

## Лекція №11

Статистична обробка неперервних величин.

1. Аналіз вибірки неперервних величин.
2. Гістограма частот та відносних частот.
3. Емпірична функція розподілу.
4. Числові характеристики для інтервального розподілу.

### **1. Аналіз вибірки неперервних величин.**

Зробити аналіз вибірки з неперервної випадкової величини вказаним у попередній лекції способом неможливо, незважаючи на формальне існування набору конкретних значень. Справа в тому, що ймовірність появи будь-якого значення неперервної випадкової величини завжди дорівнює нулю, тим часом як існує певна ймовірність попадання цієї величини в заданий інтервал. Тому безглуздо складати таблицю розподілу для всіх значень неперервної величини, але є сенс розглянути частоту попадання цих значень в деякий інтервал.

Спочатку проводять ранжування вибірки. Як правило (для неперервної величини), конкретні значення вибірки рідко приймають цілочислові значення. Зазвичай поява таких значень свідчить про низьку точність вимірів. Наприклад, врожайність пшениці, виміряна в центнерах з гектара, знаходиться в результаті зважування, тому визначається з точністю, яку допускають застосовані терези. Отже, в результаті ранжування отримують  $n$  значень випадкової величини, серед яких іноді зустрічаються цілочислові.

Знаходять найближче ціле число зліва від найменшого зі значень вибірки та найближче ціле число справа від найбільшого зі значень вибірки.

Знайдені цілі числа утворюють діапазон зміни вибірки. Наприклад, якщо найменшим значенням вибірки було число 33,2, а найбільшим – 75,6, то діапазон зміни буде становити 33–76. Знайдений діапазон розбивають на окремі інтервали, задаючись кроком  $H$  інтервалу, під яким розуміють різницю між найбільшим та найменшим значеннями країв інтервалу. Крок інтервалу вибирають однаковим для всіх інтервалів. Керуючись специфікою отриманої

вибірки, крок інтервалу вибирають таким, щоб можна було вважати, що у кожному з інтервалів спостерігається рівномірний розподіл.

Обчислюють число  $m_i$  значень випадкової величини у кожному з інтервалів, яке називається частотою попадання випадкової величини в заданий, після чого знаходять відносну частоту  $\omega_i = \frac{m_i}{n}$  попадання випадкової величини в заданий інтервал.

Приклад 1. Для знаходження середньої врожайності пшениці на різних ділянках поля було зроблено 20 пробних покосів по 1 м<sup>2</sup> кожен. Після обмолоту врожайність кожної проби була перерахована у центнери з гектара. Провести статистичну обробку вибірки з наступних даних:

39,1; 39,8; 39,3; 40,6; 38,7; 44,2; 42,3; 38,6; 40,7; 37,6; 41,7; 40,8; 39,2; 39,7; 38,8; 39,4; 40,7; 41,3; 40,2; 40,5.

Розв'язання: Проводимо ранжування отриманої вибірки: 37,6; 38,6; 38,7; 38,8; 39,1; 39,2; 39,3; 39,4; 39,7; 39,8; 40,2; 40,5; 40,6; 40,7; 40,7; 40,8; 41,3; 41,7; 42,3; 44,2;  $n = 20$ .

Складаємо діапазон зміни значень вибірки: 37–45. Різниця в діапазоні складає 8 одиниць, тому розіб'ємо діапазон на 4 інтервали ( $h = 2$ ). Отримаємо інтервали: 1) 37–39; 2) 39–41; 3) 41–43; 4) 43–45.

Знаходимо частоту  $n_i$  попадання випадкової величини в кожен з інтервалів:  $n_1 = 4$ ,  $n_2 = 12$ ,  $n_3 = 3$ ,  $n_4 = 1$ .

Знаходимо відносну частоту  $\omega_i$  попадання випадкової величини в кожен з інтервалів:  $\frac{n_1}{n} = 0,2$ ,  $\frac{n_2}{n} = 0,6$ ,  $\frac{n_3}{n} = 0,15$ ,  $\frac{n_4}{n} = 0,05$ .

