

**МІНІСТЕРСТВО ОСВІТИ І НАУКИ УКРАЇНИ**  
**Західноукраїнський національний університет**  
**Факультет комп'ютерних інформаційних технологій**  
**Кафедра комп'ютерної інженерії**

**ЦИМБАЛЮК Дмитро Вікторович**

**«Економетричні моделі локалізації українських  
ІТ компаній / Econometric models of localization of  
Ukrainian IT companies»**

спеціальність: 123 - Комп'ютерна інженерія  
освітньо-професійна програма - Комп'ютерна інженерія  
Кваліфікаційна робота

Виконав студент групи КІм-22  
Д.В. Цимбалюк

---

Науковий керівник:  
д.т.н., проф. О.М. Березький

---

Кваліфікаційну роботу допущено  
до захисту:

" \_\_\_\_ " \_\_\_\_\_ 20\_\_ р.

Завідувач кафедри  
\_\_\_\_\_ О. М. Березький

**Тернопіль – 2021**

Західноукраїнський національний університет  
Факультет комп'ютерних інформаційних технологій  
Кафедра комп'ютерної інженерії  
Освітній ступінь «магістр»  
спеціальність: 123 – Комп'ютерна інженерія  
освітньо-професійна програма – Комп'ютерна інженерія

ЗАТВЕРДЖУЮ

Завідувач кафедри

\_\_\_\_\_ О.М. Березький

“ \_\_\_\_ ” \_\_\_\_\_ 20\_\_ р.

**ЗАВДАННЯ**  
**НА КВАЛІФІКАЦІЙНУ РОБОТУ СТУДЕНТУ**

Цимбалюку Дмитру Вікторовичу

---

(прізвище, ім'я, по батькові)

1. Тема кваліфікаційної роботи «Економетричні моделі локалізації українських ІТ компаній / Econometric models of localization of Ukrainian IT companies»  
керівник роботи д.т.н., проф. О.М. Березький  
затверджені наказом по університету від 10 грудня 2020 р. № 167.
2. Строк подання студентом закінченої кваліфікаційної роботи 6 грудня 2021 року
3. Вихідні дані до кваліфікаційної роботи  
Об'єкт дослідження – процес аналізу українських ІТ компаній.  
Предмет дослідження – моделі просторової економетрики.
4. Основні питання, які потрібно розробити
  - проаналізувати розвиток ІТ-індустрії в Україні;
  - провести аналіз економетричних моделей аналізу даних;
  - проаналізувати програмні засоби реалізації економетричних моделей;
  - розробити алгоритм побудови матриці просторових ваг для областей України;
  - розробити алгоритм обчислення показників просторової кореляції;
  - розробити алгоритм кластеризації областей України за кількістю ІТ-компаній;
  - здійснити програмну реалізацію розроблених алгоритмів;
  - провести тестування розроблених алгоритмів.
5. Перелік графічного матеріалу у роботі:

- кореляційний і регресійний аналіз;
- матриці просторових ваг;
- основні коефіцієнти для дослідження просторових ефектів;
- алгоритм побудови моделі просторового лагу;
- програмна система просторової економетрики;
- комп'ютерні експерименти;
- висновки.

#### 6. Консультанти розділів кваліфікаційної роботи

| Розділ         | Прізвище, ініціали та посада консультанта | Підпис, дата   |                  |
|----------------|---|----------------|------------------|
|                |   | Завдання видав | Завдання прийняв |
| Нормо-контроль | Мельник Г.М., доцент                      |                |                  |

7. Дата видачі завдання 26 жовтня 2020 р.

#### КАЛЕНДАРНИЙ ПЛАН

| № з/п | Назва етапів кваліфікаційної роботи            | Строк виконання етапів кваліфікаційної роботи | Примітка |
|-------|--|---|----------|
| 1     | Перший розділ                                  | 26.11.2020 – 1.04.2021                        |          |
| 2     | Другий розділ                                  | 2.04.2021 – 4.07.2021                         |          |
| 3     | Третій розділ                                  | 5.07.2021 – 26.11.2021                        |          |
| 4     | Нормоконтроль, попередній захист               | 27.11.2021 – 3.12.2021                        |          |
| 5     | Подача в бібліотеку ЗУНУ електронного варіанту | 6.12.2021                                     |          |
| 6     | Захист   | 20.12 – 30.12.2021                            |          |

Студент \_\_\_\_\_ Цимбалюк Д. В.  
( підпис )

Керівник роботи \_\_\_\_\_ д.т.н., професор, О.М. Березький  
( підпис )

# ЗМІСТ

|  |  |
|--|--|
| ВСТУП.....   | 6                                      |
| 1 АНАЛІЗ ЕКОНОМЕТРИЧНИХ МОДЕЛЕЙ ОЦІНКИ РОЗТАШУВАННЯ УКРАЇНСЬКИХ ІТ-КОМПАНІЙ..... | 9                                      |
| 1.1 Розвиток ІТ-індустрії в Україні.....   | 9                                      |
| 1.2 Економетричні моделі аналізу даних.....                                      | 22                                     |
| 1.3 Аналіз програмних засобів реалізації економетричних моделей .....            | 28                                     |
| 1.4 Постановка задач дослідження.....  | 36                                     |
| 1.5 Висновки до розділу 1 .....  | 37                                     |
| 2 АЛГОРИТМИ ПОБУДОВИ МОДЕЛЕЙ ОЦІНКИ РОЗТАШУВАННЯ УКРАЇНСЬКИХ ІТ-КОМПАНІЙ.....    | 38                                     |
| 2.1 Алгоритми побудови матриці просторових ваг.....                              | 38                                     |
| 2.2 Алгоритми обчислення показників просторової кореляції.....                   | 44                                     |
| 2.3 Побудова просторових моделей.....  | 50                                     |
| 2.4 Висновки до розділу 2 .....  | 56                                     |
| 3 КОМП'ЮТЕРНЕ МОДЕЛЮВАННЯ АЛГОРИТМІВ ПРОСТОРОВОЇ ЕКОНОМЕТРИКИ..                  | 57                                     |
| 3.1 Вибір програмних засобів та системні вимоги .....                            | 57                                     |
| 3.2 Структура програмної системи .....   | 61                                     |
| 3.3 Модуль побудови матриці просторових ваг .....                                | 64                                     |
| 3.4 Модуль отримання показників просторової автокореляції.....                   | 65                                     |
| 3.6 Висновки до розділу 3 .....  | 73                                     |
| ВИСНОВКИ .....   | 74                                     |
| СПИСОК ВИКОРИСТАНИХ ДЖЕРЕЛ.....  | 75                                     |
| Додаток А .....  | <b>Ошибка! Закладка не определена.</b> |
| Таблиці статистичних даних.....  | <b>Ошибка! Закладка не определена.</b> |
| Додаток Б.....   | <b>Ошибка! Закладка не определена.</b> |
| Матриці просторових ваг. Матриця граничних сусідів. Матриця відстаней.           | <b>Ошибка! Закладка не определена.</b> |
| Додаток В.....   | <b>Ошибка! Закладка не определена.</b> |
| Довідка про використання .....   | <b>Ошибка! Закладка не определена.</b> |
| Додаток Г .....  | <b>Ошибка! Закладка не определена.</b> |
| Світлокопії виданих публікацій.....  | <b>Ошибка! Закладка не определена.</b> |

## ВСТУП

Актуальність роботи. Аналіз даних за останній період часу набуває неабиякої популярності. Це зв'язано з появою великої кількості інформації (зібраних даних). Поряд з аналізом даних набув популярності дослідницький аналіз просторових даних. Це процес дослідження з метою виявлення значних характеристик просторових даних, він включає явні методи та процедури, що беруть до уваги просторовий контекст географічних даних. Аналітики досліджують просторові закономірності, асоціації та процеси в одновимірних та багатовимірних даних, використовуючи інструменти динамічної та інтерактивної візуалізації, щоб потенційно виявити раніше невідомі асоціації та сформувані гіпотези. Хоча концепція та загальна процедура аналізу просторових даних відносно мало змінилися з кінця 1990-х років, багато атрибутів наборів просторових даних змінилися. Зокрема, різко збільшилися розмір та частота, з якою просторові дані стають загальнодоступними. Таким чином, нещодавній швидкий сплеск даних в даний час висуває підвищені вимоги до програмних додатків для обробки великих багатовимірних просторових наборів даних [1]. Тому тема кваліфікаційної роботи є актуальною.

Мета і завдання дослідження. Метою кваліфікаційної роботи є розробка алгоритмів побудови економетричних моделей локалізації українських ІТ компаній і програмування модуля побудови економетричних моделей.

