

## ВИКОРИСТАННЯ ГІС-ТЕХНОЛОГІЙ ПРИ СОЦІОЛОГІЧНИХ ДОСЛІДЖЕННЯХ

Купач Д. В.  
ТОВ «Арт-мастер»

Мета доповіді – представлення методики ГІС-обробки соціологічних даних (результатів опитувань, голосувань), на прикладі дослідження просторових закономірностей розподілу голосів виборців на виборах до Верховної Ради України 26 березня 2006р.

Вихідні дані взяті з сайту ЦВК <http://www.cvk.gov.ua/>

Першим кроком було представлення даних у вигляді тематичних карт. Карта регіонів дозволила отримати візуальну оцінку результатів по 27 об'єктах, а ось огляд карти виборчих округів засвідчив, що візуальний аналіз 225 об'єктів викликає ускладнення навіть у досвідчених спеціалістів.

Наступними діями були інтерполяція даних для відображення їх на карті у вигляді ізоліній та формування бази даних для статистичного аналізу.

Для середньомасштабних досліджень інтерполяція проводилася за даними по виборчих округах методом триангуляції. В такому випадку чітко відображаються місцеві особливості кожного регіону. Для більш дрібного масштабу було застосовано метод інтерполяції даних по регіонах. Метод відомий з давніх-давен, зупинятися варто, хіба що, на деяких його особливостях стосовно даного дослідження.

Вихідні дані ті ж, що і для побудови тематичної карти - база даних, де кількість записів дорівнює кількості регіонів (у нашому випадку - 27). Для кожного запису існує кількісне значення показника, що досліджується. Будується палетка, регулярна мережа рівновіддалених контрольних точок (КТ), що рівномірно покриває всю карту території. Кожна КТ – представлена в базі даних записом з необхідною кількістю полів. Значення кожного поля – фактичний (задокументований), або вирахований показник. Відстань між точками (шаг мережі) мусить бути як найменшою. Але якщо всі сусідні точки потрапляють в межі одного регіону, то інтерполяція втрачає сенс. Для України доцільно використовувати шаг 80-100 км, що дозволяє розташувати по всій території понад 100 точок. Кількість КТ в регіоні повинна бути пропорційна його площі. Один з засобів перевірки – вирахування коефіцієнту кореляції між значенням площі регіону та кількістю КТ в регіоні.



Мал. 1. Система контрольних точок

Кожна точка, первинно, набуває значення кількісного показника того регіону, в якому вона розташована. Далі, для кожної КТ вираховується інтегральне значення, в якому враховано значення всіх сусідніх КТ, що знаходяться на відстані шагу мережі. В

наведеному прикладі це середнє арифметичне, але можливі варіанти (напр. з уточнюючими коефіцієнтами для центральних точок).

$$P_{fin} := (Pr + P1 + P2 + \dots Pn) / n;$$

$P_{fin}$  – остаточне значення показника в точці.

$Pr$  – значення в регіоні, до якого належить точка

$P1 + P2 + \dots Pn$  – значення в регіонах для (сусідніх) точок

$n$  – кількість сусідніх точок + 1

Після розрахунків отримуємо мережу точок, кожна з яких характеризується парою координат та значенням. Фактично – базу даних карти гіпотези.

Власне, гіпотетичним є припущення, що формування показника у суміжній зоні двох сусідніх регіонів відбувається під впливом одних і тих самих чинників. Ми вважаємо, що у даному випадку, коли досліджується результат волевиявлення людей у межах однієї держави, то таке припущення правомірне.

Для перевірки відповідності фактичних та розрахованих за даною методикою даних було вираховано коефіцієнт парної кореляції між показниками двох серій карт:

1) побудована за інтерпольованими даними ЦВК по виборчих округах (225 записів даних).

2) побудована за даними ЦВК по регіонах (27 записів даних)

До кожної серії увійшли карти розподілу показників 5 партій, що за результатами виборів пройшли до Верховної ради

Зайняте партією місце	Коефіцієнт кореляції
1	0.97
2	0.95
3	0.96
4	0.86
5	0.85

Це є свідченням істотної відповідності даних карти-гіпотези до фактичних даних, та підтверджує правомірність методу.

