

Фаховий практикум.

4 курс, 1 семестр.

Аналіз вибіркової сукупності

План роботи

- ❑ Створення виборки. Види виборок
 - ❑ Обробка виборки. Поняття варіаційного ряду. Аналіз на наявність артефактів
 - ❑ Основні параметри виборки
-

Створення виборки

- *Генеральною сукупністю* називають сукупність усіх можливих спостережень, які могли б бути проведені у відповідності з метою дослідження
 - Кінцева
 - Незкінченна
-

Навіщо потрібна виборка?

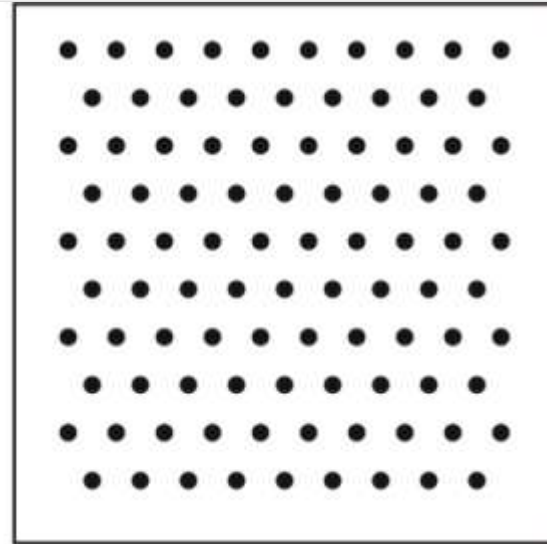
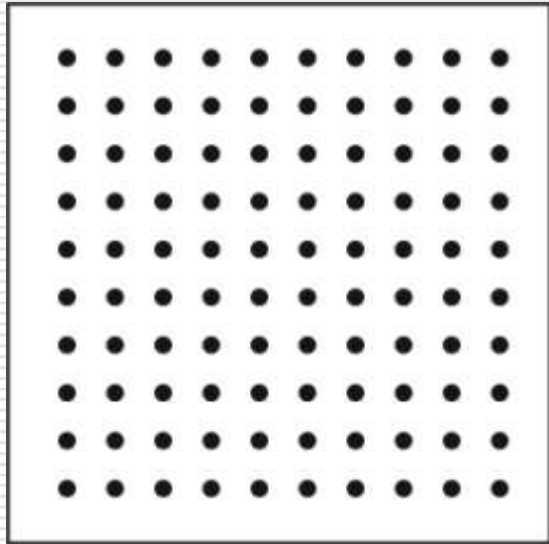
□ Об'єм виборки:

$$N = \frac{\sigma^2}{m^2}$$

де m^2 – помилка середнього арифметичного; σ^2 – середнє квадратичне відхилення

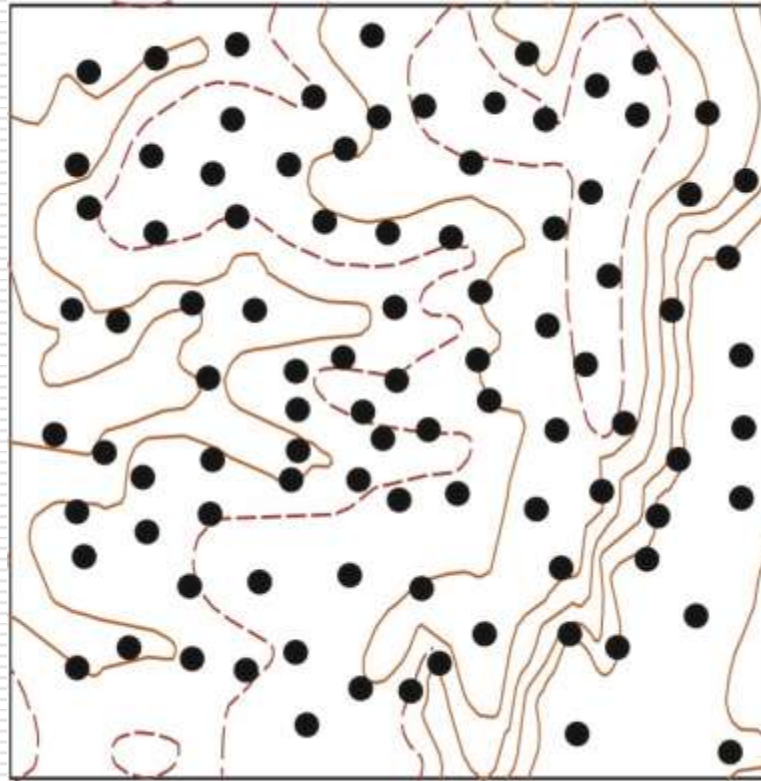
Правила створення виборок

□ Систематична виборка.



