

## Лекція 21. Інтервальний статистичний розподіл вибірки та його числові характеристики

### План

1. Інтервальний статистичний розподіл вибірки
2. Гістограма частот та відносних частот.
3. Емпірична функція  $F^*(x)$
4. Медіана та мода вибірки
5.  $\bar{x}_B, D_B, \sigma_B$  для інтервального статистичного розподілу вибірки

### 1. Інтервальний статистичний розподіл вибірки

Перелік часткових інтервалів і відповідних їм частот, або відносних частот, називають *інтервальним статистичним розподілом вибірки*.

У табличній формі цей розподіл має такий вигляд:

$h$	$x_1 - x_2$	$x_2 - x_3$	$x_3 - x_4$	...	$x_{k-1} - x_k$
$n_i$	$n_1$	$n_2$	$n_3$	...	$N_k$
$W_i$	$W_1$	$W_2$	$W_3$	...	$W_k$

Тут  $h = x_i - x_{i-1}$  є довжиною часткового  $i$ -го інтервалу. Як правило, цей інтервал береться однаковим.

Інтервальний статистичний розподіл вибірки можна подати графічно у вигляді гістограми частот або відносних частот, а також, як і для дискретного статистичного розподілу, емпіричною функцією  $F^*(x)$  (комулятою).

### 2. Гістограма частот та відносних частот.

Гістограма частот являє собою фігуру, яка складається з прямокутників, кожний з яких має основу  $h$  і висоту  $n_i \frac{1}{h}$ .

Гістограма відносних частот є фігурою, що складається з прямокутників, кожний з яких має основу завдовжки  $h$  і висоту, що дорівнює  $W_i \frac{1}{h}$ .

**Приклад 1.** За заданим інтервальним статистичним розподілом вибірки

$h = 8$	0-8	8-16	16-24	24-32	32-40	40-48
$n_i$	10	15	20	25	20	10
$W_i$	0,1	0,15	0,2	0,25	0,2	0,1

потрібно побудувати гістограму частот і відносних частот.

**Розв'язання.** Гістограми частот і відносних частот наведені на рис. 17, 18.

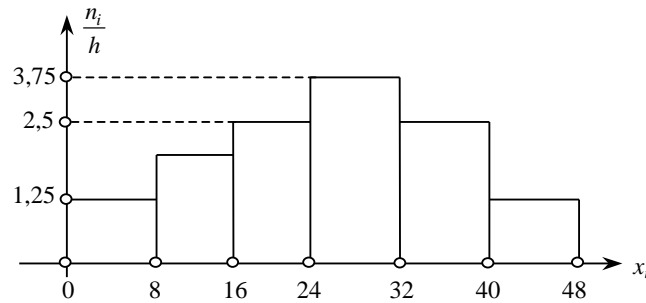


Рис. 17

Площа гістограми частот  $S = \sum h \frac{n_i}{h} = \sum n_i = n = 100$ .

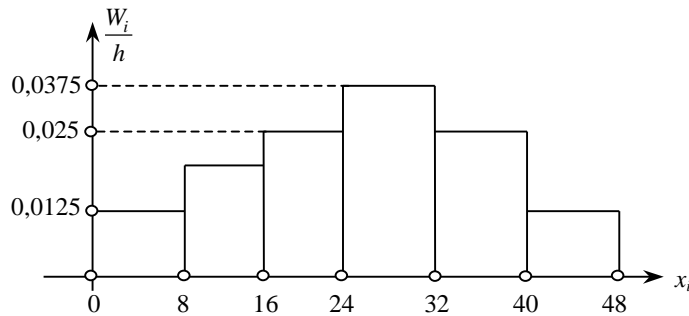


Рис. 18

Площа гістограми відносних частот

$$S = \sum h \frac{W_i}{h} = \sum W_i = 1.$$

**3. Емпірична функція  $F^*(x)$  (комулята).** При побудові комуляти  $F^*(x)$  для інтервального статистичного розподілу вибірки за основу береться припущення, що ознака на кожному частинному інтервалі має рівномірну щільність імовірностей. Тому комулята матиме вигляд ламаної лінії, яка зростає на кожному частковому інтервалі і наближається до одиниці.

Аналогом емпіричної функції  $F^*(x)$  у теорії ймовірностей є інтегральна функція  $F(x) = P(X < x)$ .