Запишемо отримані дані в таблицю:

$h = 2$	37–39	39–41	41–43	43–45
$n_i$	4	12	3	1
$\omega_i$	0,20	0,60	0,15	0,05

## 2. Гістограма частот та відносних частот.

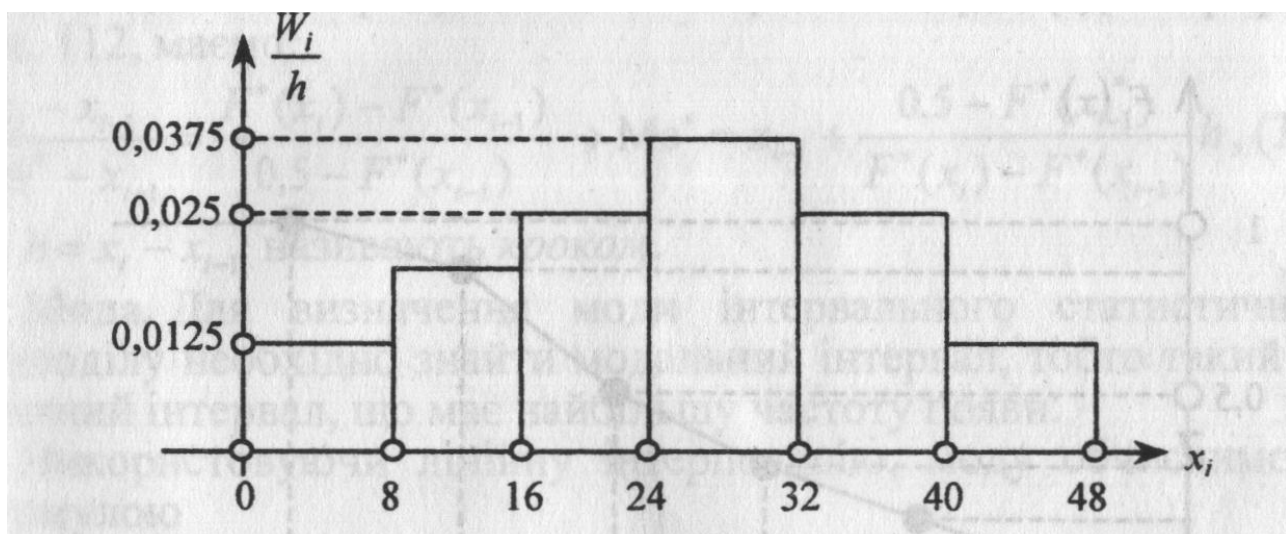
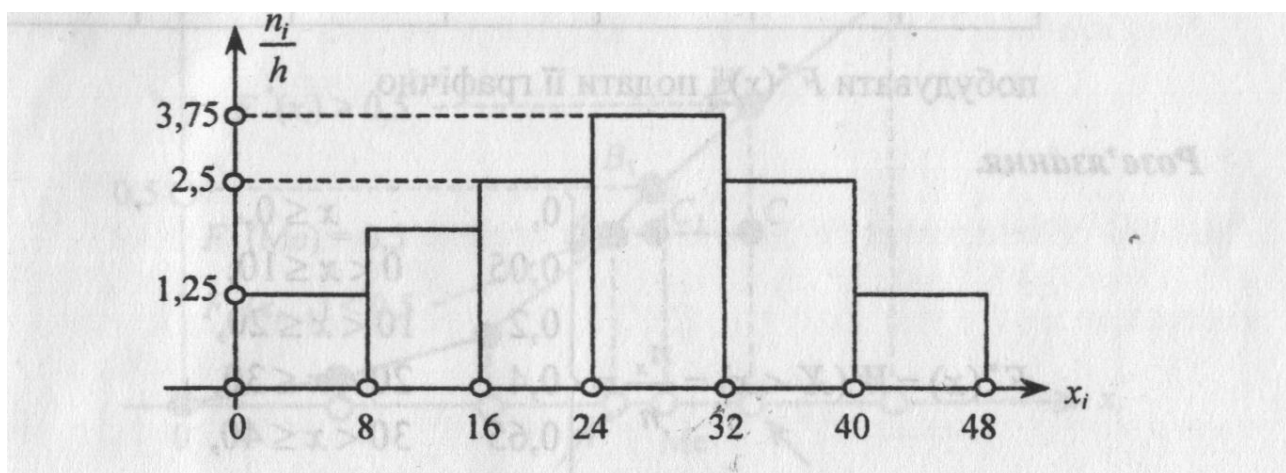
**Озн. 1:** Графіки залежності частоти та відносної частоти від значень інтервалу називається гістограмою. Гістограма частот являє собою фігуру, яка складається з прямокутників, кожен з яких має основу  $h$  і висоту  $\frac{n_i}{h}$ .

