Основні поняття математичної статистики (МС)

1. Вступ
2. Генеральна і вибіркова сукупності
3. Статистичний розподіл вибірки
4. Полігон та гістограма частот
5. Емпірична функція розподілу
6. Числові характеристики статистичного розподілу вибірки

СПИСОК РЕКОМЕНДОВАНОЇ ЛІТЕРАТУРИ

Руденко В.М. Математична статистика. - К.: Центр учбової літератури, 2017. - 304 с.

Гладунський В.Н. Математика для економістів: означення, формули, приклади. Навч. посібник. - Львів, 2013, 632 с.

Лозовий Б.Н., Пушак Я.С. Теорія ймовірностей і елементи математичної статистики. Навч. посібник. – К.: Ліра-К, 2018. – 276с.

Копич І.М., Сороківський В.М., Теорія ймовірностей та математична статистика. Навч. посібник. – К.: Ліра-К, 2018 – 382с.

1. Вступ

Математична статистика (МС) – розділ математики, який вивчає закономірності масових явищ, займається розробкою математичних методів аналізу таких явищ на підставі статистичних даних (результатів спостережень) для подальшого прогнозування поведінки масових випадкових явищ

Теоретичною базою для МС є ТЙ

2. Генеральна і вибіркова сукупності (вихідні поняття МС)

Генеральна сукупність (ГС) – множина усіх реально існуючих однорідних елементів, які вивчають з точки зору їх розподілу за певною ознакою. Елемент множини при цьому називають елементом сукупності

N – кількість усіх елементів ГС (обсяг ГС)

Наприклад:

1. Множина людей України за віком
(кількісна ознака - вік)

2. Множина акціонерів банків Укр. за прибутком
(кількісна ознака - прибуток)

Множина людей України за віком (кількісна ознака – вік – це випадкова величина)

Кількісна ознака ГС – це випадкова величина (ВВ Х)

Можливі значення ВВ Х { $x_1 \; x_2 \; ... \; x_i \; ... \; x_n$ } – це можливі значення ознаки, розподіл якої ми вивчаємо

Повна характеристика ВВ (а також і кількісної ознаки ГС) – це її закон розподілу (його наз. теоретичним ЗР)

Тобто, щоб отримати повну характеристику певної кількісної ознаки ГС, треба знати значення цієї ознаки для усіх без винятку елементів ГС – таке можливо тільки в результаті суцільного обстеження всієї ГС (в окремих ситуаціях так і роблять)

Наприклад **Перепис населення країни**

Вибіркова сукупність (вибірка) – будь-яка підмножина випадково обраних елементів із генеральної сукупності

Якщо { $x_1 \quad x_2 \ \dots \ x_i \ \dots \ x_n$ } – вибірка з ГС, то числа $x_1 \quad x_2 \ \dots \ x_i \ \dots \ x_n$ називають варіантами вибірки (варіанта вибірки)

n - кількість усіх елементів вибірки (обсяг вибірки), його значення завжди відоме (**n<<N**)

Правила формування вибірки

1. **Повторна (елемент ГС після дослідження повертається в ГС)**
2. **Безповторна (елемент ГС після дослідження не повертається в ГС) (На практиці, як правило, користуються безповторними вибірками)**
3. **Вибірка повинна бути репрезентативною (представницькою), тобто вибірка повинна правильно представляти ГС – це забезпечує випадковий відбір елементів з ГС до вибірки**

Основна задача МС – вивчення властивостей ГС за даними вибірки.

Оскільки кількісну ознаку ГС можна інтерпретувати як випадкову величину, то для вирішення цієї задачі будемо використовувати ймовірнісні методи

3.Статистичний розподіл вибірки -

основна характеристика вибірки (аналог закону розподілу випадкової величини в теорії ймовірностей)

1. Дискретний
2. Інтервальний (неперервний)

1. Дискретний статистичний розподіл вибірки

Нехай маємо вибірку з генеральної сукупності

$$x_1 \;\; x_2 \; ... \; x_i \; ... \; x_n$$

Якщо записати цю вибірку у вигляді зростаючої послідовності, то отримаємо *варіаційний ряд*.

Деякі значення варіант можуть повторюватись, тоді варіаційний ряд записують у вигляді таблиці, де вказують значення варіант(x_i) та їх частот (m_i). (*Частотою* варіанти називають число появи цієї варіанти у вибірці).