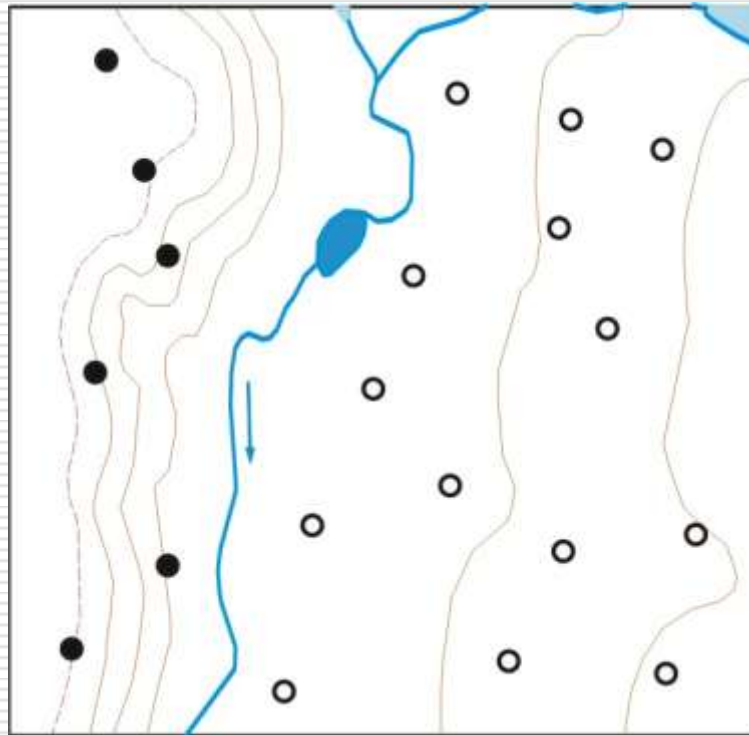
Правила створення виборок

□ *Проста випадкова виборка.*



Правила створення виборок

□ Стратифікована виборка



Обробка вибіркової сукупності

- Випадкові величини, що представлені низкою кількісних показників, утворюють **статистичну (вибіркову) сукупність**. Кожний член цієї сукупності називають **варіантою**, або **датою**. Число варіант у сукупності називають **об'ємом сукупності**.
-

Обробка вибіркової сукупності

- **Варіаційний ряд** – розміщення елементів виборки в порядку зростання (убування).
 - Різниця між найменшим і найбільшим значеннями складає **розмах варіювання, або амплітуду мінливості**
-

