Лекція 1. ЕЛЕМЕНТИ МАТЕМАТИЧНОЇ СТАТИСТИКИ

1.1. Введення в математичну статистику. Основні поняття. Задачі математичної статистики

Математична статистика – це розділ теорії ймовірностей, який займається встановленням закономірностей випадкових явищ або процесів на основі систематизованої обробки випадкових даних експериментальних вимірювань.

Основні задачі, які розв'язує математична статистика:

1) вказати способи збору та групування (якщо даних дуже багато) статистичних відомостей;

2) визначити закон розподілу випадкової величини або системи випадкових величин за статистичними даними;

3) визначити невідомі параметри розподілу;

4) перевірити правдоподібність припущень про закон розподілу випадкової величини, про значення параметра, який оцінюють, або про форму зв'язку між випадковими величинами.

1.2. Генеральна та вибіркова сукупність та їх властивості. Поняття про вибірковий метод

Множину однорідних об'єктів називають *статистичною сукупністю.*

Генеральною сукупністю називається вся сукупність об'єктів, які спостерігаються, або сукупність усіх спостережень за одним об'єктом. Інакше, генеральною називають сукупність об'єктів, відносно яких проводиться дослідження та з яких можна зробити вибірку.

Кількість об'єктів, які вивчаються, називається об'ємом *вибіркової сукупності.*

Кожний окремий об'єкт генеральної сукупності називається її *елементом*.

Вибірковою сукупністю, або вибіркою, називаються окремі вибрані елементи генеральної сукупності, які використовуються для подальшої статистичної обробки даних.

Кількість елементів у вибірковій сукупності називається *об'ємом (обсягом) вибірки*.

Вибіркова сукупність, яка повністю описує генеральну і цілком їй відповідає, називається *представницькою або репрезентативною*.

Розрізняють *повторні та безповторні вибірки.*

Вибірка називається *повторною*, якщо після вибору кожний елемент повертається до генеральної сукупності та може бути вибраний повторно.

Вибірка називається *безповторною*, якщо після вибору кожний елемент не повертається до генеральної сукупності та у зв'язку з цим його повторний вибір є неможливим.

1.3. Способи відбору статистичного матеріалу

1. Простий випадковий відбір.

2. Типовий відбір.

3. Механічний відбір.

4. Серійний відбір.

1.4. Статистичні показники вибірки

Статистичні показники, що розкривають властивості вибірки, можна представити такими основними групами:

– *емпіричними розподілами* (варіаційними, атрибутивними, ранжированими);

– *вибірковими показниками* (мірами центральної тенденції і мінливості);

– *кореляційно-регресійними показниками* (коефіцієнтами кореляції, регресії).

1.5. Варіаційні ряди та статистичні розподіли

Варіанта – це числова характеристика ознаки, для дослідження якої робиться вибірка.

Порядковий номер варіанти (значення ознаки) називається *рангом*.

Ряд значень ознаки (варіант), розміщених у порядку зростання (спадання) з відповідними їм вагами, називають *варіаційним рядом* (рядом розподілу).

В якості ваг виступають частоти або частості.

Нехай у вибірці обсягу n з варіантами $x_1, x_2, \ldots, x_i$ ознака X прийняла значення $x_1 - n_1$ раз, значення $x_2 - n_2$ раз, ..., значення $x_i - n_i$ раз, тобто кожна варіанта x_i вибірки може бути спостереженою n_i раз $(n_i \geq 1)$. Число n_i називають *частотою варіанти* x_i.

Частота (n_i) показує, скільки разів зустрічається та чи інша варіанта (значення ознаки) в статистичній сукупності.

Ряд $n_1, n_2, \ldots, n_i$ називається *рядом частот*. Сума усіх частот повинна дорівнювати *об'єму вибірки*.

$$n = \sum_{i=1}^{m} n_i ,$$

де m – кількість варіант, які відрізняються числовим значенням, n – обсяг вибірки.

Статистичний розподіл – це математична модель об'єктів реальності у вигляді співвідношень *значень* змінної X, що характеризує властивості вибірки, до частот їх появи.

Приклад 1.1. Систематизувати результати виконання випадковою вибіркою студентів тестових завдань: 3, 4, 4, 4, 3, 2, 4, 4, 5, 1 (обсяг вибірки $n = 10$).