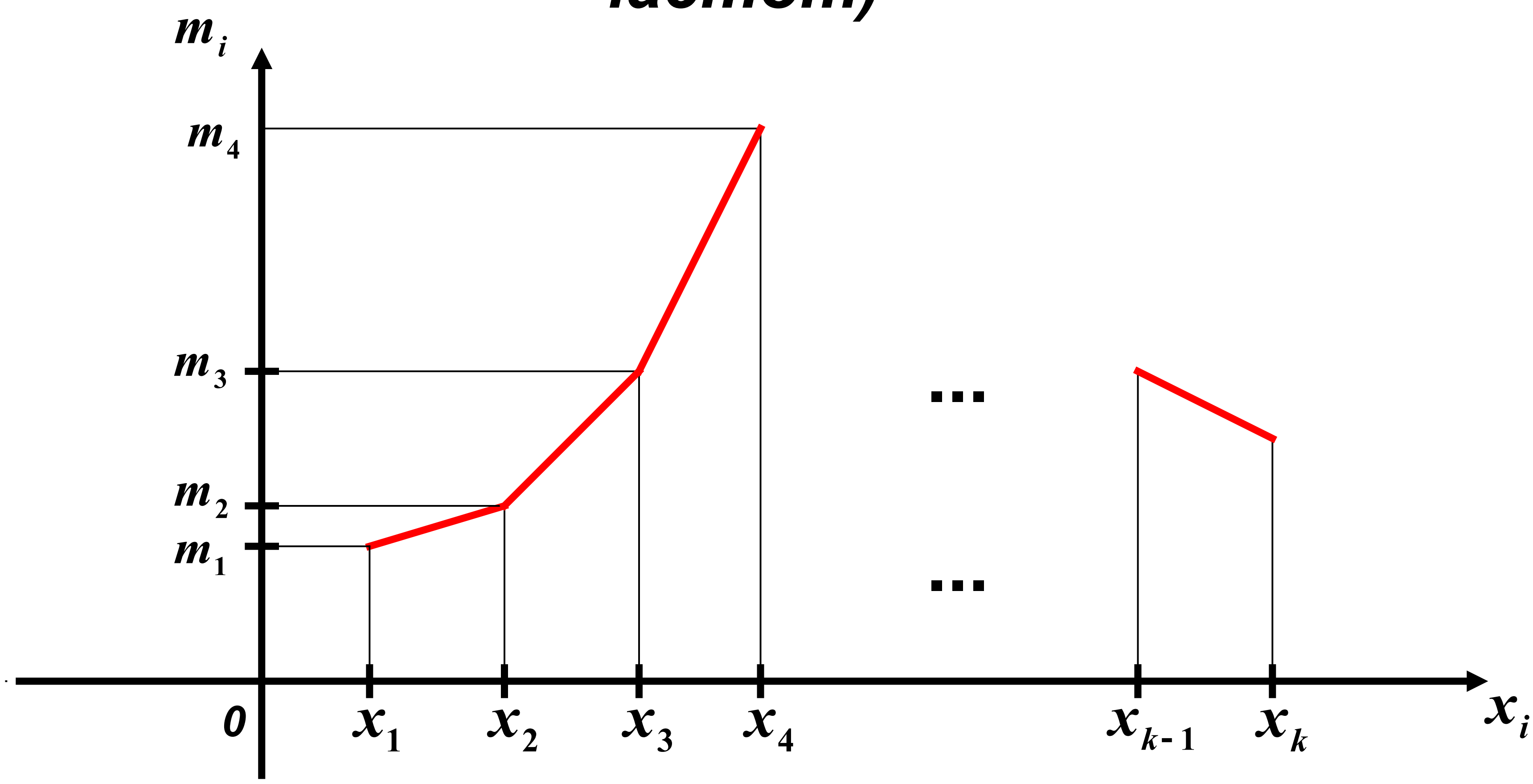
Відповідність між варіантами та їх частотами називають дискретним статистичним розподілом частот вибірки

Дискретний статистичний розподіл частот вибірки

x_i	x_1	x_2	...	x_k
m_i	m_1	m_2	...	m_k

$$\sum_{i=1}^{k} m_i = n$$

Полігон частот (графічне зображення дискретного статистичного ряду частот)



У дискретному статистичному розподілі вибірки замість частот можна вказувати відносні частоти.

