

# ГРАФОВІ ЙМОВІРНІСНІ МОДЕЛІ СТАТИСТИЧНИЙ АНАЛІЗ ДАНИХ КВАНТИЛІ

Сумський державний університет

# Статистичні ряди

Імовірнісні  
основи  
обробки  
даних

Статистичні  
ряди

Гістограма  
та полігон  
частот

Квантілі

Ящик з  
вусами

Графік  
 $Q - Q$

- **Статистичним рядом** називається сукупність пар  $i \Rightarrow x_i$ , отриманих в результаті експерименту. Зазвичай статистичні ряди представляють у вигляді таблиці, в першому стовпці якої стоїть номер дослідів ( $i$ ), а в другому — спостережуване значення випадкової величини  $x_i$ , яке називається **варіантою**.

# Статистичні ряди

Імовірнісні  
основи  
обробки  
даних

Статистичні  
ряди

Гістограма  
та полігон  
частот

Квантілі

Ящик з  
вусами

Графік  
 $Q - Q$

- **Статистичним рядом** називається сукупність пар  $i \Rightarrow x_i$ , отриманих в результаті експерименту. Зазвичай статистичні ряди представляють у вигляді таблиці, в першому стовпці якої стоїть номер досліду ( $i$ ), а в другому — спостережуване значення випадкової величини  $x_i$ , яке називається **варіантою**.

| Індекс<br>$i$ | Варианта<br>$x_i$ |
|---------------|-------------------|
| 1             | $x_1$             |
| 2             | $x_2$             |
| ...           | ...               |
| $n$           | $x_n$             |



# Статистичні ряди

Імовірнісні  
основи  
обробки  
даних

Статистичні  
ряди

Гістограма  
та полігон  
частот

Квантілі

Ящик з  
вусами

Графік  
 $Q - Q$

**Розмахом вибірки** називають різницю між найбільшою і найменшою варіантами вибірки:

$$R = x_{max} - x_{min}.$$

# Статистичні ряди

Якщо одна і та сама варіанта зустрічається у вибірці кілька разів, то статистичний ряд зручніше записувати у вигляді наступної таблиці

Таблиця 2

| Індекс<br>$i$ | Варіанта<br>$x_i$ | Частота<br>$n_i$ | Относит.<br>частота<br>$\bar{n}_i$ |
|---------------|-------------------|------------------|------------------------------------|
| 1             | $x_1$             | $n_1$            | $\bar{n}_1$                        |
| 2             | $x_2$             | $n_2$            | $\bar{n}_2$                        |
| ...           | ...               | ...              | ...                                |
| $k$           | $x_k$             | $n_k$            | $\bar{n}_k$                        |

# Статистичні ряди

Імовірнісні  
основи  
обробки  
даних

Статистичні  
ряди

Гістограма  
та полігон  
частот

Квантілі

Ящик з  
вусами

Графік  
 $Q - Q$

- Частотою  $n_i (i = \overline{1, k})$  варіанти  $x_i$  називається число повторень варіанти  $x_i$  у вибірці, причому

$$\sum_{i=1}^k n_i = n.$$

# Статистичні ряди

Імовірнісні  
основи  
обробки  
даних

Статистичні  
ряди

Гістограма  
та полігон  
частот

Квантілі

Ящик з  
вусами

Графік  
 $Q - Q$

- Частотою  $n_i (i = \overline{1, k})$  варіанти  $x_i$  називається число повторень варіанти  $x_i$  у вибірці, причому

$$\sum_{i=1}^k n_i = n.$$

- Відносною частотою або вагою  $\bar{n}_i (i = \overline{1, k})$  варіанти  $x_i$  називається відношення частоти варіанти до об'єму вибірки  $n$ , тобто

$$\bar{n}_i = \frac{n_i}{n}$$

$$\sum_{i=1}^k \bar{n}_i = 1.$$

# Статистичні ряди

Імовірнісні  
основи  
обробки  
даних

Статистичні  
ряди

Гістограма  
та полігон  
частот

Квантілі

Ящик з  
вусами

Графік  
 $Q - Q$

При великій кількості спостережень простий статистичний ряд перестає бути зручною формою запису статистичних даних. Для додання йому більшої компактності і наочності статистичний матеріал піддають додатковій обробці — будують варіаційні ряди або груповані варіаційні ряди.



# Статистичні ряди

Імовірнісні  
основи  
обробки  
даних

Статистичні  
ряди

Гістограма  
та полігон  
частот

Квантілі

Ящик з  
вусами

Графік  
 $Q - Q$

- Варіаційним рядом називається упорядкована сукупність варіант  $x_i (i = \overline{1, k})$  з відповідними їм частотами  $n_i$  або відносними частотами  $\bar{n}_i$ .

# Статистичні ряди

Імовірнісні  
основи  
обробки  
даних

Статистичні  
ряди

Гістограма  
та полігон  
частот

Квантілі

Ящик з  
вусами

Графік  
 $Q - Q$

- Варіаційним рядом називається упорядкована сукупність варіант  $x_i (i = \overline{1, k})$  з відповідними їм частотами  $n_i$  або відносними частотами  $\bar{n}_i$ .
- Для побудови групованого варіаційного ряду інтервал зміни спостережуваних значень випадкової величини  $[x_{min}; x_{max}]$  розбивають на  $N$  інтервалів, що не пересікаються (їх називають частковими інтервалами або розрядами).

