

ЧИСЛОВІ ХАРАКТЕРИСТИКИ РЯДІВ ДАНИХ

Поняття статистичного ряду та потреба у стислому описі

Статистичний ряд - це впорядкована за певною ознакою послідовність результатів спостереження за об'єктами або явищами

Це первинна форма систематизації даних, яка дозволяє перейти від хаотичного набору фактів до структурованої моделі, придатної для подальшої обробки в обчислювальних алгоритмах

Приклад формування статистичного ряду

1. Первинні («сирі») дані. Було проведено 10 випробувань швидкодії алгоритму, і отримано такі результати (у секундах):

12, 15, 12, 18, 15, 20, 15, 12, 18, 15

Це ще не ряд, а лише набір несистематизованих спостережень.

2. Варіаційний ряд (ранжований)

Першим кроком обробки є ранжування - розташування значень у порядку зростання: 12, 12, 12, 15, 15, 15, 15, 18, 18, 20

3. Дискретний статистичний ряд. Для подальшого аналізу дані доцільно представити у вигляді таблиці частот

Час розрахунку, сек	Частота	Відносна частота
12	3	0,3 (30%)
15	4	0,4 (40%)
18	2	0,2 (20%)
20	1	0,1 (10%)
Разом	10	1,0 (100%)

Основні види рядів залежно від характеру даних та мети дослідження:

1. Ряди розподілу. Групування одиниць сукупності за певною ознакою:

- *Атрибутивні* - ряди, побудовані за якісними ознаками, що не мають числового вираження (наприклад, спеціалізація ПЗ: інженерне, статистичне, символічне);
- *Варіаційні* - ряди, де групування відбувається за кількісною ознакою (наприклад, час виконання розрахунку в MathCAD).

Варіаційні ряди, своєю чергою, бувають:

Дискретні - ознака набуває лише окремих цілих значень

Інтервальні - значення ознаки задаються у вигляді певних інтервалів

2. Ряди динаміки (часові ряди)

Послідовність значень показника, що змінюються у часі

Перетворення дискретного ряду в інтервальний

Групування даних за діапазонами. Це необхідно у випадках, коли кількість спостережень велика, а значення ознаки варіюються у широких межах.

Якщо отримано не 10, а 100 результатів замірів часу розрахунку, де мінімальне значення $x_{\min} = 10$ сек, а максимальне $x_{\max} = 30$ сек.

Визначення розмаху (R): $R = x_{\max} - x_{\min} = 30 - 10 = 20$ сек

Вибір кількості інтервалів (k): За формулою Стерджеса

$k = 1 + \log_2(n)$, для $n=100$: $k = 1 + \log_2(100) = 8$ інтервалів.

Розрахунок кроку (ширини) інтервалу: $h = \frac{R}{k} = \frac{20}{8} = 2.5$ сек

Приклад інтервального статистичного ряду

Інтервал часу (сек)	Частота	Середина інтервалу
[10.0;12.5)	12	11.25
[12.5;15.0)	25	13.75
[15.0;17.5)	30	16.25
[17.5;20.0)	18	18.75
...	...	...
Разом	100	

На основі такого ряду будується гістограма, яка дозволяє візуально оцінити закон розподілу (наприклад, нормальний розподіл)

Інтервальні дані дозволяють будувати більш складні ймовірнісні моделі, де оцінюється ймовірність потрапляння параметра в конкретний числовий проміжок

При обчисленні середніх характеристик для такого ряду використовують середини інтервалів (x'_i) , що дозволяє ефективно працювати з великими масивами , наприклад, в MathCAD

Класифікація характеристик за функціональним призначенням

Основні групи показників:

- **Міри центральної тенденції** - визначають «центр» розподілу, навколо якого групуються дані (середнє, медіана, мода)
- **Міри розсіювання (варіації)** - вказують на ступінь неоднорідності даних та їхнє відхилення від центру (дисперсія, стандартне відхилення)
- **Характеристики форми розподілу** - описують симетрію та гостровершинність кривої розподілу (асиметрія, ексцес)

Функції обчислення вибірових характеристик в MathCAD

max(A)	Максимальне значення
min(A)	Мінімальне значення
sort(A)	Побудова варіаційного ряду
var(A)	Оцінка дисперсії
mean(A)	Середнє значення вибірки
stdev(A)	Стандартне відхилення
median(A)	Медіана
mode(A)	мода (найпоширеніше вибірове значення)
cvar(A,B)	Коваріація вибірки
corr(A,B)	Коефіцієнт кореляції

Введення та виведення даних в MathCAD

READPRN — читання даних в матрицю з текстового файлу

WRITEPRN — запис даних в текстовий файл

APPENDPRN — дозапис даних в існуючий текстовий файл

```
x := READPRN("c:\mnv\labx.txt")
```

x =		0
	0	2.235
	1	2.453
	2	3.761
	3	4.123
	4	5.345
	5	6.766
	6	6.9
	7	7.12
	8	7.12
	9	7.765
	10	8.65
	11	8.653

```
x := READPRN("c:\mnv\labx.txt")
```

$x^T =$		0	1	2	3	4	5	6	7	8
	0	2.235	2.453	3.761	4.123	5.345	6.766	6.9	7.12	...

Моделювання виборок зі стандартних розподілів

$d^*(x, par)$ – Щільність ймовірності

$p^*(x, par)$ – Функція розподілу

$q^*(P, par)$ – Квантиль розподілу

$r^*(M, par)$ – вектор M незалежних чисел, розподілених за обраним законом

x – Значення випадкової величини;

P – Значення ймовірності;

par – Список параметрів розподілу

Приклади функцій:

$rnorm(x, \mu, \sigma)$ - нормальний розподіл (μ - очікуване значення, σ - стандартне відхилення)

$qchisq(x, d)$ - χ^2 – розподіл (d - число ступенів свободи)

$rt(x, d)$ - розподіл Стюдента (d - число ступенів свободи)

$rF(x, d1, d2)$ - розподіл Стюдента ($d1, d2$ - числа ступенів свободи)

Побудова гістограмми та полігонів

$N := 100$ $k := 20$ +

$x := \text{morm}(N, 50, 2)$

$$x^T =$$

	0	1	2	3	4	5	6	7
0	49.122	48.641	49.053	48.097	46.629	50.087	49.759	...

$\text{left} := \min(x) = 44.198$

$m := \text{mode}(x)$

$\text{right} := \max(x) = 56.096$

$\text{mediana} := \text{median}(x) = 49.86$

$h := \frac{\text{right} - \text{left}}{k}$

$\text{disp} := \text{var}(x) = 3.819$

$h = 0.595$

$\text{srednee} := \text{mean}(x) = 49.699$

$i := 0..k$

$p_i := \text{left} + h \cdot i$

$f := \frac{\text{hist}(p, x)}{N \cdot h}$