Відносною частотою варіанти називають відношення частоти появи цієї варіанти до загального об'єму вибірки:

$$w_i = \frac{m_i}{n}$$

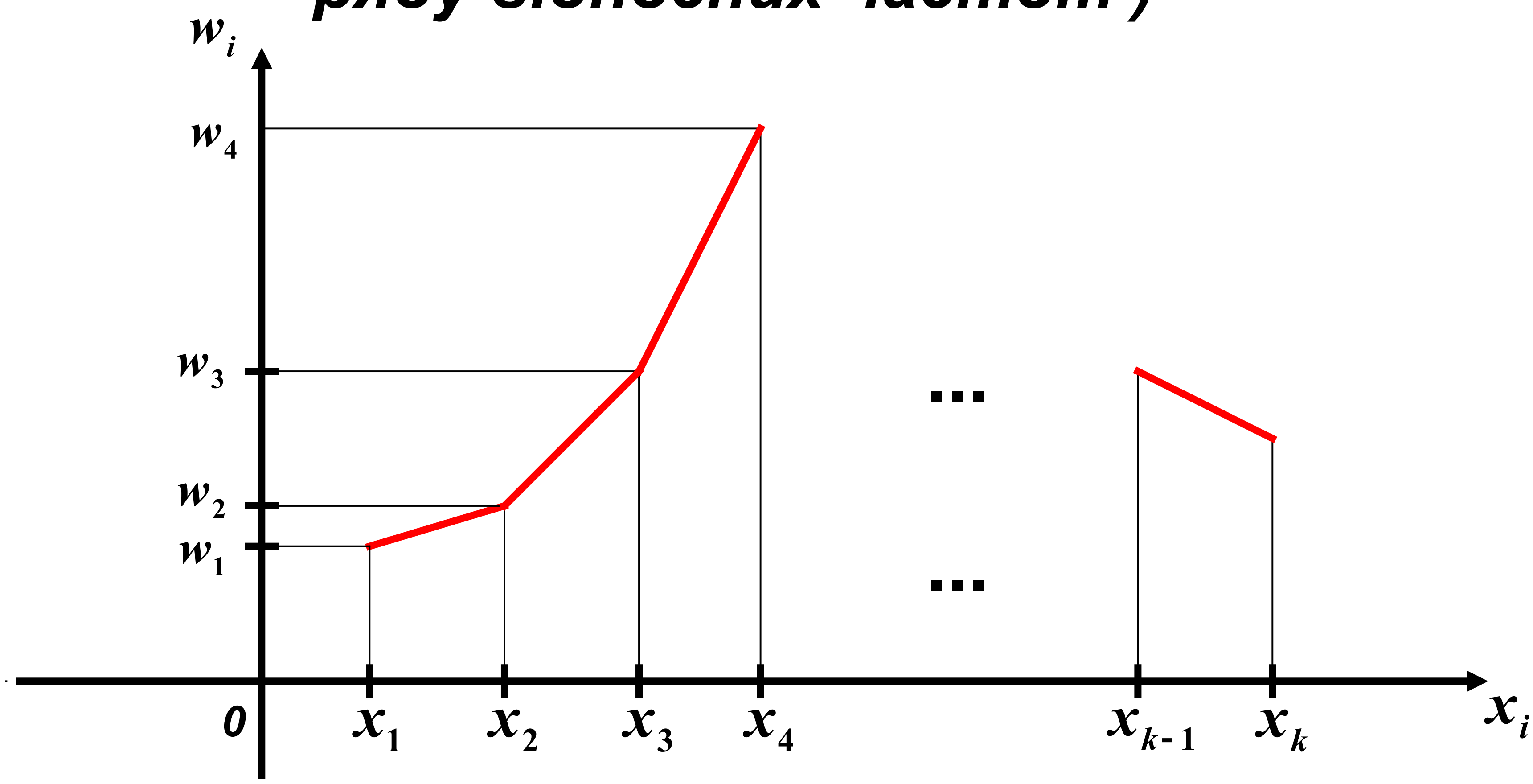
Тоді ми отримаємо дискретний статистичний розподіл відносних частот вибірки

Дискретний статистичний розподіл відносних частот вибірки

x_i	x_1	x_2	...	x_k
w_i	w_1	w_2	...	w_k

$$\sum_{i=1}^{k} w_i = 1$$

Полігон відносних частот (графічне зображення дискретного статистичного ряду відносних частот)