#### 4. Медіана та мода вибірки

**Медіана.** Для визначення медіани інтервального статистичного розподілу вибірки необхідно визначити медіанний частковий інтервал. Якщо, наприклад, на  $i$ -му інтервалі  $[x_{i-1} - x_i]$   $F^*(x_{i-1}) < 0,5$  і  $F^*(x_i) > 0,5$ , то, беручи до уваги, що досліджувана ознака  $X$  є неперервною і при цьому  $F^*(x)$  є неспадною функцією, всередині інтервалу  $[x_{i-1} - x_i]$  неодмінно існує таке значення  $X = Me$ , де  $F^*(Me) = 0,5$ .

**Мода.** Для визначення моди інтервального статистичного розподілу необхідно знайти модальний інтервал, тобто такий частинний інтервал, що має найбільшу частоту появи.

Використовуючи лінійну інтерполяцію, моду обчислимо за формулою

$$Mo^* = x_{i-1} + \frac{n_{Mo} - n_{Mo-1}}{2n_{Mo} - n_{Mo-1} - n_{Mo+1}} h \quad (21.1)$$

де  $x_{i-1}$  — початок модального інтервалу;

- $h$  — довжина, або крок, часткового інтервалу;
- $n_{Mo}$  — частота модального інтервалу;
- $n_{Mo-1}$  — частота домодального інтервалу;
- $n_{Mo+1}$  — частота післямодального інтервалу.

**Приклад 2.** За заданим інтервальним статистичним розподілом вибірки

$h = 4$	0-4	4-8	8-12	12-16	16-20	20-24
$n_i$	6	14	20	25	30	5

побудувати гістограму частот і  $F^*(x)$ . Визначити  $Mo^*$ ,  $Me^*$ .

**Розв'язання.** Гістограма частот зображена на рис. 19.

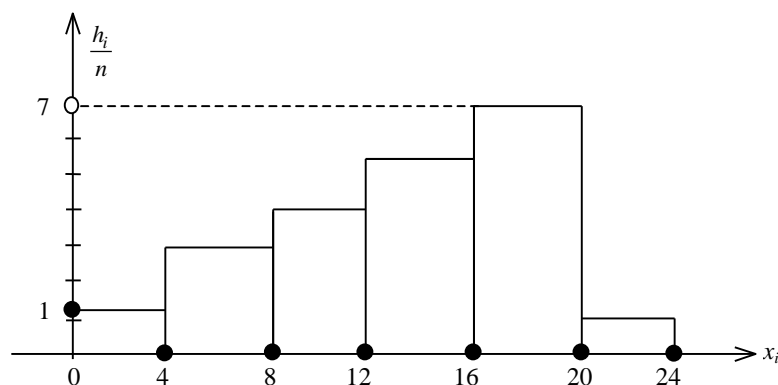


Рис. 19

Графік  $F^*(x)$  зображено на рис. 20.

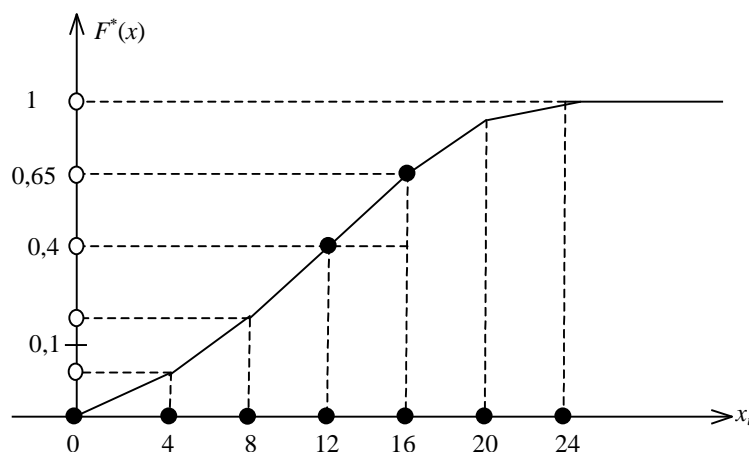


Рис. 20

З рис. 19 визначається модальний інтервал, який дорівнює 16-20.