# Статистичні ряди

Імовірнісні  
основи  
обробки  
даних

Статистичні  
ряди

Гістограма  
та полігон  
частот

Квантілі

Ящик з  
вусами

Графік  
 $Q - Q$

Число інтервалів залежить від об'єму вибірки і визначається за формулою Стерджеса



$$N = 1 + 3.32 \log n$$

# Статистичні ряди

Імовірнісні  
основи  
обробки  
даних

Статистичні  
ряди

Гістограма  
та полігон  
частот

Квантілі

Ящик з  
вусами

Графік  
 $Q - Q$

Число інтервалів залежить від об'єму вибірки і визначається за формулою Стерджеса



$$N = 1 + 3.32 \log n$$



$$N = 1 + 1.44 \ln n$$

# Статистичні ряди

Імовірнісні  
основи  
обробки  
даних

Статистичні  
ряди

Гістограма  
та полігон  
частот

Квантілі

Ящик з  
вусами

Графік  
 $Q - Q$

Число інтервалів залежить від об'єму вибірки і визначається за формулою Стерджеса



$$N = 1 + 3.32 \log n$$



$$N = 1 + 1.44 \ln n$$



$$N \geq [1 + 3.32 \log n] + 1$$

квадратні дужки позначають цілу частину числа.

# Статистичні ряди

Імовірнісні  
основи  
обробки  
даних

Статистичні  
ряди

Гістограма  
та полігон  
частот

Квантілі

Ящик з  
вусами

Графік  
 $Q - Q$

- Розбиття на мале число інтервалів може призвести до невірних статистичними висновків. Відповідно до цієї формули, наприклад, необхідно брати не менше 8 інтервалів на 100 спостережень.

# Статистичні ряди

Імовірнісні  
основи  
обробки  
даних

Статистичні  
ряди

Гістограма  
та полігон  
частот

Квантілі

Ящик з  
вусами

Графік  
Q – Q

- Розбиття на мале число інтервалів може призвести до невірних статистичними висновків. Відповідно до цієї формули, наприклад, необхідно брати не менше 8 інтервалів на 100 спостережень.
- Інтервали повинні бути однакової довжини

$$\Delta = \frac{R}{N} = \frac{x_{max} - x_{min}}{N}$$

# Статистичні ряди

Імовірнісні  
основи  
обробки  
даних

Статистичні  
ряди

Гістограма  
та полігон  
частот

Квантілі

Ящик з  
вусами

Графік  
 $Q - Q$

- Частотою  $n_i (i = \overline{1, N})$  інтервалу  $(u_i; u_{i+1}]$  називається число варіант  $x_i$ , що потрапили в цей інтервал, причому

$$\sum_{i=1}^N n_i = n.$$



# Статистичні ряди

Імовірнісні  
основи  
обробки  
даних

Статистичні  
ряди

Гістограма  
та полігон  
частот

Квантілі

Ящик з  
вусами

Графік  
 $Q - Q$

- Частотою  $n_i (i = \overline{1, N})$  інтервалу  $(u_i; u_{i+1}]$  називається число варіант  $x_i$ , що потрапили в цей інтервал, причому

$$\sum_{i=1}^N n_i = n.$$

- При групуванні спостережених значень за розрядами виникає питання про те, до якого інтервалу віднести значення, що знаходиться на межі двох розрядів. В цих випадках вважають, що дане значення належить до лівого інтервалу.

# Статистичні ряди

Імовірнісні  
основи  
обробки  
даних

Статистичні  
ряди

Гістограма  
та полігон  
частот

Квантілі

Ящик з  
вусами

Графік  
 $Q - Q$

- **Частотою**  $n_i (i = \overline{1, N})$  інтервалу  $(u_i; u_{i+1}]$  називається число варіант  $x_i$ , що потрапили в цей інтервал, причому

$$\sum_{i=1}^N n_i = n.$$

- При групуванні спостережених значень за розрядами виникає питання про те, до якого інтервалу віднести значення, що знаходиться на межі двох розрядів. В цих випадках вважають, що дане значення належить до лівого інтервалу.
- **Відносною частотою** або **вагою**  $\bar{n}_i (i = \overline{1, N})$  інтервалу  $(u_i; u_{i+1}]$  називається відношення частоти інтервалу до об'єму вибірки  $n$ , тобто

$$\bar{n}_i = \frac{n_i}{n}$$

# Статистичні ряди

Імовірнісні  
основи  
обробки  
даних

Статистичні  
ряди

Гістограма  
та полігон  
частот

Квантілі

Ящик з  
вусами

Графік  
 $Q - Q$

- **Накопиченою відносною частотою**  $w_i (i = \overline{1, N})$  інтервалу  $(u_i; u_{i+1}]$  називається сума відносних частот перших  $i$  інтервалів, тобто

$$w_i = \sum_{j=1}^i \bar{n}_j.$$

# Статистичні ряди

Імовірнісні  
основи  
обробки  
даних

Статистичні  
ряди

Гістограма  
та полігон  
частот

Квантілі

Ящик з  
вусами

Графік  
 $Q - Q$

- **Накопиченою відносною частотою**  $w_i (i = \overline{1, N})$  інтервалу  $(u_i; u_{i+1}]$  називається сума відносних частот перших  $i$  інтервалів, тобто

$$w_i = \sum_{j=1}^i \bar{n}_j.$$

- **Групованим варіаційним рядом** називається впорядкована сукупність інтервалів з відповідними їм частотами  $n_i$ , відносними частотами  $\bar{n}_i$  і накопиченими відносними частотами  $w_i$ .