Табл. 1.1. Систематизація результатів виконання тестових завдань (X)

Первинні дані		Варіаційний ряд	Варіанти X	Частоти варіант
i, ранг	x_i	x_i^*	x_i	n_i
1	3	1	1	1
2	4	2	2	1
3	4	3	3	2
4	4	3		
5	3	4	4	5
6	2	4		
7	4	4		
8	4	4		
9	5	4		
10	1	5	5	1

Статистичний розподіл можна класифікувати за ознакою типів вимірювань змінної на: варіаційні, ранжировані та атрибутивні (рис. 1.1).

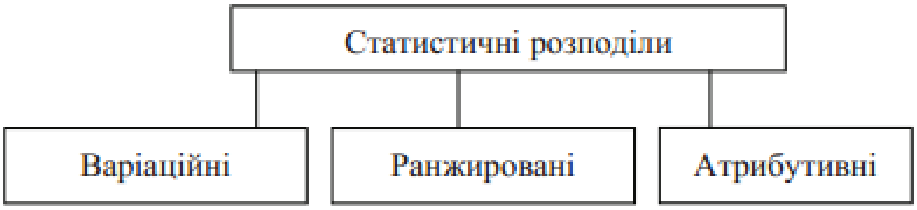

Рис. 1.1. Класифікація статистичних розподілів за типами вимірювань

Основні *види* статистичних розподілів такі: диференціальні та інтегральні, які можуть складатися з абсолютних і відносних частот (рис. 1.2).

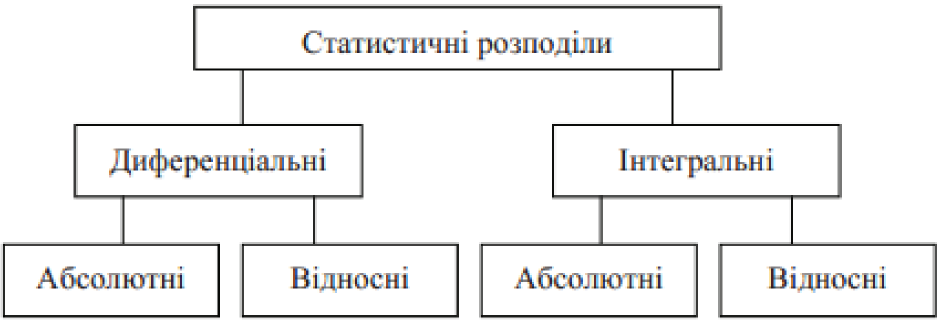

Рис. 1.2. Основні види статистичних розподілів

Диференціальні розподіли представляють значення частот окремо (тобто диференціально) для кожної варіанти x_i змінної X.

Диференціальні абсолютні частоти (або просто частота варіанти) – це кількість об'єктів n_i з однаковими значенням варіанти x_i змінної X.

Диференціальні відносні частоти (або частість, або просто відносна частота варіанти) – це відношення частоти n_i тієї чи іншої варіанти до суми всіх частот ряду n, тобто $\omega_i = \frac{n_i}{n}$. Сума всіх відносних частот дорівнює одиниці.

Інтегральні розподіли («накопичені частоти» або «кумулятивні частоти») формуються як доданки попередніх диференціальних частот. Вони визначають сумарні частоти для варіанти, що не перевищує значення x_i змінної X.

Інтегральні абсолютні частоти (*накопичені частоти*) $N_m = \sum_{i=1}^{m} n_i$ - це накопичена сума диференціальних абсолютних частот від 1-ї до m-ї варіанти.

Інтегральні відносні частоти (накопичені відносні частоти) $W_m = \sum_{i=1}^{m} \omega_i$ - це накопичена сума диференціальних відносних частот від 1-ї до m-ї варіанти.

Приклад 1.2. Розрахувати диференційні та інтегральні розподіли виконаних студентами завдань із прикладу 1.1. (обсяг вибірки $n = 10$).

