Міністерство освіти і науки України
Львівський національний університет імені Івана Франка

В.С.ГРИЦЕВИЧ

МАТЕМАТИЧНІ МЕТОДИ В СУСПІЛЬНІЙ ГЕОГРАФІЇ

Тексти лекцій для студентів заочної форми навчання

Львів 2013

Рецензенти:

докт. геогр. наук, проф. *М.С.Дністрянський*
(Львівський національний університет імені Івана Франка)
канд. геогр. наук. доц. *Я.Б.Хомин*
(Львівський національний університет імені Івана Франка)

Науковий редактор д.г.н. професор Шаблій О.І.

Рекомендовано до друку
Вченою радою географічного факультету
Львівського національного університету імені Івана Франка.
Протокол №3 від 17.04.2013

Грицевич В.С.

Математичні методи в суспільній географії: тексти лекцій для студентів заочної форми навчання. –Львів: Малий видавничий центр. Лабораторія тематичного картографування географічного факультету, 2013. -48 с.

У тексті лекцій подані теоретичні основи математичних методів, які потрібні студентам для обробки даних і отримання нової інформації при виконанні суспільно-географічних досліджень, зокрема при написанні курсових і дипломних робіт.

Для бакалаврів заочної форми навчання спеціальності “географія”.

Структура навчальної дисципліни

Назви змістових модулів і тем	Заочна форма навчання					
	Усього годин	у тому числі				
		л	п	лаб	інд	ср
1	2	3	4	5	6	7
Тема 1. Головні поняття математичної логіки	18	2				14
Тема 2. Математико-логічний аналіз понять	18	2				14
Тема 3. Теорія графів у суспільній географії	18	2	2			20
Тема 4. Географічні поля та їхня кореляція	18	2				14
Тема 5. Взаємна та просторова регресія географічних полів	18	2	2			20
Тема 6. Багатовимірна таксономізація в суспільній географії	24	2				14
Тема 7. Кількісні методи часового прогнозування.	30	2				30
Усього годин	144	14	4			126

Тема 1
ГОЛОВНІ ПОНЯТТЯ МАТЕМАТИЧНОЇ ЛОГІКИ

План:
1. Висловлювання та їхня істинність
2. Математико-логічні операції
3. Предикати

1. Висловлювання та їхня істинність.

Людська мова організована у формі речень. З їхньою допомогою науковець виконує змістовний аналіз, формулює отримані результати. Серед багатоманіття речень нашої мови особливе значення для науки має такий широкий їх клас як *висловлювання*.

Висловлювання – це речення, істинність якого можна встановити. Розглянемо з цього приводу такі групи прикладів.

1. Речення, які не є висловлюваннями.

— Давайте гарну пісню заспіваєм.

— Як тебе не любити, Києве мій !

2. Висловлювання, які є хибними.

— Дон впадає в Балтійське море.

— Австрія межує з Вірменією.

3. Висловлювання, які є істинними.

— Варшава є столицею Польщі.

— Хмельницька область входить до складу Поділля.

Існує розділ математики – алгебра логіки, в якому вивчаються операції та закони логічного мислення. Особлива цінність цього розділу для географії суспільства полягає в тому, що він дає змогу моделювати дефініції суспільно-географічних понять і аналізувати ці поняття.

Звичайна математика оперує з безмежною кількістю числових значень – цілих, раціональних (дробових), дійсних. На відміну від неї алгебра логіки оперує лише з двома логічними значеннями: «*хиба*» та «*істина*». В подальшому ми будемо для зручності позначати їх $\boldsymbol{x}$ та $\boldsymbol{i}$ відповідно.

2. Математико-логічні операції.

Нехай $A, B, C, \dots$ – прості висловлювання. На їх основі можна побудувати складні висловлювання, використовуючи операції алгебри логіки. Далі розглянемо такі операції: заперечення, кон’юнкція, диз’юнкція, імплікація, еквіваленція.

Заперечення. Нехай A – просте висловлювання. Операція заперечення позначається $\overline{A}$ (читається «не A»), Це логічна операція з одним операндом, істинність результату якої є протилежною до істинності операнда .

Дію заперечення можна зобразити такою табличкою істинності:

A	$\overline{A}$
х	***і***
і	***х***

Кон'юнкція. Нехай A, B – два простих висловлювання. Операція кон'юнкції позначається $A \wedge B$ (читається «A і B»). Це логічна операція з двома операндами. Її результат істинний лише тоді, коли істинними є обидва операнди і хибний у всіх інших випадках.

Дію кон'юнкції можна зобразити такою табличкою істинності:

A	B	$A \wedge B$
х	***х***	***х***
х	***і***	***х***
і	***х***	***х***
і	***і***	***і***

Диз'юнкція. Нехай A, B – два простих висловлювання. Операція диз'юнкції позначається $A \vee B$ (читається «A або B»). Це логічна операція з двома операндами. Її результат хибний лише тоді, коли хибними є обидва операнди та істинний у всіх інших випадках.

Дію диз'юнкції можна зобразити такою табличкою істинності:

A	B	$A \vee B$
х	***х***	***х***
х	***і***	***і***
і	***х***	***і***
і	***і***	***і***

Імплікація. Нехай A, B – два простих висловлювання. Операція імплікації позначається $A \Rightarrow B$ (читається «з A випливає B»). Це логічна операція з двома операндами. Її результат хибний лише тоді, коли перший операнд є істинним, а другий хибним та істинний у всіх інших випадках.

Дію імплікації можна зобразити такою табличкою істинності:

A	B	$A \Rightarrow B$
x	***x***	***i***
x	***i***	***i***
i	***x***	***x***
i	***i***	***i***

Імплікацію можна виразити через попередні операції так:

$$A \Rightarrow B \sim \overline{A} \vee B .$$

Еквіваленція. Нехай A, B – два простих висловлювання. Операція еквіваленції позначається $A \equiv B$ (читається «A еквівалентне B»). Це логічна операція з двома операндами. Її результат істинний тоді, коли обидва операнди мають однакову істинність і хибний коли їхня істинність протилежна.

Дію еквіваленції можна зобразити такою табличкою істинності:

A	B	$A \equiv B$
x	***x***	***i***
x	***i***	***x***
i	***x***	***x***
i	***i***	***i***

Еквіваленцію можна виразити через попередні операції так:

$$A \equiv B \sim (A \Rightarrow B) \wedge (B \Rightarrow A) \sim (A \wedge B) \vee (\overline{A} \wedge \overline{B}).$$

3. Предикати

Знання операцій математичної логіки дає змогу розглянути такий важливий клас функцій, як *предикати*.

Предикат – це функція, яка може бути визначена на будь-якій множині, але її значеннями обов'язково є логічні значення «*хиба **x***» або «*істина **i***». У географії предикати виявляються дуже зручними для моделювання різноманітних геопросторових відношень. Оскільки така функція може залежати від різної кількості аргументів (1, 2, 3 і т.д.), то відповідний предикат прийнято називати одномісним, двомісним, тримісним і т.д.

Розглянемо декілька найпростіших прикладів застосування предикатів у суспільній географії.

1. Одномісні предикати. Нехай $F_с(X)$, $F_{смт}(X)$, $F_м(X)$ - три функції, визначені на множині всіх поселень. $F_с(X)$ набуває істинного значення лише тоді, коли X є селом, $F_{смт}(X)$ - лише тоді, коли X є селищем міського типу, а $F_м(X)$ набуває істинного значення лише тоді, коли X є містом. При таких домовленностях коренями рівняння $F_{смт}(X)=i$ є всі селища міського типу. Крім, цього, очевидно, що названі три предикати перебувають у такому співвідношенні:

$$F_с(X) \vee F_{смт}(X) \vee F_м(X) = i$$

2. Двомісні предикати. Нехай $F(A,X)$ - функція, визначена на множині всіх країн. Вважатимемо, що $F(A,X)$ набуває істинного значення лише тоді, коли країна X є сусідом країни A (межує з нею). Наприклад,

$$F("Україна","Польща") = i\,,$$
$$F("Франція","Австрія") = x\,.$$

Зрозуміло, що в цьому випадку коренями предикативного рівняння $F(A,X)=i$ є всі сусіди першого порядку для країни A.

Тестові завдання

1. *Яка операція математичної логіки дає істинний результат лише тоді, коли обидва операнди істинні ?*

1. кон'юнкція; 2. диз'юнкція; 3. імплікація; 4. еквіваленція; 5. заперечення.

2. *Яка операція математичної логіки дає хибний результат лише тоді, коли обидва операнди хибні ?*

1. кон'юнкція; 2. заперечення; 3. диз'юнкція; 4. імплікація; 5. еквіваленція.

3. *Яка операція математичної логіки дає хибний результат лише тоді, коли перший операнд істинний, а другий* - *хибний ?*

1. заперечення; 2. кон'юнкція; 3. диз'юнкція; 4. імплікація; 5. еквіваленція.

4. *Яка операція математичної логіки дає хибний результат лише тоді, коли операнди мають різну істинність ?*

1. кон'юнкція; 2. диз'юнкція; 3. заперечення; 4. імплікація; 5. еквіваленція.

Тема 2

МАТЕМАТИКО-ЛОГІЧНИЙ АНАЛІЗ ПОНЯТЬ

План:

1. Моделювання дефініцій понять
2. Аналіз змісту поняття.
3. Аналіз обсягу поняття. Діаграма Венна

1. Моделювання дефініцій понять.

Аналіз поняттєво-термінологічного апарату є необхідним етапом серйозного суспільно-географічного дослідження і зустрічається в більшості курсових, дипломних, магістерських та дисертаційних роботах. Такий аналіз має ще більшу цінність, якщо він виконаний з використанням засобів логіко-математичного моделювання.

Поняття є формою мислення, яка відображає об’єкти, явища та зв’язки між ними через їхні суттєві і специфічні ознаки.