2. Інтервальний статистичний розподіл вибірки

При великому об’ємі вибірки ($n \geq 30$) користуватися дискретним статистичним розподілом незручно. У цьому випадку розмах вибірки поділяють на k рівних інтервалів довжиною h. Потім для кожного інтервала підраховують частоту m_i' або відносну частоту w_i'

Частотою m_i' називають суму частот варіант, які потрапили в i-й інтервал

Відносною частотою w_i' називають суму відносних частот варіант, які потрапили в i-й інтервал

Відповідність між інтервалами та їх частотами (або відносними частотами) називають інтервальним статистичним розподілом вибірки

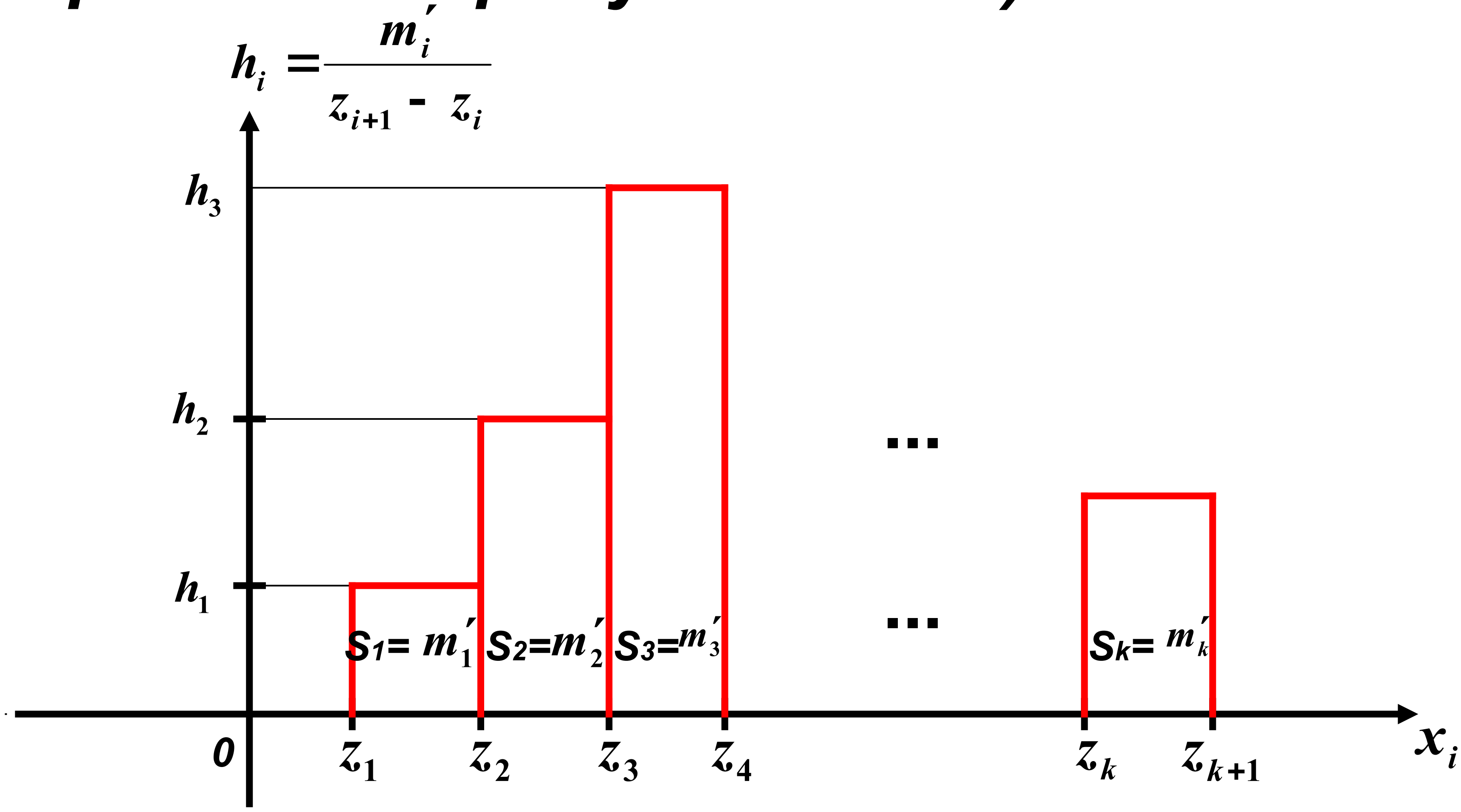
Ці розподіли також представляють у вигляді таблиць