У відповідності із поставленою метою кваліфікаційна робота включає розв'язки таких задач:

- проведення аналізу розвитку ІТ-індустрії в Україні;
- проведення аналізу економетричних моделей аналізу даних;
- аналіз програмних засобів реалізації економетричних моделей;
- розробка алгоритмів побудови матриці просторових ваг для областей України;
- розробка алгоритмів обчислення показників просторової кореляції;

розробка алгоритмів кластеризації областей України за кількістю ІТ-компаній;

здійснення комп'ютерного моделювання алгоритмів просторової економетрики;

розробка модуля побудови матриці просторових ваг;

розробка модуля отримання показників просторової автокореляції;

розробка модуля тестування просторових моделей.

Об'єкт дослідження – процес аналізу українських ІТ компаній.

Предмет дослідження – моделі просторової економетрії.

Методи досліджень. У кваліфікаційній роботі використано методи просторової економетрики, математичну статистику, статистику, кореляційний і регресійний аналіз, метод найменших квадратів, метод максимальної правдоподібності.

Наукова новизна одержаних результатів полягає в розробці алгоритмів просторової економетрики для аналізу локалізації українських ІТ компаній.

Практичне значення одержаних результатів полягає в розробці програмної системи просторового аналізу, модуля отримання показників просторової автокореляції.

Публікації результатів досліджень. За результатами досліджень опубліковані двоє тез доповідей V науково-практичної конференції молодих вчених і студентів «Інтелектуальні комп'ютерні системи та мережі» (2 грудня 2021 р., м. Тернопіль, Західноукраїнський національний університет) [2, 3]:

Березька К. М., Цимбалюк Д. В., Іванов Ю. М. Застосування програмного додатку *gwrsofMapper* в просторовій економетриці. *Інтелектуальні комп'ютерні системи та мережі* : тези доп. V Наук.-практ. конф. молодих вчених і студентів (2 груд. 2021 р.). Тернопіль : ЗУНУ, 2021. С.

Цимбалюк Д. В., Іванов Ю. М. Програмна система оцінки локалізації українських ІТ-компаній. *Інтелектуальні комп'ютерні системи та мережі* : тези доп. V Наук.-практ. конф. молодих вчених і студентів (2 груд. 2021 р.). Тернопіль : ЗУНУ, 2021. С.

Кваліфікаційна робота складається із трьох розділів, висновків, списку використаної літератури та додатків [4, 5].

У першому розділі здійснено огляд розвитку ІТ індустрії в Україні і Тернопільській області, проаналізовано економетричні моделі аналізу даних, зроблено аналіз програмних засобів реалізації економетричних моделей, здійснено аналіз завдання на кваліфікаційну роботу та постановку задач.

У другому розділі розроблено алгоритми побудови матриці просторових ваг для областей України, обчислення показників просторової кореляції, здійснено опис просторових моделей, розроблено алгоритм побудови моделі просторового лагу.

У третьому розділі спроектована і реалізована програмна система.

У додатку приведено світлокопії виданих публікацій, статистичні таблиці, матриці просторових ваг, довідка про використання.

# 1 АНАЛІЗ ЕКОНОМЕТРИЧНИХ МОДЕЛЕЙ ОЦІНКИ РОЗТАШУВАННЯ УКРАЇНСЬКИХ ІТ-КОМПАНІЙ

## 1.1 Розвиток ІТ-індустрії в Україні

В Україні стрімко зросла за останні роки ІТ-індустрія. Україна стала головним гравцем на світовій технологічній арені. Пул ІТ-талентів в Україні наближається до 200000 розробників програмного забезпечення, технологічні стартапи швидко розвиваються, потік інвестицій збільшується, а кількість компаній, які співпрацюють з українськими постачальниками послуг з розробки програмного забезпечення, зростає.

Якість українських технічних навичок визнана у всьому світі та визнана багатьма галузевими нагородами. За рейтингом компанії А. Т. Kearney Україна входить в топ-20 за показником розташування послуг. Вона була визнана «Напрямом року по Offshoring» за версією GSA; дедалі більше українських ІТ-аутсорсингових компаній входять до списку Global Outsourcing 100 за версією IAOP; багато постачальників включені до рейтингу Software 500, Inc. 5000 та інших галузевих рейтингів [6].

Значна частина успіху України впливає з її великих ресурсів. Це найбільша країна, яка повністю лежить в Європі з населенням майже 43 мільйони та робочою силою 20 мільйонів людей. Країна має вигідне географічне положення, оскільки на заході межує з Польщею, Румунією, Словаччиною, Угорщиною та Молдовою; Білорусь та Росія на півночі та сході; Болгарія, Грузія та Туреччина на південь від Чорного моря. Для європейських підприємств географічне розташування та близькість часового поясу значно спрощують співпрацю. Більше того, європейське мислення та подібна ділова культура усувають вузькі місця у спілкуванні.

Основними драйверами розвитку ІТ-компаній в Україні на думку автора [7] є наступні:



1. Експорт комп'ютерних та інформаційних послуг з України: продаж послуг із розроблення програмного забезпечення та різноманітних проектів [8].

Український ІТ ринок зарекомендував себе на світовому просторі надійним та якісним. Оплата праці є помірною, тому з кожним роком попит та розміри експорту комп'ютерних та інформаційних послуг рівномірно зростають, як показано на рисунку 1.1. ІТ галузь за останні декілька років показує стабільне зростання на 11-26% щороку, хоч в цілому спостерігається падіння експорту та економіки України [9-11].

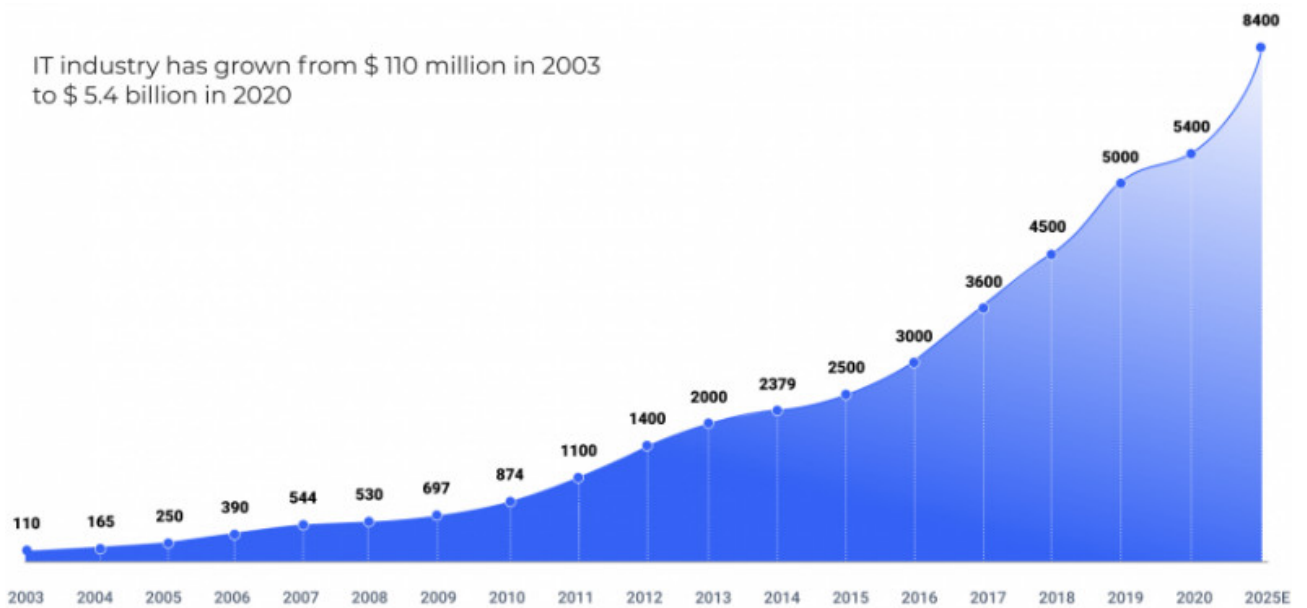


Рисунок 1.1 – Експорт комп'ютерних та інформаційних послуг has grown from

Поряд з зростанням доходу в ІТ-галузі (Industry revenue), відбувається зростання кількості ІТ-спеціалістів (рисунок 1.2) [12].