Вивчення понять здійснюють за їх головними атрибутами – *змістом* та *обсягом*, а вивчення системних чи мережевих зв’язків проводять за відношеннями, в яких перебувають поняття.

Зміст поняття визначається сукупністю ознак, які використовуються для дефініції поняття, а також способом їх поєднання.

Під обсягом поняття розуміємо сукупність об’єктів, шо володіють ознаками, закладеними в змісті.

Розширення змісту поняття природно веде до звуження сукупності об’єктів, що задовольняють дефініцію, тому тому залежність між змістом та обсягом поняття має обернений характер.

Моделювання дефініції поняття полягає в представленні цієї дефініції за допомогою математико-логічної фунції (предикату). Така модель утворює нову (складну, комбіновану) ознаку, яка вирізняє ту сферу, на яку поширюється дефініція. Цю нову ознаку можна далі використовувати як базову для дефініювання інших понять.

Розглянемо для прикладу поняття територіального господарського комплексу. Під територіальним господарським комплексом (ТГК) розуміємо сукупність територіально зосереджених елементів виробничої сфери або сфери послуг.

Зробимо такі позначення:

се – сукупність елементів, яка є аргументом предиката,

ТЗ – ознака, яка позначає територіальну зосередженість,

ВС – ознака, яка позначає належність до виробничої сфери,

СП – ознака, яка позначає належність до сфери послуг.

Тоді, логіко-математичну формулу дефініції територіального господарського комплексу можна записати так:

$$ТГК(се) = ТЗ(се) \wedge [ВС(се) \vee СП(се)], \text{ або коротше } ТГК = ТЗ \wedge (ВС \vee СП).$$

2. Аналіз змісту поняття.

Ми аналізуємо зміст поняття за допомогою математико-логічної моделі його дефініції, яка поєднує ознаки поняття. Така модель записується у вигляді відповідної математико-логічної формули. До цієї моделі входять змінні, що позначають ознаки, а також математико-логічні операції. Змінні можуть входити в прямому або запереченому вигляді, а при розгляді операцій можна, не втрачаючи загальності, обмежитись операціями кон'юнкції та диз'юнкції.

Якщо в дефініції ознаки поєднуються шляхом кон'юнкції, то зміст поняття розширюється. Навпаки, якщо ознаки поєднані за допомогою диз'юнкції, то зміст поняття звужується (у крайньому випадку, коли ознака диз'юнктивно поєднана зі своїм запереченням, зміст є взагалі пустим, бо така конструкція описує універсальну множину об'єктів).

Нехай A, B, C - змінні, що позначають ознаки дефініювання. Спочатку розглянемо операцію кон'юнкції, тобто логічного множення. Логіко-математична формула $A \wedge B$ означає, що ознаки A та B повинні справджуватись спільно, тобто дефінійоване таким чином поняття володіє як ознакою A так і ознакою B. Саме тому ми кажемо, що це поняття є змістовно багатшим (у нього ширший зміст) порівняно з поняттями, що описуються окремо ознаками A чи B.

Інакше виглядає справа при застосуванні операції диз'юнкції, тобто логічного додавання. Логіко-математична формула $A \vee B$ означає, що дефінійоване так поняття дозволяє певну варіантність, бо застосування ознак A та B має альтернативний характер. Тому, це поняття містить більше невизначеності, ніж поняття, дефінійовані окремо ознаками A чи B. Отже воно є змістовно біднішим, тобто у нього вужчий зміст ніж у понять на основі окремих ознак A чи B.

3. Аналіз обсягу поняття. Діаграма Венна.

Тепер окреслимо обсяг поняття, тобто покажемо ті логічні класи об'єктів, які описуються дефініцією поняття.

Діаграма Венна є наочним графічним зображенням логічних класів обсягу поняття. Її будують на основі логіко-математичної формули, використовуючи стандартні графічні шаблони. Вигляд цих шаблонів залежить від кількості ознак, які входять до дефініції поняття. Позначимо через A, B, C, D логічні змінні, що відповідають ознакам дефініювання.

1. Шаблон для однієї ознаки A:

$\overline{A}$	
A	

2. Шаблон для двох ознак A, B:

	$\overline{B}$	B
$\overline{A}$		
A		

3. Шаблон для трьох ознак A, B, C:

	$\overline{B}$	B	
$\overline{A}$			$\overline{C}$
A			
A			C
$\overline{A}$			

4. Шаблон для чотирьох ознак A, B, C, D:

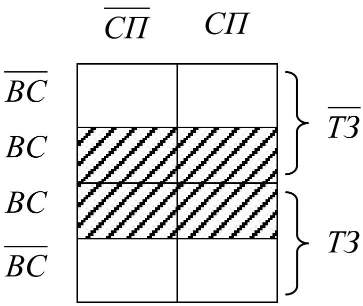
	$\overline{B}$	B	B	$\overline{B}$	
$\overline{A}$					$\overline{C}$
A					
A					C
$\overline{A}$					
	$\overline{D}$		D		

Для більшої кількості ознак (5,6,7 і т.д.) шаблони будуються аналогічно, але вони мають громіздкіший вигляд.

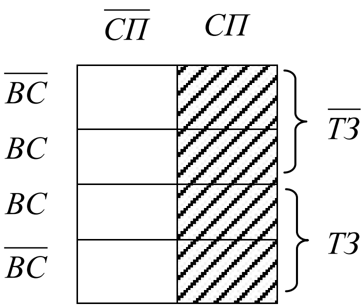
Розглянемо застосування діаграми Венна для аналізу обсягу поняття. Для зручності та простоти розуміння покажемо це на конкретному прикладі. В п.1. ми побудували логічну формулу для поняття територіального господарського комплексу. Вона мала вигляд $ТГК = ТЗ \wedge (ВС \vee СП)$. Бачимо, що ця формула містить три ознаки: *ТЗ*, *ВС* і *СП*. Отже, для побудови діаграми Венна, будемо використовувати третій шаблон.

Щоб коректно зберегти послідовність операцій формули почнемо її розгляд із найбільш внутрішньої частини, тобто $ВС \vee СП$. Тут присутні дві ознаки *ВС* і *СП*, поєднані за допомогою операції диз'юнкції, яка моделює у висловлюваннях частку «або». Це означає, що для цієї частини формули на діаграмі потрібно позначити ті області, для яких може справджуватися або знака *ВС*, або ознака *СП*.

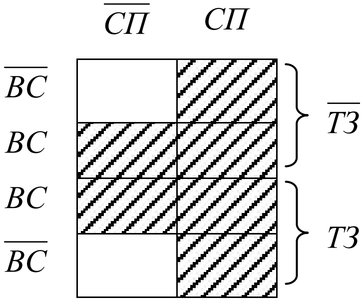
Ознака *ВС* справджується в такій області:



Ознака *СП* справджується в такій області:

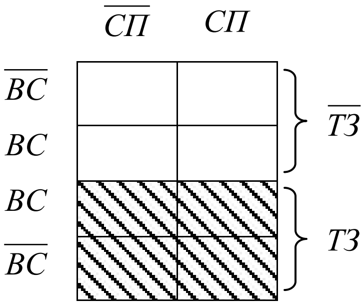


Тому, графічно вираз $ВС \vee СП$ виглядає так:

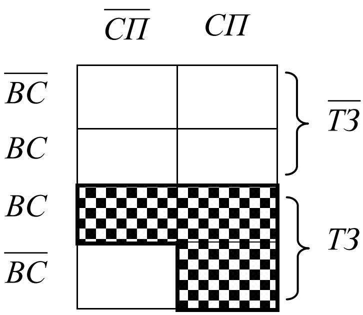


Тепер розглянемо логіко-математичну формулу в цілому. Вона є кон'юнкцією, що поєднує змінну *ТЗ* із вже розглянутим виразом *(ВС∨СП)*. Кон'юнкція є логічним множенням, яке моделює у висловлюваннях частку «і». Це означає, що остаточна графічна схема повинна одночасно поєднувати властивості ознак *ТЗ* і *(ВС∨СП)*.

Ознака *ТЗ* справджується в такій області:



Якщо взяти спільну частину останніх двох графічних зображень, то отримаємо остаточний вигляд діаграми Венна:



На підставі побудованої діаграми Венна робимо висновок, що обсяг поняття «територіальний господарський комплекс» складається з трьох конкретних логічних блоків. Усі три блоки передбачають територіальну зосередженість елементів комплексу і всі блоки вимагають належність елементів комплексу хоча-би до одної з господарських сфер: виробничої чи сфери послуг.

Тестові завдання

1. *Як залежить обсяг поняття від його змісту ?*

1. прямо; 2. квадратично; 3. обернено; 4. ніяк; 5; географічно.

2. *Як змінюється зміст поняття при кон'юнкції ознак, що входять до його дефініції ?*

1. стрибкоподібно; 2. розширюється; 3. звужується; 4. не змінюється; 5. коливається.

3. *Як змінюється зміст поняття при диз'юнкції ознак, що входять до його дефініції ?*

1. розширюється; 2. стрибкоподібно; 3. звужується; 4. незмінюється; 5. коливається.

4. *Як змінюється обсяг поняття при кон'юнкції ознак, що входять до його дефініції ?*

1. розширюється; 2. звужується; 3. стрибкоподібно; 4. не змінюється; 5. коливається.

5. *Як змінюється обсяг поняття при диз'юнкції ознак, що входять до його дефініції ?*

1. розширюється; 2. звужується; 3. стрибкоподібно; 4. не змінюється; 5. коливається.

6. *Що позначає одна комірка діаграми Венна ?*

1. зв’язок між змістом та обсягом; 2. частину змісту поняття; 3. частину обсягу поняття; 4. зв'язок між поняттями; 5. відношення між поняттями.