Інтервальний статистичний розподіл частот

$(z_i, z_{i+1}]$	$(z_1, z_2]$	$(z_2, z_3]$	...	$(z_k, z_{k+1}]$
m_i'	m_1'	m_2'	...	m_k'

$$\sum_{i=1}^{k} m_i' = n$$

Гістограма частот (графічне зображення інтервального ряду частот)



$$S_Г = S_1 + S_2 + ... + S_k = \sum m_i' = n \ (\textit{обсяг вибірки})$$

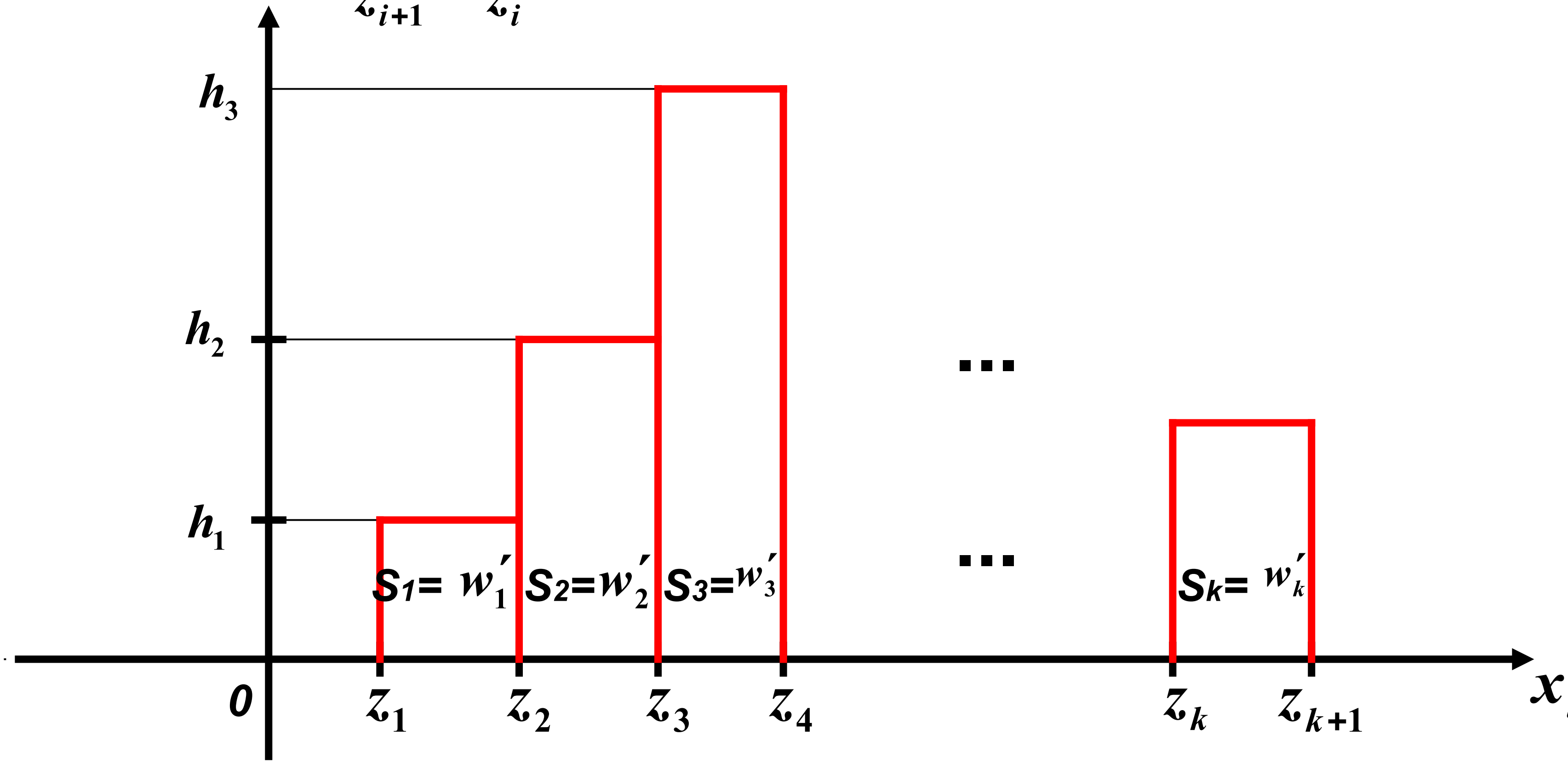
Інтервальний статистичний розподіл відносних частот