# Статистичні ряди

Імовірнісні  
основи  
обробки  
даних

Статистичні  
ряди

Гістограма  
та полігон  
частот

Квантілі

Ящик з  
вусами

Графік  
 $Q - Q$

| Індекс<br>$i$  | Інтервал<br>$(u_i; u_{i+1}]$ | Частота<br>$n_i$ | Относит.<br>частота<br>$\bar{n}_i$ | Накопл.<br>относит.<br>частота<br>$w_i$ |
|----------------|------------------------------|------------------|------------------------------------|---|
| 1              | $[u_1; u_2]$                 | $n_1$            | $\bar{n}_1$                        | $w_1 = \bar{n}_1$                       |
| 2              | $(u_2; u_3]$                 | $n_2$            | $\bar{n}_2$                        | $w_2 = \bar{n}_1 + \bar{n}_2$           |
| ...            | ...                          | ...              | ...                                | ...                                     |
| $N$            | $(u_N; u_{N+1}]$             | $n_N$            | $\bar{n}_N$                        | $w_N = 1$                               |
| $\sum_{i=1}^N$ |                              | $n$              | 1                                  |   |

# Статистичні ряди

Імовірнісні  
основи  
обробки  
даних

## Розглянемо приклад

|      |      |      |       |      |      |      |      |      |       |
|------|------|------|-------|------|------|------|------|------|-------|
| 2,88 | 2,78 | 4,90 | 4,41  | 4,86 | 4,46 | 4,76 | 4,48 | 4,71 | 4,70  |
| 2,94 | 5,37 | 7,48 | -3,32 | 5,79 | 8,55 | 8,27 | 5,65 | 7,23 | 7,95  |
| 2,95 | 2,44 | 7,89 | 2,45  | 5,90 | 2,45 | 2,67 | 2,50 | 2,67 | 2,51  |
| 5,16 | 4,40 | 9,12 | 5,52  | 1,56 | 8,46 | 1,34 | 5,69 | 9,57 | -1,07 |
| 5,20 | 4,99 | 9,00 | 8,47  | 6,55 | 2,88 | 6,78 | 5,72 | 6,10 | 0,13  |
| 4,23 | 5,15 | 6,39 | 4,39  | 6,56 | 5,78 | 6,85 | 4,40 | 6,23 | 0,56  |
| 4,23 | 2,99 | 6,46 | 6,88  | 9,63 | 4,22 | 3,58 | 6,57 | 5,83 | 9,35  |
| 4,33 | 3,24 | 9,97 | 6,99  | 5,22 | 8,93 | 3,69 | 6,58 | 7,09 | 5,68  |
| 4,38 | 3,27 | 7,19 | 1,73  | 5,29 | 1,96 | 3,71 | 1,99 | 2,31 | 2,30  |
| 5,67 | 3,90 | 7,38 | 3,94  | 5,33 | 3,98 | 3,79 | 4,08 | 4,12 | 4,12  |

Статистичні  
ряди

Гістограма  
та полігон  
частот

Квантілі

Ящик з  
вусами

Графік  
 $Q - Q$

# Статистичні ряди

Імовірнісні  
основи  
обробки  
даних

Статистичні  
ряди

Гістограма  
та полігон  
частот

Квантілі

Ящик з  
вусами

Графік  
 $Q - Q$

Потрібно скласти групований варіаційний ряд для вибіркової сукупності значень випадкової величини  $X$ , розбивши вибірку на

$$N = 10$$

рівних інтервалів.

Дана вибірка має об'єм

$$n = 10.$$

Визначимо інтервал зміни випадкової величини  $X$ . Для цього в таблиці знаходимо максимальний і мінімальний елементи:

$$x_{max} = 9,97 \quad x_{min} = -3,32$$

# Статистичні ряди

Імовірнісні  
основи  
обробки  
даних

- Визначимо розмах вибірки:

$$R = x_{max} - x_{min} = 13,29$$

Статистичні  
ряди

Гістограма  
та полігон  
частот

Квантілі

Ящик з  
вусами

Графік  
 $Q - Q$



# Статистичні ряди

Імовірнісні  
основи  
обробки  
даних

Статистичні  
ряди

Гістограма  
та полігон  
частот

Квантілі

Ящик з  
вусами

Графік  
 $Q-Q$

- Визначимо розмах вибірки:

$$R = x_{max} - x_{min} = 13,29$$

- Для зручності подальшої обробки статистичних даних іноді слід округлити  $x_{max} - x_{min}$  до найближчих цілих чисел

$$x_{max}^0 = 10 \quad x_{min}^0 = -4$$

# Статистичні ряди

Імовірнісні  
основи  
обробки  
даних

Статистичні  
ряди

Гістограма  
та полігон  
частот

Квантілі

Ящик з  
вусами

Графік  
 $Q - Q$

- Визначимо розмах вибірки:

$$R = x_{max} - x_{min} = 13,29$$

- Для зручності подальшої обробки статистичних даних іноді слід округлити  $x_{max} - x_{min}$  до найближчих цілих чисел

$$x_{max}^0 = 10 \quad x_{min}^0 = -4$$

- Тоді новий розмах

$$R^0 = x_{max}^0 - x_{min}^0 = 14$$

# Статистичні ряди

Імовірнісні  
основи  
обробки  
даних

Статистичні  
ряди

Гістограма  
та полігон  
частот

Квантілі

Ящик з  
вусами

Графік  
 $Q - Q$

Далі можна розбити вибірку  
на  $N = 10$  рівних інтервалів,  
довжина кожного часткового  
інтервалу дорівнює

$$\Delta = \frac{R^0}{N} = \frac{14}{10} = 1,4$$

Інтервали наведені у другому  
стовбчику таблиці.