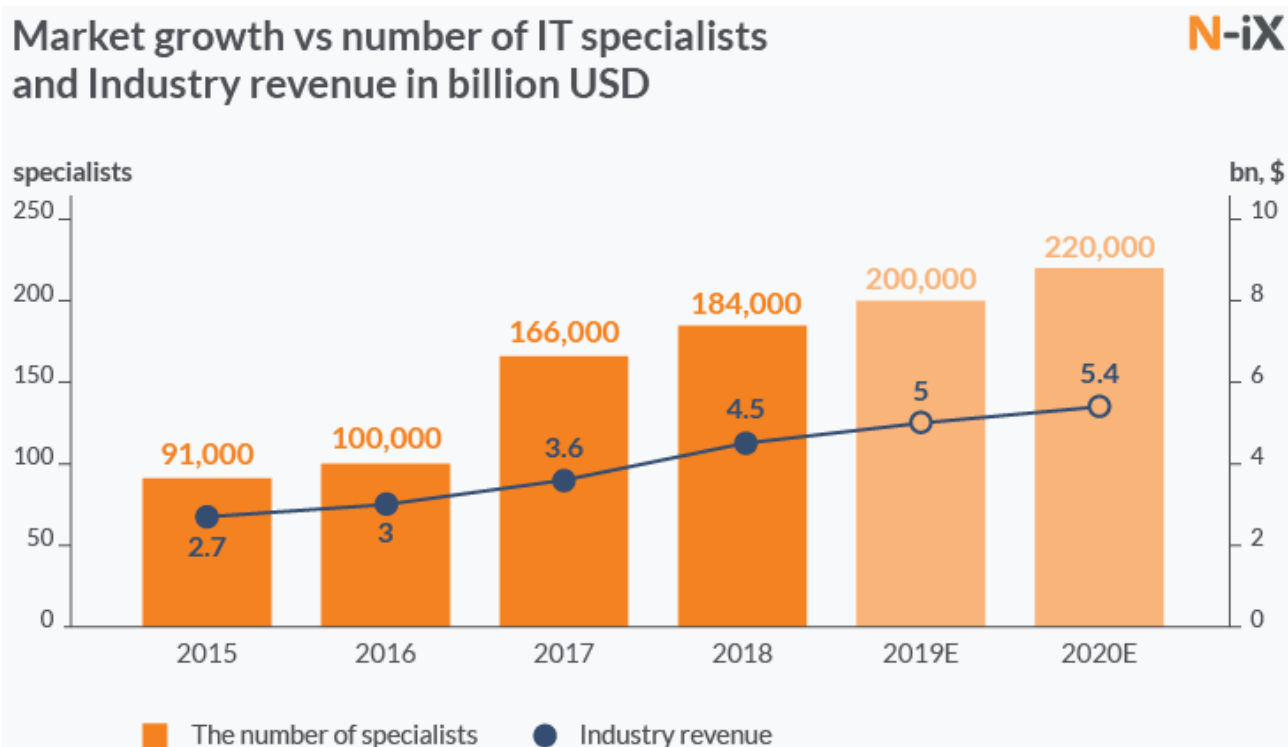
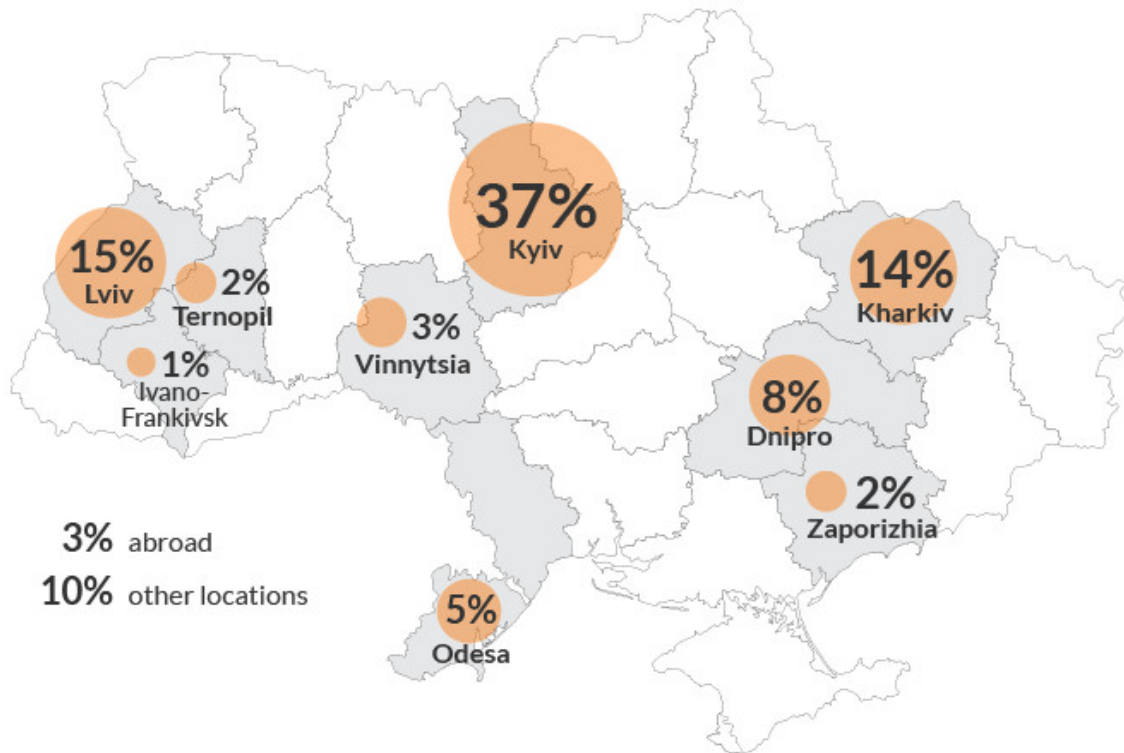


Рисунок 1.2 – Зростання ринку порівняно з кількістю ІТ-спеціалістів та доходом галузі у млрд. дол.

2. Наступним драйвером [7] є те, що 30% світових ІТ-компаній, мають офіси в Україні. Відомо, що в нашій країні працює більше чотирьох тисяч технологічних ІТ-компаній, більше 110 R&D центрів всесвітньо відомих міжнародних компаній та більше 1600 компаній, що займаються ІТ-послугами [11].

Найбільші українські ІТ-центри з кількістю ІТ-спеціалістів приведені на рисунку 1.3 [12].

3. Третім драйвером, на думку автора [7] є те, що працівники ІТ-компаній наповнюють місцеві та регіональні бюджети. Єдиний податок, сплачений ФОПами (в ІТ галузі широко застосовують модель співпраці зі спеціалістами на умовах ФОП), повністю залишається в місцевих бюджетах за місцем реєстрації, ПДФО на 75% розподіляється серед місцевих бюджетів різного рівня, а податок на прибуток на 90% надходить до центрального бюджету. Один працівник ІТ-сфери дає надходження до державного бюджету вищі, ніж в середньому по країні. Податок на доходи фізичних осіб, сплачений працівниками ІТ-галузі, у 3,7 разів більший, ніж середній по країні.



### Number of software developers in Ukrainian IT cities

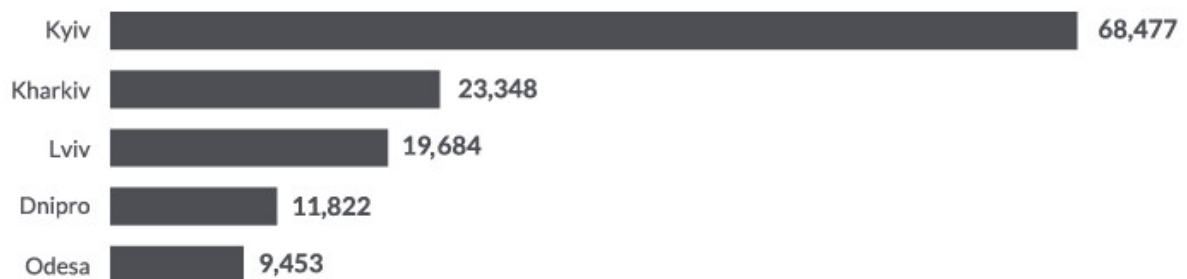


Рисунок 1.3 – Найбільші українські ІТ-центри з кількістю розробників програмного забезпечення (2020)

4. Наступним рушієм розвитку ІТ є зростання внутрішнього попиту на ІТ-послуги [7, 13]. Велика кількість галузей: виробництво, медицина, освіта, банківський сектор, агропромисловий комплекс та інші, впроваджують у свою діяльність все більшу частину інформаційних технологій. З кожним роком проходить автоматизація певних процесів. Появилися роботизовані виробничі

лінії, роботи-прибиральники, роботи-хірурги, застосовується технологія розпізнавання обличчя на охоронних пунктах різних закладів, використовуються дрони та безпілотники для контролю врожаю та оцінки якості ґрунтів.

5. На ІТ впливає стимулювання з боку держави розвитку ІТ-компаній, впровадження пільг для ІТ-бізнесу (зміни в податковому законодавстві для 3-ї групи ФОП – 3% доходу та обрано додатково сплату ПДВ або 5% доходу – якщо ПДВ не сплачують), акцентування уряду на пріоритетних напрямках ІТ-сектору, комплексний підхід та розв'язування буденних проблем ІТ-компаній, серед яких: забезпечення роботою іноземців, розвиток кваліфікованих працівників, скасування застарілих законів, розбудова екосистеми та створення технологічних парків [7].

Виділимо ІТ-компанії України за кількістю працівників. Приведемо першу десятку [14].