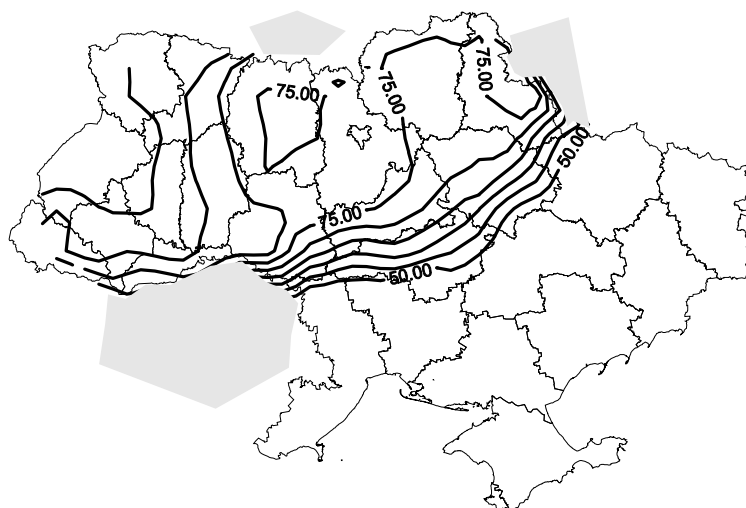
Даний метод приваблює тим, що на розподіл кількісних показників впливають не стільки внутрішньорегіональні особливості (в таких випадках слід проводити середньо- та крупномасштабні дослідження), а більш глобальні, ті, що проявляються в межах країни.

В ході дослідження базу даних по КТ було розширено до кількох десятків полів з кількісними даними.

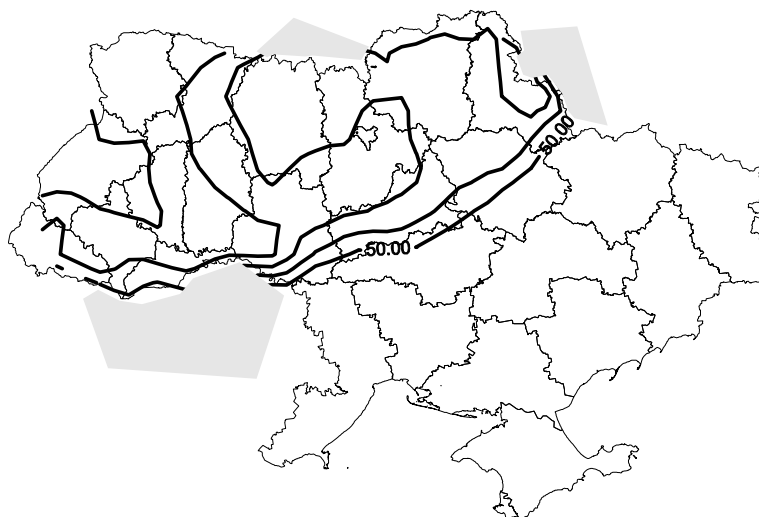
1. Дані по результатам президентських виборів
2. Відсоток голосів виборців відданих за партії, що зайняли перші 10 місць
  - a. По виборчих округах
  - b. По регіонах
3. Відсоток голосів відданих за партії та блоки, що не пройшли
4. Показники статистичних функцій (кореляція, ентропія)

На даний момент формування бази продовжується.

