**Лекція 1. Основні поняття використання статистичних методів у суспільній географії. Моделювання, як метод пізнання, в суспільній географії.**

План лекції

1. Визначення математичних методів та моделювання.
2. Поняття «модель». Класифікація моделей: натурні, аналогові, математичні.
3. Рівні математизації географічних об’єктів
4. Місце методів ідеалізації, формалізації, математичних методів, методів моделювання в методології суспільно- географічного дослідження.

**1. Визначення математичних методів та моделювання.**

**Математичні методи** – універсальні методи, використовуються на всіх стадіях дослідження – від первинної обробки до узагальнення. Різні розділи математики, найчастіше теорія ймовірностей і математична статистика (факторний аналіз, кластер – аналіз, тренд – аналіз тощо), мат-логіка і теорія множин, диференційні рівняння (динаміка, еволюція систем), матпрограмування, теорія графів.

**Моделювання** - загальнонауковий метод, що використовується у процесі проведення дослідження розвитку та розміщення соціальних об'єктів, процесів і явищ. При застосуванні цього методу об'єкт дослідження вивчається не безпосередньо, а за допомогою посередника - ідеальної (уявної), матеріальної (предметної) моделі чи особливої форми ідеально-матеріальної моделі, якою може бути, наприклад, мала (картосхема).

**Метою моделювання** є здобуття, обробка, представлення і використання інформації про об'єкти, які взаємодіють між собою і зовнішнім середовищем; а модель тут виступає як засіб пізнання властивостей і закономірностей поведінки об'єкту. Основним призначенням моделі в задачах управління є прогноз реакції об’єкту на керуючі впливи. Крім того, моделі використовуються для дослідження об’єкта, аналізу його чутливості.

**Властивості моделей:** цілеспрямованість; скінченність; спрощеність; повнота; адекватність. **Задача моделювання** полягає в тому, що для заданого об’єкта потрібно підібрати такий опис, який у повній мірі відображав би оригінал з точки зору заданої мети моделювання.

**2. Поняття «модель». Класифікація моделей: натурні, аналогові, математичні**

**Модель** – це об’єкт-замінник, створений з метою відтворення при певних умовах суттєвих властивостей об’єкта-оригіналу. Модель може бути представлена фізичним об’єктом, подібним до оригіналу, або описом об’єкта у вигляді математичних формул, тексту, комп’ютерної програми.

Вирізняють натурні, аналогові, математичні моделі. **Математичною моделлю** називається сукупність математичних співвідношень, рівнянь, нерівностей, що описують основні закономірності, властиві досліджуваному процесу, об'єкту або системі. **Математична модель** ставить у відповідність модельованому фізичному процесу систему математичних співвідношень, вирішення якої дозволяє отримати відповідь на питання про поведінку об'єкту без створення фізичної моделі, яка часто є дорогою і малоефективною. **Натурна модель** в природознавстві, медико-біологічних напрямках, біотехнології, медицині, ветеринарії, сільському господарстві - використання ізольованих органів і тканин, культури клітин, штучних середовищ і умов. Натурні моделі застосовуються і в мистецтві (натурники), в біоніці та в ін. Наприклад анатомія в малюнках сформувалася в результаті анатомічних досліджень художниками і скульпторами як предмет мистецтва.

**Аналогова модель** представляє досліджуваний об'єкт аналогом, який веде себе як реальний об'єкт, але не виглядає як такий. Аналогова модель явно більш простий і ефективний спосіб сприйняття і прояву складних взаємозв’язків структури великої організації, ніж, скажімо, складання переліку взаємозв'язків всіх працівників.

**3. Рівні математизації географічних об’єктів**

1. Рівень параметризації географічних об’єктів. - методологічне та методичне обгрунтування кількісної, напівкількісної, напівякісної та якісної оцінки географічних параметрів, тобто, їх цифрового представлення для подальшого кількісного аналізу.

2. Рівень емпіричних моделей - узагальнення географічних параметрів окремих об’єктів з екстраполяцією у часі та просторі. Для цього часто використовуються статистичні моделі та методи. Емпіричні моделі мають обмежене використання, але створюють інформаційні умови для побудови концепцій та теорій географічних процесів. При побудові емпіричних моделей використовується метод індукції.

3. Рівень теоретичних моделей – побудова моделей географічних процесів та об’єктів, виходячи з теоретичних уявлень про них. Такі моделі часто називаються концептуальними. При їх створенні здебільшого застосовується метод дедукції. Використання теоретичних моделей свідчить про зрілість науки.

**4. Місце методів ідеалізації, формалізації, математичних методів, методів моделювання в методології суспільно- географічного дослідження**

Найпоширенішим є метод моделювання (стоїть на 1 місці), на другому математичні методи (засновані на математичних законах, імовірність на приклад), на третьому методи формалізації (формули с-г), методи ідеалізації найменьш поширені, т.я. не дають комплексної оцінки с/г явищ.

Методи ідеалізації - створення ідеального образу досліджуваного об’єкту, відсіювання другорядних та несуттєвих ознак, визначення основних (імперативних) закономірностей та співвідношень між ознаками (параметрами) тощо. Для цього використовуються методи аналізу та синтезу.

Методи формалізації – створення однозначного опису образу об’єкта за допомогою методів формальної логіки та математики. В ідеальному випадку результатом є система математичних рівнянь у загальному вигляді.

Математичні методи – фрагментарний опис та оцінка співвідношення параметрів об’єкту, висвітлення його часткових та локальних особливостей, первинна обробка географічної інформації.

Математичне моделювання - створення цілісної системної моделі об’єкту з урахуванням всіх його суттєвих часткових та локальних особливостей, його взаємодії з навколишнім середовищем, закономірностей розвитку та функціонування.

**Лекція 2. Специфіка розвитку природних і соціальних об’єктів. Випадковий характер організації географічного середовища.**

**План лекції**

1. Організація географічного середовища. Локальні та інтегральні процеси.
2. Багатофакторність географічних процесів
3. Складність інтерпретації причинно-наслідкових зв’язків в соціогеосистемах
4. Методологічні підходи у дослідженні географічних об’єктів
5. Поняття про ймовірність
6. Особливості географічної інформації
7. ***Організація географічного середовища. Локальні та інтегральні процеси.***

Організацію географічного середовища складають географічні, соціально-економічні об’єкти, його складають також поля (силові, інерціальні).

Потенціал природних ресурсів складають:

А) технології, люди, засоби виробництва, транспорт.

Б) Локальні та інтегральні процеси, які розвиваються за фізичними, хімічними законами (відкритими Вернадським), більшість з з них розвивається детерміновано – незалежно.

**Локальні процеси** - фізичні, хімічні, фізико-хімічні процеси, що мають детерміновані механізми (асоціація – дисоціація, окислення – відновлення, сорбція – десорбція, гідратація - дегідратація тощо) незалежно від умов; може змінюватися швидкість і напрям процесів, але механізми залишаються незмінними; точно описуються детермінованими рівняннями.

**2. Інтегральні процеси** - всі інші процеси в географічному середовищі, які складаються з локальних процесів, але мають принципово випадковий характер в залежності від конкретних умов, діючих факторів тощо. Образно говорячи, локальні процеси – це незмінні цеглини світоустрію, інтегральні процеси створюють різноманіття «архітектурних» форм буття і коректно описуються у термінах теорії ймовірностей.

1. **Багатофакторність географічних процесів**

Багатофакторність географічних процесів полягає у наступному:

1. Відсутність даних про невідомі фактори зумовлює невизначеність опису об’єкту, із зростанням кількості неконтрольованих факторів точність представлення об’єкту зменшується.

2. Діючі фактори змінюються у просторі і часі, закономірності їхнього сполучення і загальної дії невідомі, тому ознаки та параметри географічних факторів принципово мають випадковий характер – їх неможливо точно описати або передбачити. Тому географічні процеси є випадковими.

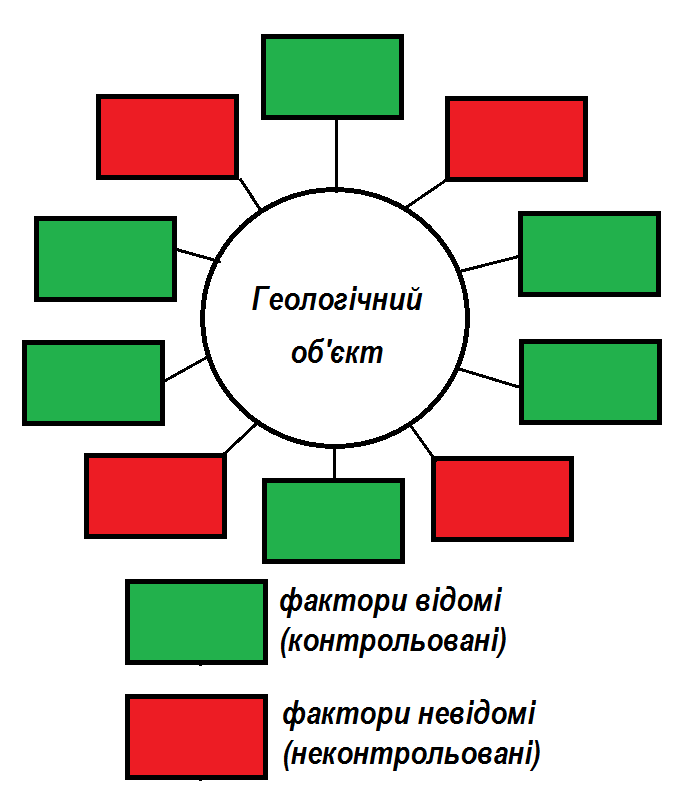


Рис. 1. Багатофакторність географічних процесів

1. **Складність інтерпретації причинно-наслідкових зв’язків в соціогеосистемах**

Складність інтерпретації причинно-наслідкових зв’язків у геосистемах

Зумовлена особливостями розвитку та функціонування геосистем:

- елементи (підсистеми) одночасно мають різні типи зв’язків;

- динамічність зв’язків у просторі та часі, що характерно для геосистем під антропогенним тиском;

- принципово випадковий характер процесів у геосистемах – в одних і тих же умовах отримуються різні результати;

- невпинний розвиток або взаємна адаптація геосистем – зміни параметрів у часі;

- обмежена кількість контрольних точок;

- мінливість геосистем у часі і просторі.

1. **Типи причинно-наслідкових зв’язків у геосистемах**

1. Прямі зв’язки – безпосередньо між елементами (підсистемами) одного ієрархічного рівня – найбільш поширений варіант.