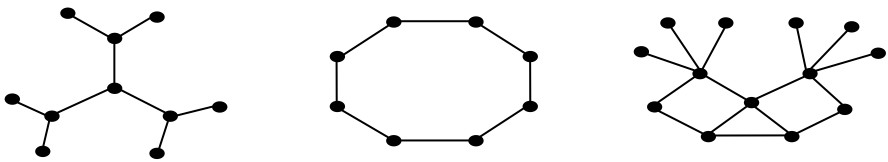
Тема 3
ТЕОРІЯ ГРАФІВ У СУСПІЛЬНІЙ ГЕОГРАФІЇ

План

1. Поняття графа. Види графів
2. Числові характеристики графів
3. Міри центральності графів

1. Поняття графа. Види графів

Графом називають сукупність вузлів (вершин), з'єднаних ребрами. Граф є дуже цінною математичною абстракцією, яка дає змогу здійснювати широкий клас моделювання в географії суспільства. Графами моделюють структуру найрізноманітніших суспільних систем та мереж. Графи використовують для моделювання лінійних об'єктів, транспортних та інших комунікацій. Графи можна застосувати до моделювання структури та границь територіального устрою і в багатьох інших ситуаціях.



Приклади графів

Розглянемо декілька найважливіших понять, пов'язаних з графами.

Суміжність вузлів. Два вузли графа називаються суміжними, якщо в графі існує ребро, яке їх з'єднує.

Суміжність ребер. Два ребра графа називаються суміжними, якщо вони мають у графі спільний вузол.

Інцидентність ребер та вузлів. Ребро називається інцидентним вузлу, якщо цей вузол є одним з кінців ребра.

Ланцюг – це послідовність ребер, у якій кінець одного ребра співпадає з початком наступного.

Цикл – це замкнутий ланцюг, тобто ланцюг, у якому початок і кінець співпадають.

Топологічна відстань між двома вузлами графа – це кількість ребер найкоротшого ланцюга, який з'єднує ці вузли.

Ребро графа називається *орієнтованим*, якщо воно має напрямок (графічно зображається стрілкою).

Розглянемо деякі важливі види графів.

Граф називається *деревом*, якщо він не має жодною циклу. У деревах кількість ребер на одиницю менша кількості вузлів. Для фіксованої множини вузлів кожне дерево є мінімальним графом, що з'єднує всі вузли.

Граф називається *повнозв'язним*, якщо у ньому є ребра між усіма парами вузлів. Для фіксованої множини вузлів повнозв'язний граф є максимальним графом.

Граф називається *плоским*, якщо його можна нарисувати на площині без взаємоперетинів ребер. Навпаки, граф називається *неплоским*, якщо його неможливо нарисувати на площині без взаємоперетинів ребер.

Граф називається *зв'язним*, якщо він не має ізольованих компонент. Навпаки, граф називається *незв'язним*, якщо в нього є ізольовані компоненти.

Граф називається *орієнтованим*, якщо в нього всі ребра орієнтовані. Навпаки, граф називається *неорієнтованим*, якщо в нього всі ребра неорієнтовані.

2. Числові характеристики графів

Абсолютні числові характеристики. Перелічимо головні абсолютні числові характеристики графа, що найчастіше зустрічаються в суспільно-географічних дослідженнях.

V - кількість вузлів графа;

R - кількість ребер графа;

C - цикломатичне число графа, кількість його незалежних циклів (для дерева $C = 0$);

K - кількість компонент зв'язності графа (для зв'язного графа $K = 1$);

G - кількість граней плоского графа (гранню плоского графа вважаємо частину площини, обмежену ребрами, плоский зв'язний граф має C внутрішніх граней і одну зовнішню, тобто $G = C + 1$);

D - діаметр графа, тобто найбільша топологічна відстань між двома вузлами в графі.

Для плоских графів справджуються такі співвідношення:

$$R - V + K = C,$$
$$V - R + G = 2,$$
$$G - C + K = 2.$$

На рівні окремих вузлів важливою характеристикою є *степінь* вузла, тобто кількість ребер, які йому інцидентні. Очевидно, що сума степеней всіх вузлів графа рівна подвоєній кількості його ребер.

Відносні числові характеристики. Відносні характеристики графа можуть відображати різні його аспекти. Тут розглянемо такі важливі для географії суспільства аспекти, як зв'язаність і компактність. У практиці суспільно-географічних досліджень широко використовують три міри зв'язаності графа: α-індекс, β-індекс і φ-індекс, а також міру компактності графа: π-індекс.

В основу побудови α-індекса покладена така ідея, що чим більше незалежних циклів має граф (при фіксованій кількості вузлів), тим сильніше він зв'язаний. Тому цей індекс обчислюють як відношення фактичного цикломатичного числа графа до максимально можливого цикломатичного числа при фіксованій кількості вузлів. Отже:

$$\alpha = \frac{C}{C_{\max}} = \frac{R - V + K}{2 \cdot V - 5}.$$

α-індекс може приймати значення від 0 до 1, причому крайні значення є досяжними.

В основу побудови β-індекса покладена дуже проста ідея, що чим більше ребер має граф (при фіксованій кількості вузлів), тим сильніше він зв'язаний. Тому цей індекс обчислюють як відношення кількості ребер до кількості вузлів. Тобто:

$$\beta = \frac{R}{V}.$$

β-індекс може примати значення від 0 до 3, причому крайні значення для скінчених зв'язаних графів недосяжні.

В основу побудови φ-індекса також покладена ідея, що чим більше ребер має граф (при фіксованій кількості вузлів), тим сильніше він зв'язаний. Тому цей індекс обчислюють як відношення фактичної кількості ребер графа до їх максимально можливої кількості при фіксованій кількості вузлів. Отже:

$$\alpha = \frac{R}{R_{\max}} = \frac{R}{3 \cdot (V - 2)}.$$

α-індекс може приймати значення від 0 до 1, причому нульове значення для зв'язаних графів недосяжне.

В основу побудови міри компактності, тобто π-індекса покладена така ідея, що чим більше ребер має граф при фіксованому його діаметрі, тим він компактніший (можна сказати й навпаки: чим менший діаметр має граф при фіксованій кількості ребер, тим він компактніший). Тому:

$$\pi = \frac{R}{D}.$$

π-індекс може приймати значення від 0 до R, причому нульове значення для зв'язаних графів недосяжне.

3. Міри центральності графів

У суспільній географії часто виникає завдання визначити центральний вузол деякого графа. Для великих і складних комунікаційних графів об'єктивне визначення центрального вузла може бути проблемою, тому потрібно мати алгоритм вирішення цього завдання. У теорії графів є відповідні засоби, які називаються мірами центральності графів.

Спочатку будуємо матрицю топологічних відстаней між вузлами $D = \{d_{jl}\}$.

Матриця топологічних відстаней це квадратна матриця D, яка має стільки рядків і стовпців, скільки є вузлів, і елементами якої є топологічні відстані d_{jl} між усіма парами вузлів j та l на графі. Ця матриця є симетричною відносно головної діагоналі. На діагоналі матриці стоять нулі.

На основі матриці топологічних відстаней обчислюємо індекс доступності та число Кеніга для кожного вузла графа.

Індекс доступності $S_j = \sum_l d_{jl}$ (сума чисел у кожному рядку матриці).

Число Кеніга $K_j = \max_l(d_{jl})$ (найбільше число у кожному рядку графа).

Кожну з цих мір можна використати для визначення центрального вузла.

За індексом доступності вузол вважають центральним, якщо він має найменший індекс доступності.

За числом Кеніга вузол вважають центральним, якщо він має найменше число Кеніга.

Знаючи центральний вузол графа можна оцінити положення інших вузлів відносно нього. Для цього існують такі міри:

- міра d_{jc} ієрархічного положення вузла j (показує порядок сусідства);
- оцінка $\frac{S_j}{S_c}$ положення вузла j.

Тестові завдання

1. *Що є спільним для суміжних вузлів графа ?*

1. грань; 2. цикл; 3. вузол; 4. ребро; 5. ланцюг.

2. *Що є спільним для суміжних ребер графа ?*

1. ланцюг; 2. цикл;3. грань; 4. вузол; 5. ребро.

3. *У якому відношенні перебуває вузол графа до ребра, якщо він є одним з його кінців ?*

1. інцидентності; 2. близькості; 3. сусідства; 4. суміжності; 5. панібратства.

4. *Як називається ланцюг графа, в якому початок і кінець співпадають ?*

1. дерево; 2. кільце; 3. грань; 4. цикл; 5. діаметр.

5. *Скільки циклів має граф-дерево ?*

1. жодного; 2. один; 3. два; 4. стільки скільки вершин; 5. стільки скільки ребер.

6. *Який граф завжди розпадається на незв'язні частини при вилученні одного ребра ?*

1. плоский; 2. цикл; 3. дерево; 4. повнозв'язний; 5. неорієнтований.

7. *Який граф ніколи не розпадається на незв'язні частини при вилученні одного ребра ?*

1. плоский; 2. дерево; 3. лінійний; 4. повнозв'язний; 5. незв'язний.

8. *Що таке цикломатичне число графа ?*

1. довжина найбільшого циклу; 2. кількість незалежних циклів; 3. кількість залежних циклів; 4. кількість усіх циклів; 5. кількість граней.

9. *Яка з наведених мір характеризує зв'язаність графа ?*

1. кількість вузлів; 2. альфа-індекс; 3. індекс доступності; 4. діаметр; 5. число Кеніга.

10. *Яка з наведених мір характеризує зв'язаність графа ?*

1. бета-індекс; 2. діаметр; 3. кількість вузлів; 4. індекс доступності; 5. число Кеніга.

11. *Яка з наведених мір характеризує зв'язаність графа ?*

1. число Кеніга; 2. індекс доступності; 3. діаметр; 4. кількість вузлів; 5. фі-індекс.

12. *В яких межах може змінюватись альфа-індекс ?*

1. 0 - 1; 2. 0 - 2; 3. 0 - 3; 4. - 1 - +1; 5. 0 -100.

13. *В яких межах може змінюватись фі-індекс ?*

1. 0 - 2; 2. 0 - 3; 3. 0 - 1; 4. -1 - +1; 5. 0 -100.