$(z_i, z_{i+1}]$	$(z_1, z_2]$	$(z_2, z_3]$	...	$(z_k, z_{k+1}]$
w_i'	w_1'	w_2'	...	w_k'

$$\sum_{i=1}^{k} w_i' = 1$$

Гістограма відносних частот

$$h_i = \frac{w_i'}{z_{i+1} - z_i}$$



$$S_{\Gamma} = S_1 + S_2 + \ldots + S_k = \sum w_i' = 1$$

Зауваження

$$\left(z_i\ , z_{i+1}\right]$$

Якщо значення варіанти потрапить на межу інтервалу, то його включають до попереднього інтервалу

4.Полігон та гістограма частот –графічне представлення вибірки

1. ***Полігон частот*** ***(графічне зображення дискретного статистичного розподілу частот вибірки)***
2. ***Полігон відносних частот*** ***(графічне зображення дискретного статистичного розподілу відносних частот вибірки)***
3. ***Гістограма частот*** ***(графічне зображення інтервального статистичного розподілу частот вибірки)***
4. ***Гістограма відносних частот*** ***(графічне зображення інтервального статистичного розподілу відносних частот вибірки)***

5. Емпірична функція розподілу - функція розподілу вибірки (будується дослідним шляхом за даними вибірки)

Емпіричною функцією розподілу називають функцію

$$F^*(x) = \sum_{x_i < x} W_i$$

Емпірична функція розподілу – це функція накопичених відносних частот, тобто

$$F^*(x) = W(X < x)$$

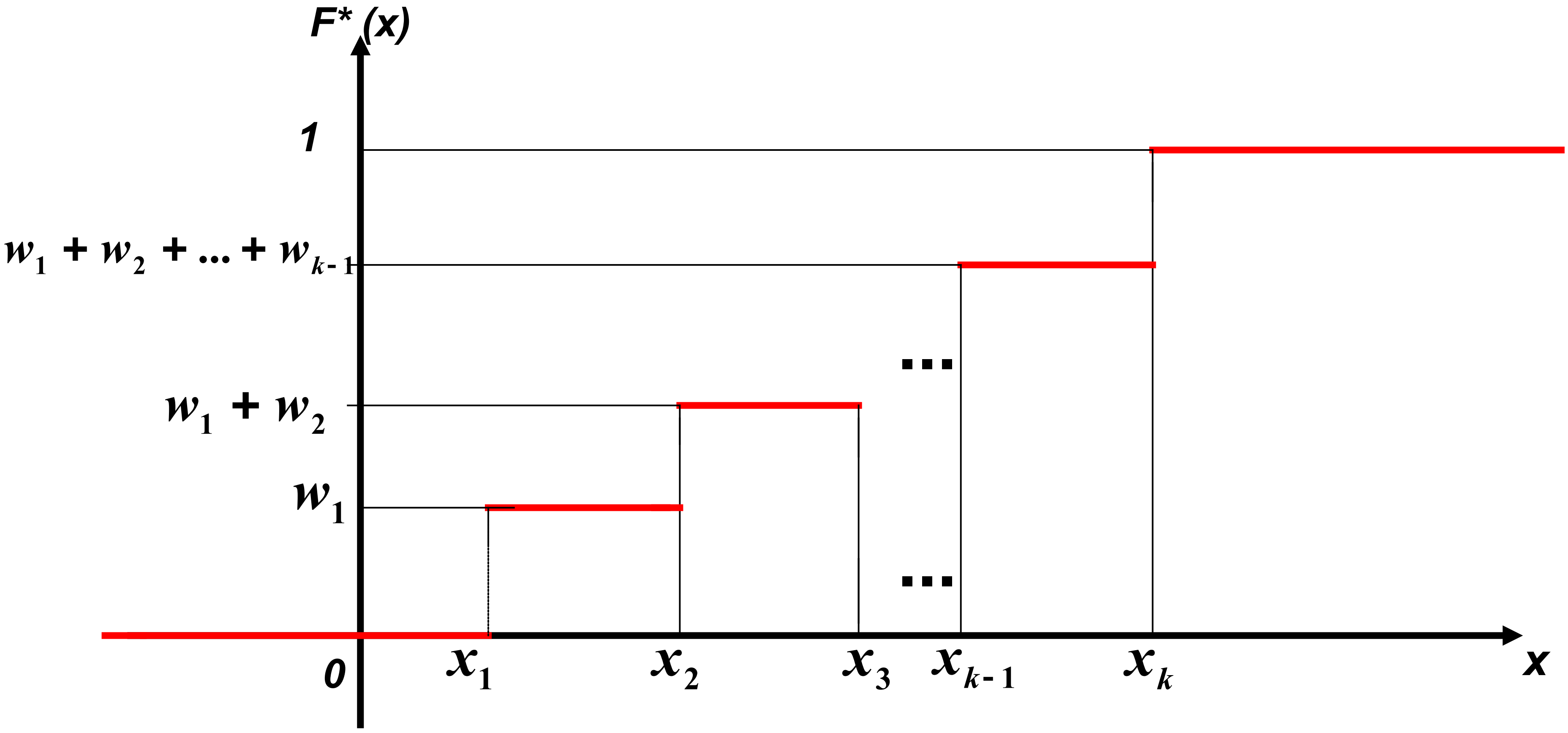
Дискретний статистичний розподіл відносних частот вибірки