На 1-му місці – Ерам. Це найбільша ІТ-компанія України, яка входить в міжнародну групу Ерам. Офіси є у Києві, Львові, Харкові, Дніпрі, Вінниці. В них працює більше 10 тис. спеціалістів. Дохід за 2020 рік – 10,44 млрд грн, прибуток – 942 млн грн.

В Україні Ерам представлена двома ТОВ: "Елам Системз" та "Елам Рішення". Проте основний грошовий потік зосереджено на першій компанії.

2-е місце обіймає львівська компанія SoftServe. Це ІТ корпорація. SoftServe має більше десяти юридичних осіб в Україні. Засновниками її є два випускники факультету прикладної математики Львівського національного університету ім. Івана Франка Тарас Кицмей і Ярослав Любінець. Офіси компанії є в Києві, Харкові, Львові, Дніпрі, Рівному, Чернівцях, Івано-Франківську. Компанія налічує майже 8 тис. працівників.

Бізнес-представництва знаходяться в США, Великобританії, Німеччині та Нідерландах. Офіційний дохід компанії в п'ять разів менший, ніж в Ерам.

На 3-му місці – GlobalLogic.

Ця міжнародна корпорація представлена в Україні юридичною особою ТОВ "Глобаллоджик Україна". Її дохід за 2020 р. становив 5,5 млрд. грн. Це – більше, ніж в два рази перевищує дохід SoftServe.

Україна є однією з найбільших локацій для GlobalLogic. Офіси компанії знаходяться в Києві, Харкові, Львові та Миколаєві. Компанія на початок 2021 нараховувала 5700 чол. Також центри розробок знаходяться в Європі, Індії, США і Аргентині.

На час написання роботи власником GlobalLogic є Канадський пенсійний фонд та швейцарський інвестиційний фонд Partners Group. 31 березня 2021 японська Hitachi оголосила про покупку американської IT-компанії GlobalLogic за \$9,6 млрд. Угода поки не закрита.

Четверте місце обіймає компанія Luxoft. Її засновник – Дмитро Лоцинін (2000 р., Москва). На початку діяльності Luxoft був дочірнім підприємством російської IT-компанії IBS, з 2019 Luxoft у власності американського IT-гіганта DXC Technology Co за \$2 млрд. Представництва Luxoft в Україні знаходяться в Києві, Одесі та Дніпрі. Компанія нараховує 3500 працівників.

Юридичною особою Luxoft в Україні є ТОВ "Люксофт-Україна", з доходом 180 млн грн, а прибутком – 3,8 млн грн.

На п'ятій сходинці компанія – Ciklum (Сіклум). Засновником цієї міжнародної компанії в Києві був датчанин Торбен Майгаард (2002 р.). З 2015 року власником є американський мільярдер Джордж Сорос.

Компанія розміщується в Лондоні. Філіали є у Великобританії, Швейцарії, Данії, Білорусії, Пакистані та Україні. В Україні працюють більше 2700 чоловік.

За 2020 рік ТОВ "Сіклум" отримала 3,2 млрд. грн. доходу та 49,6 млн. грн. прибутку.

На шостому місці – компанія NIX. Це є харківська компанія. Нарховує вона на початок 2021 р. 2400 спеціалістів. Власник – Віктор Шальнев. Дохід ТОВ "Нікс Солюшенс Лтд" – 525 млн. грн. NIX працює на ринки Північної та Південної Америки, ЄС та Азії.

На сьомій сходинці компанія – EVOPLAY (2003 р. заснування). Компанія спочатку була лише розробником онлайн-ігор. Зараз EVOPLAY розробляє ПЗ для ігрової математики, ігрових платформ, платіжних рішень, підтримки користувачів, аналітики та фінансів, залучення та управління трафіком.

Власник компанії – Максим Криппа (екс-кандидат в депутати Київради від партії "Самопоміч"), партнер – американський підприємець з українським корінням Максим Поляков.

В Україні в EVOPLAY працює 2200 працівників. Дохід в 2020 році – 4,7 млн. грн.

Восьме місце обіймає компанія американця Євгена Голанда DataArt (1997 р. заснування). У 1998 році з компанії була виділена окрема структура – Port.ru, яка створила поштову службу Mail.Ru.

В Україні в DataArt працює більше 2000 спеціалістів. Компанія має 3,4 млн. грн. доходу за 2020 рік та 0,2 млн. грн. прибутку.

На дев'ятому місці компанія – Inforpulse (1991 рік заснування). Засновники – українські підприємці Олексій Сігов і Андрій Анісімов. З 2007 року Inforpulse перебуває у власності великого норвезького ІТ-холдингу Evry. У 2019 році Evry об'єдналась з фінською компанією Tieto і є частиною об'єднаної TietoEvry – компанії-лідера ІТ-ринку Скандинавії.

Спочатку компанія орієнтувалася на Україну, але з середини 90-х років почала працювати на Європу, надавати послуги аутсорсингу.

Штат компанії налічує більше 2000 спеціалістів. Основні продакшн-офіси перебувають в Україні. Головний – у Києві, а також є офіси в Одесі, Харкові, Львові, Вінниці, Житомирі та Чернігові.

Дохід ТОВ "Інфопульс Україна" склав 2,4 млрд грн, чистий прибуток – 420 млн грн.

10-е місце має компанія Ajax Systems (засновник – Олександр Конотопський). Ця ІТ-компанія України розробляє продукти, серед яких – бездротові системи безпеки з функціями "розумного дому".

Виробництво Аїах розташовано в Україні. Компанія розробила екосистему: датчики руху, відчинення, захисту від затоплення та пожежні датчики, вуличні й домашні сирени, тривожна кнопка з функцією керування приладами, розумні розетки та реле.

Рейтинг українських ІТ-компаній та основні відомості про них приведені в таблиці А.1 (додаток А).

В Україні нараховується 212 547 ФОПів, які працюють у ІТ-сфері [1]. За 2020 рік їхня кількість збільшилася на 29 111 осіб, або на 16%, але темпи приросту порівняно з 2018-2020 роками дещо впали. За відповідні роки кількість ФОПів в українському ІТ збільшувалася на 20-22% щороку.

На рисунку 1.4 приведено розподіл ІТ-ФОПів за регіонами [1].

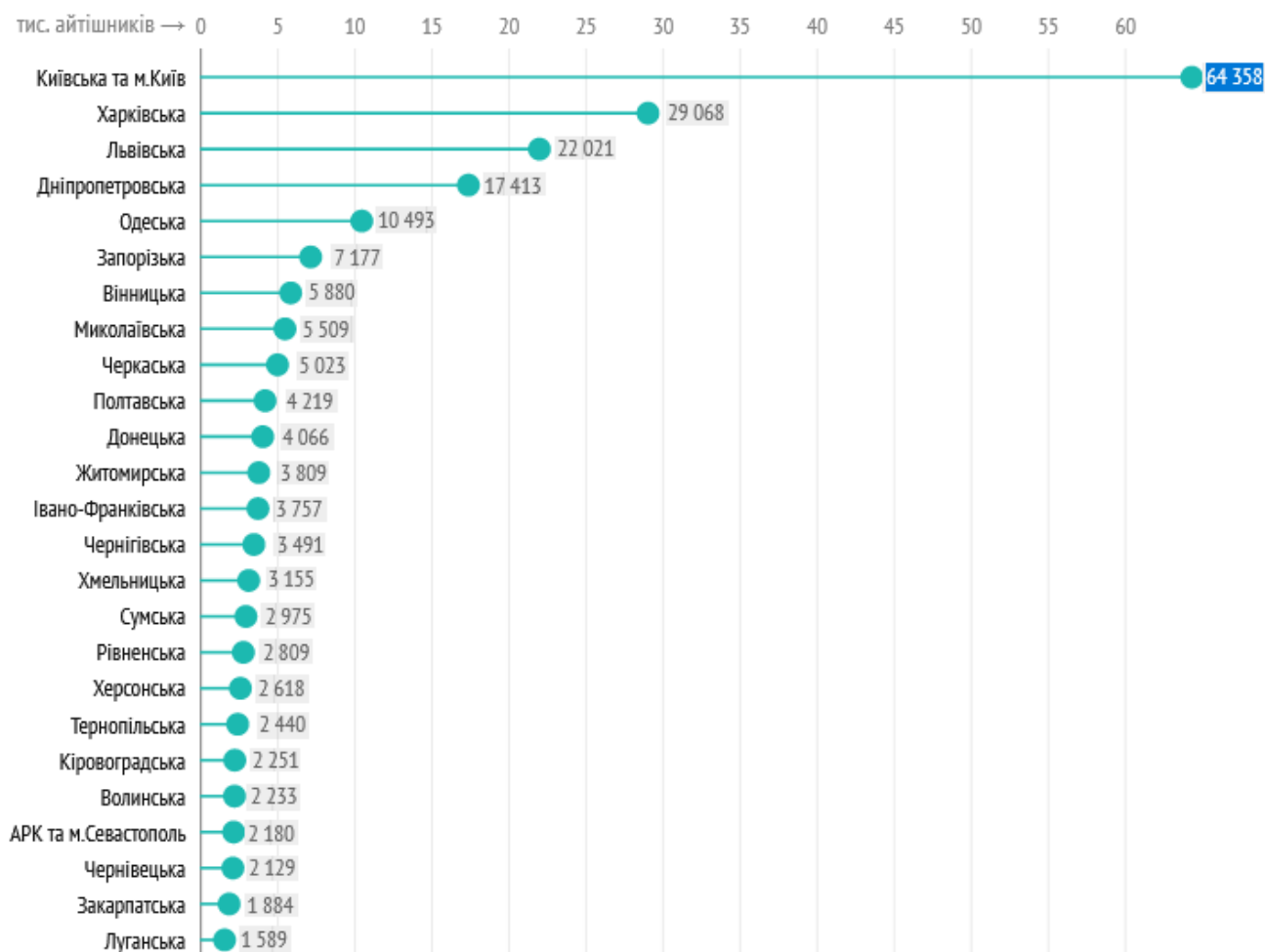


Рисунок 1.4 – Розподіл ІТ-ФОПів за регіонами (2020 р.)