# Статистичні ряди

Імовірнісні  
основи  
обробки  
даних

Статистичні  
ряди

Гістограма  
та полігон  
частот

Квантілі

Ящик з  
вусами

Графік  
 $Q - Q$

Далі можна розбити вибірку на  $N = 10$  рівних інтервалів, довжина кожного часткового інтервалу дорівнює

$$\Delta = \frac{R^0}{N} = \frac{14}{10} = 1,4$$

Інтервали наведені у другому стовбчику таблиці.

| Індекс<br>$i$ | Інтервал<br>( $u_i$ ; $u_{i+1}$ ] |
|---------------|-----------------------------------|
| 1             | [ -4,0; -2,6 ]                    |
| 2             | ( -2,6; -1,2 ]                    |
| 3             | ( -1,2; 0,2 ]                     |
| 4             | ( 0,2; 1,6 ]                      |
| 5             | ( 1,6; 3,0 ]                      |
| 6             | ( 3,0; 4,4 ]                      |
| 7             | ( 4,4; 5,8 ]                      |
| 8             | ( 5,8; 7,2 ]                      |
| 9             | ( 7,2; 8,6 ]                      |
| 10            | ( 8,6; 10,0 ]                     |

# Статистичні ряди

Імовірнісні  
основи  
обробки  
даних

Статистичні  
ряди

Гістограма  
та полігон  
частот

Квантілі

Ящик з  
вусами

Графік  
 $Q - Q$

Знайдемо кількість варіант, що потрапили в кожний частковий інтервал розбиття, і заповнимо третій стовчик таблиці. Сума всіх частот повинна бути

$$n = 100.$$

Далі знаходимо відносні частоти і накопичені відносні частоти (четвертий і п'ятий стовпці таблиці). Варіаційний ряд представляється у вигляді таблиці

# Статистичні ряди

Імовірнісні  
основи  
обробки  
даних

Статистичні  
ряди

Гістограма  
та полігон  
частот

Квантілі

Ящик з  
вусами

Графік  
 $Q - Q$

| Індекс<br>$i$ | Інтервал<br>( $u_i; u_{i+1}$ ] | Частота<br>$n_i$ | Относит.<br>частота<br>$\bar{n}_i$ | Накопл.<br>относит.<br>частота<br>$w_i$ |
|---------------|--------------------------------|------------------|------------------------------------|---|
| 1             | [-4,0; -2,6]                   | 1                | 0,01                               | 0,01                                    |
| 2             | (-2,6; -1,2]                   | 0                | 0                                  | 0,01                                    |
| 3             | (-1,2; 0,2]                    | 2                | 0,02                               | 0,03                                    |
| 4             | (0,2; 1,6]                     | 3                | 0,03                               | 0,06                                    |
| 5             | (1,6; 3,0]                     | 18               | 0,18                               | 0,24                                    |
| 6             | (3,0; 4,4]                     | 20               | 0,2                                | 0,44                                    |
| 7             | (4,4; 5,8]                     | 24               | 0,24                               | 0,68                                    |
| 8             | (5,8; 7,2]                     | 16               | 0,16                               | 0,84                                    |
| 9             | (7,2; 8,6]                     | 9                | 0,09                               | 0,93                                    |
| 10            | (8,6; 10,0]                    | 7                | 0,07                               | 1                                       |
| Сумма         |                                | 100              | 1                                  |   |

# Статистичні ряди

Імовірнісні  
основи  
обробки  
даних

Статистичні  
ряди

Гістограма  
та полігон  
частот

Квантілі

Ящик з  
вусами

Графік  
 $Q - Q$

З таблиці видно, що дана вибірка має одну ізольовану точку

$$x_{min} = -3,32$$

віддалену від групи інших експериментальних точок.

У такому випадку можна вважати цю ізольовану точку аномальним спостереженням, грубою помилкою вимірювання і видалити її з вибірки, тоді обсяг вибірки зменшиться і дорівнюватиме  $n = 99$ . Змінюються також і вибіркові характеристики

# Статистичні ряди

Імовірнісні  
основи  
обробки  
даних

Статистичні  
ряди

Гістограма  
та полігон  
частот

Квантілі

Ящик з  
вусами

Графік  
 $Q - Q$

За вбудованими функціями EXCEL  $n = 100$

$$\bar{x} = \frac{1}{n} \cdot \sum_{i=1}^n x_i = \frac{1}{100} \cdot 491,2900 = 4,9129.$$



# Статистичні ряди

Імовірнісні  
основи  
обробки  
даних

Статистичні  
ряди

Гістограма  
та полігон  
частот

Квантілі

Ящик з  
вусами

Графік  
 $Q - Q$

За вбудованими функціями EXCEL  $n = 100$

$$S^2 = \frac{1}{n} \cdot \sum_{i=1}^n x_i^2 - \bar{x}^2 = \frac{1}{100} \cdot 2985,1739 - 4,9129^2 = 5,7152,$$

$$S_0^2 = \frac{1}{n-1} \cdot \sum_{i=1}^n (x_i - \bar{x})^2 = \frac{n \cdot S^2}{n-1} = \frac{100 \cdot 5,7152}{99} = 5,7729.$$

$$\tilde{\sigma}_X = \sqrt{\tilde{D}_X} = \sqrt{S_0^2} = S_0 = 2,4027.$$

# Статистичні ряди

Імовірнісні  
основи  
обробки  
даних

За вбудованими функціями EXCEL  $n = 99$

$$\bar{x} = \frac{1}{n} \cdot \sum_{i=1}^n x_i = 4,9961,$$

$$S_0^2 = \frac{1}{n-1} \cdot \sum_{i=1}^n (x_i - \bar{x})^2 = 5,1332,$$

$$S_0 = \sqrt{S_0^2} = \sqrt{5,1332} = 2,2657.$$

Статистичні  
ряди

Гістограма  
та полігон  
частот

Квантілі

Ящик з  
вусами

Графік  
 $Q - Q$

# Статистичні ряди

Імовірнісні  
основи  
обробки  
даних

Статистичні  
ряди

Гістограма  
та полігон  
частот

Квантілі

Ящик з  
вусами

Графік  
 $Q - Q$

Перевірка гіпотези про аномальність спостереження проводиться наступним чином:  
значення визнається аномальним і викидається з вибірки обсягу  $n$ , якщо воно не відповідає даній **довірчій імовірності** або заданій **квантілі**.