Аналіз на артефакти

□ *критерій τ*

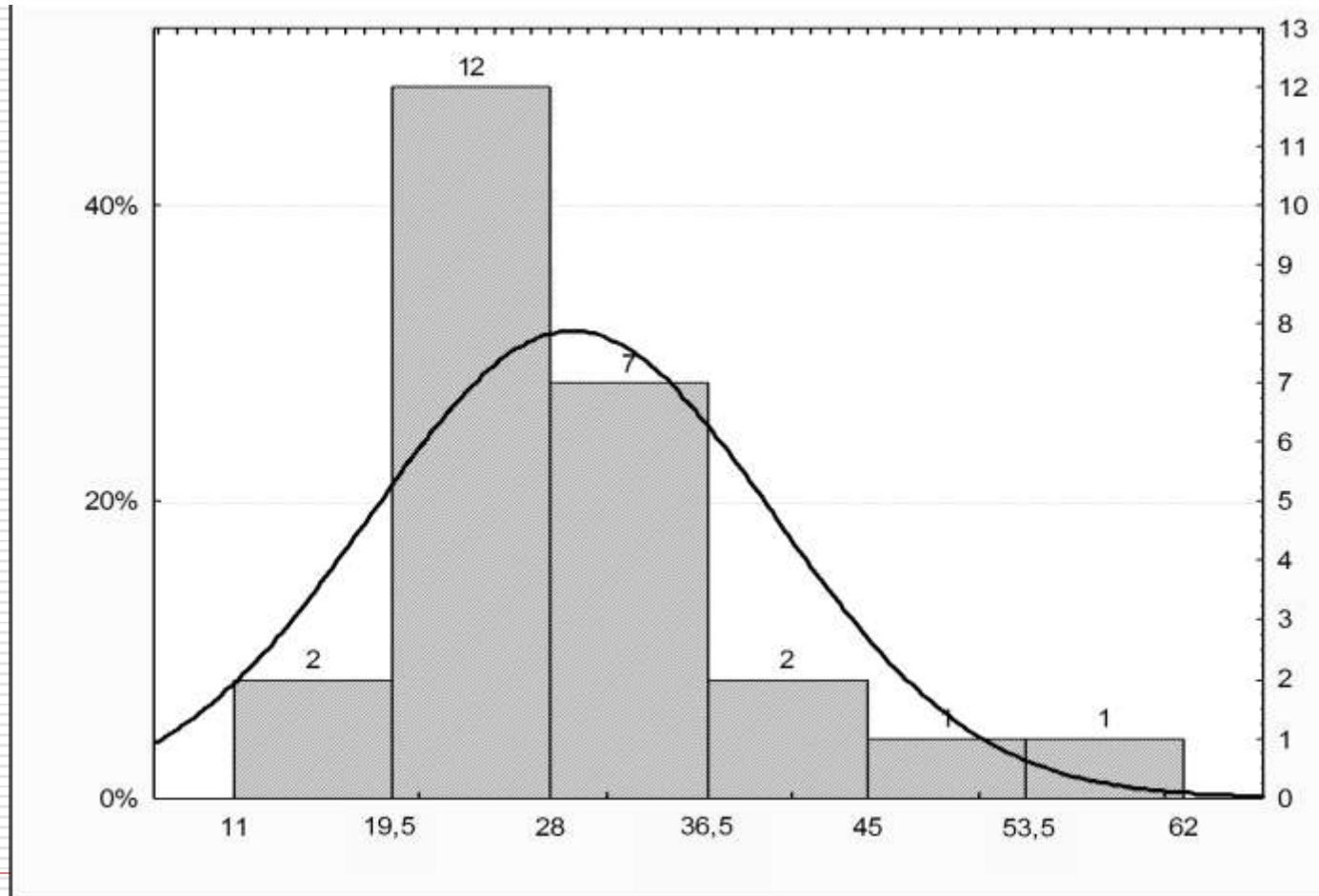
□ Для найменшого значення

$$\tau_1 = (x_2 - x_1) / (x_{n-1} - x_1)$$

□ Для найбільшого значення

$$\tau_n = (x_n - x_{n-1}) / (x_n - x_2)$$

Аналіз виборки. Гістограма



Гістограма: оптимальна кількість інтервалів

$$k = 1 + 3,3 \lg N$$

Число класів

$$i = (x_{\max} - x_{\min}) / k$$

Класовий інтервал

Параметри виборки

- ❑ **непараметричні**, тобто значення яких не залежить від частотного розподілу (мода, медіана);
 - ❑ **параметричні**, значення яких тісно пов'язане з формою розподілу частот точніші (середні величини: арифметичне, гармонічне, квадратичне, кубічне, геометричне).
-

Мода

- *Мода (Mo)* представляє собою варіанту, що найбільш часто зустрічається у варіаційному ряді.

$$Mo = x_m + i \left(\frac{f_2 - f_1}{2f_2 - f_1 - f_3} \right)$$

- де x_m – менша межа модального класа; i – класовий інтервал; f_1 – частота класа, що передує модальному; f_2 – частота модального класа; f_3 – частота класа, наступного за модальним.
-

Медіана (Me)

- Медіана (Me) є середньою варіантою в ранжованому варіаційному ряді, яка ділить його на дві рівні за числом варіант частини.

$$Me = x_{Me} + i \left(\frac{0,5N - \sum f}{f_{Mo}} \right)$$

- де x_{Me} – початок класа, в якому знаходиться медіана; N – об'єм виборки; $\sum f$ – сума частот усіх класів, що передують модальному; f_{Mo} – частота модального класа.
-

Середнє арифметичне (M , x)

- є величиною, сума позитивних і від'ємних відхилень від якої дорівнює нулю

$$M = \frac{\sum x}{N}$$

- где $\sum x$ - сума всіх варіант сукупності, N – її об'єм.
-

Помилка середнього арифметичного

$$m_M = \sigma / \sqrt{N}$$

- *Достовірність* – це апріорне переконання в здійсненності деякого явища, що виключає всякий сумнів.
-

Оцінка достовірності

□ Критерій Стюдента

$$t_{\phi} = M / m_M$$

Розрахунковий критерій Стюдента t_{ϕ} зіставляють з його табличним значенням t_T . У випадку, якщо $t_{\phi} > t_T$, то значення показника є достовірним.

Для t_{ϕ} необхідно знати значення таких показників, як число ступенів свободи ν та рівень вірогідності, або довірчу вірогідність P .

Оцінка достовірності

- Числом ступенів свободи ν вважається число незадежних відхилень окремих варіант від середнього.

$$\nu = N - 1$$

Вірогідність, з якою встановлюються можливі значення змінної величини, отримала назву **довірчої вірогідності p** .

Значення: 0,95, 0,99 і 0,999

рівень значущості α - відсоток числа варіант, значення яких не підтверджують шукану закономірність. Це величина, зворотна p

Різноманітність ознак

- *Середнє квадратичне (стандартне) відхилення, або сигма (σ), показує ступінь розсіяння значень статистичної сукупності навколо середнього значення.*

$$\sigma = \sqrt{\frac{\sum (x_i - M)^2}{N - 1}}$$

де x_i – індивідуальна варіанта сукупності; $x_i - M$ – відхилення від середнього індивідуальних варіант;

Помилка стандартного відхилення

□ Визначається за формулою:

$$m_{\sigma} = \sigma / \sqrt{2(N-1)}$$

Достовірність визначається за допомогою критерію Стюдента

Точність обчислення помилок середнього квадратичного відхилення і середнього арифметичного можна перевірити приблизно за допомогою співвідношення: $\sigma/M = 0,70711$ Якщо співвідношення виявиться близьким до 0,7, то отримані результати обчислень слід вважати репрезентативними