2. Опосередкований зв’язок – між елементами (підсистемами) А і Б через елемент В – досить поширений варіант.

3. Непрямий зв’язок - між елементами (підсистемами) різних геологічних систем або різних ієрархічних рівнів однієї геосистеми

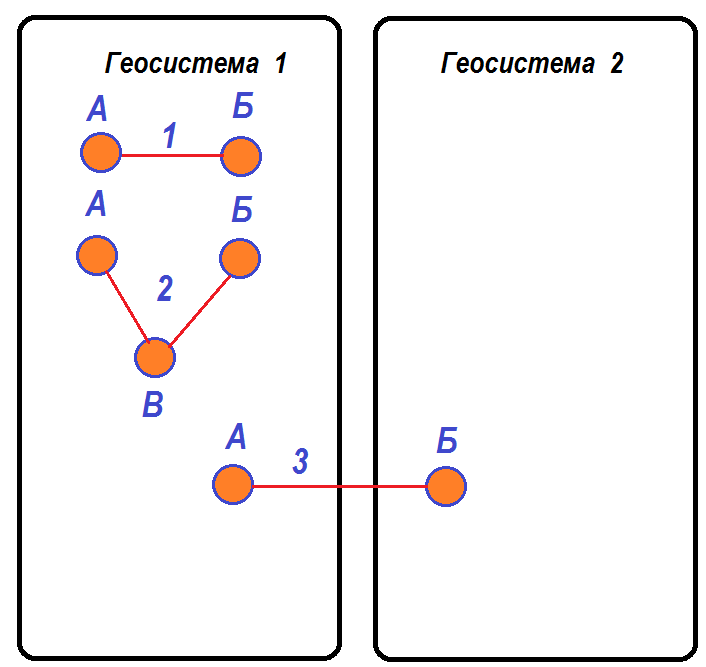


Рис. 3. Типи причинно-наслідкових зв’язків у геосистемах

1. **Методологічні підходи у дослідженні географічних об’єктів**

**1. Географічний (просторовий) підхід –** досліджуваний об’єкт перш за все слід розглядати у просторовому аспекті – побудова просторових моделей, виявлення просторових закономірностей тощо

2**. Геосистемний підхід -** досліджуваний об’єкт розглядається як цілісна система, що складається із взаємодіючих підсистем, елементів тощо.

**3. Синергетичний підхід –** враховуються синергічні ефекти взаємодії геосистем та їх підсистем; еволюція геосистем є нелінійним процесом саморозвитку.

**4. Інформаційний підхід –** в основі всіх процесів у літосфері є інформаційний обмін між геосистемами та їх підсистемами (елементами), який узагальнює всі різноманітні процеси речовинного та енергетичного обміну.

**5. Історичний підхід** – дослідження розвитку геосистем у часі, тобто, дослідження процесів їх розвитку, а не тільки стану на сьогоднішній момент – виявлення часових трендів розвитку, відновлення історії розвитку тощо.

1. ***Поняття про ймовірність***

***Ймовірність*** – числова характеристика можливості того, що випадкова подія відбудеться в умовах, які можуть бути відтворені необмежену кількість разів.

***Довірча ймовірність*** – ймовірність накриття довірчим інтервалом істинного значення параметра. Довжина довірчого інтервалу пропорційна величині довірчої ймовірності. Найбільш вживані стандартні значення довірчої ймовірності наступні: Р = 0.9; 0.95; 0.99; 0.999.

Випадкову подію можна передбачити лише з деякою ймовірністю.

Ймовірність події А позначається Р(А). Якщо здійснювати різноманітні випробування, то можна констатувати, що різні випадкові події можуть мати різну можливість появи. Ймовірність неможливої події и дорівнює нулю, Р(Ц) = 0. Ймовірність достовірної події V дорівнює одиниці, Р(У) = 1.

Отже, ймовірність Р(А) будь-якої випадкової події А знаходиться між нулем і одиницею: 0<Р(А) <1. Інколи події можна вважати рівноможливими, якщо за умовами випробувань відсутні підстави вважати деякі з них більш можливими, аніж будь-які інші. Якщо декілька подій: 1) утворюють повну групу; 2) несумісні; 3) рі-вноможливі, то вони мають назву "випадки".

Звичайна скалярна величина характеризується тільки числовим значенням.

На відміну від неї випадкова величина має дві характеристики – числове значення і ймовірність цього значення. Головним атрибутом випадкової величини є **ймовірність – кількісна оцінка можливості наставання певної події в серії експериментів.**

Емпірична ймовірність визначається за результатами повторюваного багато разів експерименту як відношення частоти очікуваної події до кількості всіх повторень експерименту.

Онтологічна концепція ймовірності - ймовірність є об’єктивно існуючим атрибутом матерії незалежно від накопиченого знання.

Епістомологічна концепція – ймовірність є результатом неможливості врахування всіх факторів впливу на експеримент і тому є відображення недосконалості знань.

1. ***Особливості географічної інформації***

Види географічної інформації:

1. Вербальна (словесна) - за відсутності формалізованих і однозначно трактованих понять є неоднозначною, а також має інформаційну надлишковість.

2. Графічна – рисунки, схеми профілі, карти тощо. Особливе значення мають карти, але при використанні різних методів інтерполяції містять методичну похибку, яка заважає коректно виконувати порівняльний аналіз.

3. Кількісна – таблиці, матриці, відомості тощо, що містять кількісні оцінки різних властивостей (параметри) ознак геосистем. В залежності від суті ознаки використовуються такі кількісні шкали:

Для напівякісних (напівкількісних) оцінок:

- номінальна;

- порядкова;

Для кількісних оцінок:

- інтервальна;

- відносна.

**Лекція 3. Сутність одновимірних статистичних моделей.**

***Одновимірні статистичні моделі***

Містять одну випадкову величину, в географії це кількісні оцінки поодиноких ознак географічних об’єктів.

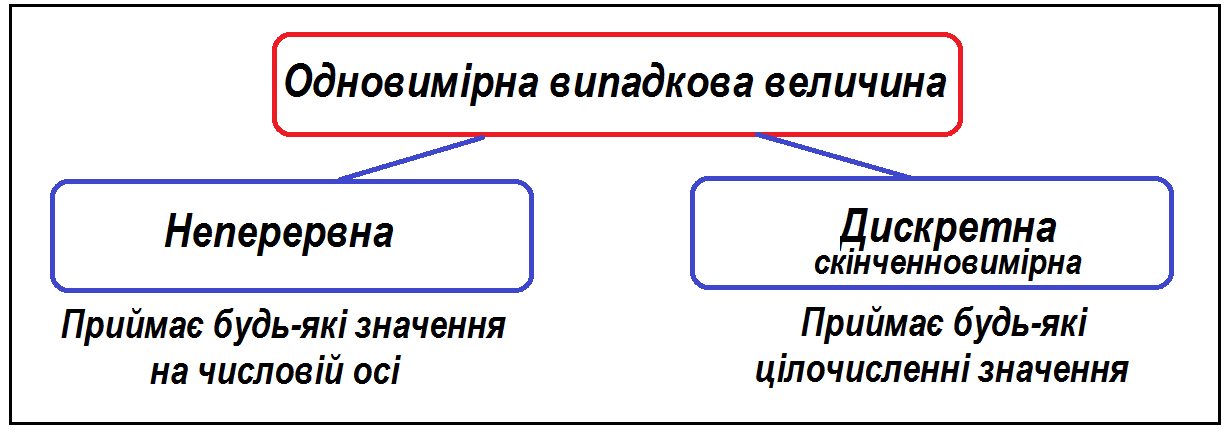


Рис. 4. Одновимірна випадкова величина

***Групування значень випадкової величини***

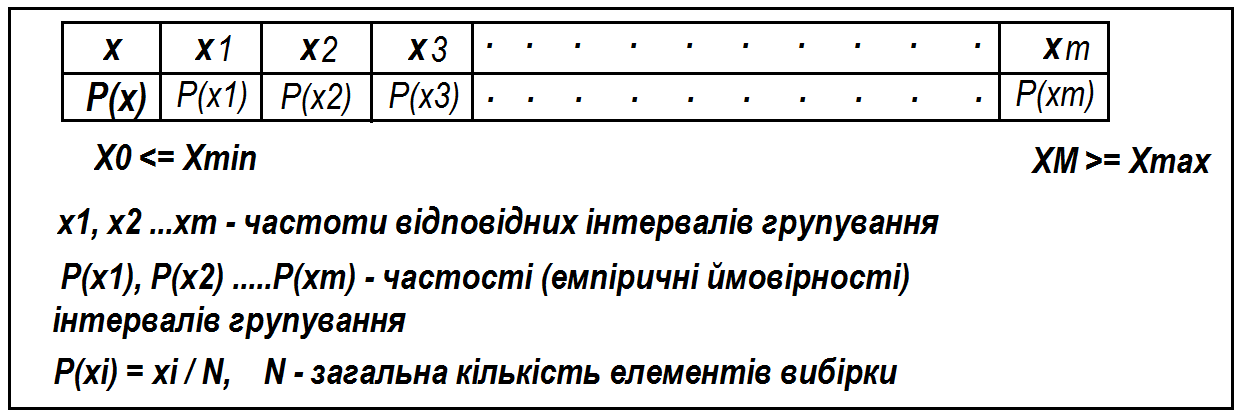


Рис. 5. Одновимірна випадкова величина

***Функція розподілу випадкової величини***

Випадкова величина має **дві характеристики**:

1. Числове значення Хі.
2. Ймовірність числового значення Р(Хі).

Функція розподілу відображає зв’язок між цими характеристиками:

Р(Хі) = f(Хі)

Розрізняють два різновиди функції розподілу:

- функція щільності розподілу;

- інтегральна функція розподілу.