**Озн. 2:** Гістограма відносних частот є фігурою, яка складається з прямокутників, кожен з яких має основу  $h$  і висоту  $\frac{\omega_i}{h}$ .

**Приклад 2.** За даним інтервальним статистичним розподілом вибірки побудувати гістограму частот і відносних частот.

$h = 8$	0–8	8–16	16–24	24–32	32–40	40–48
$n_i$	10	15	20	25	20	10
$\omega_i$	0,1	0,15	0,2	0,25	0,2	0,1

**Розв'язання:** Побудуємо гістограми частот та відносних частот:



### 3. Емпірична функція розподілу (кумулята).

**Озн. 3:** Емпірична функція розподілу –це нагромаджу вальна для кожного інтервалу функція:  $F^*(x) = \omega(X \leq x)$  (1)

При побудові кумуляти для інтервального статистичного розподілу вибірки за основу береться припущення, що ознака на кожному частинному інтервалі має рівномірну щільність ймовірностей. Тому кумулята матиме вигляд ламаної лінії, яка зростає на кожному інтервалі і наближається до одиниці.

Приклад 3. За даним інтервальним статистичним розподілом вибірки побудувати емпіричну функцію розподілу.

$h = 10$	0–10	10–20	20–30	30–40	40–50	50–60
$n_i$	5	15	20	25	30	5

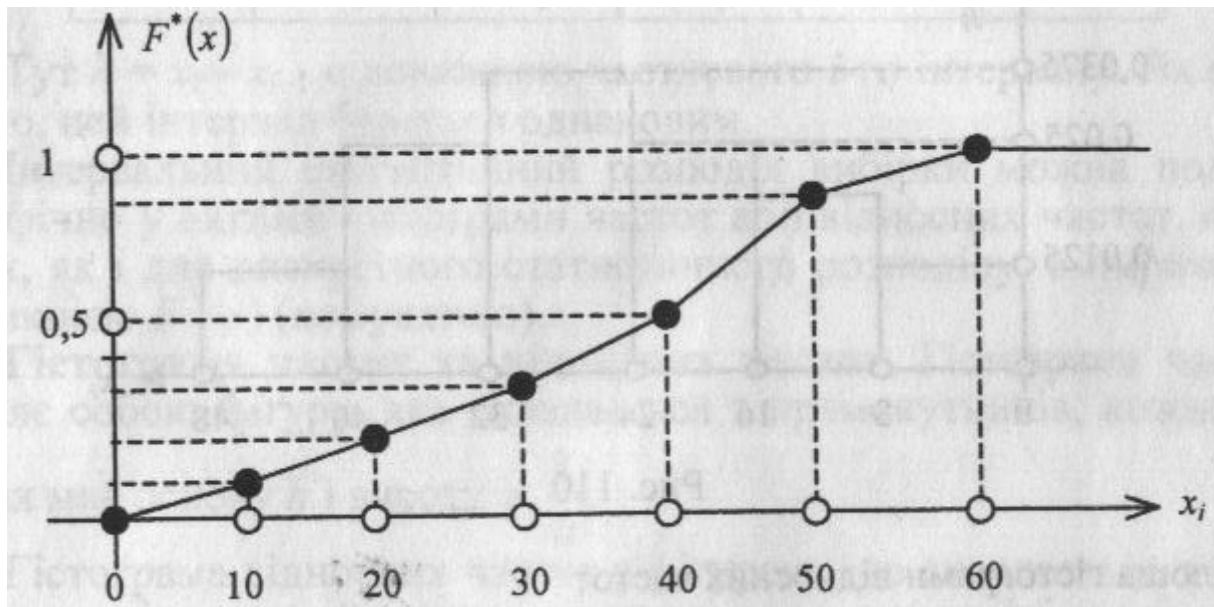
Розв'язання: Доповнимо заданий розподіл двома рядками  $\omega_i$  та  $F^*(x)$ .