# Статистичні ряди

Імовірнісні  
основи  
обробки  
даних

Статистичні  
ряди

Гістограма  
та полігон  
частот

Квантілі

Ящик з  
вусами

Графік  
 $Q - Q$

Коли умова виконується, то точку

$$x_{min} = -3,32$$

можна з вибірки виключити. Відповідно в таблиці можна виключити два перших інтервали.

Зауважимо, що число інтервалів тепер виявилося 8, що відповідає умові:

$$N \geq [1 + 3,32 \cdot \lg n] + 1 = [1 + 3,32 \cdot \lg 99] + 1 = 8.$$

В іншому випадку число інтервалів довелося б збільшити.

# Статистичні ряди

Імовірнісні  
основи  
обробки  
даних

Статистичні  
ряди

Гістограма  
та полігон  
частот

Квантілі

Ящик з  
вусами

Графік  
 $Q - Q$

| Індекс<br>$i$ | Інтервал<br>( $u_i$ ; $u_{i+1}$ ] | Частота<br>$n_i$ | Относит.<br>частота<br>$\bar{n}_i$ | Накопл.<br>относит.<br>частота<br>$w_i$ |
|---------------|-----------------------------------|------------------|------------------------------------|---|
| 1             | ( -1,2; 0,2 ]                     | 2                | 0,0202                             | 0,0202                                  |
| 2             | ( 0,2; 1,6 ]                      | 3                | 0,0303                             | 0,0505                                  |
| 3             | ( 1,6; 3,0 ]                      | 18               | 0,1818                             | 0,2323                                  |
| 4             | ( 3,0; 4,4 ]                      | 20               | 0,2020                             | 0,4343                                  |
| 5             | ( 4,4; 5,8 ]                      | 24               | 0,2424                             | 0,6768                                  |
| 6             | ( 5,8; 7,2 ]                      | 16               | 0,1616                             | 0,8384                                  |
| 7             | ( 7,2; 8,6 ]                      | 9                | 0,0909                             | 0,9293                                  |
| 8             | ( 8,6; 10,0 ]                     | 7                | 0,0707                             | 1,0000                                  |
| Сумма         |                                   | 99               | 1,0000                             |   |

# Гістограма та полігон частот

Імовірнісні  
основи  
обробки  
даних

Статистичні  
ряди

Гістограма  
та полігон  
частот

Квантілі

Ящик з  
вусами

Графік  
 $Q - Q$

Нехай  $\{x_i\}$  вибірка обсягу  $n$ , що містить  $k$  різних варіант, з генеральної сукупності випадкової величини  $X$  з невідомою густиною імовірності  $f(x)$ .

Наближенням (оцінкою) невідомої густини імовірності можуть служити гістограма або полігон відносних частот.

**Гістограма і полігон відносних частот** служать для геометричного зображення асоційованого варіаційного ряду.

# Гістограма та полігон частот

Імовірнісні  
основи  
обробки  
даних

Статистичні  
ряди

Гістограма  
та полігон  
частот

Квантілі

Ящик з  
вусами

Графік  
 $Q - Q$

Гістограма відносних частот представляється у вигляді прямокутників, що примикають один до одного, з основами

$$\Delta = \frac{R}{N}$$

рівними ширині інтервалів груп, і висотами

$$h_i = \frac{\bar{n}_i}{\Delta}$$

Для гістограми відносних частот площа фігури відповідає сумі ймовірностей і дорівнює 1. Площа будь-якого прямокутника гістограми дорівнює ймовірності влучення значень розглянутої випадкової величини в інтервал, що відповідає основи прямокутника.

# Гістограма та полігон частот

Імовірнісні  
основи  
обробки  
даних

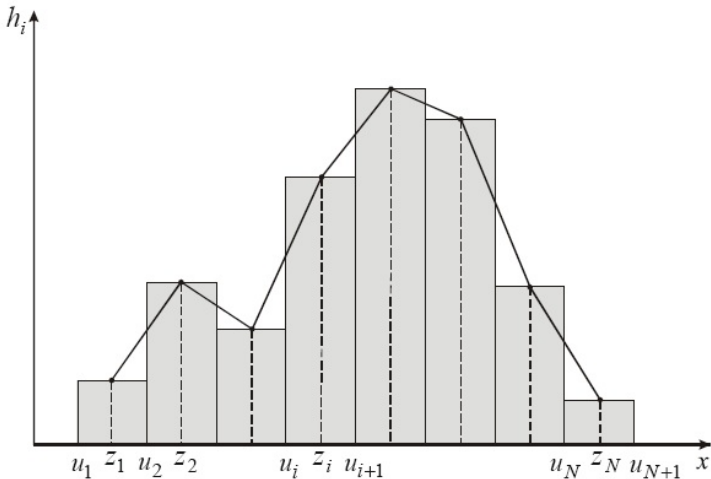
Статистичні  
ряди

Гістограма  
та полігон  
частот

Квантілі

Ящик з  
вусами

Графік  
 $Q - Q$





# Гістограма та полігон частот

Імовірнісні  
основи  
обробки  
даних

Статистичні  
ряди

Гістограма  
та полігон  
частот

Квантілі

Ящик з  
вусами

Графік  
 $Q - Q$

**Полігоном** відносних частот називається ламана, що з'єднує точки середини інтервалів.

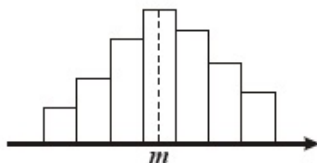
При збільшенні обсягу вибірки і зменшенні довжин інтервалів гістограма і полігон відносних частот наближаються до графіку невідомої функції  $f(x)$  – густини імовірності сукупності.