14. *Як називається граф, що має дві компоненти зв'язності ?*

1. зв'язний; 2. незв'язний; 3. дерево; 4. цикл; 5. повнозв'язний.

15. *Скільки граней є у плоского графа, що має вигляд трикутника ?*

1. дві; 2. одна; 3. нуль; 4. три; 5. довільна кількість.

16. *Що таке діаметр графа ?*

1. кількість ребер; 2. найменша відстань між вузлами; 3. найдовше

ребро в графі; 4. найбільша відстань між вузлами; 5. найкоротше ребро в графі.

17. *Скільки рядків має матриця топологічних відстаней графа ?*

1. кількість ребер; 2. кількість циклів; 3. кількість вершин; 4. діаметр графа; 5. цикломатичне число.

18. *Скільки стовпців має матриця топологічних відстаней графа ?*

1. кількість ребер; 2. діаметр графа; 3. кількість вершин; 4. цикломатичне число; 5. кількість циклів.

19. *Скільки дорівнює цикломатичне число для графа-дерева ?*

1. нуль; 2. один; 3. два; 4. кількість вершин; 5. кількість ребер.

20. *Кількість ребер, які інцидентні вершині, це*

1. діаметр вершини; 2. суміжність вершини; 3. орієнтація вершини; 4. зв'язність вершини; 5. степінь вершини.

21. *Яка з наведених мір характеризує центральність вузла графа ?*

1. цикломатичне число; 2. число Кеніга; 3. альфа-індекс; 4. бета-індекс; 5. діаметр.

22. *Яка з наведених мір характеризує центральність вузла графа ?*

1. індекс доступності; 2. альфа-індекс; 3. бета-індекс; 4. діаметр; 5. цикломатичне число.

23. *Що характеризує топологічна відстань від центрального вузла ?*

1. зв'язність; 2. циклічність; 3. орієнтованість; 4. складність; 5. ієрархічне положення.

24. *Яке значенння індекса доступності відповідає центральному вузлу графа ?*

1. нульове; 2. найменше; 3. середнє; 4. найбільше; 5. довільне.

26. *Яке значення числа Кеніга відповідає центральному вузлу графа ?*

1. довільне; 2. нульове; 3. найменше; 4. найбільше; 5. середнє.

27. *Що характеризує відношення індекса доступності вузла графа до індекса доступності його центрального вузла ?*

1. циклічність; 2. складність; 3. оцінку положення; 4. зв'язність; 5. суміжність.

28. *Як називається сума всіх чисел у рядку матриці топологічних відстаней графа ?*

1. цикломатичне число; 2. число Кеніга; 3. індекс доступності; 4. міра зв'язності; 5. діаметр.

29. *Як називається найбільше число у рядку матриці топологічних відстаней графа ?*

1. число Кеніга; 2. цикломатичне число; 3. індекс доступності; 4. міра зв'язності; 5. діаметр.

Тема 4

ГЕОГРАФІЧНІ ПОЛЯ ТА ЇХНЯ КОРЕЛЯЦІЯ

План:

1. Поняття про географічне поле
2. Кореляційний з’язок географічних полів
   2.1. Часовий кореляційний з’язок
   2.2. Просторовий кореляційний з’язок

1. Поняття про географічне поле

Поняття географічного поля природно виникає з багатьох міркувань. У географії суспільства вивчають об’єкти, їхні властивості та відношення трьох головних геометричних типів: точково зосереджені, лінійно зосереджені і просторово поширені. Розглянемо два розуміння географічного поля: субстанціональне та математичне, які логічно узгоджені між собою.

При субстанціональному розумінні *географічне поле* – це просторово поширене суспільно-географічне явище. Зрозуміло, що, для коректності, кожне таке явище потрібно розглядати в регіоні відповідного таксономічного рівня. Прикладами просторово поширених явищ у географії суспільства можуть бути: сільське розселення, вирощування картоплі, середня освіта, первинна медико-санітарна допомога, розміщення будівельної промисловості, електропостачання, телефонний зв'язок і багато інших.

При математичному розумінні *географічне поле* – це математична функція, про особливості якої потрібно сказати окремо. Відомо, що кожна математична функція має область визначення, область значень і закон відповідності між ними. Областю визначення, тобто носієм математичної функції, як географічного поля, є ареал поширення певного явища, областю значень – діапазон чисел, які характеризують поле, а закон відповідності кожній точці з області визначення ставить у відповідність число з області значень, яке характеризує міру інтенсивності явища в цій точці. Такий підхід дає змогу розглядати відповідні математичні функції (або їхні сукупності) як модель географічного поля.

Така модель має загальновідомий вигляд: $f(x, y)$, де x, y - координати точок з області визначення функції, тобто з ареалу поширення явища, f - закон відповідності, який кожній точці x, y ставить у відповідність значення $f(x, y)$, що характеризує інтенсивність відповідного поля в цій точці.

Географічне поле можна зображати картографічно. Для просторово-поширених явищ найчастіше використовують такі картографічні способи, як спосіб ізоліній, спосіб картограм, крапковий спосіб та деякі інші.

Під **декомпозицією** об'єкта розуміємо його логічний поділ на такі підоб'єкти, які при зворотному логічному поєднанні (композиції, агрегації) утворюють вихідний об'єкт. При цьому важливо, щоб отримані підоб'єкти були у певному розумінні "простішими" за вихідний об'єкт. В такому випадку дослідження складного об'єкта зводиться з одного боку до ефективнішого вивчення "простіших" підоб'єктів, а з другого боку до ефективного вивчення тих зв'язків між ними, які необхідні для зворотної композиції (агрегації) вихідного об'єкта.

Трендова декомпозиція полягає у математичному поділі досліджуваного поля на такі складові частини, які виділяються на підставі певних модельних припущень. Найчастіше вона вживається у формах просторового та часового трендів. Складові частини трендів можуть виділятися на основі лінійних, квадратичних, поліноміальних, сплайнових, гармонійних, спеціальних та багатьох інших моделей і їх поєднань. Ці складові перебувають між собою як правило в адитивних відношеннях, що дає змогу легко компонувати цілісне явище з окремих його частин.

Особливе значення має трендова декомпозиція при вивченні географічних полів. Трендова декомпозиція географічного поля полягає в його розкладенні на трендову та залишкову частини. Математично розклад виглядає так:

$$w(x, y) = w_{тр}(x, y) + w_{зал}(x, y) .$$

Трендова частина $w_{тр}(x, y)$ географічного поля показує (схоплює) загальну тенденцію геопросторової мінливості поля $w(x, y)$. Тренд будують, виходячи з певних модельних припущень, тому змістовно він завжди виражає певну емпіричну чи теоретичну геопросторову закономірність.

Залишкову частину географічного поля обчислюють так: $w_{зал}(x, y) = w(x, y) - w_{тр}(x, y)$. Вона показує місцеві (локальні) особливості мінливості поля $w(x, y)$ і змістовно виражає геопросторовий залишок (дефект закономірності), який в рамках прийнятої трендової моделі вважають випадковим.

Отже, суспільно-географічне вивчення реального географічного поля $w(x,y)$, декомпонується на вивчення його тренду $w_{тр}(x,y)$ та глобальних (в рамках конкретного об’єкта) факторів, що його зумовлюють, а також на вивчення залишку $w_{зал}(x,y)$ і відповідних йому локальних факторів.

2. Кореляційний зв’язок географічних полів

2.1. Часовий кореляційний зв'язок. Розглянемо два нестаціонарних географічних поля $u(x,y,t)$ і $v(x,y,t)$. В кожній точці території вони характеризуються певною часовою динамікою. Це дає змогу в кожній такій точці визначити кореляційний зв'язок між динамікою $u(x,y,t)$ і динамікою $v(x,y,t)$. Нехай спостереження за цими географічними полями здійснювалися в n моментів часу $t_1,t_2,...,t_n$. Тоді отримаємо часові масиви спостережень $u(x,y,t_1),u(x,y,t_2),...,u(x,y,t_n)$ і $v(x,y,t_1),v(x,y,t_2),...,v(x,y,t_n)$. На основі цих масивів обчислюємо коефіцієнт кореляції між ними в довільній точці (x,y):

$$r_{uv}(x,y)=\frac{S_{uv}(x,y)}{\sqrt{S_{uu}(x,y)\cdot S_{vv}(x,y)}},$$

де $$S_{uu}(x,y)=\sum_{k=1}^{n}[u(x,y,t_k)-\bar{u}(x,y)]^2,$$

$$S_{vv}(x,y)=\sum_{k=1}^{n}[v(x,y,t_k)-\bar{v}(x,y)]^2,$$

$$S_{uv}=\sum_{k=1}^{n}[(u(x,y,t_k)-\bar{u}(x,y)]\cdot[v(x,y,t_k)-\bar{v}(x,y)],$$

середні величини $\bar{u}(x,y)=\frac{1}{n}\cdot\sum_{k=1}^{n}u(x,y,t_k)$, $\bar{v}(x,y)=\frac{1}{n}\cdot\sum_{k=1}^{n}v(x,y,t_k)$.

Важливо відзначити, що величина $r_{uv}(x,y)$ є просторово розподіленою, бо визначається в кожній точці території. Отже, $r_{uv}(x,y)$ - це географічне поле часового кореляційного зв’язку.

Географічне поле $r_{uv}(x,y)$ можна відобразити картографічно. Якщо кореляція розрахована в достатньо великій кількості точок

(x, y), то для її відображення доцільно скористатися картографічним способом ізоліній. Отримана картосхема покаже на території розташування ареалів трьох видів: ареалів прямого часового зв'язку між $u(x, y, t)$ і $v(x, y, t)$, ареалів оберненого часового зв'язку між ними і ареалів відсутності часового кореляційного зв'язку між цими полями.