***Функція щільності розподілу***

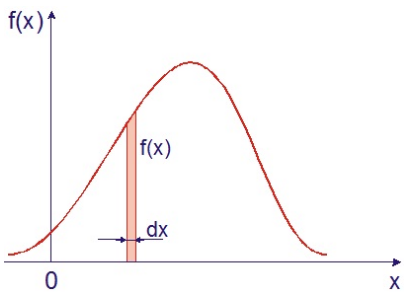


Рис. 6. Функція щільності розподілу

***Інтегральна функція розподілу***

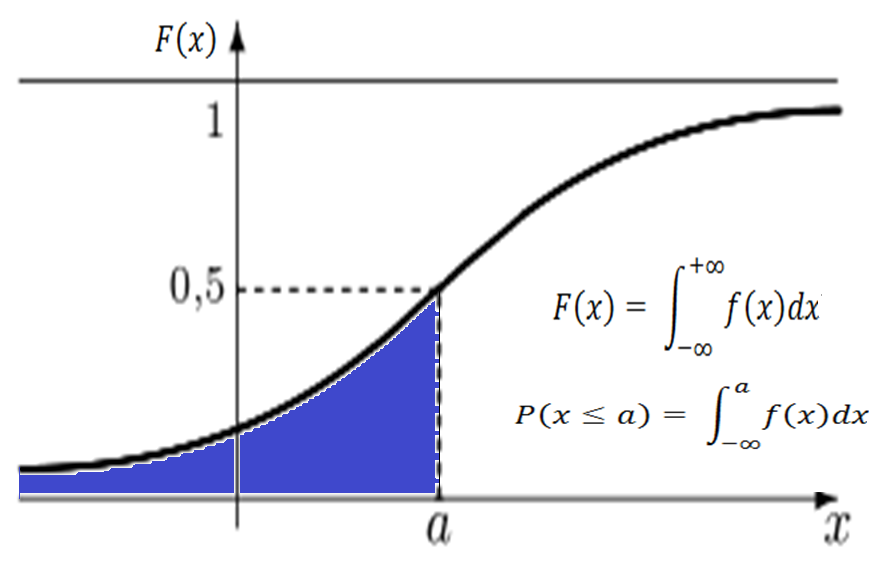


Рис. 7. Інтегральна функція розподілу

***Числові характеристики розподілу***

|  |  |
| --- | --- |
|  | **Центр розподілу**:  - математичне очікування;  - мода;  - медіана.  **2. Розсіювання:**  - дисперсія;  - середньоквадратичне відхилення;  - асиметрія;  - ексцес. |

Рис. 8. Числові характеристики розподілу

***Оцінки центру розподілу***

**Математичне очікування (Mx)**(середнє значення)

|  |  |  |
| --- | --- | --- |
|  |  |  |
|  |  |  |
|  |  |  |

***Оцінки центру розподілу***

**Мода (Mo) –** значення з найбільшою ймовірністю, якщо розподіл змішаний, може бути кілька модальних значень.

**Медіана (Me) -** середнє по порядку значення у ранжированій (впорядкованій за зростанням або зменшенням значень) сукупності.

***Центральні моменти розподілу***

Аналогія функції щільності розподілу з рівновагим станом механічної системи

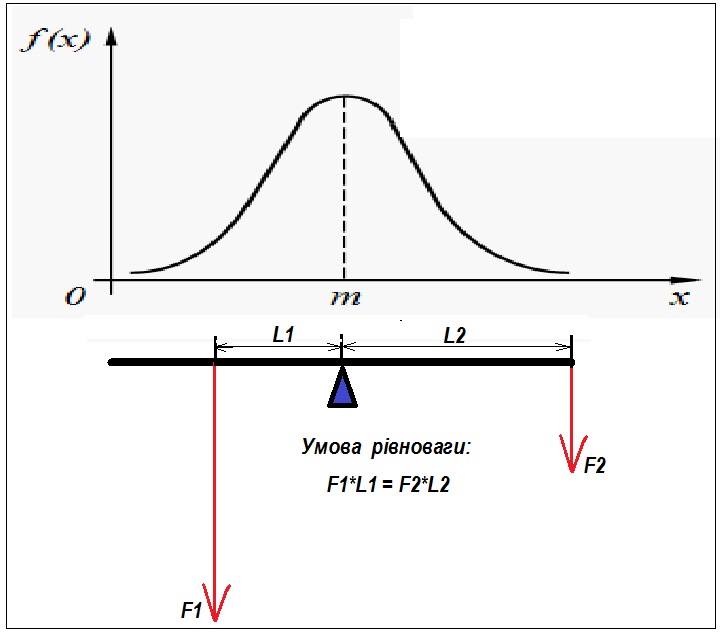
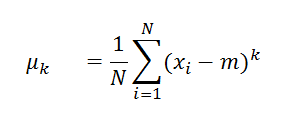


Рис. 9. Центральні моменти розподілу

***Центральні моменти розподілу***



***Кількісні оцінки розподілу***

Дисперсія (D) D = μ2

Середньоквадратичне відхилення (σ)

σ=

Коефіцієнт варіації (V) V = σ/m

Коефіцієнт асиметрії (As) As = μ3/

Коефіцієнт ексцесу (Е) E = μ4/ - 3

***Коефіцієнт асиметрії***

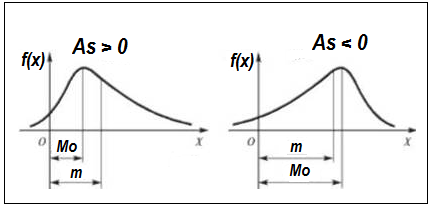


Рис. 10. Коефіцієнт асиметрії

***Коефіцієнт ексцесу***

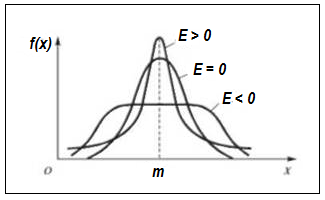


Рис. 11. Коефіцієнт асиметрії

***Нормальний закон розподілу***

|  |  |
| --- | --- |
|  | Характеристики:  M=Mo=Me  As = 0  E = 0 |

Рис. 12. Нормальний закон розподілу

***Логнормальний закон розподілу***

|  |  |
| --- | --- |
|  | Трансформація у нормальний закон через  логарифмування:  x = ln у |

Рис. 13. Логнормальний закон розподілу

***Правило “трьох сигм”***

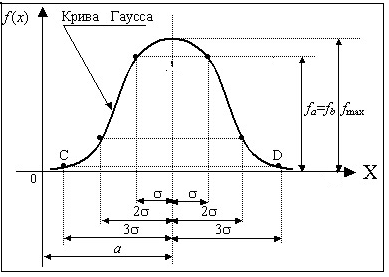


Рис. 14. Правило “трьох сигм”

***Функція Лапласа***

|  |  |
| --- | --- |
|  | Функція Лапласа дійсна для нормованої випадкової величини:  σ  Величина σ виступає як масштабний коефіцієнт числової осі |