Відомо, що офіційні статистичні дані погано показують внутрішню міграцію, коли ІТ-спеціаліст зареєстрований в одному регіоні чи населеному пункті в Україні, а насправді живе та працює в іншому. На рисунку 1.5 приведена фактична кількість ІТ-спеціалістів за даними «Портрета» [1]. Це відкалібровані дані з «Портрета ІТ-спеціаліста», які відповідають місцю роботи та проживанню ФОПа, а не місцю його реєстрації.

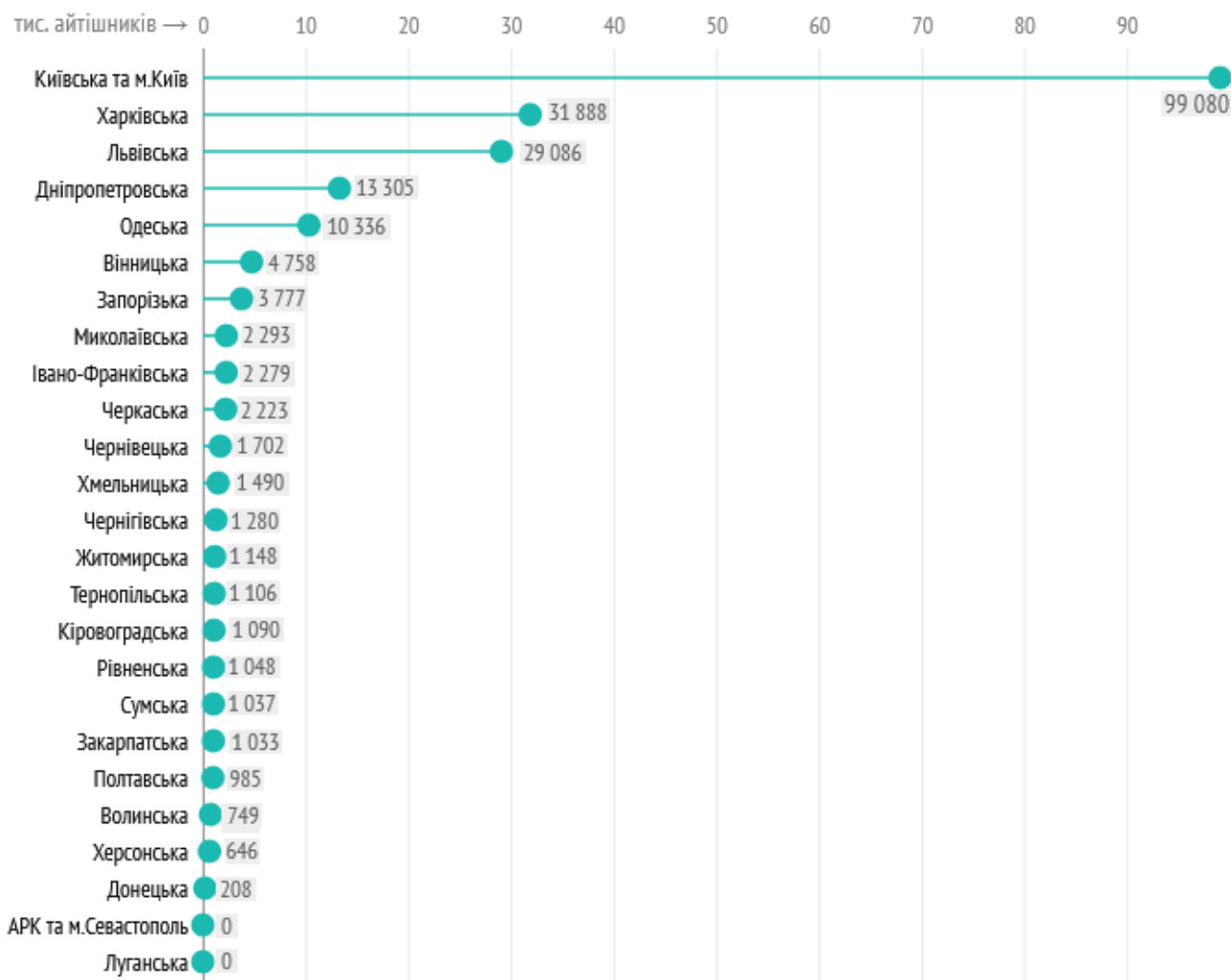


Рисунок 1.5 – Фактична кількість ІТ-спеціалістів за даними «Портрета» (2020 р.)

Як видно реально працюючих ФОПів у Київській, Львівській та Харківській областях набагато більше, ніж зареєстрованих. Майже третина ФОПів, що працюють в Києві та області, приїхали з інших регіонів. В Львівській області таких є близько чверті, в Харківській області – 9%.



В єдиній Одеській області практично відсутня різниця між кількістю зареєстрованих ФОПів і фактичною кількістю ІТ-спеціалістів.

Решта областей є регіонами-донорами для трьох перших. Найбільше мають від'ємний баланс між зареєстрованими та реально працюючими ІТ-ФОПами, в них зареєстровано відчутно більше ІТ-спеціалістів, ніж реально працює такі області: Дніпропетровська, Донецька, Запорізька, Миколаївська та Полтавська. Найчастіше ІТ-спеціалісти виїжджають з Донецької, Полтавської, Херсонської, Житомирської, Волинської, Рівненської, Сумської та Чернігівської областей: більш як 60% ІТ-спеціалістів, які зареєстровані у цих регіонах насправді працюють в інших містах України.

Звернемо увагу на ІТ-сферу міста Тернополя [15]. В місто зайшло більше 20 ІТ-компаній, в них працює більше тисячі спеціалістів.

Середні зарплати програмістів у Тернополі (медіани, грудень-2020):

Junior – \$850;

Middle – \$1050;

Senior – недостатньо даних.

ІТ-компанії Тернополя в основному займаються аутсорсингом.

Приведемо найбільш відомих працедавців міста.

На 1-му місці ELEKS. Ця компанія дає хороший старт для початківців. Фірма нараховує у Тернополі близько 120 спеціалістів. Студенти нашого вузу мають можливість навчатися на безоплатних курсах на конкурсній основі. Після закінчення курсів 60% - 80% працевлаштовуються в компанії. ELEKS проводить курси за мовами: SQL, Java, C++, C#, Front End; для тестувальників програмного забезпечення (інженерів забезпечення якості), проектних менеджерів та бізнес аналітиків.

На 2-му місці Forte Group. Це – американська компанія основними напрямками якої є розробка, тестування та підтримка програмного забезпечення для фірм американського ринку. Forte Group нараховує у Тернополі 105 спеціалістів. Розробки компанії направлені на сферу освіти, фінансів, юриспруденції, електронну комерцію, охорону здоров'я, телекомунікації.

Технології розробки: Java, Oracle ATG, .NET, Python, Front-End, PHP, iOS, Android, IoT. Для початківців існує відділ Forte Knowledge, на базі якого проходять безкоштовні курси для початківців по тестуванню ПЗ, Android/iOS розробці, Talent Management.

Ці дві компанії, за результатами голосування читачів DOU, визнані кращими працедавцями м. Тернополя.

Третю сходинку обіймає MagneticOne. Ця українська компанія нараховує близько сотні спеціалістів у Тернополі. Продуктова компанія: розробляють продукти для електронної комерції, CMS та CRM-систем, муніципальних органів, мобільні додатки та інші. В компанії є можливість пройти стажування на певній посаді протягом одного місяця, що дозволяє потрапити на роботу в разі наявності відповідних відкритих вакансій.