За виглядом гістограми та полігоном частот можна висунути гіпотезу про вигляд розподілу генеральної сукупності.

# Гістограма та полігон частот

## Нормальний розподіл (розподіл Гаусса)

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-m)^2}{2\sigma^2}}$$



Імовірнісні  
основи  
обробки  
даних

Статистичні  
ряди

Гістограма  
та полігон  
частот

Квантілі

Ящик з  
вусами

Графік  
 $Q - Q$

# Гістограма та полігон частот

Імовірнісні  
основи  
обробки  
даних

## Рівномірний розподіл

$$f(x) = \begin{cases} \frac{1}{b-a}, & x \in [a; b], \\ 0, & x \notin [a; b] \end{cases}$$



Статистичні  
ряди

Гістограма  
та полігон  
частот

Квантілі

Ящик з  
вусами

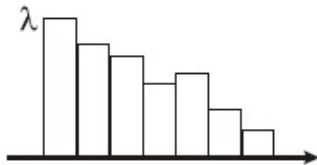
Графік  
 $Q-Q$

# Гістограма та полігон частот

Імовірнісні  
основи  
обробки  
даних

## Експоненційний розподіл

$$f(x) = \begin{cases} \lambda \cdot e^{-\lambda x}, & x > 0, \\ 0, & x \leq 0 \end{cases}$$



Статистичні  
ряди

Гістограма  
та полігон  
частот

Квантілі

Ящик з  
вусами

Графік  
Q — Q

# Гістограма та полігон частот

Імовірнісні  
основи  
обробки  
даних

Для зручності побудови полігону відносних частот в третій стовбчик додають середину інтервалу

| Індекс<br>$i$ | Інтервал<br>( $u_i; u_{i+1}$ ] | Середина<br>інтервала<br>$z_i$ | Относит.<br>частота<br>$\bar{n}_i$ | Висота<br>прямоуг.<br>$h_i$ |
|---------------|--------------------------------|--------------------------------|------------------------------------|-----------------------------|
| 1             | ( -1,2; 0,2 ]                  | -0,5                           | 0,0202                             | 0,0144                      |
| 2             | ( 0,2; 1,6 ]                   | 0,9                            | 0,0303                             | 0,0216                      |
| 3             | ( 1,6; 3,0 ]                   | 2,3                            | 0,1818                             | 0,1299                      |
| 4             | ( 3,0; 4,4 ]                   | 3,7                            | 0,2020                             | 0,1443                      |
| 5             | ( 4,4; 5,8 ]                   | 5,1                            | 0,2424                             | 0,1732                      |
| 6             | ( 5,8; 7,2 ]                   | 6,5                            | 0,1616                             | 0,1154                      |
| 7             | ( 7,2; 8,6 ]                   | 7,9                            | 0,0909                             | 0,0649                      |
| 8             | ( 8,6; 10,0 ]                  | 9,3                            | 0,0707                             | 0,0505                      |
| Сумма         |                                |                                | 1,0000                             | 1,0000                      |