Рис. 15. Функція Лапласа

***Візуалізація розподілу географічних даних***

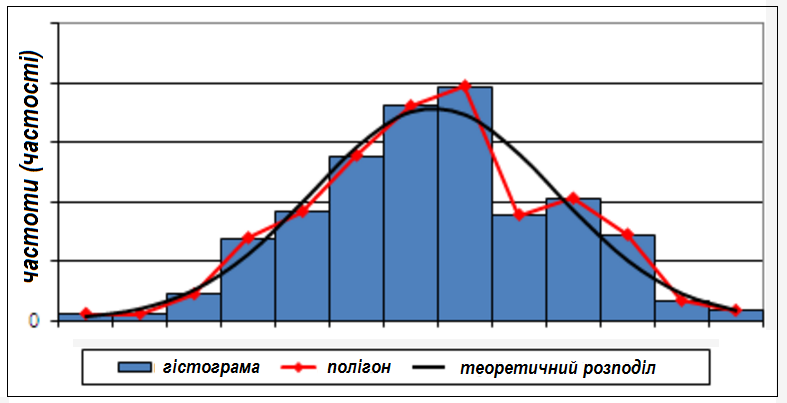
**

Рис. 16. Візуалізація розподілу географічних даних

**Лекція 4. Оцінка параметрів суспільно-географічних об’єктів**

***Вибірковий метод***

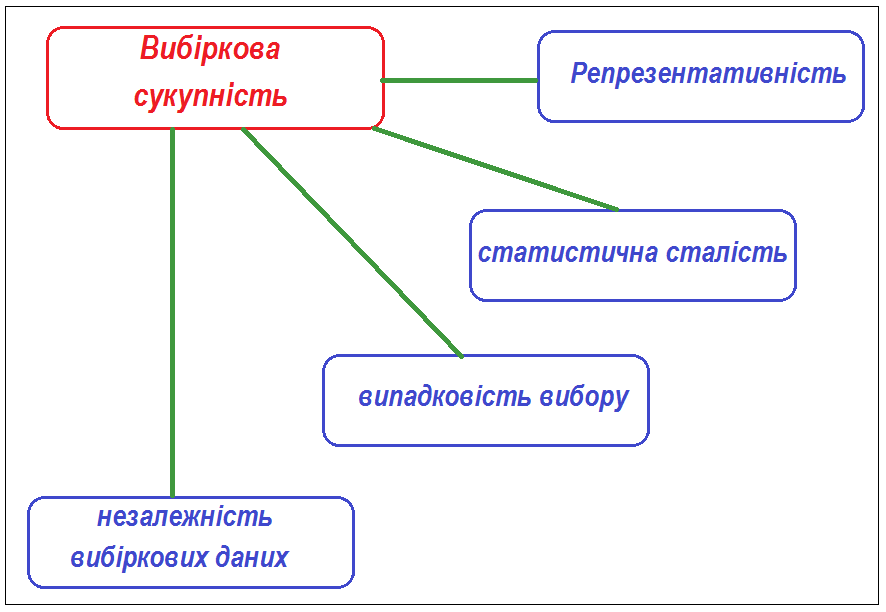


Рис. 17. Сутність вибіркового методу

***Оцінка параметрів географічних об’єктів***

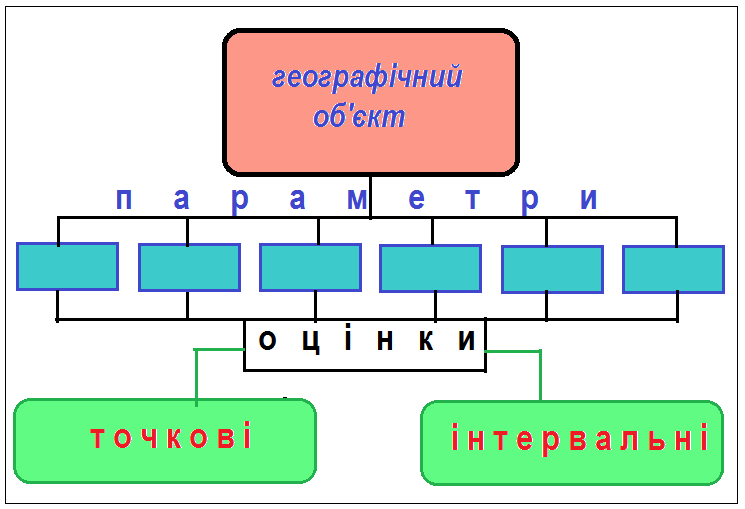


Рис. 18. Оцінка параметрів географічних об’єктів

***Вимоги до точкових оцінок***

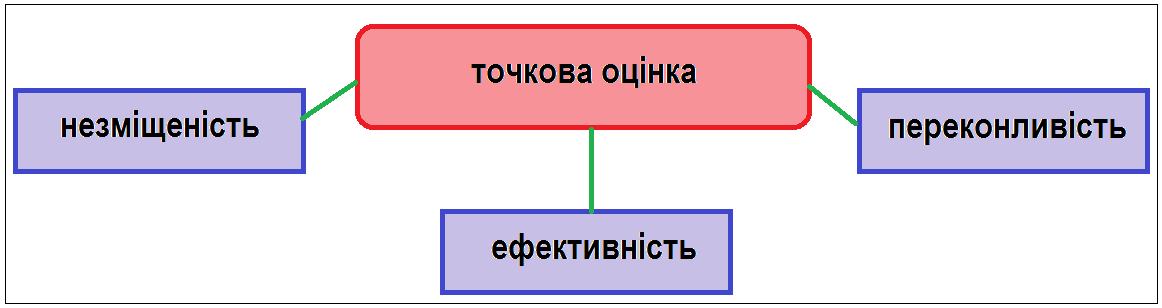


Рис. 18. Вимоги до точкових оцінок

Точність, надійність та достовірність точкових оцінок встановити неможливо

***Інтервальна оцінка***

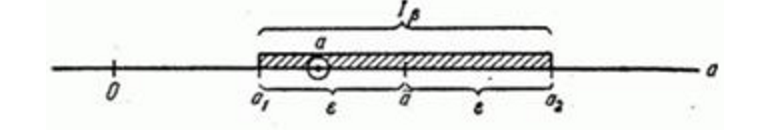
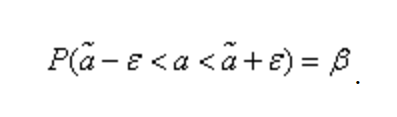


Рис. 19. Інтервальна оцінка



**Правило 3 сигм** – при довжині інтервалу:

**± 3σ** - покриває 99,7% можливих значень (можливість виходу за межі інтервалу 0,3%);

**± 2σ -**  покриває 95,5% (за межами 4,5%);

**± σ** - покриває 68,3% (за межами 31,7%).

Це правило дійсне для нормального закону розподілу випадкової величини, якщо емпіричний розподіл значимо відрізняється, випадкова величина перетворюється нелінійними перетвореннями так, щоб її розподіл відповідав нормальному закону.

***Перевірка статистичних гіпотез***

**Типи задач в географії, що зводяться до перевірки статистичних гіпотез:**

1. Рівність вибіркових середніх значень.

2. Рівність вибіркових дисперсій.

3. Однорідність вибіркової сукупності.

**Критерії перевірки:**

Параметричні

Непараметричні

***Перевірка статистичних гіпотез***

|  |  |  |
| --- | --- | --- |
| **Прийнята**  **гіпотеза** | **Вірна гіпотеза** | |
| Н0 | Н1 |
| Н0 | правильно | Помилка 2 роду  (ймовірність прийняття невірної гіпотези  β = 1 – α |
| Н1 | Помилка 1 роду  (рівень значимості α – ймовірність відхилення правильної гіпотези) | правильно |

***Метод порівняння довірчих інтервалів***

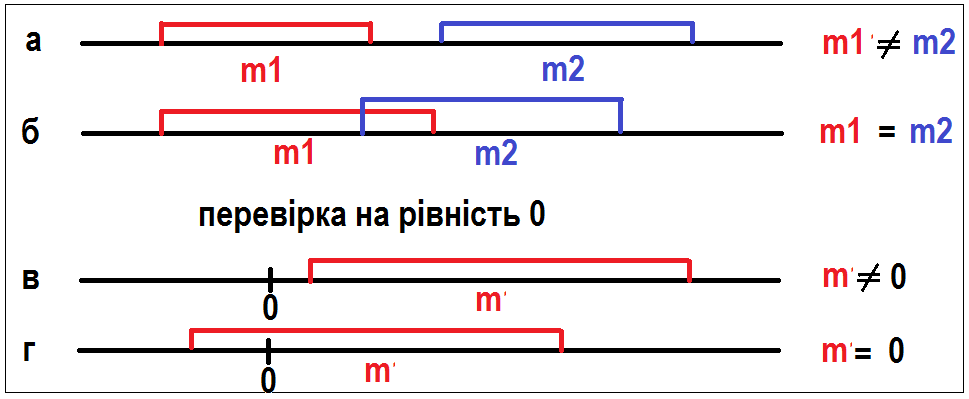


Рис. 19. Метод порівняння довірчих інтервалів

**Лекція 6. Двовимірні статистичні моделі**

***Двовимірні статистичні моделі***

Містять дві випадкові величини (компоненти)

**Два підходи**:

1. Дві одновимірні випадкові величини, які аналізуються сукупно:

|**Х**|= х1, х2, …….хm; |**Y**|= y1, y2, …….ym

2. Одна двовимірна випадкова величина, що містить дві компоненти:

z1 = x1,y1; z2 = x2,y2;………..zm = xm,ym

|**Z**|= z1, z2, …….zm = |**x,y**|

В залежності від типу складових двовимірні випадкові величини можуть бути:

* неперервними;
* дискретними;
* змішаними.

Задачі дослідження двовимірних випадкових величин (двовимірні статистичні моделі, двовимірний статистичний аналіз):

1. Встановити зв’язок між компонентами.
2. Якщо є, встановити форму зв’язку.
3. Описати лінійний або нелінійний зв’язок рівнянням.
4. Дати оцінку тісноти зв’язку.

Основний метод дослідження – **регресійно-кореляційний аналіз**

**Перевірка наявності зв’язку**

Форма хмаринки точок у кореляційному полі

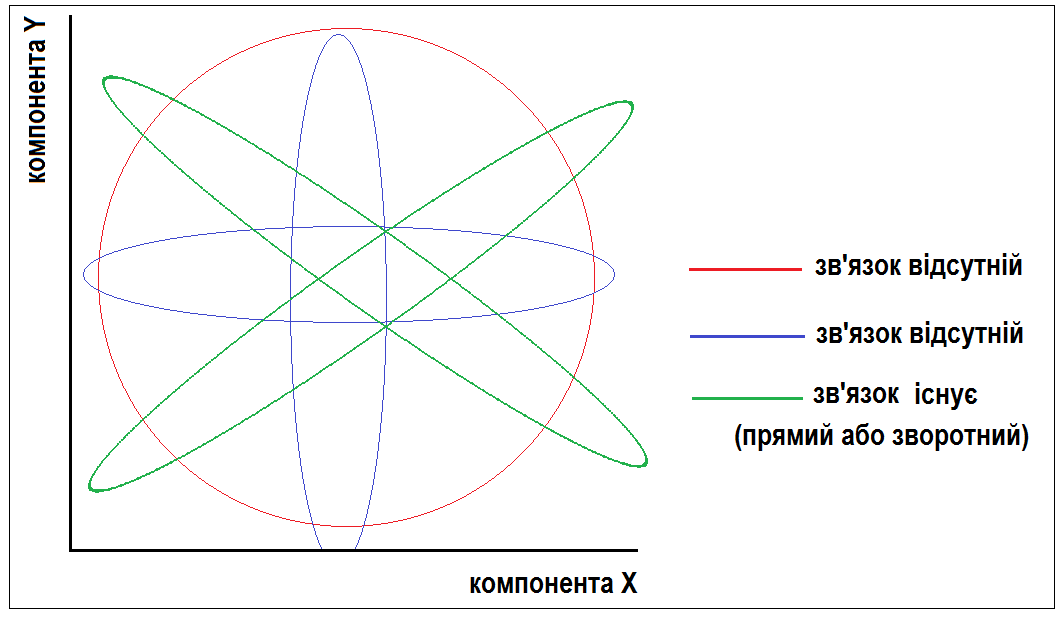
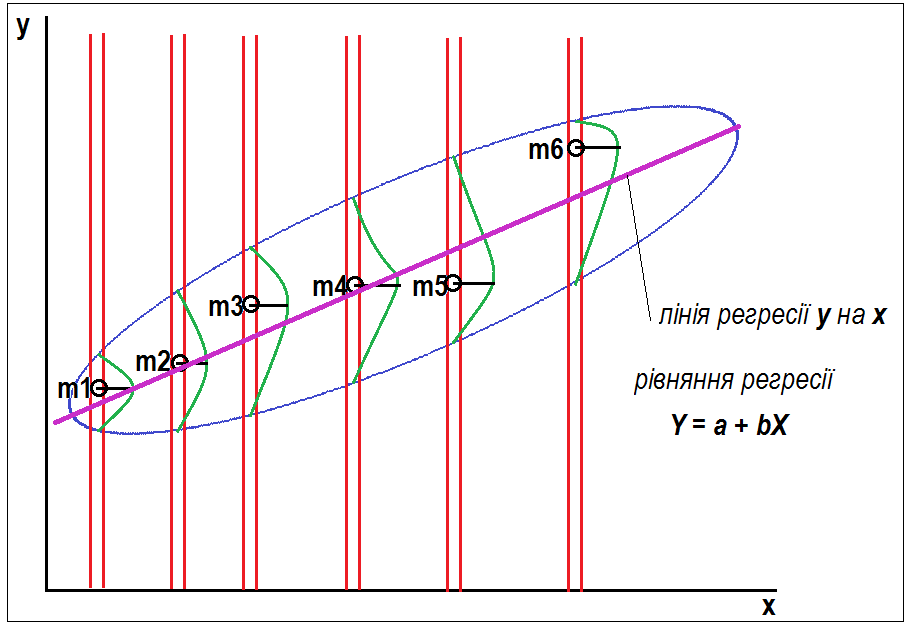