Застосовуючи (21.1) і беручи до уваги, що  $n_{Mo} = 30$ ,  $n_{Mo-1} = 25$ ,  $n_{Mo+1} = 5$ ,  $h = 4$ ,  $x_{i-1} = 16$ , дістанемо

$$Mo^* = x_{i-1} + \frac{n_{Mo} - n_{Mo-1}}{2n_{Mo} - n_{Mo-1} - n_{Mo+1}} h;$$

$$Mo^* = 16 + \frac{30 - 25}{60 - 25 - 5} \cdot 4 = 16 + \frac{5}{30} \cdot 4 = 16,17.$$

Отже,  $Mo^* = 16,17$ .

З графіка  $F^*(x)$  визначається медіанний інтервал, який дорівнює 12-16. Беручи до уваги, що  $F(12) = 0,4$ ,  $F(16) = 0,65$ ,  $h = 4$  дістанемо:

$$Me^* = x_{i-1} + \frac{0,5 - F^*(x_{i-1})}{F^*(x_i) - F^*(x_{i-1})} h = 12 + \frac{0,5 - 0,4}{0,65 - 0,4} \cdot 4 = 12 + \frac{0,1}{0,25} \cdot 4 = 13,6.$$

Отже,  $Me^* = 13,6$ .

### 5. $\bar{x}_B, D_B, \sigma_B$ для інтервального статистичного розподілу вибірки

Для визначення  $\bar{x}_B, D_B, \sigma_B$  перейдемо від інтервального розподілу до дискретного, варіантами якого є середина часткових інтервалів

$x_i^* = x_{i-1} + \frac{h}{2} = x_i - \frac{h}{2}$  і який має такий вигляд:

$x_i^* = x_i - \frac{h}{2} = x_{i-1} + \frac{h}{2}$	$x_1^*$	$x_2^*$	$x_3^*$	...	$x_k^*$
$h_i$	$h_1$	$h_2$	$h_3$	...	$h_k$

Тоді  $\bar{x}_B, D_B, \sigma_B$  обчислюються за формулами:

$$\bar{x}_B = \frac{\sum x_i^* n_i}{h} \quad (21.2)$$

$$D_B = \frac{\sum (x_i^*)^2 n_i}{h} - (\bar{x}_B)^2; \quad (21.3)$$

$$\sigma_B = \sqrt{D_B} \quad (21.4)$$

**Приклад.** За заданим інтервальним статистичним розподілом вибірки, в якому наведено розподіл маси новонароджених  $x_i$ ,

$X = x_i, \text{ кг}$	1—1,2	1,2—1,4	1,4—1,6	1,6—1,8	1,8—2	1,8—2	2—2,2	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

обчислити  $\bar{x}_B, D_B, \sigma_B$ .

**Розв'язання.** Побудуємо дискретний статистичний розподіл за заданим інтервальним. Оскільки  $h = 0,2$ , то дістанемо:

$x_i^* = x_i - \frac{h}{2} = x_{i-1} + \frac{h}{2}$	1,1	1,3	1,5	1,7	1,9	2,1	2,3	2,5	2,7	2,9	3,1
---	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----

$h_i$	5	12	18	22	36	24	19	15	11	9	2
-------	---	----	----	----	----	----	----	----	----	---	---

Беручи до уваги те, що  $n = 173$ , дістанемо:

$$\bar{x}_B = \frac{\sum x_1^* n_i}{n} = \frac{5,5 + 15,6 + 27 + 37,4 + 68,4 + 50,4 + 43,7}{173} + \frac{37,5 + 29,7 + 26,1 + 6,2}{173} = \frac{347,5}{173} \approx 2,008671 \text{ кг.}$$

Отже,  $\bar{x}_B = 2,008671 \text{ кг.}$

$$\frac{\sum (x_1^*)^2 n_i}{n} = \frac{6,05 + 20,29 + 40,5 + 63,58 + 129,96 + 105,84 + 100,51}{173} + \frac{93,75 + 80,19 + 75,69 + 19,22}{173} = \frac{735,58}{173} = 4,251908.$$

$$D_B = \frac{\sum (x_1^*)^2 n_i}{n} - (\bar{x}_B)^2 = 4,251908 - (2,008671)^2 = 4,251908 - 4,034759 = 0,217149.$$

$$D_B = 0,217149.$$

$$\sigma_B = \sqrt{D_B} = \sqrt{0,217149} \approx 0,466.$$

Отже,  $\sigma_B = 0,466 \text{ кг.}$