На четвертому місці компанія – Crowdin. Ця українська компанія має близько 40 спеціалістів у Тернополі. Як і попередня компанія, вона є теж продуктовою компанією. Серед продуктів: розробка платформи для управління локалізацією в технологічних компаніях. Клієнти: Microsoft, Prestashop, Kickstarter, Xiaomi, Huawei, Citrix, Reddit, OWASP. Технології: LEMP, AWS. Для початківців існує безкоштовна програма Crowdin Space, що включає навчання/стажування для студентів технічних вузів.

На п'ятій сходинці – Yaware. Це невелика Українська компанія, що налічує 30 спеціалістів у Тернополі. Вона є продуктовою компанією. Працівники створюють програмне забезпечення для обліку і підвищення ефективності використання часу, що дозволяє керівникам компаній і співробітникам-аналітикам ефективно використовувати робочий час. Цей онлайн-продукт використовується в більш ніж 27-и країнах світу.

Для початківців існує можливість стажування протягом одного місяця в разі наявності відповідних відкритих вакансій. Крім цього, за допомогою продукту компанії Yaware – Yaware Courses, надається можливість використання системи онлайн-навчання для працівників.

Аріко (в минулому SSolutions) знаходиться на шостому місці [16]. Це – українська компанія, що нараховує 45 спеціалістів в Україні. Її локації у Тернополі, Києві, Івано-Франківську. Її засновники – Микола та Йосип Стрілецькі (випускники ТНЕУ). Фірма займається розробкою JavaScript додатків для венчурних стартапів і малих підприємств, використовують сучасні технології JavaScript – Node.js, React, ReactNative, Meteor. Також практикують веб-розробки, iOS і Android розробки, QA. Для початківців проводяться курси, діє JSSolutions Academy.

Orange35 – це українська компанія, що налічує 18 спеціалістів у Тернополі. Займається розробкою індивідуальних сайтів, в тому числі для Magento, Wordpress та PrestaShop, надає професійну технічну підтримку для всіх типів сайтів. Основні напрямки: Magento eCommerce Web Development, Magento Extensions, Wordpress Development і Custom Website development. Можливості для початківців в плані IT немає.

Восьме місце займає швейцарська компанія Virtido. Ця невелика фірма нараховує 15 спеціалістів у Тернополі, 40 – в Україні. Вона є дочірньою компанією від однойменної у Швейцарії. Virtido надає IT та адміністративні послуги. Спеціалісти розробляють ПЗ, використовуючи .NET, Python, C#, PHP, JS. MS SQL, MySQL; розробляють сайти як на основі існуючих CMS, так і нестандартних. Крім IT спеціалістів компанія має команду німецько- та англо-мовників, які займаються наповненням сайтів. Працівники, що прийшли на роботу, двічі на рік мають можливість пройти стажування, де пропонуються задачі, пов'язані з роботою з клієнтами.

На дев'ятому місці компанія – Unicorn Systems. Це – чеська компанія з головним офісом у Празі. Займається розробкою власних програмних продуктів, а також виконують проекти для компаній в енергетиці, банківському та страховому ринку, виробництві для клієнтів з Європи. Технології: в основному Java.

Можливості для початківців: курси з Java.

Десяту сходинку займає польська аутсорсингова компанія Future Processing. Компанія надає послуги офшорного програмування для клієнтів із Західної

Європи і Скандинавії. Технології компанії: Microsoft .NET, Microsoft SharePoint, Хмара (Microsoft Azure, Amazon Web Services) на мобільних, веб, настільних та вбудованих платформах.

І на одинадцятій позиції одна з компаній групи MagneticOne Group – MagneticOne Mobile. Це – українська компанія, що нараховує 10 спеціалістів у Тернополі. Поле діяльності: розробка мобільних додатків на базі Android та iOS. Реалізовує проекти у сфері цифрової трансформації бізнесу, наприклад, у сферах Business, Retail, Social Projects, Sales та ін. Найуспішніші з проектів: Business Card Reader та Call Tracker. Для початківців передбачено безкоштовне стажування з подальшим працевлаштуванням.

Крім вище приведених компаній в Тернопіль зайшла Mindy Supports – міжнародна компанія, яка здійснює аутсорсингові послуги, послуги клієнтської підтримки для клієнтів малого і середнього бізнесу на ринках ЄС і Америки.

Невелика компанія ХАРДСОФТ (10-50 співробітників). Розробник веб-додатків, ремонт комп'ютерів і периферійного устаткування.

Intltech – компанія, що надає послуги із цифрового маркетингу та розробки сайтів. Орієнтується компанія на ринок Канади. Займається: розробкою та наповненням інтернет ресурсів, просуванням сайтів у пошукових системах (SEO), маркетингом, Call-center & Customer support, контекстною рекламою (PPC).

«Business Evolution» – компанія, що займається створенням мобільних застосунків, спеціалізованого програмного забезпечення, послугами ІТ-консалтингу.

Крім ІТ-компаній у м. Тернополі існують наступні спільноти:

– центр інформаційних технологій. Сформований при ТДТУ ім. І. Пулюя. До складу Центру входять Академія Cisco Systems, Сектор розробки та впровадження інформаційних систем, Група ліцензування програмних продуктів та ін.;

– GDG Ternopil. Це – спільнота розробників, які слідкують за технологіями Google. В спільноті регулярно проходять зустрічі, які присвячені Android та іншим технологіям та напрямкам (.Net, iOS, Machine Learning, QA);

– TernopilJS. Це – спільнота шанувальників Javascript. Засновник спільноти – компанія Аріко (JSSolutions). Періодично проводять зустрічі, готують безкоштовні курси для початківців. Спілкування, в основному, проходять в Telegram чаті.

Elixir Club – це зустрічі розробників тернопільської Elixir спільноти, спеціалістів, що цікавляться мовою програмування Elixir.

В підрозділі описано розвиток ІТ-індустрії в Україні, приведено головні ІТ-компанії України, ІТ-компанії, що існують в м. Тернополі, а також ІТ-спільноти м. Тернополя.

## 1.2 Економетричні моделі аналізу даних

Розділ математики, в який входять методи, що обробляють дані незалежно від їх природи, називається аналізом даних [17].

Аналіз даних послідовно і логічно тлумачить зібрані дані (наприклад, відповіді респондентів), перетворює на статистичні форми, які потім використовуються для прийняття маркетингових та управлінських рішень.

Аналіз даних включає такі етапи:

- отримання даних;
- обробка даних;
- аналіз даних;
- інтерпретація результатів.

Аналіз даних можна віднести до математичної статистики, але він не обов'язково використовує імовірнісні моделі в описі досліджуваних об'єктів, явищ та процесів. Цей розділ математики обробляє як кількісні, так і якісні дані.

Аналіз даних також включає в себе наступні розділи:

- розвідувальний аналіз;
- кореляційний аналіз;

- дисперсійний аналіз;
- регресійний аналіз;
- коваріаційний аналіз;
- дискримінантний аналіз;
- кластерний аналіз;
- аналіз часових рядів.

Розділи: кореляційний аналіз, дисперсійний аналіз, регресійний аналіз, аналіз часових рядів входять в дисципліну «Економетрика». Моделі, що входять у ці розділи називаються економетричними моделями аналізу даних.

Кореляційний аналіз – це статистичне дослідження стохастичної залежності (взаємозв'язку) між випадковими величинами. Для дослідження використовують або два набори даних ( $X$  та  $Y$ ) або декілька наборів даних ( $Y, X_1, X_2, \dots, X_n$ ).

Кореляційний аналіз має наступні завдання:

- оцінити за вибірковими даними коефіцієнти кореляції;
- здійснити перевірку значимості вибіркових коефіцієнтів кореляції або кореляційного відношення;
- оцінити при наявності зв'язку його близькість до лінійного;
- побудувати довірчі інтервали для коефіцієнтів кореляції.

При дослідженні двох наборів даних  $X$  та  $Y$  проходить оцінювання коефіцієнта парної кореляції  $r = r_{yx}$ . Він дає можливість кількісно оцінити щільність зв'язку між  $Y$  та  $X$ . Він розраховується за формулою:

$$r = \frac{\text{cov}(x, y)}{\sqrt{\text{var}(x) \text{var}(y)}} = \frac{\text{cov}(x, y)}{\sigma_x \sigma_y} = \frac{\frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^2 \cdot \frac{1}{n} \sum_{i=1}^n (y_i - \bar{y})^2}},$$

де  $\text{cov}(x, y)$  – коефіцієнт коваріації між  $x$  та  $y$ ,  $\text{var}(x)$  – дисперсія змінної  $x$ ,  $\text{var}(y)$  – дисперсія змінної  $y$ .