Рис. 20. Перевірка наявності зв’язку

**Лекція 7. Регресійно-кореляційний аналіз та його застосування.**

***Регресійний аналіз***

Формування та аналіз умовних сукупностей



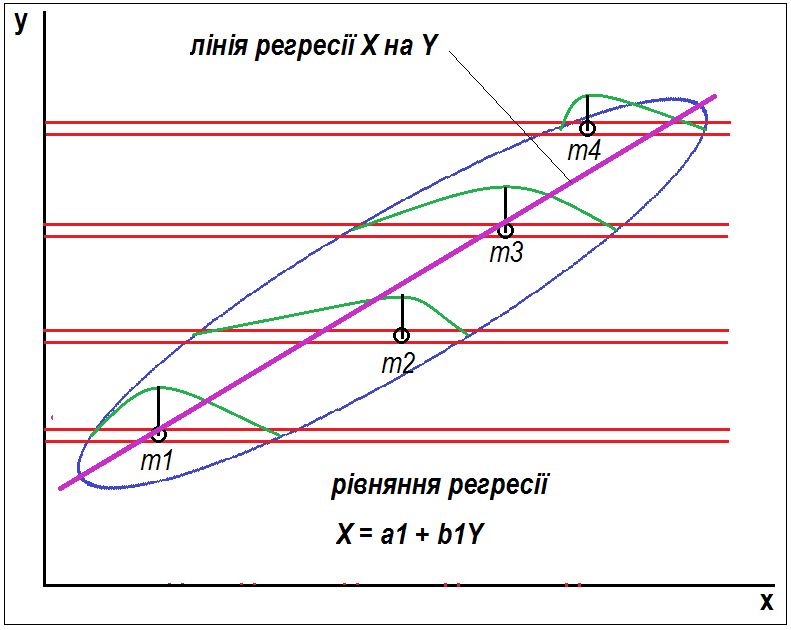


Рис. 21. Регресійний аналіз

***Нелінійні рівняння регресії***

* Y = a + bX + cX2 - другого ступеня
* Y = a + bX + cX2 + dX3 - третього ступеня
* Y = a + bX + cX2 + dX3 + eX4 - четвертого ступеня
* Y = a + bX + cX2 + dX3 + eX4 + fX5  п’ятого ступеня

Аналогічно записуються нелінійні рівняння регресії

Х на У, але коефіцієнти рівнянь мають індекси 1 (наприклад, a1, b1, c1, d1, e1, f1)

***Довірчий інтервал лінії регресії***

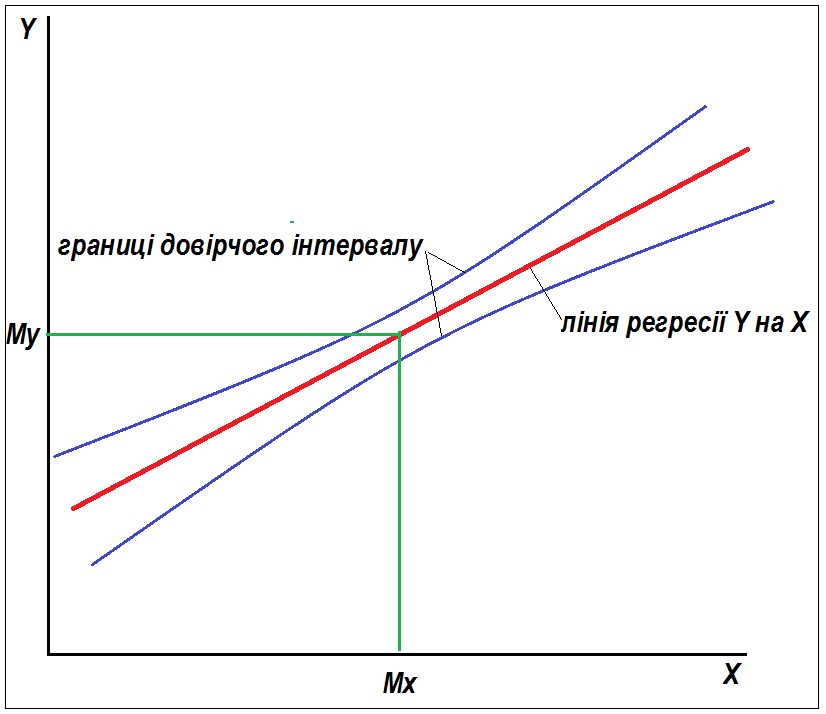


Рис. 22. Довірчий інтервал лінії регресії

***Поняття про кореляцію***

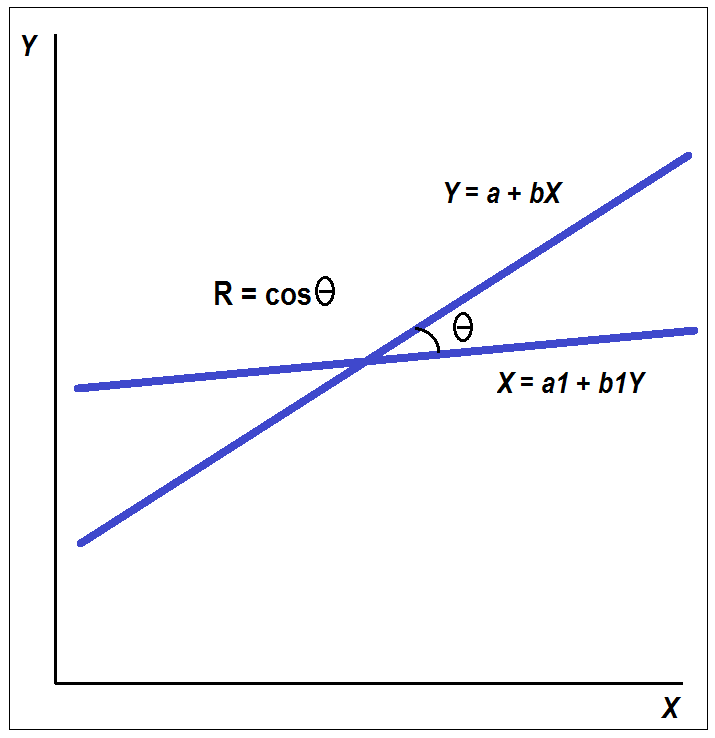


Рис. 22. Поняття про кореляцію

***Коефіцієнт кореляції***

Коефіцієнт кореляції – показник тісноти зв’язку x та y:

-1

***Інтерпретація коефіцієнту кореляції***

1. r = 0 – довірчий інтервал для r покриває значення 0 - кореляційний зв’язок відсутній.

2. |r|= 1 – зв’язок між X та Y функціональний (однозначний).

3. r < 0 – зв’язок зворотний.

4. r > 0 - зв’язок прямий.

r може розглядатися як показник сили зв’язку між X та Y лише при виконанні таких умов:

* розподіл X та Y відповідає нормальному закону;
* зв’язок між X та Y лінійний.

Для дослідження нелінійних зв’язків застосовується інший показник – кореляційне відношення.

***Застосування регресійно-кореляційного аналізу***

1. Дослідження взаємозв’язку між різними властивостями соціогеосистем.
2. Опис зв’язку між різними властивостями соціогеосистем.
3. Узагальнення накопиченого фактичного матеріалу, виявлення та опис нових закономірностей функціонування соціогеосистем.
4. Побудова емпіричних та теоретичних моделей соціогеосистем.

***Використання графів для інтерпретації кореляційної матриці***

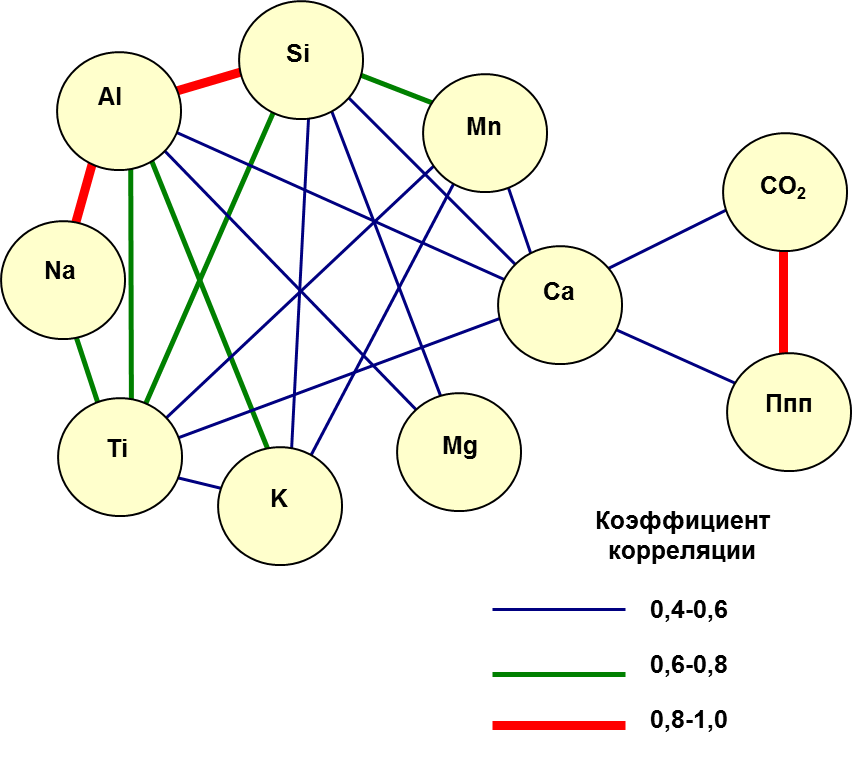


Рис. 23. Використання графів для інтерпретації кореляційної матриці

**Лекція 8. Багатовимірний статистичний аналіз**

***Багатовимірні статистичні моделі***

Багатовимірні статистичні моделі містять більше двох випадкових величин. Вони є розвитком двовимірних статистичних моделей і утворені сукупністю спряжених випадкових величин (випадкових багатовимірних випадкових векторів), які складають матрицю вихідних даних (ВД) розміром *n*×*k* *,* де *n –* число спостережень (випадкових векторів – об’єктів, *k -* число ознак (параметрів об’єктів):



***Властивості матриці ВД***

Вважається, що спостереження (рядки матриці ВД) є незалежними (одна з вимог вибіркового методу), тобто, їх можна переставляти в матриці будь-яким способом.

Стовпчики матриці (ознаки, параметри) можуть бути взаємозалежними. Багатовимірний аналіз застосовується як для спостережень, так і для ознак (параметрів)

***Багатовимірний кореляційний аналіз***

Для дослідження зв’язків між параметрами об’єктів використовуються такі показники:

1. Парний коефіцієнт кореляції (як у двовимірному аналізі).
2. Множинний коефіцієнт кореляції.
3. Частковий коефіцієнт кореляції, який відображає зв’язок між двома ознаками без впливу інших ознак.