2.2. Просторовий кореляційний зв'язок. Розглянемо два стаціонарних географічних поля $u(x, y)$ і $v(x, y)$. Кожне з цих полів в довільній точці (x, y) має напрямок (азимут) найшвидшого зростання (він називається градієнтом). Якщо для $u(x, y)$ і $v(x, y)$ такі напрямки співпадають, то це означає, що в напрямку зростанням одного поля друге також максимально зростає і можемо вважати, що в точці (x, y) ці поля мають найвищу пряму кореляцію. Якщо градієнти $u(x, y)$ і $v(x, y)$ мають протилежні напрямки, то в напрямку зростання одного поля друге максимально спадає тому вважаємо що ці поля в точці (x, y) мають найвищу від'ємну кореляцію. Якщо ж градієнти цих полів у точці (x, y) взаємно перпендикулярні, то це означає, що в напрямку зростання одного поля друге поле взагалі не змінюється, тобто точкова кореляція між $u(x, y)$ і $v(x, y)$ є нульовою.

Функція $r_{uv}(x, y)$ також моделює географічне поле – поле просторового кореляційного зв'язку між заданими полями.

Формулу для просторової кореляції можна представити у вигляді

$$r_{uv}(x, y) = \cos(\varphi_{uv}),$$

де φ_{uv} - кут між градієнтами (напрямками максимального зростання) полів $u(x, y)$ і $v(x, y)$ в точці (x, y). Інколи цей кут можна визначити без обчислень візуально на основі наявних картосхем полів $u(x, y)$ та $v(x, y)$, виконаних попередньо способом ізоліній, хоча такий спосіб є чисто «ручним» і досить трудомістким.

Географічне поле просторової кореляції $r_{uv}(x, y)$ також можна відобразити картографічно. Якщо кореляція розрахована в достатньо великій кількості точок (x, y), то для її відображення можна скористатися картографічним способом ізоліній. Отримана картосхема покаже на території розташування ареалів трьох видів: ареалів прямого просторового зв'язку між $u(x, y)$ і $v(x, y)$, ареалів оберненого просторового зв'язку між ними і ареалів відсутності просторового кореляційного зв'язку між цими полями.

Тестові завдання

1. *В яких межах може змінюватися коефіцієнт кореляції ?*

1. 0 - 1; 2. -1 - 1; 3. 0 - 100; 4. -100 - 100; 5. У довільних.

2. *Для яких географічних полів розраховують часовий коефіцієнт кореляції ?*

1. осесиметричних; 2. неосесиметричних; 3. стаціонарних; 4. нестаціонарних; 5. векторних.

3. *Для яких географічних полів розраховують просторовий коефіцієнт кореляції ?*

1. осесиметричних; 2. неосесиметричних; 3. стаціонарних; 4. нестаціонарних; 5. векторних.

4. *Яка частина географічного поля показує загальні особливості його мінливості ?*

1. нестаціонарна; 2. картографічна; 3. залишкова; 4. трендова; 5. стаціонарна.

5. *Яка частина географічного поля показує місцеві особливості його мінливості ?*

1. стаціонарна; 2. залишкова; 3. трендова; 4. нестаціонарна. 5. картографічна.

Тема 5
ВЗАЄМНА ТА ПРОСТОРОВА РЕГРЕСІЯ ГЕОГРАФІЧНИХ ПОЛІВ

План
1. Взаємна регресія
2. Просторова регресія

1. Взаємна регресія двох географічних полів.

Якщо між двома географічними полями встановлено наявність кореляційного зв'язку, то наступним кроком їхнього дослідження є побудова регресійного зв'язку між ними, тобто фактично моделювання залежності одного поля від іншого.

Нехай $u(x, y)$ - факторне географічне поле, а $w(x, y)$ - результуюче географічне поле. Розглянемо лінійну регресію між ними, рівняння якої має вигляд:

$$w(x, y) = a_0 + a_1 \cdot u(x, y) .$$

Побудова такого рівняння полягає у визначені коефіцієнтів a_0, a_1. Інформаційною базою для їх визначення служить масив спостережень за полями $u(x, y)$ та $w(x, y)$. Ці спостереження можна записати у вигляді такої таблиці:

№ спостереження	Точка спостереження x_i, y_i	Значення факторного поля $u(x_i, y_i)$	Значення результуючого поля $w(x_i, y_i)$
1	x_1, y_1	u_1	w_1
2	x_2, y_2	u_2	w_2
…	…	…	…
m	x_m, y_m	u_m	w_m

Провідним методом визначення коефіцієнтів a_0, a_1 є метод найменших квадратів. Після його застосування, для a_0, a_1 отримуються та формули:

$$a_1 = \frac{S_{uw}}{S_{uu}}, \quad a_0 = \overline{w} - a_0 \cdot \overline{u} ,$$

де $S_{uw} = \sum_{i=1}^{m} (u_i - \overline{u}) \cdot (w_i - \overline{w})$, $S_{uu} = \sum_{i=1}^{m} (u_i - \overline{u})^2$, $\overline{u} = \sum_{i=1}^{m} u_i$, $\overline{w} = \sum_{i=1}^{m} w_i$.

Крім лінійної регресії в практиці моделювання можуть зустрітися інші форми регресійного зв'язку. Загалом їх безліч. Серед найуживаніших нелінійних зв'язків назвемо такі:

- квадратичний: $w(x, y) = a_0 + a_1 \cdot u(x, y) + a_2 \cdot u^2(x, y)$;
- гіперболічний (у двох формах):

$$w(x, y) = a_0 + \frac{a_1}{u(x, y)} \text{ і } w(x, y) = \frac{1}{a_0 + a_1 \cdot u(x, y)} ;$$

- степеневий: $w(x, y) = a_0 \cdot [u(x, y)]^{a_1}$;
- експоненціальний: $w(x, y) = a_0 \cdot e^{a_1 \cdot u(x,y)}$.

У випадку нелінійного зв'язку, для визначення його коефіцієнтів відповідне регресійне рівняння перетворюють до лінійного вигляду (всі названі рівняння дозволяють таке перетворення), або хоча-би до вигляду, лінійного відносно шуканих коефіцієнтів. Далі застосовують один із варіантів методу найменших квадратів. Якщо перетворення до лінійного вигляду неможливе, або недоцільне, тоді застосовують один з числових методів розв'язання нелінійних рівнянь, що виникають у результаті безпосереднього застосування методу найменших квадратів.

2. Просторова регресія.

Просторова регресія є одним з видів трендових географічних полів, коли здійснюється трендова декомпозиція досліджуваного географічного поля.

Просторова регресія дає змогу вивчити загальні особливості та просторові тенденції розміщення суспільно-географічного явища. Її застосовують тоді, коли досліджуване явище можна розглядати як географічне поле, що є спостережуваним на визначеній території. Нехай географічне поле $w(x, y)$ відоме в m точках (x_i, y_i), $i = 1, \ldots, m$. Цю інформацію можна записати у вигляді таблиці

Номер точки	Координата x	Координата y	Інтенсивність поля w
1	x_1	y_1	w_1
2	x_2	y_2	w_2
$\vdots$	$\vdots$	$\vdots$	$\vdots$
m	x_m	y_m	w_m

Таблиця служить інформаційною базою для визначення коефіцієнтів регресійного рівняння. Для аналізу географічного поля можна скористатися різними формами рівняння просторової регресії: лінійним, білінійним, квадратичним, рівнянням вищих порядків.

Розглянемо детальніше лінійне рівняння просторової регресії. Воно має вигляд:

$$w_{mp}(x, y) = a_0 + a_1 \cdot x + a_2 \cdot y,$$

де a_0, a_1, a_2 – коефіцієнти рівняння.

Побудова цього рівняння полягає у визначенні коефіцієнтів a_0, a_1, a_2.

Якщо скористатись методом найменших квадратів, то ці коефіцієнти виражаються такими формулами:

$$a_1 = \frac{S_{wx} \cdot S_{yy} - S_{xy} \cdot S_{wy}}{\Delta}, \quad a_2 = \frac{S_{xx} \cdot S_{wy} - S_{wx} \cdot S_{xy}}{\Delta},$$

$$a_0 = \overline{w} - a_1 \cdot \overline{x} - a_2 \cdot \overline{y},$$

де дискримінант $\Delta = S_{xx} \cdot S_{yy} - S_{xy}^2$,

середні величини $\overline{x} = \frac{1}{m} \cdot \sum_{i=1}^{m} x_i$, $\overline{y} = \frac{1}{m} \cdot \sum_{i=1}^{m} y_i$, $\overline{w} = \frac{1}{m} \cdot \sum_{i=1}^{m} w_i$,

суми $S_{xx} = \sum_{i=1}^{m} (x_i - \overline{x})^2$, $S_{yy} = \sum_{i=1}^{m} (y_i - \overline{y})^2$, $S_{xy} = \sum_{i=1}^{m} (x_i - \overline{x})(y_i - \overline{y})$,

$$S_{wx} = \sum_{i=1}^{m} (w_i - \overline{w})(x_i - \overline{x}), \quad S_{wy} = \sum_{i=1}^{m} (w_i - \overline{w})(y_i - \overline{y}).$$

При картографічному зображенні лінійного регресійного поля способом ізоліній отримується сім'я паралельних прямих ліній. Лінійне регресійне поле вказує на загальний напрямок територіальної мінливості досліджуваного явища.

Білінійне регресійне рівняння має вигляд:

$$w_{mp}(x, y) = a_0 + a_1 \cdot x + a_2 \cdot y + a_3 \cdot xy \,.$$

Квадратичне регресійне рівняння має вигляд:

$$w_{mp}(x, y) = a_0 + a_1 \cdot x + a_2 \cdot y + a_3 \cdot xy + a_4 \cdot x^2 + a_5 \cdot y^2 \,.$$

Побудова цього рівняння полягає у визначенні шести коефіцієнтів $a_0, a_1, a_2, a_3, a_4, a_5$. При картографічному зображенні квадратичного регресійного поля способом ізоліній отримується сім'я кривих другого порядку, зокрема це може бути сім'я концентричних еліпсів. Квадратичне регресійне поле вказує не лише на загальний напрямок територіальної мінливості досліджуваного явища, але і на просторову тенденцію локалізації його епіцентра.

