

Тема II. Угрупування результатів спостережень

1. Таблиці та ряди розподілу.
2. Класифікація ознак.
3. Побудова варіаційних рядів.
4. Графіки розподілу.

1. Таблиці та ряди розподілу

Результати дослідження фіксуються зазвичай в первинних документах - протоколах дослідів, польових щоденниках, робочих журналах і т.п. Зібраний фактичний матеріал потім піддається статистичній обробці. Мета обробки - витяг з маси фактів укладеної в них інформації, отримання на підставі проведеного дослідження об'єктивних і переконливих висновків.

Перший крок на шляху статистичної обробки - угрупування зібраних даних відповідно до завдань дослідження і тими умовами, в яких воно проводилося. Найбільш раціональною формою угрупування служать статистичні таблиці. У них зазвичай зводяться результати масових спостережень. Статистичні таблиці бувають складні і прості, і їх будова залежить від того, за якими ознаками і за якою їх кількістю групується матеріал, а також від завдань, які вирішуються угрупуванням зібраного матеріалу. Приклад порівняно простий угрупування: результати П.М. Константинова (1955), отримані в досвіді по випробуванню врожайності ячменю і вівса в умовах нечорноземної смуги Російської Федерації:

Таблиця 1

Культура	Врожайність зерна по рокам досвіду (ц/га)							Середня урожайність
	1928	1929	1930	1931	1932	1933	1934	
Ячмінь	7,70	9,00	9,40	7,40	7,40	10,90	8,0	8,54
Овес	8,26	7,22	8,43	5,57	6,35	8,00	9,13	7,57

Прикладом складних таблиць, що ілюструють залежність однієї з варіюючих ознак від змін іншої, служать кореляційні таблиці, а також таблиці дисперсійних комплексів.

Найбільш просту форму статистичного угруповання представляють ряди розподілу, які будуються на основі операції ранжирування, тобто шляхом розташування варіант (окремих числових значень варіюючої ознаки) в зростаючому або спадному порядку. Наприклад, є такий сукупність 20 вимірювань ознаки: 2, 5, 3, 6, 4, 7, 4, 5, 6, 6, 5, 9, 5, 6, 10, 8, 12, 9, 7, 6.

Видно, що ознака варіює від 2 до 12 одиниць. Розташуємо цю сукупність у зростаючому порядку:

2, 3, 4, 4, 5, 5, 5, 5, 6, 6, 6, 6, 6, 7, 7, 8, 9, 9, 10, 12.

Вийде ранжирований ряд значень ознак.

При розподілі членів сукупності в ряд переслідуються певні цілі. Одна з них - розкриття закономірності варіювання досліджуваної ознаки. Тому до рядів розподілу пред'являються певні вимоги:

1) вони повинні бути легкодоступні для огляду;

2) добре ілюструвати закономірність варіювання. Ранжирований нами ряд сам по собі погано задовольняє цим вимогам. Якщо, ті ж варіанти розташувати у вигляді подвійного ряду, враховуючи їх повторюваність в загальному строю, сукупність розподілиться таким чином:

Таблиця 2

Варіанти	2	3	4	5	6	7	8	9	10	11	12
Повторюваність варіант (р)	1	1	2	4	5	2	1	2	1	0	1

Такий упорядкований ряд розподілу, в якому вказана повторюваність варіант, що належать до даної сукупності, називається варіаційним рядом. Повторюваність варіант в сукупності називають вагами або частотами.

У біометрії ознаки позначають прописними буквами $X, Y, Z \dots$ їх числові значення - малими x_1, x_2, x_3, \dots або y_1, y_2, y_3, \dots . Їх частоти позначаються

латинською буквою p . Загальна кількість варіант, що входять до складу даної сукупності, називають обсягом сукупності і позначають буквами n або N .

$$\sum p = n \text{ [1].}$$

Відносна частота або частість обчислюється як частка від ділення p / n і виражається в частках або відсотках.

$$\sum \frac{p}{n} = 1 \text{ [2].}$$

Заміна абсолютних значень ознаки (частот) частостей полегшує зіставлення одного варіаційного ряду з іншим. І робить більш виразними характерні риси варіювання.