Для відображення попарних зв’язків між ознаками розраховується матриця парних коефіцієнтів кореляції, яка є квадратною і симетричною відносно головної діагоналі:



***Багатовимірний регресійний аналіз***

Багатофакторна залежність описується рівнянням множинної лінійної (або нелінійної) регресії:

*y = a*1*x*1 + *a*2*x*2 + ⋅⋅⋅ + *akxk* + *b,* де

*a* – часткові коефіцієнти регресії;

*b -* вільний член;

*y -* залежна ознака;

*x*1, *x*2, ⋅⋅⋅ *xk* - незалежні ознаки.

Коефіцієнти рівняння регресії знаходять методом найменших квадратів (МНК), або через вибіркові статистичні параметри.

***Оцінка коефіцієнтів рівняння множинної регресії через вибіркові статистики***

Для знаходження коефіцієнтів рівняння множинної регресії вибіркові дані нормують , тобто, заміняють величинами:

У такому випадку рівняння множинної регресії має вигляд:

де *А* – нормовані коефіцієнти регресії

Якщо підставити нормовані значення, то отримаємо часто вживане рівняння множинної регресії у такому вигляді:



коефіцієнти рівняння можна знайти із системи лінійних рівнянь, побудованої із коефіцієнтів кореляції



***Аналіз точності рівняння множинної регресії***

Відхилення *δ* від рівняння множинної регресії отримують порівнянням фактичних *у* і розрахункових *у*р значень:

Далі розраховується дисперсія фактичних значень *у*

(σy2 ) та відхилень (σδ2 ), за якими визначається множинний коефіцієнт кореляції:



Похибка рівняння множинної регресії дорівнює *tσ*δ , де t залежить від довірчої ймовірності

**Лекція 9. Методи багатовимірного статистичного аналізу, особливості їх застосування.**

***Задачі розпізнавання образів***

Основна мета - порівняння нового об’єкту з існуючими прототипами (еталонами) для того, щоб перенести на нього властивості та ознаки добре вивченого прототипу. Для цього використовується багатовимірний простір, утворений множиною ознак (осей простору).

Формула Евкліда для двовимірного простору (площини):

формально поширюється на простір будь-якої вимірності:

Де К – вимірність багатовимірного простору.

***Властивості багатовимірного простору***

Кожний прототип займає у багатовимірному просторі свою «комірку» – власну область прототипу. Комірки (класи багатовимірної класифікації) розділяються гіперповерхня-ми і вони можуть не всі бути заповнені прототипами.

Для визначення належності контрольного об’єкту до певного класу по кожній координаті розраховується вирішальна функція , яка має екстремальне значення в середині класу. Також встановлюється вирішальне правило, за яким, якщо контрольний об’єкт потрапляє на границю класів, він відноситься або до старшого, або до молодшого класу.

Відстань Евкліда у багатовимірному просторі характеризує ступінь подібності або відмінності контрольного об’єкту відносно існуючих прототипів.

***Загальний алгоритм розпізнавання образів***

1. Вибір прототипів – еталонних об’єктів.
2. Вибір ознак (параметрів) для порівняння об’єктів.
3. Вибір методу розпізнавання (зокрема, вирішальної функції).
4. Вибір вирішального правила.
5. Класифікація (розпізнавання) контрольного об’єкту.

***Дискримінантний аналіз***

**Дискримінантний аналіз** – один з методів розпізнавання образів. За його допомогою вирішуються такі задачі:

1. Інтерпретація різниці між класами – наскільки ефективно відрізняти класи за даним набором ознак; які з ознак найбільш значущі для впевненого розрізнення класів.
2. Кваліфікація об’єктів – як віднести кожний об’єкт до одного з класів за найкращим набором ознак.

***Кластерний аналіз***

**Кластерний аналіз** – метод багатовимірного багаторівневого (ієрархічного) покрокового групування (класифікації) вихідної множини об’єктів (або ознак). Початкова ідея методу – по крокам знаходження у матриці парних коефіцієнтів кореляції блоків з найбільш тісним кореляційним зв’язком. Потім у якості дистанційних коефіцієнтів у багатовимірному просторі замість коефіцінту кореляції стали використовувати інші показники, зокрема відстань Евкліда, Манхеттенську відстань тощо.

***Дендрограма***

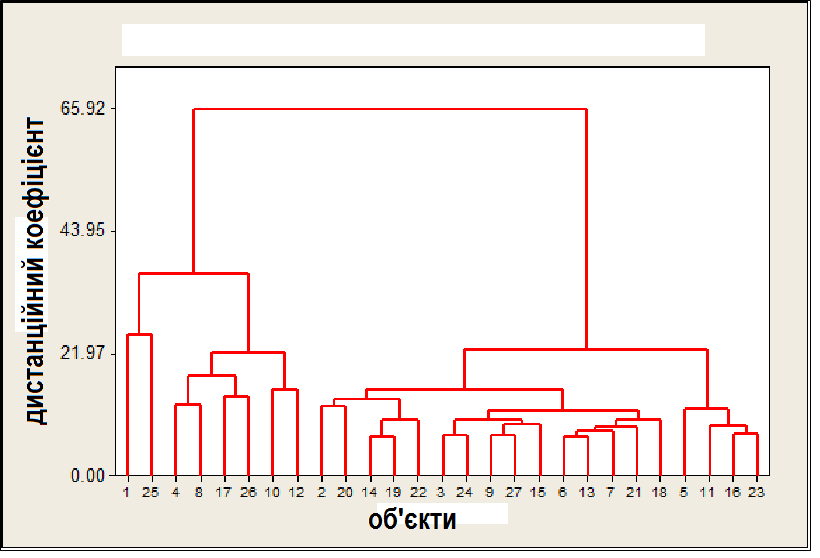


Рис. 24. Кластерна дендрограма

***Недоліки кластерного аналізу***

* В залежності від вибраного методу групування і визначення дистанційного коефіцієнту має неоднозначні результати;
* Не визначаються чіткі кількісні критерії виділення груп;
* Часто використовується в пошуковому сенсі для початкового уявлення про можливе групування об’єктів (для уточнення необхідно використовувати більш строгі методи).

***Факторний аналіз (ФА)***

**Факторний аналіз -** група методів для дослідження взаємовідношень в заданій множині об’єктів або ознак. Їх загальна властивість – наявність гіпотетичних ознак (об’єктів), які називаються факторами і містять всю суттєву інформацію, що зберігається у великій множині заданих параметрів чи об’єктів.

Наявність сильних внутрішніх взаємозв’язків в матриці ВД дає можливість без істотної втрати інформації замінити p заданих параметрів k гіпотетичними параметрами як комбінаціями заданих параметрів (k<p). У такому розумінні ФА є ефективним методом згортки (представлення у більш компактній формі) інформації, що й було його первинним завданням. Але виявилося, що семантичний аналіз гіпотетичних факторів, як побічний ефект, дає значно більше корисної інформації, ніж згортка інформації.

Це перетворило ФА в дуже ефективний інструмент дослідження інформації, що міститься в матриці ВД.

***Згортка інформації***

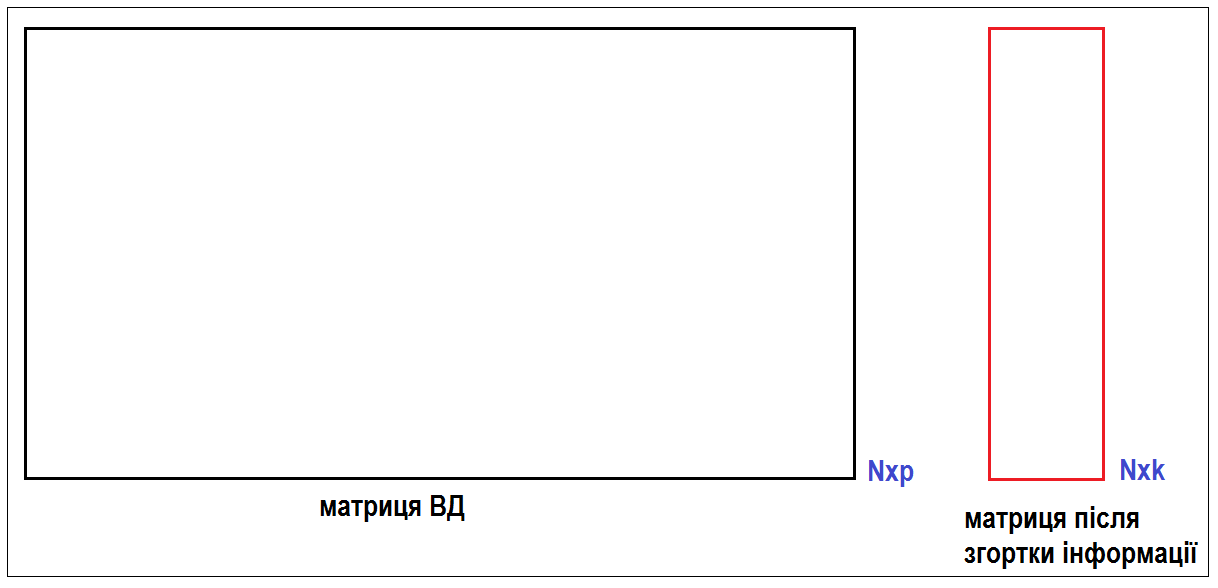


Рис. 25. Згортка інформації при факторному аналізі

***Модифікації факторного аналізу***

Вусі методи факторного аналізу поділяються на дві великі групи:

1. **R – модифікація** – аналіз взаємозв’язків між параметрами (стовпчиками матриці ВД). Далі будемо розглядати переважно цю модифікацію ФА
2. **Q – модифікація** - аналіз взаємозв’язків між об’єктами (рядками матриці ВД)

***Використання ФА***

1. **Описовий варіант** – відображення аналізу взаємозв’язків між даними однієї сукупності, яка не розглядається як вибірка і результати не розповсюджуються на генеральну сукупність. Це детермінований випадок.
2. **Модельний варіант** – досліджувана сукупність розглядається як вибірка з генеральної сукупності і результати розповсюджуються на неї. Це стохастичний випадок.

***R – модифікація факторного аналізу***

Модель ФА можна розглядати на основі 2 підходів:

* в термінах натуральних параметрів;
* в термінах дисперсії та коваріації натуральних параметрів.

Ці підходи висвітлюють різні особливості матриці ВД.

***ФА в термінах натуральних параметрів***

**Основне рівняння ФА** в матричній формі має вигляд:

де - матриця ВД;

- матриця значень факторів;

- матриця значень факторних навантажень;

- матриця відхилень («відгуків»);

*k* – число гіпотетичних факторів.