Тестові завдання

1. *Скільки коефіцієнтів має рівняння лінійної взаємної регресії двох географічних полів ?*

1. жодного; 2. один; 3. два; 4. три; 5. чотири.

2. *Скільки коефіцієнтів має рівняння квадратичної взаємної регресії двох географічних полів ?*

1. жодного; 2. один; 3. два; 4. три; 5. чотири.

3. *Скільки коефіцієнтів має рівняння лінійної просторової регресії географічного поля ?*

1. жодного; 2. один; 3. два; 4. три; 5. чотири.

4. *Скільки коефіцієнтів має рівняння квадратичної просторової регресії географічного поля ?*

1. два; 2. три; 3. чотири; 4. п’ять; 5. шість.

Тема 6
БАГАТОВИМІРНА ТАКСОНОМІЗАЦІЯ В СУСПІЛЬНІЙ ГЕОГРАФІЇ

План:
1. Багатовимірна характеристика об’єкта
2. Вибір елементів та їхніх ознак
3. Методи стандартизації ознак
4. Методи метризації елементів
5. Таксономізація об’єктів. Метод дерева поєднань

1. Багатовимірна характеристика об’єкта

Суть багатовимірного аналізу полягає у вивченні суспільно-географічного об’єкта у сукупності взаємопов’язаних ознак. При суспільно-географічному дослідженні територіального об’єкта, що складається з багатьох територіальних елементів, на певному етапі виникає завдання виявлення та вивчення територіальних суспільних утворень, формувань, структур.

Нехай досліджуваний територіальний об’єкт складається з M елементів, кожен з яких характеризується N ознаками. Розглянемо матрицю даних, числа якої характеризують M елементів територіального об’єкта за N ознаками. Матриця даних це прямокутна матриця, яка має стільки рядків, скільки є елементів і стільки стовпців, скільки є ознак. Елементами матриці даних є значення кожної ознаки на кожному об’єкті. Розглянемо таку матрицю X розміром $M \times N$, яка характеризує M об’єктів за N ознаками:

$$X = \begin{pmatrix} x_{11} & x_{12} & \cdots & x_{1N} \\ x_{21} & x_{22} & \cdots & x_{2N} \\ \vdots & \vdots & \ddots & \vdots \\ x_{M1} & x_{M2} & \cdots & x_{MN} \end{pmatrix}.$$

У цій матриці x_{ij} - це значення j-ї ознаки на i-му елементі.

Кожен рядок матриці даних можна розуміти як координати точки в N-вимірному просторі ознак. В такому просторі, елементи об’єкта виглядають як сукупність точок. Взаємне розташування цих точок може бути близьким і далеким. Якщо точки в N-вимірному просторі ознак розташовані близько, то це означає, що елементи об'єкта мають подібні ознаки, тобто характеризуються деякою ознаковою спільністю. Якщо ж точки в N-вимірному просторі ознак розташовані далеко одна від одної, то це означає, що відповідні елементи об'єкта істотно відрізняються як кількісно так і якісно.

Багатовимірний таксономічний аналіз є важливим методом вивчення територіальних суспільних комплексів і широко використовується в практиці суспільно-географічних досліджень. Він дає змогу встановити ієрархічну класифікаційну структуру і вийти на науково-обгрунтовану типізацію територіальних елементів. Таксономізація елементів передбачає, що їх точкові образи в багатовимірному просторі ознак утворюють певні групи, скупчення, таксони, і їх потрібно виявити. Метою таксономічного аналізу є розпізнання прихованих структур даних. При цьому виникає нова інформація і нове знання.

Особлива цінність багатовимірного таксономічного аналізу для географів та картографів випливає з можливості врахувати будь-яку кількість характеристик, тобто об'єктивно здійснювати багатовимірний аналіз та інтегральне картографічне районування.

2. Вибір елементів та їхніх ознак

Для отримання ефективних результатів потрібно правильно підійти до вибору територіальних елементів, а також ознак, що їх характеризують. Помилки на цьому етапі можуть привести до того, що результати багатовимірного аналізу буде неможливо інтерпретувати (а весь багатовимірний аналіз якраз і здійснюють заради наступної географічної інтерпретації та отримання змістовних суспільно-географічних висновків).

До вибору елементів висувають певні вимоги. Елементи повинні:

1. бути просторово локалізованими;
2. належати до одного ієрархічного рівня;
3. відповідати масштабу дослідження.

До вибору ознак також висувають певні вимоги. Ознаки повинні:

1. відображати суттєві риси елементів;
2. характеризувати елементи з різних боків;
3. мати розкид навколо середнього значення.

Традиційно ми вимагаємо, щоб усі ознаки, взяті для багатовимірного аналізу були відносними, а не абсолютними.

3. Методи стандартизації ознак

Потреба у стандартизації викликана тим, що досліджувані ознаки як правило є різновимірними та різномасштабними.

На основі матриці вхідних даних $X = \{x_{ij}\}$ обчислюють матрицю стандартизованих даних $\hat{X} = \{\hat{x}_{ij}\}$. Є різні методи стандартизації. За методом, який називають методом нормалізації:

$$\hat{x}_{ij} = \frac{x_{ij} - \bar{x}_j}{\sigma_j},$$

де $\bar{x}_j = \frac{1}{M}\sum_{i=1}^{M} x_{ij}$ - середнє значення j-го стовпця матриці даних,

$\sigma_j = \sqrt{\frac{1}{M}\sum_{i=1}^{M}(x_{ij} - \bar{x}_j)^2}$ - середньоквадратичне відхилення в j-му стовпці матриці даних.

Знайдені таким методом нормалізовані дані завжди задовольняють наступні дві умови, які можна використовувати для контролю правильності обчислень:

1. $\sum_{i=1}^{M} \hat{x}_{ij} = 0$, (сума чисел у кожному стовпці нормалізованої матриці рівна нулю).
2. $\sum_{i=1}^{M} \hat{x}_{ij}^{\ 2} = M$, (сума квадр. чисел у кожному стовпці нормал. матриці рівна кількості територіальних елементів).

4. Методи метризації елементів об'єкта

Метризація полягає у визначенні матриці таксономічних відстаней $D = \{d_{ik}\}, \quad i, k = 1,...,M$ між усіма територіальними елементами в багатовимірному просторі ознак.

Матриця таксономічних відстаней – це квадратна матриця, яка має стільки рядків і стовпців, скільки є елементів у досліджуваному об'єкті. Числами матриці є відстані у багатовимірному просторі ознак між точками-образами елементів об'єкта.

Матриця таксономічних відстаней має такий загальний вигляд:

$$D = \begin{pmatrix} 0 & d_{12} & \cdots & d_{1M} \\ d_{21} & 0 & \cdots & d_{2M} \\ \vdots & \vdots & \ddots & \vdots \\ d_{M1} & d_{M2} & \cdots & 0 \end{pmatrix}$$

Існують різні методи знаходження таксономічних відстаней. Розглянемо найбільш уживаний. Цей метод називається Евклідовою метрикою, він випливає з формули Піфагора для N-вимірного простору і в цьому випадку:

$$d_{ik} = \sqrt{\sum_{j=1}^{N} (\hat{x}_{ij} - \hat{x}_{kj})^2} \ .$$

З цієї формули випливає, що практично, для кожної пари територіальних елементів,обчислення виконується для двох відповідних рядків нормалізованої матриці даних, а сумування здійснюється вздовж рядка по всіх ознаках.

5. Таксономізація об'єктів. Метод дерева поєднань

Метризація полягає у визначенні матриці таксономічних відстаней $D = \{d_{ik}\}, \quad i, k = 1,..., M$ між усіма територіальними елементами в багатовимірному просторі ознак.

Матриця таксономічних відстаней – це квадратна матриця, яка має стільки рядків і стовпців, скільки є елементів у досліджуваному об'єкті. Числами матриці є відстані у багатовимірному просторі ознак між точками-образами елементів об'єкта

d_{ik} - таксономічна відстань між елементами з номерами i та k.

Матриця таксономічних відстаней має такі властивості:

1. Усі числа матриці додатні, крім діагональних, а діагональні рівні нулю. Символічно: $d_{ik} > 0$, $d_{ii} = 0$.

2. Матриця симетрична відносно головної діагоналі. Символічно: $d_{ki} = d_{ik}$.

3. Числа матриці задовольняють нерівність трикутника. Символічно: $d_{il} + d_{lk} \ge d_{ik}$.

Отже, матриця таксономічних відстаней має такий загальний вигляд:

$$D = \begin{pmatrix} 0 & d_{12} & \cdots & d_{1M} \\ d_{21} & 0 & \cdots & d_{2M} \\ \vdots & \vdots & \ddots & \vdots \\ d_{M1} & d_{M2} & \cdots & 0 \end{pmatrix}.$$

Існують різні методи знаходження таксономічних відстаней. Розглянемо найбільш уживаний. Цей метод називається Евклідовою метрикою, він випливає з формули Піфагора для N-вимірного простору і в цьому випадку:

$$d_{ik} = \sqrt{\sum_{j=1}^{N} (\hat{x}_{ij} - \hat{x}_{kj})^2}\ .$$

З цієї формули випливає, що для кожної пари територіальних елементів, обчислення виконується для двох відповідних рядків нормалізованої матриці даних, а сумування здійснюється вздовж рядка по всіх ознаках.

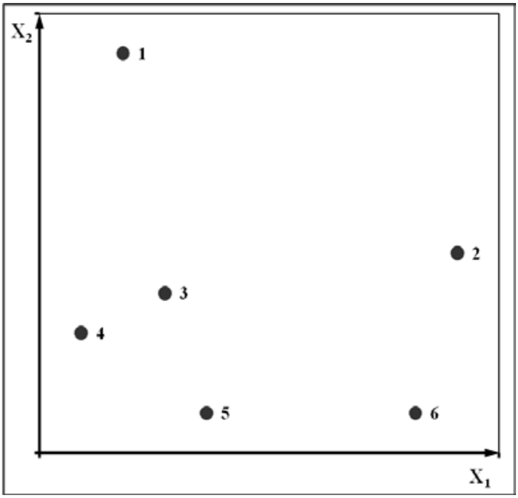
Приступимо до таксономізації.