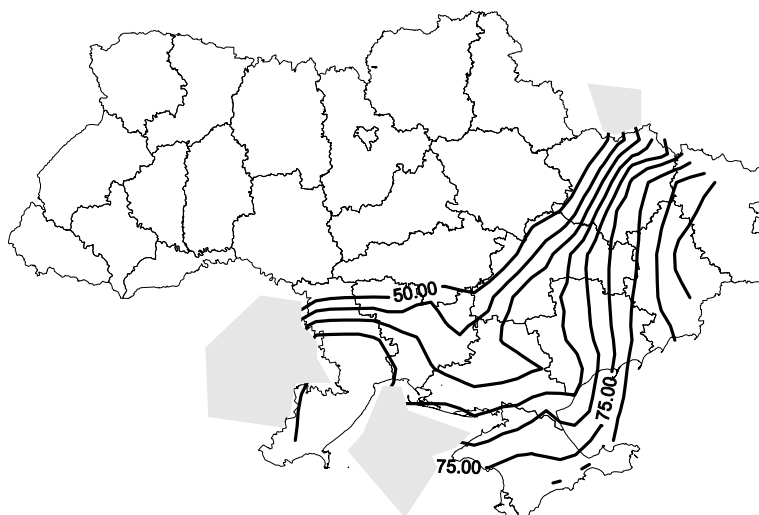
Для ілюстрації пропонується кілька карт



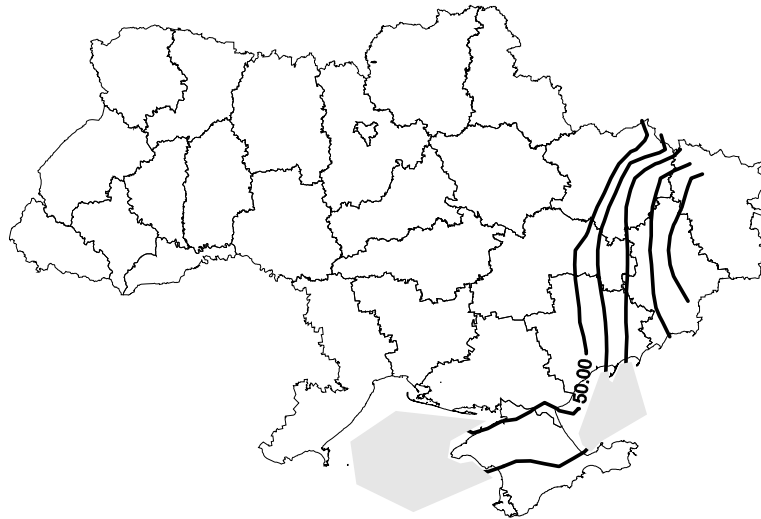
Мал. 2. Результати виборів президента України 2004 р. Розподіл голосів виборців, що проголосували за В. Ющенка (> 50%)



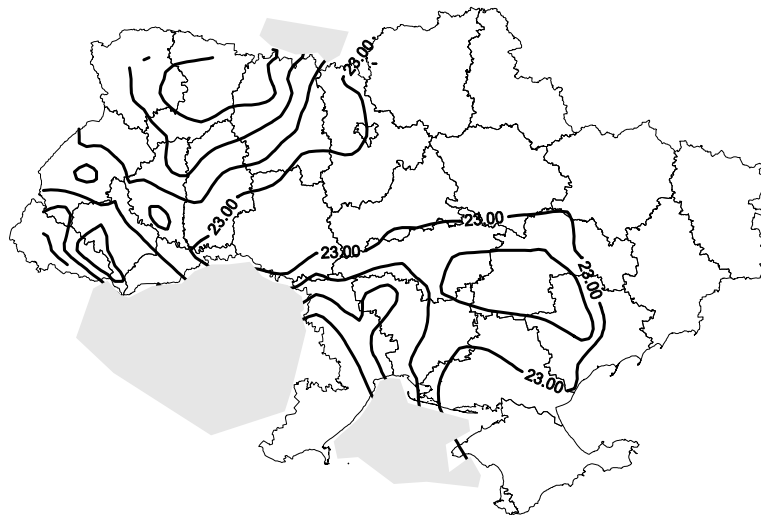
Мал. 3. Результати виборів до ВР України 2006 р. Розподіл голосів виборців, що проголосували за БЮТ, БНУ, СПУ (в сумі > 50%)



Мал. 4. Результати виборів президента України 2004 р. Розподіл голосів виборців, що проголосували за В. Януковича (> 50%)



Мал. 5. Результати виборів до ВР України 2006 р. Розподіл голосів виборців, що проголосували за ПР (> 50%)



Мал. 6. Результати виборів до ВР України 2006 р. Розподіл голосів виборців, чії голоси були віддані за партії, що не пройшли до ВР

Закономірності розподілу голосів виборців обох блоків 2004 року повністю збереглися і у 2006. Коефіцієнти кореляції в обох випадках дорівнюють 0.98. А ось кількісно обидва блоки втратили. Значний відсоток голосів виборців відійшов до партій та блоків, що не подолали поріг 3%.

Негативна кореляція при порівнянні карт помітна і без врахування коефіцієнту.

Запропонована методика (інтерполяція) дозволяє перейти до способу представлення даних методом ізоліній та унаочнити інформацію, що на тематичній карті не проявляється. Особливо цікавими для аналізу є лінії екстремуму, своєрідні «гальвеги» та «вододіли».

Використання БД по контрольних точках дозволяє ширше використовувати методики статистичного аналізу. В нашому випадку збір даних проводиться по вибірці з 113 КТ, що значно точніше ніж традиційне використання 27 регіонів.

Ефективними можуть стати статистичні дослідження – кореляційний та діахронічний види аналізу, пошук однорідності, районування, тощо.

Найбільш цікавою видається перспектива побудови моделі на підставі даних самих різних виборів, референдумів, опитувань за багато років.