$h = 10$	0–10	10–20	20–30	30–40	40–50	50–60
$n_i$	5	15	20	25	30	5
$\omega_i$	0,05	0,15	0,20	0,25	0,30	0,05
$F^*(x)$	0,05	0,20	0,40	0,65	0,95	1

Тоді кумуляту можна подати у вигляді:

$$F^*(x) = \omega(X \leq x) = \begin{cases} 0, & x \leq 0 \\ 0,05, & 0 \leq x \leq 10 \\ 0,2, & 10 \leq x \leq 20 \\ 0,4, & 20 \leq x \leq 30 \\ 0,65, & 30 \leq x \leq 40 \\ 0,95, & 40 \leq x \leq 50 \\ 1, & 50 \leq x \leq 60 \end{cases} .$$

Зобразимо графік емпіричної функції розподілу:



#### 4. Числові характеристики для інтервального розподілу.

Основними характеристиками інтервального розподілу є середнє значення, дисперсія та середнє квадратичне відхилення, які обчислюються за наступними формулами:

$$\bar{x}_e = \frac{\sum x_i^* n_i}{h} \quad (2)$$

$$\bar{D}_e = \frac{\sum x_i^{*2} n_i}{h} - \bar{x}_e^2 \quad (3)$$

$$\sigma_e = \sqrt{\bar{D}_e} \quad (4)$$

Приклад 4: За даним інтервальним статистичним розподілом вибірки, в якому наведено розподіл маси новонароджених, обчислити числові характеристики.

$X=x_i$	1-1,2	1,2-1,4	1,4-1,6	1,6-1,8	1,8-2	2-2,2	2,2-2,4	2,4-2,6	2,6-2,8	2,8-3	3-3,2
$n_i$	5	12	18	22	36	24	19	15	11	9	2

Розв'язання: Побудуємо дискретний статистичний розподіл за заданим інтервальним:

$X=x_i$	1	1,2	1,4	1,6	1,8	2	2,2	2,4	2,6	2,8	3
$n_i$	5	12	18	22	36	24	19	15	11	9	2

Згідно формул (2), (3), (4) маємо:

$$\bar{x}_g = \frac{1}{173} (5,5 + 15,6 + 27 + 37,4 + 68,4 + 50,4 + 43,7 + 37,5 + 29,7 + 26,1 + 6,2) \approx 2,01;$$

Для обчислення дисперсії обчислимо спочатку суму  $\frac{\sum x_i^{*2} n_i}{h} = 4,251908$ .

$$\text{Тоді, } \overline{D}_g = \frac{\sum x_i^{*2} n_i}{h} - \bar{x}_g^2 = 4,251908 - (2,01)^2 = 0,217149$$

$$\sigma_g = \sqrt{\overline{D}_g} = \sqrt{0,217149} \approx 0,466.$$