2. Класифікація ознак

Біологічні ознаки діляться на якісні та кількісні. До якісних належать такі ознаки, як забарвлення листя і квіток, смак і запах продуктів. Якщо мова йде про вимірювані або обчислювальні величини - це будуть кількісні ознаки.

У варіаційні ряди розподіляються тільки кількісні ознаки, а якісні ознаки зазвичай розглядають в альтернативній формі. Кількісні ознаки можуть бути рахунковими (варіюють дискретно) та мірними (варіюють безперервно).

3. Побудова варіаційних рядів

Відомо два види варіаційних рядів: безінтервальні та інтервальні. Прикладом безінтервального варіаційного ряду може служити розподіл американських вугрів ($n = 863$) за кількістю хребців (по Бергу, 1924):

Таблиця 3

Число хребців (x)	103	104	105	106	107	108	109	110	111
Число особин (p)	1	8	45	183	274	221	96	31	3

В даному випадку ознака варіює слабо. Але багато ознак варіюють у дуже широких межах і розподіл їх в безінтервальний ряд не досягає мети: ряди виходять занадто розтягнутими, погано доступним для огляду, що не відображають чітко закономірності варіювання. Наприклад, число зерен ячменю в 50 колосках:

Таблиця 4

Число зерен (x)	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27
Кількість колосків (p)	1	1	1	1	1	1	6	5	6	5	3	2	5	5	2	3	1	0	1

У таких випадках найкращий результат виходить від розподілу сукупності в інтервальний варіаційний ряд. Для цього вся варіація ознаки від мінімуму до максимуму варіанти розбивається на рівні інтервали (від і до) або класи. Потім всі варіанти розподіляють по класах і частоти p будуть частотами класів. Для обчислення середніх величин на таких дискретних варіаційних рядах використовуються не класові інтервали, а їх середні, рівні півсумі верхньої і нижньої меж класу.

Число класів залежить від завдання дослідження і характеру зібраного матеріалу. Ширина класового інтервалу позначається не тільки на характері розподілу варіант по класах, а й на точності середніх характеристик. Установка широких класових інтервалів спотворює типові риси варіювання і

веде до зниження точності числових характеристик ряду. При виборі надмірно вузьких інтервалів точність узагальнюючих числових змінних підвищується, але ряд виходить занадто розтягнутим і не дає чіткої картини варіювання. Щоб визначити величину класового інтервалу для побудови доступного для огляду варіаційного ряду Г.А. Стерджес (Sturges, 1926) рекомендує наступну формулу:

$$i = \frac{x_{max} - x_{min}}{1 + 3,32lg_n} [4] \text{ или } i = \frac{x_{max} - x_{min}}{5xlg_n}$$

Формулою [5] рекомендується користуватися при наявності в сукупності великого числа членів ($n > 100$).

Отже, розглянемо побудову варіаційного ряду для випадку з колосками ячменю: $x_{min} = 9$, $x_{max} = 27$; величина класового інтервалу буде наступною - $i = 3$ [4] і $i = 2$ [5]. Візьмемо $i = 3$. При розбивці варіації на класи межі першого класу встановлюємо так, щоб мінімальна варіанту потрапила приблизно в середину цього класу. Якщо нижня (мінімальна) межа 1-го класу дорівнює 8, то вийде 8 класів з нижніми межами, рівними 8, 11, 14, 17, 20, 23, 26, 29. Щоб однозначно вирішити питання про приналежність варіанти до якогось класу, його верхню межу зменшують на 0,1 або 0,01, що і дає необхідне розмежування класів. наприклад, таблиця 5

Таблиця 5

Класи по числу зерен в колосках ячменю	Середнє значення класів (x)	Частоти (p)
8-10,9	9,5	2
11-13,9	12,5	3
14-16,9	15,5	12
17-19,9	18,5	14
20-22,9	21,5	12
23-25,9	24,5	6
26-28,9	27,5	1
сума	-	50

В результаті виходить інтервальний варіаційний ряд з переривчастим варіюванням. При побудові варіаційного ряду не допускається подвійний облік однієї і тієї ж варіанти.