Статистичні  
ряди

Гістограма  
та полігон  
частот

Квантілі

Ящик з  
вусами

Графік  
 $Q - Q$

# Гістограма та полігон частот

Імовірнісні  
основи  
обробки  
даних

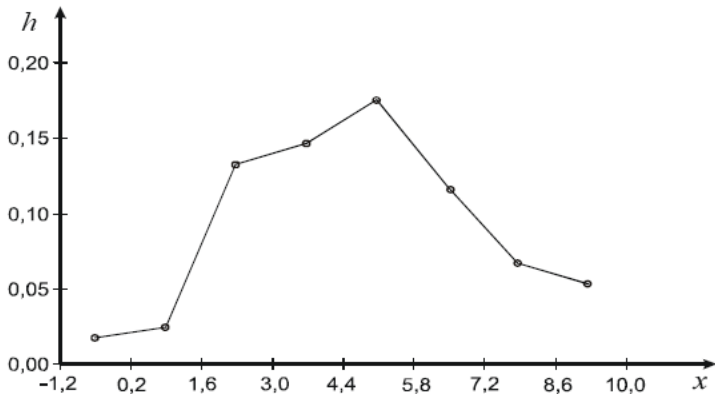
Статистичні  
ряди

Гістограма  
та полігон  
частот

Квантілі

Ящик з  
вусами

Графік  
 $Q - Q$



# Гістограма та полігон частот

Імовірнісні  
основи  
обробки  
даних

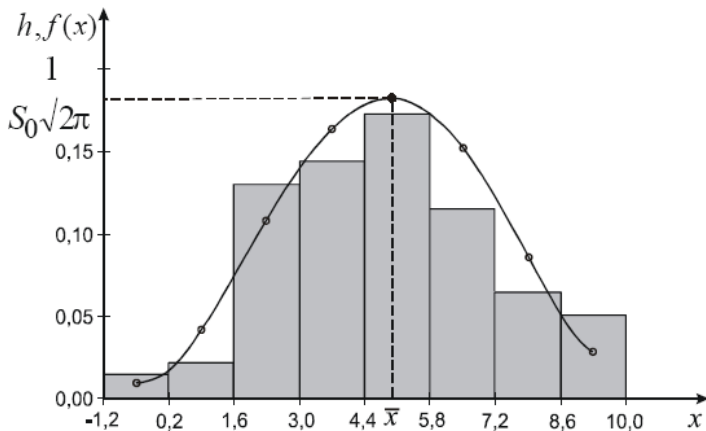
Статистичні  
ряди

Гістограма  
та полігон  
частот

Квантілі

Ящик з  
вусами

Графік  
 $Q - Q$



# Гістограма та полігон частот

Імовірнісні  
основи  
обробки  
даних

Статистичні  
ряди

Гістограма  
та полігон  
частот

Квантілі

Ящик з  
вусами

Графік  
 $Q - Q$

Гістограма і полігон відносних частот, є статистичними оцінками густини імовірностей генеральної сукупності, схожі з кривою густини імовірностей нормального закону.

На підставі цього висувається нульова гіпотеза.

Генеральна сукупність, з якої взята вибірка, розподілена за нормальним законом з параметрами

$$\bar{x} = 4,9961 \quad \sigma = 2,2657$$

тобто теоретична густина імовірності має вигляд

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-m)^2}{2\sigma^2}}$$



# Квантілі

Ймовірнісні  
основи  
обробки  
даних

Статистичні  
ряди

Гістограма  
та полігон  
частот

Квантілі

Ящик з  
вусами

Графік  
 $Q - Q$

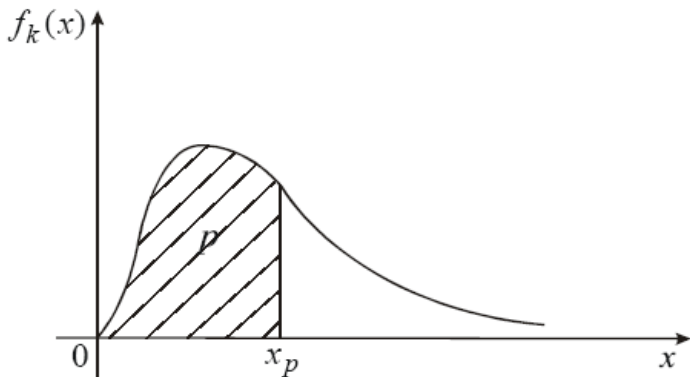
Квантиль, відповідної ймовірності  $p$ , називається таке значення  $x$ , при якому виконується співвідношення:

$$P(X < x_p) = \int_{-\infty}^{x_p} f(x) dx = p,$$

# Квантілі

Імовірнісні  
основи  
обробки  
даних

## Геометричне пояснення квантілі



Статистичні  
ряди

Гістограма  
та полігон  
частот

Квантілі

Ящик з  
вусами

Графік  
 $Q - Q$

# Квантилі

## Види квантилей

| Название квантилей | Число частей, на которые разбивается ряд |
|--------------------|--|
| Медиана            | 2  |
| Терциль            | 3  |
| Квартиль           | 4  |
| Дециль             | 10                                       |
| Процентиль         | 100                                      |

Імовірнісні  
основи  
обробки  
даних

Статистичні  
ряди

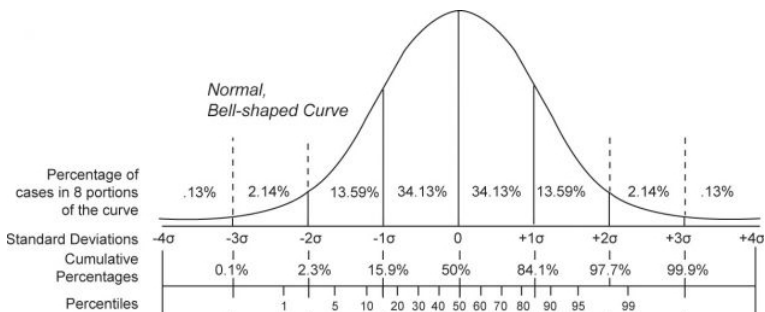
Гістограма  
та полігон  
частот

Квантілі

Ящик з  
вусами

Графік  
 $Q - Q$

## Нормальний розподіл



Normal Bell-shaped Curve – Нормальна колоколо-подібна крива

Percentage of cases in 8 portions of the curve – імовірність в % події на 8 відрізках кривої

Standard deviation – стандартне відхилення

Cumulative Percentages – кумулятивні відсотки

Percentiles – процентилі

# Квантилі

Імовірнісні  
основи  
обробки  
даних

Статистичні  
ряди

Гістограма  
та полігон  
частот

**Квантилі**

Ящик з  
вусами

Графік  
 $Q - Q$

Процентилі – це характеристика набору даних, що виражають ранги елементів масива і є показником того, який відсоток значень знаходиться нижче визначеного рівня.

Наприклад, значення 30-ї процентилі вказує, що 30% значень розташовано нижче цього рівня.