Коефіцієнт кореляції є відносною мірою зв'язку між двома факторами. Значення його лежать в межах від  $-1$  до  $1$  ( $-1 \leq r \leq 1$ ). Додатне значення коефіцієнта кореляції свідчить про існування прямого зв'язку між змінними, від'ємне – про зворотній зв'язок. Якщо коефіцієнт кореляції прямує за абсолютною величиною до  $1$ , то між змінними  $x$  та  $y$  існує сильний тісний зв'язок ( $r \rightarrow \pm 1$  – щільність зв'язку велика); а якщо коефіцієнт кореляції прямує до  $0$  ( $r \rightarrow 0$ ) – то між змінними  $x$  та  $y$  лінійного зв'язку немає.

При дослідженні декількох наборів даних ( $Y, X_1, X_2, \dots, X_n$ ) проходить оцінювання трьох видів коефіцієнтів кореляції: парної, частинної, множинної кореляції.

Значимість вибірових коефіцієнтів кореляції перевіряється за статистикою:

$$t_{emn} = \frac{r\sqrt{n-2}}{\sqrt{1-r^2}}.$$

$t_{emn}$  має розподіл Стьюдента з  $k=n-2$  ступенями вільності. Для заданого рівня значущості  $\gamma$  і  $k$  ступенів вільності знаходять табличне значення  $t_{kp}$  за таблицями розподілу Стьюдента. Якщо  $|t_{emn}| > t_{kp}$  то з заданим рівнем значущості  $\gamma$  вибіровий коефіцієнт кореляції  $r$  вважають статистично значущим, в протилежному випадку – статистично незначущим.

Дисперсійний аналіз (ANOVA, Analysis of Variations) – це сукупність статистичних методів, які призначені для оцінки впливу факторів на результат досліджень, які можна використати для наступних планувань експериментів. В дисперсійному аналізі можна досліджувати залежність кількісної ознаки від однієї чи багатьох ознак-факторів.

За допомогою дисперсійного аналізу перевіряють статистичні гіпотези про рівність середніх значень декількох нормально розподілених генеральних сукупностей. Перевірку гіпотез проводять на основі поділу загальної дисперсії

залежної змінної на дві складові: внутрішньогрупову дисперсію та міжгрупову дисперсію.

Внутрішньогрупова дисперсія – це дисперсія, зумовлена випадковою помилкою, тобто є наслідком дії неврахованих у дослідженні факторів.

Міжгрупова дисперсія – це дисперсія, зумовлена відмінністю середніх значень, і є наслідком ймовірного впливу фактору.

Чим більшою є міжгрупова дисперсія за внутрішньогрупову, тим більш значущою буде відмінність між середніми залежної змінної у групах. В такому випадку нульову гіпотезу про рівність середніх значень відкидають на обраному рівні значущості, а, отже, вплив фактору буде більшим. Значущість перевіряється за допомогою *F*-тесту.

Дисперсійний аналіз дозволяє оцінити внесок окремих факторів у загальну дисперсію результуючої змінної в тих випадках, коли побудувати регресійну модель неможливо, бо фактори є атрибутивними ознаками.

Регресійний аналіз тісно пов'язаний з кореляційним, але на відміну від останнього не вивчає чи істотний зв'язок, а проводить пошук моделі цього зв'язку, яка виражається функцією регресії.

Регресійний аналіз – це дослідження односторонніх статистичних залежностей між випадковими величинами. При цьому деякі із факторних величин можуть бути не випадковими величинами.

Задачі регресійного аналізу:

- 1) визначення форми залежностей;
- 2) знаходження функції регресії;
- 3) побудова точкових та інтервальних оцінок параметрів функції регресії;
- 4) знаходження точкових та інтервальних оцінок умовних математичних сподівань, необхідних для визначення меж, в яких із заданою надійністю будуть міститися середні значення досліджуваної величини, якщо інші пов'язані з нею величини набувають певних значень;

5) перевірка узгодженості знайденої емпіричної функції регресії спостереженим даним.



Основною метою регресійного аналізу є теоретично обґрунтований і статистично надійний точковий та інтервальний прогноз значень залежної величини або умовного математичного сподівання цієї величини.

Рівняння лінійної моделі парної регресії описується за формулою:

$$\hat{y} = a_0 + a_1x,$$

де  $a_0$  та  $a_1$  невідомі параметри, які знаходяться на основі вибіркового даних,  $a_0$  – перетин з віссю ординат,  $a_1$  – нахил прямої до осі абсцис.  $a_1$  визначає зміну результативного показника при зміні  $x$  на одиницю. Значення  $\hat{y}_i$  показують середнє значення залежної змінної  $y$  при заданому  $x_i$  у припущенні, що єдиною причиною зміни  $y$  є змінна  $x$ , а випадкова змінна  $u$  прийняла значення, рівне нулю.

Для обчислення  $a_0$  та  $a_1$  використовують систему лінійних рівнянь:

$$\begin{cases} \sum_{i=1}^n y_i = na_0 + a_1 \sum_{i=1}^n x_i; \\ \sum_{i=1}^n y_i x_i = a_0 \sum_{i=1}^n x_i + a_1 \sum_{i=1}^n x_i^2, \end{cases}$$

яка називається нормальною системою методу найменших квадратів (МНК).

У випадку, якщо регресія не є лінійною, то будують нелінійні моделі парної регресії. Відомими моделями є:

- експоненційна –  $y = \alpha\beta^x$ ;
- степенева (мультиплікативна) –  $y = \alpha x^\beta$ ;
- зворотна –  $y = \alpha + \beta \frac{1}{x}$ ;
- квадратична –  $y = \alpha + \beta x + \gamma x^2$ ;
- модифікована експонента –  $y = \alpha\beta^x + \gamma$ ;

- крива Гомперця –  $y = e^{\alpha\beta^x + \gamma}$ ;
- логістична крива –  $y = \frac{1}{\alpha\beta^x + \gamma}$ .

Для знаходження параметрів перших чотирьох моделей – проводять лінеаризацію, а потім застосовують МНК, для наступних трьох – застосовують методи нелінійного оцінювання.

Рівняння лінійної моделі багатofакторної регресії описується за формулою:

$$\hat{y} = \hat{\beta}_0 + \hat{\beta}_1 x_1 + \dots + \hat{\beta}_m x_m,$$

де  $\hat{\beta}_0, \hat{\beta}_1, \dots, \hat{\beta}_m$  – оцінки невідомих параметрів, які знаходяться з системи нормальних рівнянь МНК:

$$\begin{cases} n\hat{\beta}_0 + \hat{\beta}_1 \sum_{i=1}^n x_{1i} + \hat{\beta}_2 \sum_{i=1}^n x_{2i} + \dots + \hat{\beta}_m \sum_{i=1}^n x_{mi} = \sum_{i=1}^n y_i \\ \hat{\beta}_0 \sum_{i=1}^n x_{1i} + \hat{\beta}_1 \sum_{i=1}^n x_{1i}^2 + \hat{\beta}_2 \sum_{i=1}^n x_{2i} x_{1i} + \dots + \hat{\beta}_m \sum_{i=1}^n x_{mi} x_{1i} = \sum_{i=1}^n y_i x_{1i} \\ \text{M} \\ \hat{\beta}_0 \sum_{i=1}^n x_{mi} + \hat{\beta}_1 \sum_{i=1}^n x_{1i} x_{mi} + \hat{\beta}_2 \sum_{i=1}^n x_{2i} x_{mi} + \dots + \hat{\beta}_m \sum_{i=1}^n x_{mi}^2 = \sum_{i=1}^n y_i x_{mi} \end{cases}.$$

Аналіз часових рядів в основному проводиться для прогнозування. Досить поширеними і розвиненими є економетричні методи прогнозування. Економетричні методи поділяються на два класи: екстраполяційні та факторні методи [18]. До першого класу відносять трендові моделі, ARIMA-моделі (Бокса-Дженкінса), експоненційне згладжування та ін., що екстраполюють часові ряди. Ці методи застосовуються на короткотерміновий прогноз в умовах стабільної соціально-економічної системи, тому що для прогнозування використовуються лише дані спостережень прогнозованого показника. Для середньотермінового і довготермінового прогнозу вони є неефективними. До факторних методів

належать регресійні моделі. Вони є більш достовірними для довготермінового прогнозу, тому що для прогнозування використовуються не лише дані спостережень прогнозованого показника, але враховуються і зовнішні фактори, що впливають на модельований показник.

### 1.3 Аналіз програмних засобів реалізації економетричних моделей

Економетрія - наука, що вивчає кількісні та якісні економічні взаємозв'язки за допомогою математичних і статистичних методів і моделей.

Економетрія – ефективний інструмент наукового аналізу, дослідження та моделювання, що застосовується в професійній діяльності менеджера, економіста та фінансового аналітика.

Основними напрямками економетричного дослідження є:

- регресія;
- кореляція;
- аналіз і прогнозування часових рядів;
- статистичний і багатовимірний аналіз і ін.

Економетричне дослідження потребує складних розрахунків, реалізації алгоритмів. Застосування обчислювальної (комп'ютерної) техніки при такому дослідженні значно полегшує процес дослідження.