**Рівняння ФА**

Для будь-якого елементу матриці маємо:

де *n* – номер рядка матриці;

– номер стовпчика матриці;

– індекс фактора;

– кількість факторів.

У матричній формі:

***Графічна інтерпретація рівняння факторного аналізу***

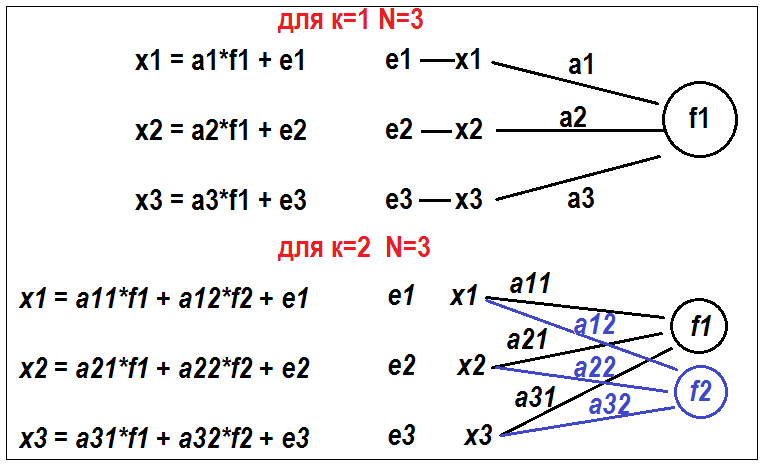


Рис. 25. Графічна інтерпретація рівняння факторного аналізу

***Змістовна інтерпретація рівняння ФА***

Рівняння ФА для будь-якого спостереження можна записати так:

або

Це звичайне рівняння множинної регресії

Рівняння множинної регресії можна записати таким чином:

- загальна частина змінної , яка корелює з іншими змінними;

- похибка вимірювання;

специфічна частина змінної , яка не корелює з іншими змінними;

***R – модифікація в термінах дисперсії***

Розглянемо останнє рівняння ФА:

**=**

В термінах дисперсії воно має вигляд:

***+= + +***

та наступну інтерпретацію:

- загальність, яка пов’язана з іншими змінними і входить в коваріацію;

- остаточна дисперсія (унікальність) , яка не пов’язана з іншими змінними і не входить в коваріацію;

- специфічність;

- дисперсія , зумовлена похибкою вимірювання.

**Лекція 10. Просторові змінні як об’єкт геоінформатики.**

***Поняття просторової змінної***

**Просторові змінні** – новий клас математичних об’єктів, в якому обов’язково є прив’язка до певної точки простору. Вони є функціями координат простору, але зберігають основні властивості випадкової величини. Їхні специфічні характеристики:

* мають розмірність (результати виміру);
* мають область існування (границі географічних об’єктів);
* мають область впливу (область визначення).

***Структура геопросторових даних***

Загальна структура геопросторових даних

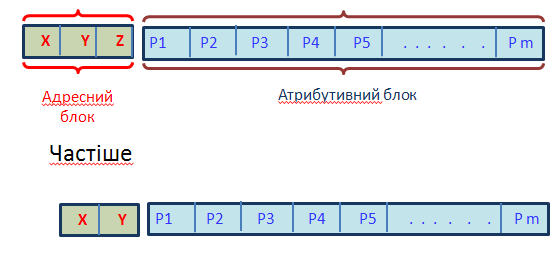


Рис. 25. Структура геопросторових даних

Координатою Z є узагальнений параметр Рзаг

**ГО як поля просторових змінних**

***Поля просторових змінних*** описують тенденції мінливості властивостей (ознак) географічних об’єктів, тобто їхню неоднорідність. Кожній точці простору відповідає певне значення ознаки ГО.

**Випадкова неоднорідність** – значення ознаки ГО в різних точках простору не корелюють (незалежні), тобто мінливість випадкова. Описується статистичними моделями.

**Закономірна неоднорідність** – значення ознаки ГО залежать від координат простору. Описується моделями просторових змінних.

***Моделі статистичні і просторових змінних***

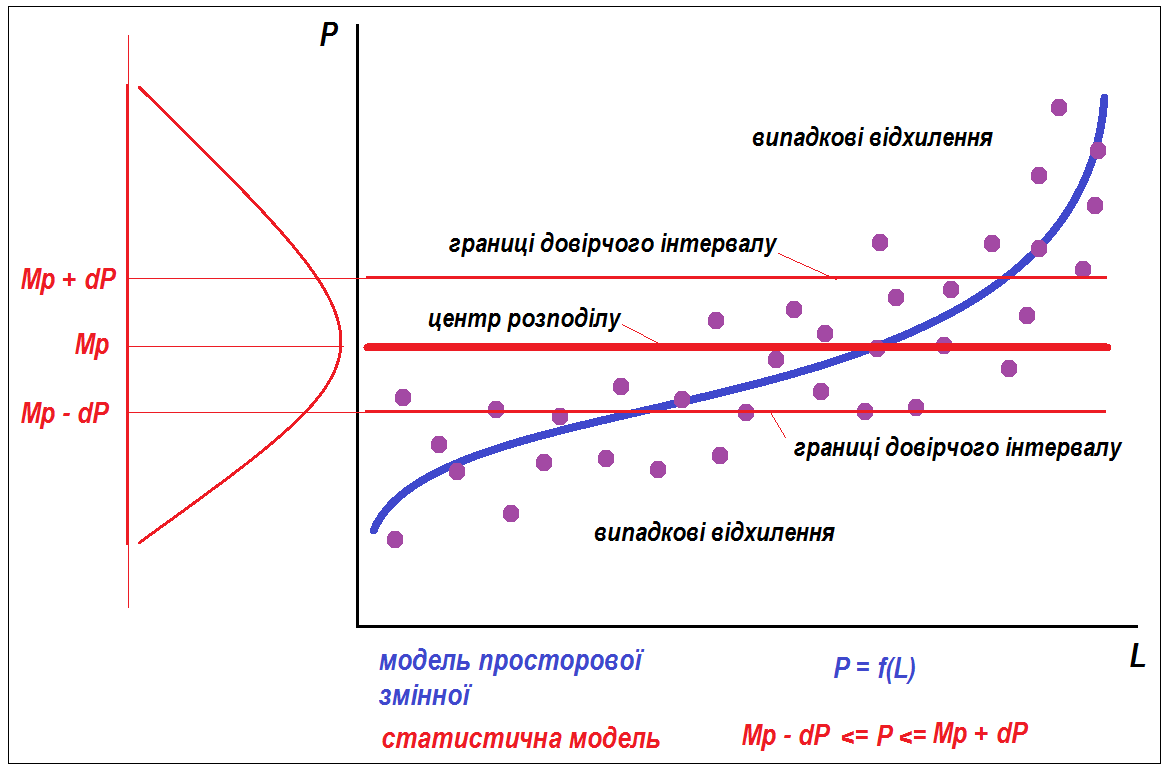


Рис. 26. Моделі статистичні і просторових змінних

***Анізотропія ГО***

**Анізотропія** – залежність значення ознаки ГО від напряму у полі просторової змінної.

**Коефіцієнт анізотропії** – відношення значень ознаки в різних напрямах:

Ка = PLx/Ply

Приклад: вплив транспортної інфраструктури на доступність населених пунктів

***Типи полів ознак ГО***

1. За природою – геодемографічні, геоекологічні, економічні, інноваційні тощо.
2. За мірністю – 1, 2, 3-вимірні, багатовимірні ознакові.
3. За характером представлення – скалярні та векторні.
4. За особливістю виміру - неперервні та дискретні.

***Неоднорідність полів ознак ГО***

Географ завжди прагне за вимірами ознак ГО в різних точках поля інтуїтивно вбачати певні просторові закономірності (структури). Але похибки вимірів, особливості формування полів ознак ГО зумовлюють принципово випадковий характер кількісних оцінок ознак. Тому їх можна розглядати як вектори, що мають у першому наближенні три компоненти: закономірну (відображає структрні особливості ГО), випадкову (відображає принципово випадковий характер полів) і похибку виміру (відображає недосконалість методів і засобів виміру).

***Оптимізація кроку опробування***

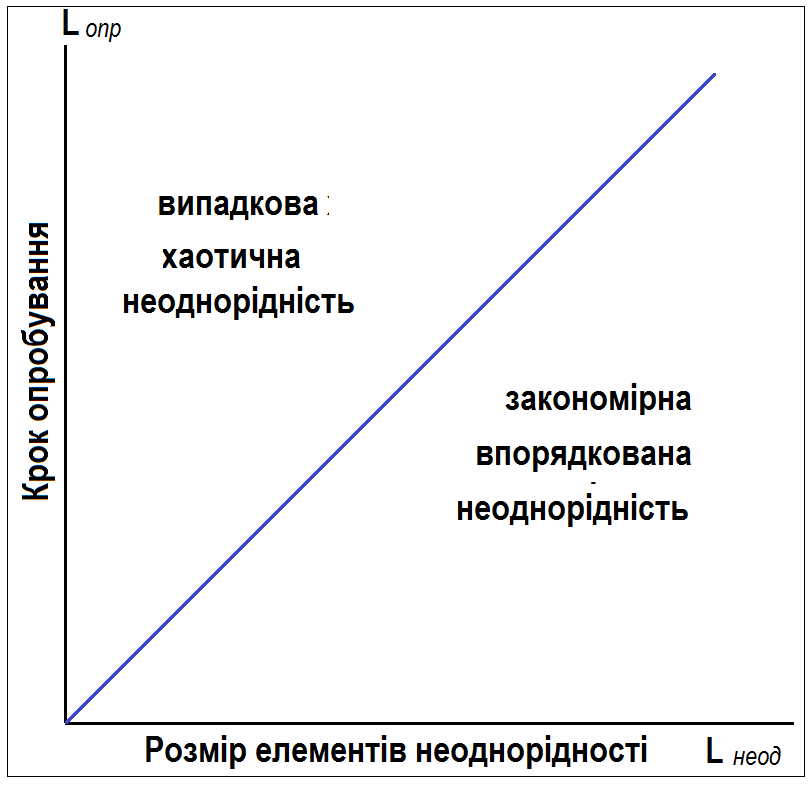


Рис. 28. Моделі статистичні і просторових змінних

***Масштабні ефекти (МЕ)***

Масштабні ефекти – залежність певного статистичного показника від співвідношення розміру елементів неоднорідності і визначальної області ознаки об’єкту.