x_i	x_1	x_2	$\cdots$	x_k
w_i	w_1	w_2	$\cdots$	w_k
Накопичені відносні частоти	0	w_1	**...**	w_1+w_2+ $+w_3+\ldots+w_{k-1}$

Емпірична ФР для дискретного статистичного розподілу відносних частот вибірки

$$F^*(x) = \begin{cases} 0, & x \le x_1 \\ w_1, & x_1 < x \le x_2 \\ w_1 + w_2, & x_2 < x \le x_3 \\ \dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots & \\ w_1 + w_2 + \dots + w_{k-1}, & x_{k-1} < x \le x_k \\ 1, & x > x_k \end{cases}$$

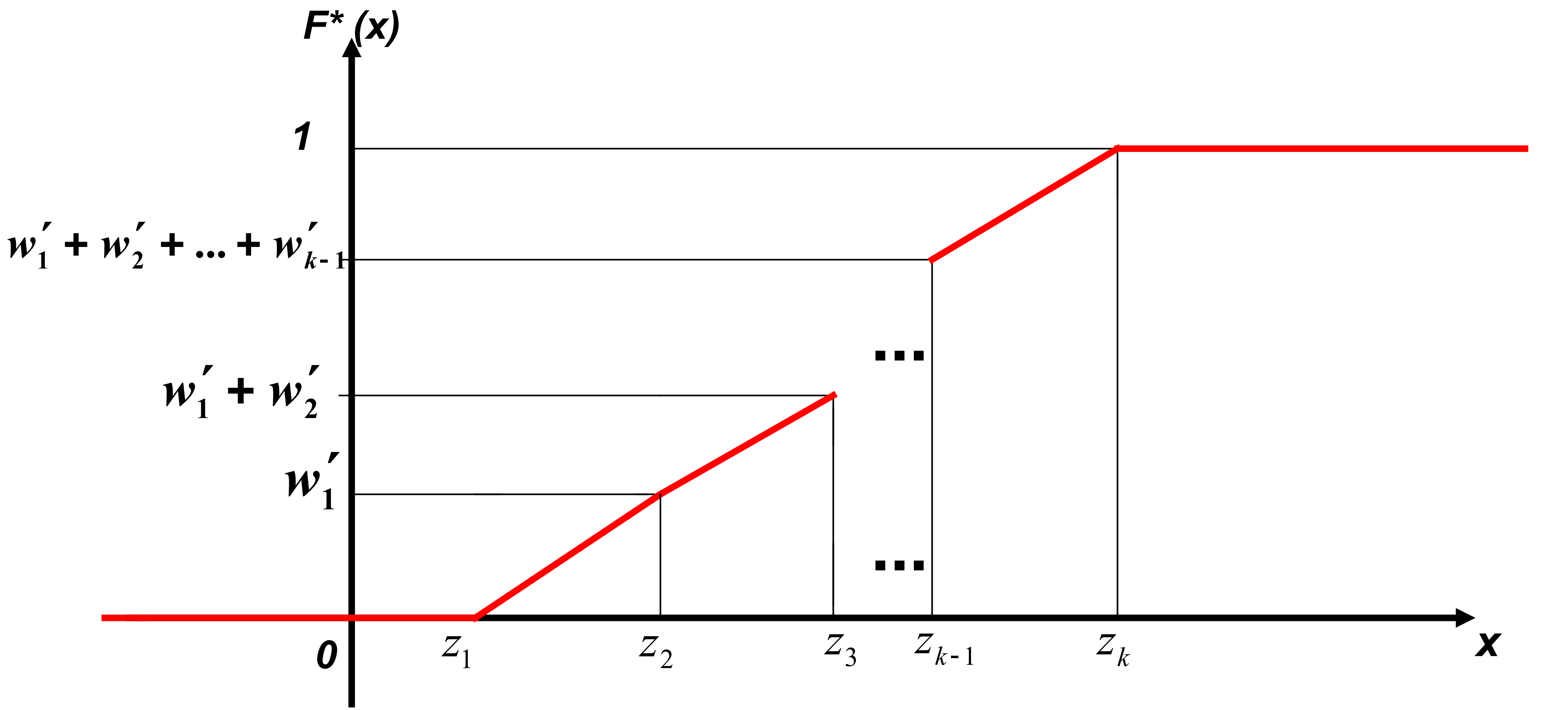
Графік емпіричної ФР для дискретного розподілу



