Тема III. Закономірності розподілу

1. Характерні риси варіювання.
2. Імовірність і її властивості.
3. Поняття випадкової величини.
4. Нормальний розподіл і його параметри.

1. Характерні риси варіювання

У розподілі емпіричних сукупностей кидається в очі одна важлива особливість - переважне накопичення варіант в центральних класах і поступове убування їх числа в міру віддалення від серединної точки варіаційного ряду. Ця особливість, яка становить одну з характерних рис варіювання біологічних ознак, - факт дуже важливий, що має широке поширення в природі (рис. 4). У самих різних випадках проявляється одна і та ж закономірність: в масі відносно однорідних одиниць (варіант) переважна більшість складають варіанти середнього розміру, і чим далі вони відхиляються від середнього рівня ознаки, тим рідше зустрічаються в даній сукупності. Отже, між значеннями ознаки та їх зустрічаємістю існує певний зв'язок. Наочним виразом зв'язку з цим служить варіаційний ряд і його графік - варіаційна крива.

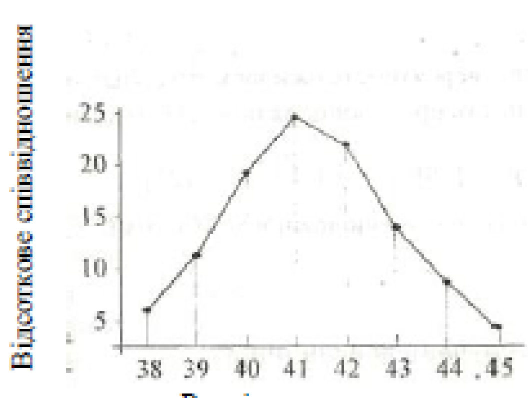

Рис. 4. Полігон процентного співвідношення в розподілі чоловічого взуття серед населення центральних областей Російської Федерації

2. Імовірність і її властивості

Припустимо, що в пологовому будинку народилося 208 хлопчиків і 200 дівчаток, всього 408 дітей. Число 208 - це абсолютна частота народжених хлопчиків, а число 200 абсолютна частота народжених дівчаток. Якщо число хлопчиків і дівчаток, які народилися в цьому пологовому будинку, віднести до загальної кількості новонароджених дітей, вийдуть відносні частоти або частости цих подій:

208/408 = 0,51 відносна частота народжених хлопчиків,

200/408 = 0,49 відносна частота народжених дівчаток.

Однак теоретично повинно народжуватися порівну хлопчиків і дівчаток. Теоретичне значення відносної частоти очікуваної події називається його ймовірністю. Подія - це той результат, який виходить при кожному випробуванні. Якщо при кожному випробуванні подія неминуче настає, вона називається достовірною. Якщо немає - неможливою. Якщо і може наступити і не може - це випадкова подія. Події, які при випробуванні в постійних умовах повторюються багаторазово, називаються масовими.

Ймовірністю називається відношення числа випадків або випадків m сприяють настанню очікуваної події A, до числа всіх можливих і несумісних в даному випробуванні результатів n, тобто

$$P(A) = \frac{m}{n} \text{ [19]}$$

Для зручності, ймовірність очікуваної події прийнято позначати через p, а ймовірність протилежної події через q, тоді

$P(A) = p$ [20], $P(\bar{A}) = q$ [21], $p + q = 1$ [22],

де $(\bar{A})$ - ймовірність протилежної події (*не A*).

3. Поняття випадкової величини

Все, що можна підрахувати або виміряти, називається величиною. Величини (або "ознаки") діляться на постійні і змінні. Постійною називається величина, яка в заданих умовах не змінює свого значення. Змінна - це така величина, яка в даних умовах здатна приймати різні числові значення.

Змінна величина називається випадковою, якщо в заданих умовах вона може приймати то одні, то інші значення.

Випадкова величина в N повторних випробуваннях може приймати самі різні значення, але в кожному окремому випробуванні вона приймає завжди тільки одне з можливих значень. Яке значення прийме випадкова величина в результаті кожного випробування, заздалегідь сказати неможливо. Тому, характеризувати випадкову величину можна лише з певною ймовірністю, тобто вказуючи ймовірність її можливих значень.

4. Нормальний розподіл і його параметри

При відкритті закону розподілу змінної випадкової величини була виведена формула залежності, графік функції якої називається кривою нормального розподілу. Її площа дорівнює 1. Чим далі від середньої арифметичної стоїть значення, тим менше ймовірність (частота) появи даного значення.

Нормальний розподіл характеризується математичним очікуванням (відповідає поняттю середньої арифметичної) і дисперсією (мірою варіювання) випадкової величини. Дисперсія служить мірою відхилення можливих значень випадкової величини X від її математичного очікування M. Випадкова величина, що належить до даної сукупності, може відхилятися від середньої арифметичної до трьох ст.

1. Види середніх і їх значення

Однією з найважливіших узагальнюючих характеристик варіюючих ознак є середня величина. Характеризуючи ту чи іншу популяцію, кажуть, наприклад, про середню продуктивність тварин або рослин, середню успішність, про середню швидкість біохімічної реакції і т.д. Значення середніх полягає в їх властивості нівелювати індивідуальні відмінності, в результаті чого виступає більш-менш стійка числова характеристика ознаки - не окремих представників, а цілої групи статистичних одиниць.