Розрізняють МЕ трьох родів:

1. МЕ ***І роду*** - мінливість математичного очікування (середнього) – зникає при досягненні елементарного об’єму.
2. МЕ ***ІІ роду*** - мінливість асиметрії.
3. МЕ ***ІІІ роду*** – мінливість ексцесу.

***Масштабний ефект І роду***

Різні розміри визначальної області: R1,R2,R3.

R4 і більші – оптимальні.

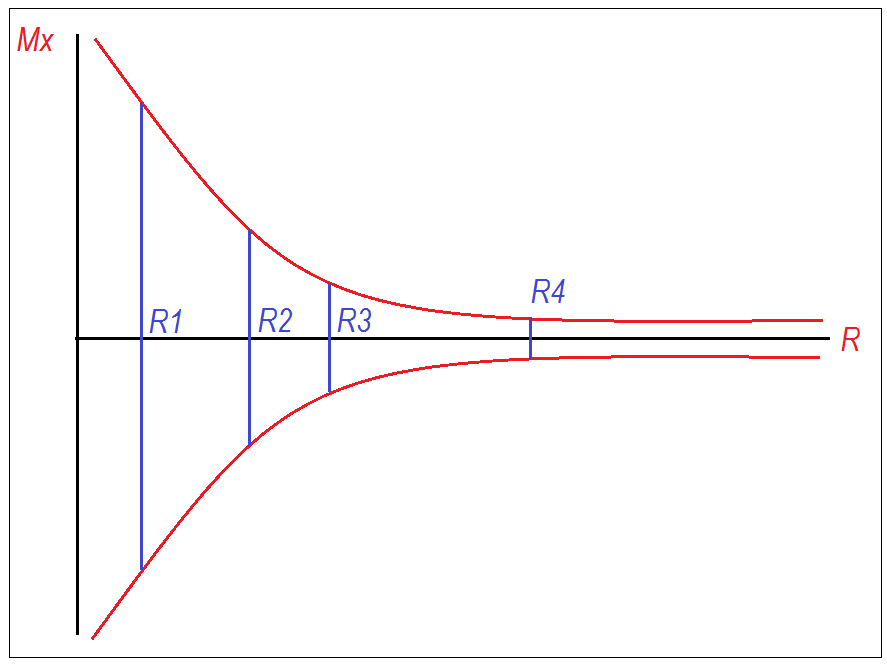


Рис. 29. Моделі статистичні і просторових змінних

**Лекція 11. Методи розділення фону та аномалій. Методи розрахунку локальних показників**

***Співвідношення випадкової та закономірної складових поля ГО***

|  |  |
| --- | --- |
|  | **Модель поля ознаки ГО**:  P=f(x,y) = dPзкнм + dPвпд + dPвим  Похибка виміру залежить від методу і технічного засобу виміру.  Границя між випадковою та закономірною складовими поля ознаки ГО залежить від співвідношення розмірів елементів неоднорідності ГО і області визначення просторової змінної.  Прийнята термінологія: закономірна складова – **ФОН**. Випадкова складова - **АНОМАЛІЯ** |

Рис. 29. Співвідношення випадкової та закономірної складових поля ГО

***Основні визначення неоднорідності***

**Однорідність ГО** за ознакою Р – незалежність Р від координат простору, або сталість Р в межах ГО.

**Неоднорідність ГО** за ознакою Р – залежність Р від координат простору, або мінливість Р в межах ГО.

**Визначальна область експерименту** (область визначення просторової змінної) – частина простору ГО, властивості якої визначають результат експерименту і границю між закономірною та випадковою складовими.

**Елементарний об’єм** – віртуальна частина визначальної області (як мінімум на порядок менша) в межах якої значення ознаки статистично стійке.

***Діаграма структурної неоднорідності ГО***

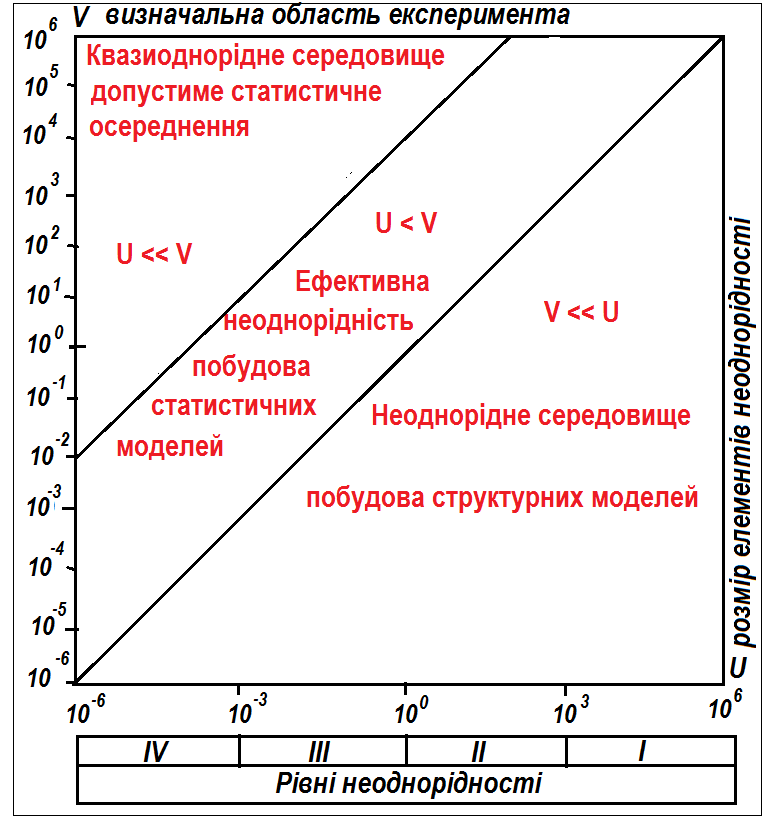


Рис. 29. Діаграма структурної неоднорідності ГО

***Класифікація структурної неоднорідності ГО***

|  |  |  |
| --- | --- | --- |
| **№** | **Характеристика неоднорідності** | **U/V** |
| 1 | ***Ультранеоднорідність –*** *середовище розглядається як однорідне із середніми значеннями параметрів* | **< 0,1** |
| 2 | ***Мікронеоднорідність –*** *(ефективна неоднорідність) проявляється як хаотична, тому використовуються статистичні моделі* | **0,1 – 1,0** |
| 3 | ***Макронеоднорідність –*** *систематична (закономірна) мінливість, завуальована випадковими флуктуаціями, застосовуються структурні моделі (просторових змінних)* | **> 1,0** |

***Тренд - аналіз***

**Тренд-аналіз** – група методів для виділення фону, тобто розкладання значення ознаки на закономірну та випадкову складові.

Використовуються два підходи:

1. **Зглажування поля** вихідних даних ковзними статистичними вікнами і розрахунок локальних середніх.
2. **Апроксимація поля** єдиною функцією координат простору.

**Властивості методів зглажування поля**

|  |  |
| --- | --- |
|  | 1. Краще враховують локальні закономірності. 2. Варіабельність результату при зміні радіусу вікна і методу обчислення локального середнього. 3. Серія карт, побудованих за різними варіантами дає можливість аналізувати особливості поля на різних рівнях узагальнення. |

Рис. 30. Властивості методів зглажування поля

***Приклад використання середнього гармонійного***

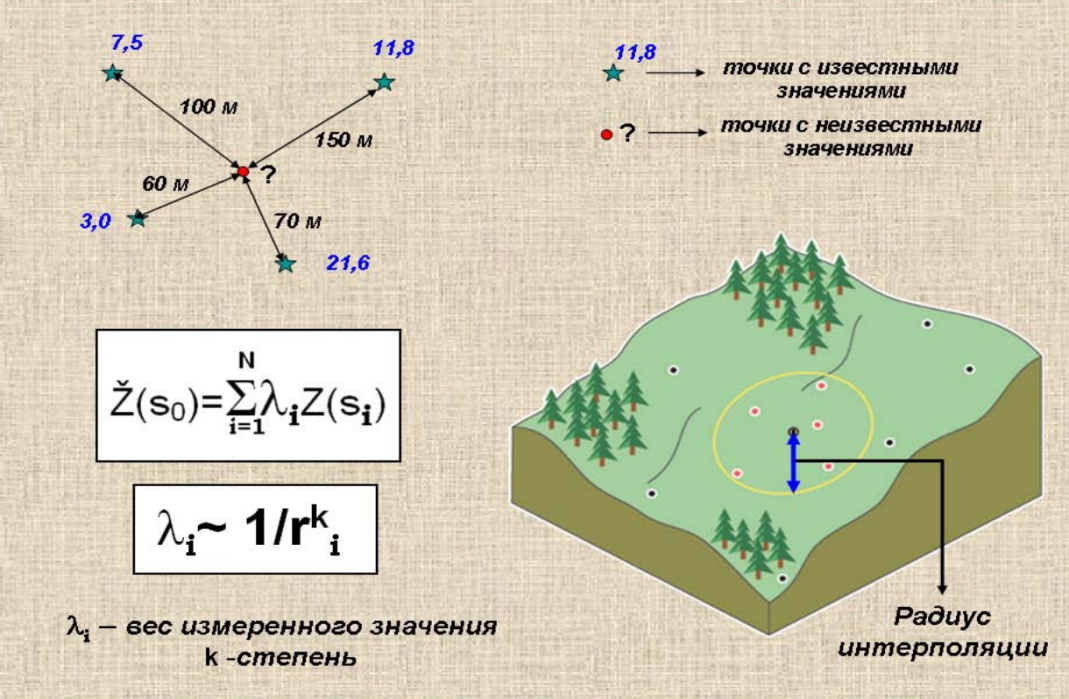


Рис. 30. Приклад використання середнього гармонійного

***Інтерпретація локального середнього в ГІС-технології***



**Лекція 12. Апроксимація поверхні тренда єдиною функцією координат простору.**

***Апроксимація поля єдиною функцією координат простору***

Отримується «жорстка» поверхня, яка відображає основні регіональні тенденції мінливості поля ознаки ГО. Локальні особливості не враховуються.

Для побудови таких моделей частіше всього використовуються алгебраїчні та тригонометричні поліноми, рівняння Лапласа.

F(x,y) = a0 + a1X + a2Y – лінійний поліном, що описує площину

F(x,y) = a0 + a1X + a2Y +a3XX + a4YY + a5XY

Це тренд-поверхня другого порядку



***Статистична поверхня: тренд–модель***