Емпірична ФР для інтервального статистичного розподілу відносних частот вибірки

$$F^*(x)=\begin{cases}0, & x \le z_1\\ w_1', & z_1 < x \le z_2\\ w_1'+w_2', & z_2 < x \le z_3\\ \dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots & \\ w_1'+w_2'+\ldots+w_{k-1}', & z_{k-1} < x \le z_k\\ 1, & x > z_k\end{cases}$$

Графік емпіричної ФР для інтервального розподілу



Емпірична функція розподілу - функція розподілу вибірки, яка будується дослідним шляхом за даними вибірки, використовується для оцінки теоретичної ФР генеральної сукупності.

Коли обсяг вибірки (*n*) стає досить великим, то *F*(x) та F(x) практично не відрізняються*

6. Числові характеристики статистичного розподілу вибірки

Числові характеристики дискретного статистичного розподілу вибірки

1. Вибіркове середне $\bar{x} = \dfrac{\sum\limits_{i=1}^{k} x_i m_i}{n}$, **де** $n = \sum\limits_{i=1}^{k} m_i'$

2. Мода (Mo)– значення варіанти вибірки, якому відповідає найбільше значення частоти (мод може бути декілька)

3. Медіана (Me) - значення варіанти вибірки, по обидві сторони якої міститься не більше ніж половина варіант (враховуючи їх частоту)

4.Вибіркова дисперсія:

$$\overline{D} = \frac{1}{n}\sum_{i=1}^{k}(x_i - \overline{x})^2 m_i = \frac{1}{n}\sum_{i=1}^{k} x_i^2 m_i - (\overline{x})^2$$

5.Вибіркове середнє квадратичне відхилення:

$$\overline{\sigma} = \sqrt{\overline{D}}.$$

6.Розмах вибірки

$$\omega = x_{max} - x_{min}$$

7. Вибірковий коефіцієнт варіації:

$$V = \frac{\overline{\sigma}}{\overline{x}}$$

Числові характеристики інтервального статистичного розподілу вибірки

1. **Вибіркове середне** $\overline{x} = \frac{\sum_{i=1}^{k} x_i' m_i'}{n}$, **де** $n = \sum_{i=1}^{k} m_i'$

а x_i' - *середина i - го інтервалу*

2. **Для визначення моди (Мо) спочатку знаходять модальний інтервал (інтервал з найбільшою частотою), а потім обчислюють моду**

3. **Для визначення медіани (Ме) спочатку знаходять медіанний інтервал (інтервал, по обидві сторони якого міститься не більше ніж половина варіант вибірки), а потім наближено обчислюють медіану (за медіану часто обирають середину медіанного інтервалу)**

4.Вибіркова дисперсія:

$$\overline{D} = \frac{1}{n}\sum_{i=1}^{k}\left(x_i' - \overline{x}\right)^2 m_i' = \frac{1}{n}\sum_{i=1}^{k} x_i'^2 m_i' - (\overline{x})^2$$

5.Вибіркове середнє квадратичне відхилення:

$$\overline{\sigma} = \sqrt{\overline{D}}.$$

6.Розмах вибірки

$$\omega = x_{max} - x_{min}$$

7. Вибірковий коефіцієнт варіації:

$$V = \frac{\overline{\sigma}}{\overline{x}}$$