2. Середня арифметична

а) Проста середня арифметична

$$\bar{x} = \frac{x_1 + x_2 + x_3 + \cdots + x_n}{n} = \frac{\sum x_1}{n} \ [6]$$

$$\bar{x} = \frac{x_1 + x_2 + x_3 + \cdots + x_n}{n} = \frac{\sum x_1}{n} \ [6]$$

б) Зважена середня

Якщо в сукупності спостережень окремі варіанти повторюються *p* раз, то середня арифметична обчислюється за формулою [7] з урахуванням повторюваності (або "ваг") окремих варіант.

$$\bar{x} = \frac{x_1p_1 + x_2p_2 + x_3p_3 + \cdots + x_np_n}{p_1 + p_2 + p_3 + \ \ldots p_n} = \frac{\sum x_1p_1}{p_1} \ [7]$$

Так як середня обчислюється з урахуванням частот або «ваг» окремих варіант, то вона називається зваженою середньою.

Таблиця 7

Кількість зерен в колоску (варіанти) *(x)*	7	8	9	10	11	12	13
Число колосків (повторюваність) *(p)*:	1	1	2	7	3	3	1

x = (7x1+ 8x1+9x2+10x7+11x3+12x3+13x1)/18 = 185/18 =10,28 (зерен)

Аналогічним способом розраховується і загальна середня *(x ̅)* з суми приватних середніх *((x_1) ̅)*.

$$\bar{x} = \frac{x_1n_1 + x_2n_2 + x_3n_3 + \cdots + x_kn_k}{n_1 + n_2 + n_3 + \ \ldots n_n} = \frac{\sum x_1n_1}{n_1} \ [8]$$

3. Скорочений спосіб обчислення середньої арифметичної (спосіб умовної середньої)

Обчислення середньої арифметичної способом зваженої середньої не завжди зручно, особливо на сукупностях великого обсягу і при наявності багатозначних чисел, коли обчислювальна робота стає особливо трудомісткою. У таких випадках простіше розрахувати середню арифметичну спрощеним способом - способом умовної середньої.

Одна з варіант, все одно якась, умовно приймається за середню величину, що позначається через *A*. Зазвичай в якості умовної середньої береться варіанти (або клас) з найбільшою або близькою до неї частотою, хоча це не обов'язково. Потім розраховують відхилення варіант (або класів) від цієї умовної середньої і знаходять середню арифметичну за такою формулою:

$$\bar{x} = A + \frac{\sum pa}{n} \ [9]$$

де А - умовна середня, *a* = *x-A* - відхилення варіанти від умовної середньої.

Наприклад, скористаємося наведеної формулою [9] для визначення середнього числа зерен в 18 колосках озимого жита:

Таблиця 8

Варіанта (x):	7	8	9	10	11	12	13
Повторюваність (p):	1	1	2	7	3	3	1
a:	-4	-3	-2	-1	0	+1	+2
pa:	-4	-3	-4	-7	0	+3	+2

$$\sum pa = +5 - 18 = -13$$

$$\bar{x} = 11 + \frac{-13}{18} = 11 - 0{,}72 = 10{,}28 \text{ (зерен)}$$

Таким же способом можна розрахувати середнє число хребців у американських вугрів. Якщо ж сукупність досить велика, то дані розрахунки простіше проводити за допомогою таблиць. Наприклад, при вивченні вмісту кальцію (в мг%) в сироватці 100 павіанів-гамадрилів отримали значення від 8,99 до 14,7. Розбивши варіацію на класи відповідно до формул [4] або [5], побудуємо таблицю і рознесемо відповідні дані експерименту (табл. 9).

Таблиця 9

Середні значення класів або класові варіанти	Частоти	Відхилення класових варіант від умовної середньої	Похідні відхилень на частоти	$a' = \frac{x - A}{i}$	*a′*
(x)	*(P)*	*(a=x- A)*	*(pa)*		
8,9	2	-2,8	-5,6	-4	-8
9,6	3	-2,1	-6,3	-3	-9
10,3	9	-1,4	-12,6	-2	-18
11,0	17	-0,7	-11,9	-1	-17
Разом	-	-	-36,4	-	-52
A = 11,7	25	0	0	0	0
12,4	23	+0,7	+ 16,1	+ 1	+23
13,1	10	+1,4	+ 14,0	+2	+20
13,8	7	+2,1	+14,72	+3	+21
14,5	4	+2,8	11,2	+4	+ 16
Разом	-	-	+56.0	-	+80
Сума	100	-	+ 19,6	-	+28

$$\bar{x} = A + \frac{\sum pa}{n} = 11{,}7 + \frac{19{,}6}{100} = 11{,}896 \approx 11{,}90 \text{ (мг \%)}$$

Якщо замість, $a = x\text{-}A$ використати $a' = \frac{x-A}{i}$, де i - величина класового інтервалу ($i = 0{,}7$) і формулу

$$\bar{x} = A + i\frac{\sum pa'}{n} \text{ [10]}$$

то результат обчислень виявиться таким же:

$$\bar{x} = 11{,}7 + 0{,}7\left(\frac{28}{100}\right) = 11{,}7 + 0{,}196 = 11{,}896 \approx 10{,}90 \text{ (мг \%)}$$

Порівнюючи перший і другий спосіб розрахунку середньої, бачимо, що другий спосіб набагато простіше.