Перед тим, як розглянути загальний випадок з довільною кількістю елементів та ознак, розглянемо демонстраційний приклад, де є 6 елементів та 2 ознаки. Кількість ознак рівна двом вибрана для того, щоб багатовимірний простір був двовимірним і його можна було зобразити на площині.

Побудуємо систему координат, у якій горизонтальна вісь представляє першу ознаку, а вертикальна вісь – другу ознаку. Зобразимо точкові образи шести елементів у цій системі координат і припустимо, що отримається такий графік, як на рисунку. Пронумеруємо точки: 1,2,3,4,5,6.

Спочатку знайдемо дві найближчі точки (у нас 3 і 4) і об’єднаємо їх у нову точку 7 з середнім розташуванням. Тепер маємо п’ять точок 1,2,5,6,7.

Знову знаходимо дві найближчі точки (тепер це 5 і 7) і об’єднуємо їх у нову точку 8 з середнім розташуванням. Тепер маємо чотири точки 1,2,6,8.



Знову знаходимо дві найближчі точки (тепер це 2 і 6) і об’єднуємо їх у нову точку 9 з середнім розташуванням. Тепер маємо три точки 1,8,9.

Знову знаходимо дві найближчі точки (тепер це 8 і 9) і об’єднуємо їх у нову точку 10 з середнім розташуванням. Тепер маємо дві точки 1,10.

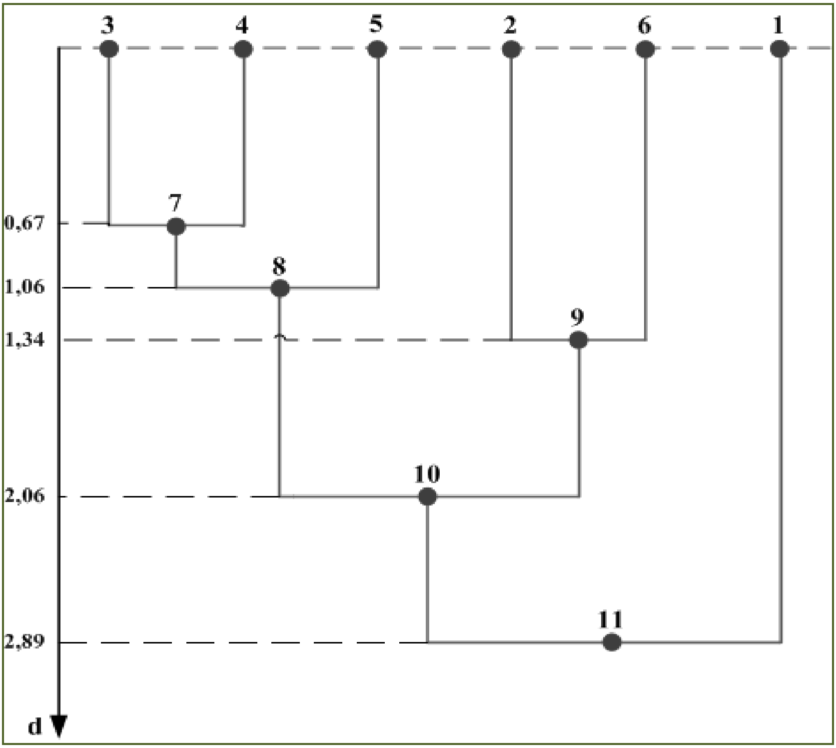
Остаточно об’єднуємо точки 1 і 10 у нову точку 11 з середнім розташуванням.

Перейдемо до побудови графа “дерево поєднань”.

Записуємо результати цього процесу в таблицю, яка містить номер кроку, номери об’єднуваних точок, відстань між точками і номер результуючої точки.

Номер кроку	Перша точка	Друга точка	Відстань між точками	Результуюча точка
1	3	4	0,67	7
2	5	7	1,06	8
3	2	6	1,34	9
4	8	9	2,06	10
5	1	10	2,89	11

На основі цієї таблиці будуємо граф "дерево поєднань". Проводимо горизонтальну лінію, на якій розміщуємо точкові образи елементів. Проводимо вертикальну вісь, яка зображає відстані між точковими образами. Далі відкладаємо точки на горизонтальній лінії, з'єднуємо їх дугами потрібної довжини і показуємо виникнення нових точок. Довжина дуги відповідає відстані між точками.



В результаті отримуємо граф "дерево поєднань". Перейдемо до аналізу графа. Для цього вибираємо деяку відстань і для цієї відстані горизонтальною лінією розрізаємо граф. Частини графа, які відпадають при розрізанні, вказують на таксони. Якщо повернутись до початкового розміщення точок, то видно, що отримані таксони помітні візуально.

Дуже важливим є питання про вибір лінії розрізу.

З одного боку, цей вибір знаходиться в руках дослідника, який має певну свободу, що повинна спиратися на професійний досвід. З другого боку, при виборі лінії розрізу, варто пам'ятати кілька правил:

1. Якщо лінію розрізу вибрати для замалого значення відстані, то кожен елемент буде окремим таксоном, що малоінформативно.

2. Якщо лінію розрізу вибрати для завеликого значення відстані, то майже всі елементи можуть попасти в один або два таксони, що також малоінформативно.

Отже, лінію розрізу потрібно вибирати для деякого середнього оптимального значення відстані.

3. З кількох можливих ліній розрізу доцільно вибрати ту, для якої таксони при картографічному зображенні є найбільш компактними, а в ідеалі утворюють цілісні територіальні угрупування.

Тепер розглянемо загальний випадок, коли кількість елементів і ознак є довільною.

Матриця	Розмір	Мінімальна додатна відстань
D	M x M	d_{ik}
D’	(M-1) x (M-1)	$d_{i'k'}$
D”	(M-2) x (M-2)	$d_{i''k''}$
…	…	…
	2 x 2	

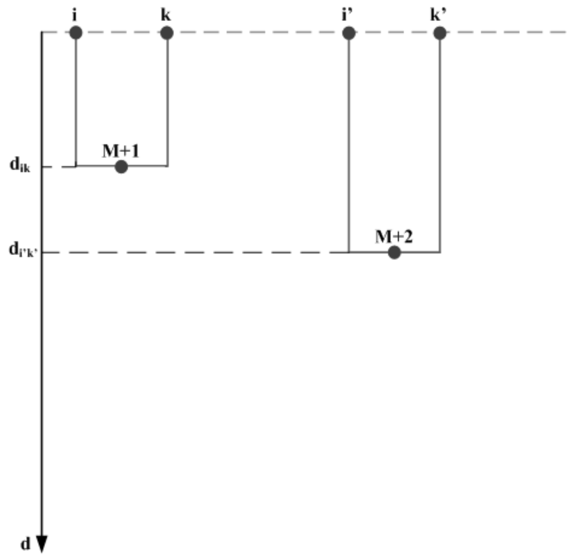
Розглядаємо матрицю таксономічних відстаней D між елементами.

Вона має розмір $M \times M$. Приступимо до таксономізації.

Спочатку знайдемо в матриці найменше додатнє число. Нехай це буде d_{ik} , тобто таксономічна відстань між елементами з номерами i та k . Викреслюємо з матриці i-й та k-й рядки, а також i-й та k-й стовпці. Дописуємо до матриці новий рядок і новий стовбець з номером $M+1$, які є півсумою викреслених. Отримаємо нову матрицю D' розміром $(M-1)\times(M-1)$.

У цій матриці знову знаходимо найменше додатнє число. Нехай це буде $d_{i'k'}$. Далі знову викреслюємо з матриці рядки та стовпці з номерами i' та k', і дописуємо до матриці новий рядок і новий стовбець з номером $M+2$, які є півсумою викреслених. У результаті отримаємо нову матрицю розміром $(M-2)\times(M-2)$.

Цю процедуру продовжуємо до тих пір, поки отримаємо матрицю розміром 2×2.



Далі переходимо до побудови графа "дерево поєднань".

На горизонтальній лінії відкладаємо номери i та k і з'єднуємо їх дугою довжиною d_{ik}. Далі відкладаємо номери i' та k' і з'єднуємо їх дугою довжиною $d_{i'k'}$. Цей процес продовжуємо до повної побудови графа.

Для географічної інтерпретації результати таксономізації відображають картографічно. Якщо вивчаються ареальні елементи, то найчастіше використовують картографічний спосіб якісного фону.

У цьому випадку ареальні елементи, що належать до одного таксону, позначають одним кольором, а різні таксони позначають різними кольорами.

У такій картосхемі дуже важливим є блок умовних позначень, бо він власне показує джерела якісної різниці між таксонами. Для цього аналізують початкову матрицю даних і роблять висновки, чим саме за сукупністю значень ознак, один таксон відрізняється від інших таксонів.

Тестові завдання

1. *Скільки стовпців має матриця даних ?*

1. кількість елементів 2. кількість рівнів 3. один. 4. кількість ознак 5. кількість рядків

2. *Скільки рядків має матриця даних ?*

1. кількість елементів 2. кількість рівнів 3. кількість ознак 4. один 5. кількість стовпців

3. *Як називається група близьких точок у багатовимірному просторі ознак ?*

1. коаліція 2. таксон 3. сім'я 4. Колектив 5. союз

4. *Якою є матриця таксономічних відстаней ?*

1. симетричною 2. одиничною 3. діагональною 4. круглою 5. трикутною

5. *Якою є матриця таксономічних відстаней ?*

1. квадратною 2. одиничною 3. діагональною 4. круглою 5. трикутною

6. *Скільки стовпців має матриця таксономічних відстаней ?*

1. кількість елементів 2. кількість рівнів 3. один. 4. кількість ознак 5. кількість рядків

7. Скільки рядків має матриця таксономічних відстаней ?