4. Графіки розподілу

Щоб надати більшу наочність закономірності варіювання ознак, варіаційні ряди прийнято зображати графічно у вигляді гістограми, або полігону, а також у вигляді кумуляти або огівни.

Графік, званий гістограмою розподілу частот, виходить, якщо в системі координат по осі абсцис відкласти межу класів, а по осі ординат - їх частоти.

У випадку з розподілом класів ячменю за кількістю зерен в колосі (розглянутий раніше) гістограма буде виглядати, як показано на рис. 1.



Рис . 1. Гістограма розподілу класів ячменю за кількістю зерен

Гістограма зображує закономірності розподілу варіант по класах варіаційного ряду, тобто при безперервному варіюванні ознаки. Прямокутники відповідають класам, а їх висота - частотам варіаційного ряду.

Якщо з середніх точок прямокутників гістограми опустити перпендикуляри на вісь абсцис, а самі точки з'єднати між собою, вийде графік дискретного варіювання, званий полігоном розподілу (рис. 2).

Полігон розподілу можна побудувати незалежно від гістограми, завдаючи на вісь абсцис середні значення класів. А коли виникає необхідність, можна полігон перетворити в гістограму.

В інших випадках графік варіаційного ряду будується в вигляді кумуляти (від слова *sumulo* - накопичувати). Для побудови кумуляти розподілу класів ячменю за кількістю зерен скористаємося даними 1 та 3 стовпців табл. 5. При цьому по осі абсцис відкладають значення класових варіант, а по осі ординат - накопичені частоти (рис. 3).

Таблиця 6

Середнє значення класів	Частоти (P)	Накопичення частоти
9,5	2	2
12,5	3	5
15,5	12	17
18,5	14	31
21,5	12	43
24,5	6	49
27,5	1	50
Сума	50	50

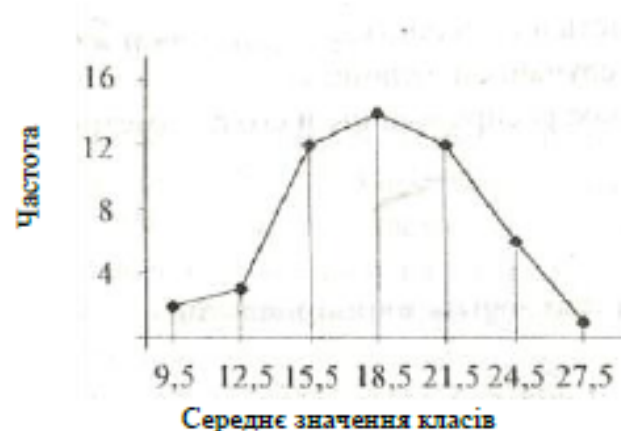


Рис. 2. Полігон розподілу класів ячменю за кількістю зерен

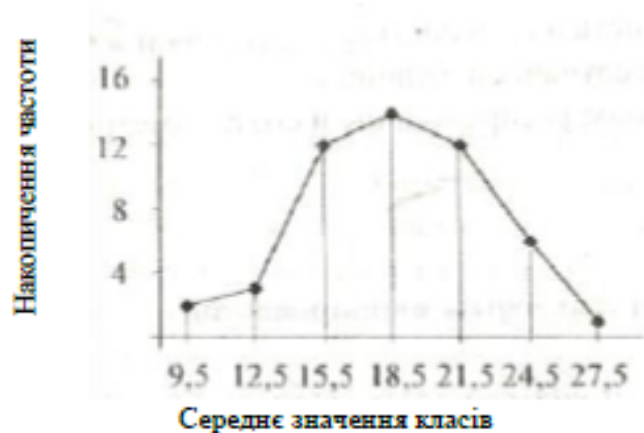


Рис. 3. Кумулята розподілу класів ячменю за кількістю зерен

Якщо ряд накопичених частот нанести на вісь абсцис, а значення варіант розташувати по осі ординат і побудувати графік, виходить огива.

Значення графіків полягає в їх наочності. Але вони не дають точної характеристики варіюючої ознаки, так як залежать від прийнятих масштабів. Точну характеристику варіюючих ознак дають статистичні показники, про які йтиметься далі.