Імовірнісні  
основи  
обробки  
даних

Г

Статистичні  
ряди

Гістограма  
та полігон  
частот

Квантілі

Ящик з  
вусами

Графік  
 $Q - Q$

# Ящик з вусами

Імовірнісні  
основи  
обробки  
даних

Статистичні  
ряди

Гістограма  
та полігон  
частот

Квантілі

Ящик з  
вусами

Графік  
 $Q - Q$



# Ящик з вусами

Імовірнісні  
основи  
обробки  
даних

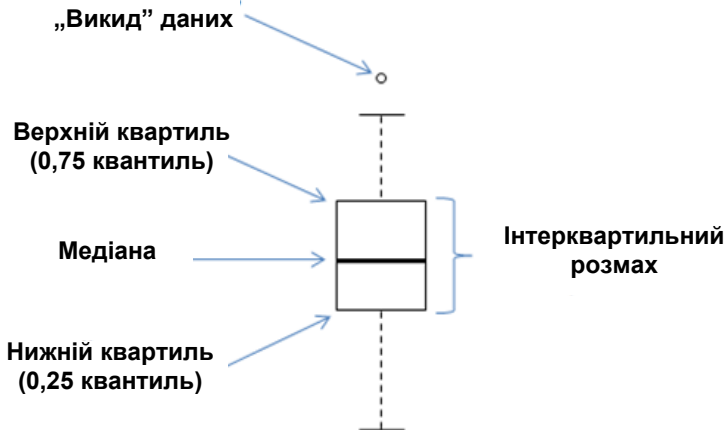
Статистичні  
ряди

Гістограма  
та полігон  
частот

Квантілі

Ящик з  
вусами

Графік  
 $Q - Q$





# Ящик з вусами

Імовірнісні  
основи  
обробки  
даних

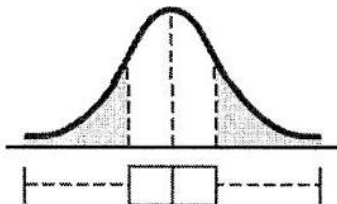
Статистичні  
ряди

Гістограма  
та полігон  
частот

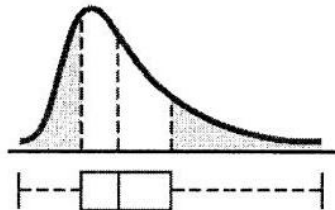
Квантілі

Ящик з  
вусами

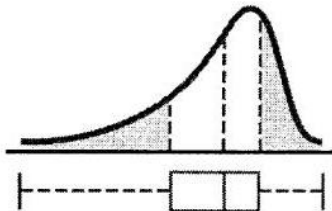
Графік  
 $Q - Q$



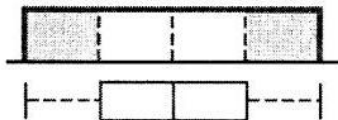
A



B



B



Г

# Ящик з вусами

Імовірнісні  
основи  
обробки  
даних

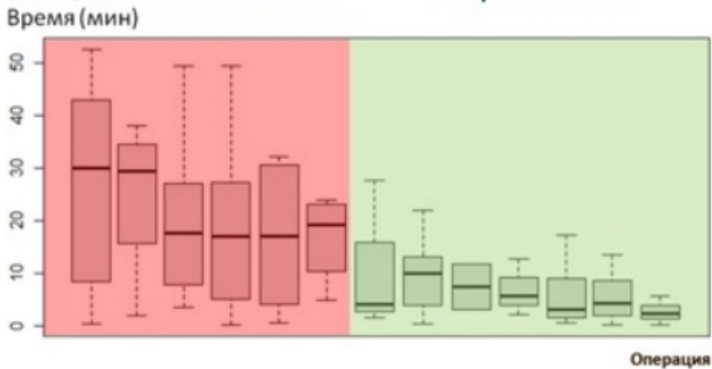
Статистичні  
ряди

Гістограма  
та полігон  
частот

Квантілі

Ящик з  
вусами

Графік  
 $Q - Q$



# Ящик з вусами

Імовірнісні  
основи  
обробки  
даних

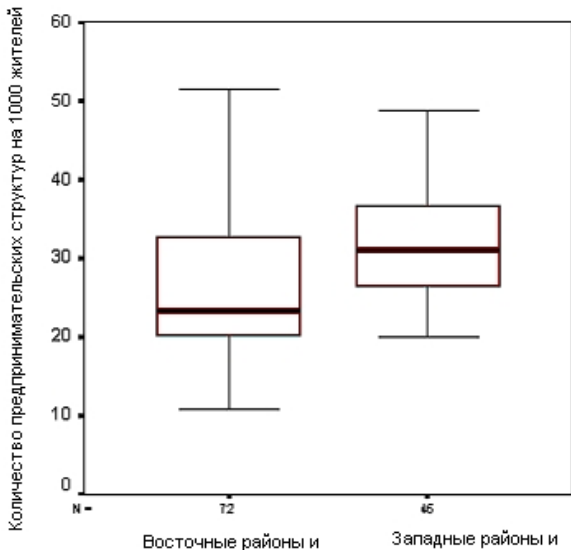
Статистичні  
ряди

Гістограма  
та полігон  
частот

Квантілі

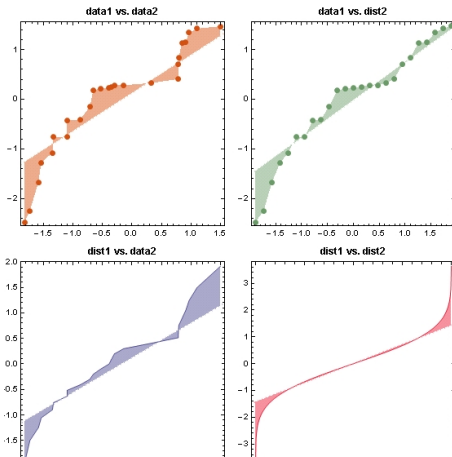
Ящик з  
вусами

Графік  
 $Q - Q$



# Графік $Q - Q$

Графік квантиль-квантиль використовується для порівняння даних з розподілами імовірностей, даних з даними, розподілів з даними і розподілів з розподілами.



Імовірнісні  
основи  
обробки  
даних

Статистичні  
ряди

Гістограма  
та полігон  
частот

Кванtilі

Ящик з  
вусами

Графік  
 $Q - Q$