1. кількість елементів 2. кількість рівнів 3. кількість ознак 4. один 5. кількість стовпців

8. *Який спосіб картографічного зображення використовують для географічної інтерпретації результатів багатовимірного аналізу ?*

1. ізоліній 2. крапковий 3. ліній руху 4. якісного фону 5. картодіаграми

Тема 7

КІЛЬКІСНІ МЕТОДИ ЧАСОВОГО ПРОГНОЗУВАННЯ

План

1. Точна лінійна часова екстраполяція
2. Точна квадратична часова екстраполяція
3. Регресійна часова лінійна екстраполяція
4. Регресійна часова квадратична екстраполяція

1. Точна лінійна часова екстраполяція

Нехай відомі значення кількісної ознаки за два моменти часу. Запишемо цю інформацію у вигляді такої таблиці:

Момент часу	t_0	t_1
Значення ознаки	w_0	w_1

Ставиться задача здійснити прогноз w для моменту часу $t > t_1$.

Для розв’язання цієї задачі існує точна лінійна формула:

$$w_{прогн}(t) = w_1 + D_{10} \cdot (t - t_1),$$

де $D_{10} = \dfrac{w_1 - w_0}{t_1 - t_0}$ - швидкість зміни показника.

Наведена формула називається точною, бо в моменти часу t_0, t_1 ця формула дає точні значення w_0, w_1.

2. Точна квадратична часова екстраполяція

Нехай відомі значення кількісної ознаки за три моменти часу. Запишемо цю інформацію у вигляді такої таблиці:

Момент часу	t_0	t_1	t_2
Значення ознаки	w_0	w_1	w_2

Ставиться задача здійснити прогноз w для моменту часу $t > t_2$.

Для розв'язання цієї задачі існує точна квадратична формула:

$$w_{прогн}(t) = w_2 + D_{21} \cdot (t - t_2) + D_{210} \cdot (t - t_2) \cdot (t - t_1),$$

де $D_{210} = \dfrac{D_{21} - D_{10}}{t_2 - t_0}$, $D_{21} = \dfrac{w_2 - w_1}{t_2 - t_1}$, $D_{10} = \dfrac{w_1 - w_0}{t_1 - t_0}$.

Наведена формула називається точною, бо в моменти часу t_0, t_1, t_2 ця формула дає точні значення w_0, w_1, w_2. Однак, слід пам'ятати, що хоча квадратична прогнозна формула є по-суті точнішою від лінійної, але вона дуже чутлива до похибок у даних, тобто малі похибки в даних можуть привести до значних похибок у прогнозі. Тому її недоцільно використовувати в тих випадках, коли відомо, що початкові дані містять навмисні чи ненавмисні похибки.

Крім точної лінійної та квадратичної екстраполяції можна аналогічно, маючи значення ознаки за більшу кількість моментів часу, побудувати формули для екстраполяцій вищих порядків. Ми не будемо цього робити з трьох причин: по-перше, такі формули достатньо громіздкі, по-друге, вони також чутливі до похибок, по-третє, при більшій кількості даних чудово працюють методи регресійної екстраполяції, про які мова піде далі.

3. Регресійна часова лінійна екстраполяція

Нехай відомі значення кількісної ознаки за багато моментів часу $t_1, t_2, ..., t_n$. Запишемо цю інформацію у вигляді такої таблиці:

Момент часу	t_1	t_2	$\cdots$	t_n
Значення ознаки	w_1	w_2	$\cdots$	w_n

Ставиться задача здійснити прогноз значень w для момента часу $t > t_n$.

Для розв'язання цієї задачі існує регресійна формула:

$$w_{прогн}(t) = \overline{w} + a \cdot (t - \bar{t}),$$

де $\bar{t} = \dfrac{1}{n} \cdot \sum_{j=1}^{n} t_j$, $\overline{w} = \dfrac{1}{n} \cdot \sum_{j=1}^{n} w_j$,

$$a = \frac{S_{tw}}{S_{tt}}, \quad S_{tt} = \sum_{j=1}^{n} (t_j - \bar{t})^2, \quad S_{tw} = \sum_{j=1}^{n} (t_j - \bar{t}) \cdot (w_j - \overline{w}).$$

Потрібно відзначити, що при обчисленні коефіцієнтів регресійної формули виконуються декілька операцій усереднення та сумування. Це дає змогу певною мірою взаємно погасити вплив можливих похибок у початкових даних. Саме тому регресійна формула має ту перевагу, що вона малочутлива до випадкових похибок у даних.

4. Регресійна часова квадратична екстраполяція

Нехай відомі значення кількісної ознаки за багато моментів часу $t_1, t_2, ..., t_n$. Запишемо цю інформацію у вигляді такої таблиці:

Момент часу	t_1	t_2	$\cdots$	t_n
Значення ознаки	w_1	w_2	$\cdots$	w_n

Ставиться задача здійснити прогноз значень ознаки w для моменту часу $t > t_n$.

Для розв'язання цієї задачі існує регресійна формула:

$$w_{прогн}(t) = \overline{w} + a \cdot (t - \bar{t})^2 + b \cdot (t - \bar{t}) + c,$$

У загальному випадку формули для коефіцієнтів цієї формули є досить складними, однак, якщо моменти часу $t_1, t_2, ..., t_n$ слідують один за одним через однакові проміжки (а така ситуація дуже часто трапляється в суспільно-географічних дослідженнях), то для визначення коефіцієнтів регресійного рівняння можна скористатися формулами:

$$a = \frac{S_{ttw}}{S_{tttt} - \frac{S_{tt}^2}{n}}, \qquad b = \frac{S_{tw}}{S_{tt}}, \qquad c = -\frac{S_{tt}}{n} \cdot a.$$

Для неоднакових проміжків між моментами часу $t_1, t_2, ..., t_n$, ці формули недійсні.

У цілому квадратична регресійна формула дає точніший прогноз ніж лінійна регресійна і вона також малочутлива до похибок даних.

Тестові завдання

1. *Скільки спостережень потрібно мати для точного лінійного прогнозу ?*

1. жодного 2. два 3. один 4. три 5. довільну кількість

2. *Скільки спостережень потрібно мати для точного квадратичного прогнозу ?*

1. два 2. довільну кількість 3. чотири 4. три 5. жодного

3. *Яка екстраполяція моделює прямолінійне зростання показника для багатьох спостережень ?*

1. точна лінійна 2. точна квадратична 3. гармонійна 4. регресійна лінійна 5. регресійна квадратична

4. *Яка екстраполяція моделює параболічне зростання показника для багатьох спостережень ?*

1. точна квадратична 2. регресійна квадратична 3. гармонійна 4. точна лінійна 5. регресійна лінійна

5. *Яка екстраполяція є чутливою до похибок у даних ?*

1. точна лінійна 2. точна квадратична 3. гармонійна 4. регресійна лінійна 5. регресійна квадратична

Рекомендована література

Базова

1. *Шаблій О. І.* Математичні методи в соціально-економічній географії. –Львів: Світ, 1994.

2. *Грицевич В.С.* Математичні моделі в демогеографії. Текст лекції. –Львів: Вид. центр ЛНУ імені Івана Франка. -2003.

3. *Голиков А. П., Черваньов І. Г., Трофимов А .М.* Математичні методи в географії. – Харків: вид-во при Харків, ун-ті. 1986.

4. *Глівенко С.*В., Соколов М.О., Теліженко О.М. Економічне прогнозування: навчальний посібник. – Суми: ВТД Університетська книга, 2004.

Допоміжна

5. Грицевич В.С. Підходи до математико-географічного вивчення суспільно-географічного положення // Наукові записки Тернопільського педуніверситету. Серія: географічна. №2, 2000. -Тернопіль, -С.46-50.

6. Грицевич В.С. Моделювання територіальних систем у суспільно-географічній науці // Вісник Львів. ун-ту. Сер. географія. Вип. 29. Ч.2. -Львів, 2003. -С.8-12.

7. Грицевич В.С. Математико-географічна оцінка господарської спеціалізації регіону // Наукові записки Тернопільського педуніверситету. Серія: географічна. №1, 2003. -Тернопіль, -С.72-76.

8. Грицевич В.С. Картографічна інтерпретація багатовимірної таксономізації в суспільно-географічних дослідженнях // Вісник геодезії та картографії. №3, 2005. –С.34-38.

9. Грицевич В.С. Математико-географічні основи менеджменту демографічною ситуацією регіону // Матер. міжнар. наук.-практ. конф. „Регіон-2007: Стратегія оптимального розвитку”. –Харків: Вид. центр ХНУ, 2007. С.28-31.

10. Грицевич В.С. Узагальнення правила Ціпфа для суспільно-географічних досліджень // Науковий вісник Волинського національного ун-ту імені Лесі Українки. №1, 2008. –Луцьк: РВВ "Вежа". –С.78-83.

11. Грицевич В.С. Моделювання функціональних закономірностей в географії суспільства // Часопис соціально-економічної географії. -Вип. 4(1). -Харків, 2008. -С.41-44.

12. Грицевич В.С. Система ієрархічних форм геопросторової організації територіальних суспільних комплексів // Часопис соціально-економічної географії. -Вип. 5(1). -Харків, 2009. -С.55-61.

ЗМІСТ

Навчально-методичне видання

Грицевич Володимир Степанович

Математичні методи в суспільній географії:
тексти лекцій для студентів заочної форми навчання
напряму підготовки 6.040104 − географія

Підготовлено до друку 19.04.2013 формат 60*84/16
Умовн. друк.арк.
Наклад 100 прим.

Малий видавничий центр
Лабораторія тематичного картографування географічного факультету,
Львівський національний університет імені Івана Франка
Україна, 79000, Львів, вул. П.Дорошенка, 41