4 \%$$

Точність дослід вважається відмінною, якщо $sx\% = 1-2 \%$; доброю – $sx\% = 2-5 \%$; задовільною – $sx\% = 5-8 \%$ і незадовільною – $sx\% > 8 \%$.

–

Довірчий інтервал на 5 %-му рівні значущості розраховується за отриманими даними:

$$\bar{x} \pm t_{05} s_{\bar{x}} = 5,11 \pm 2,78 \times 0,12 = 5,11 \pm 0,33 = 4,88 \div 5,44 \text{ т/га},$$

В агрономічних дослідженнях найчастіше використовують 95 % рівень ймовірності, тобто можливість того, що істинне значення наших статистик не попаде в цей інтервал складає 5 % (рівень значущості).

Довірчий інтервал на 5 %-му рівні значущості розраховується за отриманими даними:

$$\bar{x} \pm t_{05} s_{\bar{x}} = 5,11 \pm 2,78 \times 0,12 = 5,11 \pm 0,33 = 4,88 \div 5,44 \text{ т/га}$$

Тут за чотирьох ступенів свободи (табл. 1 додатків, $\nu = n-1 = 5-1 = 4$) величина критерію Ст'юдента $t_{05} = 2,78$.

Таким чином ми переконуємось, що довірчий інтервал $4,88 \div 5,44$ із ймовірністю 95 % містить в собі істинне значення μ , об'єднуючи в собі на 5-% рівні значущості 95 % всіх варіант вибірки.