Послідовність рішення:

1. Характер емпіричних даних відповідає умовам для розрахунку незгрупованих розподілів, оскільки діапазон варіант x_i змінної X містить всього 5 дискретних значень варіант $\{1, 2, 3, 4, 5\}$;

2. Значення варіант від 1 виконаних завдань (мінімальне) до 5 виконаних завдань (максимальне), кількість варіантів $m = 5$;

3. Диференціальні абсолютні частоти (частоти) n_i наступні:

- для $x_1 = 1$ частота $n_1 = 1$ (один об'єкт з цим значенням змінної);
- для $x_2 = 2$ частота $n_2 = 1$ (один об'єкт з цим значенням змінної);
- для $x_3 = 3$ частота $n_3 = 2$ (два об'єкта з цим значенням змінної) і т.д.

Сума всіх абсолютних частот повинна дорівнювати обсягу вибірки:

$n = \sum_{i=1}^{m} n_i$, тобто $n = \sum_{i=1}^{5} n_i = n_1 + n_2 + n_3 + n_4 + n_5 = 1 + 1 + 2 + 5 + 1 = 10$.

4. Диференціальні відносні частоти (частість, відносна частота) $\omega_i = \dfrac{n_i}{n}$ наступні:

- для $x_1 = 1$ частота $\omega_1 = n_1 / n = 1/10 = 0{,}10$ (або 10%);
- для $x_2 = 2$ частота $\omega_2 = n_2 / n = 1/10 = 0{,}10$ (або 10%);
- для $x_3 = 3$ частота $\omega_3 = n_3 / n = 2/10 = 0{,}20$ (або 20%) і т.д.

Сума всіх відносних частот має дорівнювати одиниці (або 100%) $\sum_{i=1}^{5} \omega_i = \sum_{i=1}^{5} \dfrac{n_i}{n} = 1$, тобто 0,10+0,10+0,20+0,50+0,10=1,00.

5. Інтегральні абсолютні частоти $N_i = \sum_{i=1}^{m} n_i$:

- для $x_1 = 1$ частота $N_1 = \sum_{i=1}^{1} n_i = n_1 = 1$;
- для $x_2 = 2$ частота $N_2 = \sum_{i=1}^{2} n_i = n_1 + n_2 = 1 + 1 = 2$;
- для $x_3 = 3$ частота $N_3 = \sum_{i=1}^{3} n_i = n_1 + n_2 + n_3 = 1 + 1 + 2 = 4$ і т.д.

Остання інтегральна абсолютна частота для $x_5 = 5$ дорівнюватиме обсягу вибірки:

$$N_5 = \sum_{i=1}^{5} n_i = n_1 + n_2 + n_3 + n_4 + n_5 = 1 + 1 + 2 + 5 + 1 = 10.$$

6. Інтегральні відносні частоти $W_m = \sum_{i=1}^{m} \omega_i$ наступні:

- для $x_1 = 1$ частота $W_1 = \sum_{i=1}^{1} \omega_i = \omega_1 = 0{,}10$;

- для $x_2 = 2$ частота $W_2 = \sum_{i=1}^{2} \omega_i = \omega_1 + \omega_2 = 0{,}10 + 0{,}10 = 0{,}20$;

- для $x_3 = 3$ частота $W_3 = \sum_{i=1}^{3} \omega_i = \omega_1 + \omega_2 + \omega_3 = 0{,}10 + 0{,}10 + 0{,}20 = 0{,}40$ і т.д.

Остання інтегральна відносна частота для $x_5 = 5$ дорівнюватиме одиниці $W_5 = \sum_{i=1}^{5} \omega_i = 1$ (або 100%), оскільки сума всіх диференціальних відносних частот складає 1,00 або 100%:

$$W_5 = \sum_{i=1}^{5} \omega_i = \omega_1 + \omega_2 + \omega_3 + \omega_4 + \omega_5 = 0{,}10 + 0{,}10 + 0{,}20 + 0{,}50 + 0{,}10 = 1{,}00 = 100\%.$$

Занесемо наші розрахунки до таблиці:

Табл. 1.2. Розподіли кількості виконаних завдань

Кількість виконаних завдань		Частоти:			
		диференціальні		інтегральні	
		абсолютні	відносні	абсолютні	відносні
i	x_i	n_i	$\omega_i = \frac{n_i}{n}$	$N_i = \sum_{i=1} n_i$	$W_i = \sum_{i=1} \omega_i$
1	1	1	0,10	1	0,10
2	2	1	0,10	2	0,20
3	3	2	0,20	4	0,40
4	4	5	0,50	9	0,90
5	5	1	0,10	10	1,00
	Суми	10	1,00		