Економетричний інструментарій представляє собою програмні продукти, що розроблені для рішення математичних задач економетричного дослідження. У більшості випадків економічні завдання носять комплексний характер. Через це програмні продукти економетричного аналізу не обмежуються тільки економетрикою.

Необхідною умовою сучасного статистичного аналізу даних є ефективне використання комп'ютерних програм, від функціональної повноти і алгоритмічної

продуманості яких залежить підсумкова інтерпретація результатів дослідження та надійність висновків.

Прикладні програмні засоби статистичної можуть бути професійними, спеціалізованими та напівпрофесійними.

Професійні економетричні програмні засоби мають велику кількість методів аналізу, а в популярних пакетах кількість функцій обмежена.

Спеціалізовані програмні засоби спрямовані на будь-яку область аналізу даних.

Вибираючи програмний продукт для проведення економетричного дослідження, необхідно знати, що його освоєння вимагає багато часу та навиків.

Майже всі програмні продукти схожі за інтерфейсом і набором базових функцій.

Опишемо економетричні програмні продукти.

Найпростішим і найбільш вживаним є MS EXCEL [19].

За відсутності спеціалізованих програм розрахунки можна проводити в середовищі EXCEL. При цьому необхідно добре знати алгоритми обчислення статистичних величин.

MS EXCEL широко поширений (дане ПЗ встановлено практично на всіх комп'ютерах у пакеті прикладних програм MS Office), тісна інтеграція з PowerPoint і MS Word, наявність україномовної версії та сервісного обслуговування, велика кількість літератури, зручний інтерфейс для роботи з вибірками.

Програмний засіб MS EXCEL постійно оновлюється та доповнюється. Недоліком засобу є те, що за його допомогою не завжди можна знайти рішення складної економетричної задачі.

У MS Excel неможливо побудувати якісні наукові графіки. MS Excel добре підходить для нагромадження даних, проміжного перетворення, попередніх статистичних обчислень, для побудови деяких видів діаграм. Кінцевий статистичний аналіз слід проводити в спеціальних, створених для цього програмах. Наявні макроси-доповнення для MS Excel можуть включати додаткові

статистичні функції, що в певних випадках є достатніми для звичайного застосування. Пробну версію макросів можна узяти на сайті виробника.

Побудувати модель, яка описує більш складну залежність результуючого показника від набору пояснюючих факторів, в MS EXCEL неможливо. Реалізація таких складних економетричних процедур вимагає бездоганного знання всіх обчислювальних алгоритмів. Застосування EXCEL навіть при вивченні базового курсу економетрики є мало ефективним.

Опишемо STADIA [20]. STADIA представляє собою набір актуальних ефективних методів аналізу. В програмі є повний комплект ділової та наукової графіки: функції, залежності, розподіли, діаграми розсіювання, багатовимірні діаграми, карти, поверхні обертання, сплайни, прогнози, гістограми, стовпчикові та кругові діаграми, дендрограми, установка розмірів, написів по осях, під малюнком, графічний архів, різноманітні перетворення й обчислення, імпорт/експорт даних і результатів у стандартний міжнародний формат (ASCII і DBF); розвинена екранна довідка, зрозуміла інтерпретація результатів, великий сервіс, до 64 000 чисел в матриці даних. Функції ПЗ – ділова графіка; параметричні тести (критерії Стюдента, Фішера, гістограми, розподіли, узгодження частот, послідовний аналіз); непараметричні тести (Ксі-квадрат, знаків, Вілкоксона, Колмогорова-Смірнова, Ван-дер-Вардена, Клотца, Анастрі-Бредлі, кореляція Кенделла, кореляція Спірмана); категоріальний аналіз (кростабуляція, Ксі-квадрат оцінка, коефіцієнти узгодження); дисперсійний аналіз (одно-, дво-, багатofакторний, коваріаційний); регресійний аналіз (проста регресія, множинна регресія, покрокова регресія, нелінійна регресія); аналіз часових рядів (згладжування, фільтрація, автокореляція, кроскореляція, спектральний, крос-спектральний, ARIMA-моделі); багатовимірні методи дискримінантний аналіз, кластерний аналіз, факторний аналіз). STADIA має систему контекстної екранної допомоги, включає всі необхідні статистичні функції. Зовнішній інтерфейс фактично не змінний з 1996 року. Графіки та діаграми, побудовані за допомогою STADIA, виглядають у сучасних презентаціях застаріло. До позитивних якостей

програми можна віднести російськомовний інтерфейс і наявність великої кількості літератури.

Розглянемо пакет EViews – Econometric Views (Quantitative Micro Software / IHS Markit Ltd.) [21].

EViews – статистичний пакет для аналізу економетричних даних часових рядів, аналізу та моделювання даних, побудови регресійних моделей. EViews – один із потужних пакетів побудови економетричних моделей. EViews успішно використовується при аналізі наукової інформації; фінансовому аналізі; макроекономічному прогнозуванні; моделюванні економічних процесів; прогнозуванні ситуацій на ринках і ін. Пакет EViews має широкі можливості для аналізу часових рядів. Функції пакету: проста лінійна регресія, багатофакторна регресія, нелінійна регресія, симулятивні моделі, ARIMA-моделі. Завдяки розвиненій підказці (help), пакет EViews достатньо доступний для засвоєння

На даний момент останньою версією є Eviews 12.

Програмні засоби STATISTICA і SPSS [22-23] – системи з базовим і додатковим модулем, які включають в себе набір аналітичних, статистичних або економетричних методів.

Програмні продукти STATISTICA та SPSS ефективно спрощують процес економетричного дослідження.

Програмні засоби SPSS і STATISTICA – електронні таблиці з системою меню, що орієнтовані на роботу з часовими рядами та просторовими даними.

Statistica (StatSoft, Inc.) – вдало збалансоване за співвідношенням «потужність/зручність» ПЗ. Широкий спектр функціональних алгоритмів, розвинена графіка, відповідні засоби для редагування графічних матеріалів, більше 250 статистичних функцій представляють собою функціональні можливості ПЗ. Об'ємна довідкова система ознайомлює з використовуваними алгоритмами та статистичною термінологією. ПЗ STATISTICA широко розповсюджене. Існують сайти підтримки програми, що представляють собою різного роду допоміжні та навчальні матеріали. Вбудовані функції об'єднані спеціалізованими статистичними модулями: основні статистики і таблиці,

непараметрична статистика, дисперсійний аналіз, множинна регресія, нелінійне оцінювання, аналіз часових рядів і прогнозування, кластерний аналіз, факторний аналіз, функціональний аналіз, дискримінантний, аналіз тривалості життя, канонічна кореляція, багатовимірні шкали, моделювання структурними рівняннями тощо. Пакет можна рекомендувати для різноманітних статистичних досліджень - простих або складних. Суттєвими перевагами STATISTICA є: за допомогою реалізованих в системі STATISTICA мов програмування (SCL, STATISTICA BASIC), досить просто створюються закінчені рішення, що можна вбудувати в різні обчислювальні середовища. Пакет перекладено російською мовою, можна придбати ліцензійну, русифіковану версію, наявна значна кількість літератури з детальним описом системи STATISTICA; можливе розширення користувачем бібліотеки функцій, що дають можливість вирішити багато завдань з теорії ймовірності; реалізовано обмін даними між STATISTICA і Windows-додатками; використання пакету має сенс при рішенні досить трудомістких, математично складних і громіздких в реалізації методів багатовимірного аналізу; будь-яка графічна та текстова інформація в STATISTICA може бути виведена в файл формату RTF, який відкривається і редагується в Microsoft Office Word.

В SPSS та STATISTICA є функція автоматичного формування звіту з результатами моделювання.

На сьогоднішній день останніми версіями систем є русифікована STATISTICA 13, англomовна STATISTICA 13.5, IBM SPSS Statistics 22.0.

Програмний засіб SPSS застосовують при проведенні прикладних соціальних досліджень.

Gretl – пакет прикладних програм для економетричного моделювання, розроблений у 2000 році спільноту з 15 економетристів. Програма Gretl є відкритою, вільною та безкоштовною зі стандартною громадською ліцензією GNU.

Основні можливості пакету.

- Оцінювання параметрів з допомогою МНК, методу максимальної правдоподібності, узагальненого МНК.

- Виділення сезонності часового ряду.

- Моделі часових рядів: авторегресія (ARIMA), узагальнена авторегресія умовної гетероскедастичності (GARCH), векторна авторегресія (VAR), векторна модель корекції помилок (VECM).

- Моделі з обмеженими залежними змінними, інтервальна оцінка регресії.

Сьогодні останньою версією є Gretl 1.9.92. У комплексах інструкцій пакету Gretl є приклади побудови моделей, що полегшує роботу недосвідченому користувачеві. Досвідчені ж користувачі мають можливість працювати в режимі консолі.

Prognoz Platform – комерційна інтегрована платформа російської компанії, що спеціалізується на програмному забезпеченні для бізнес-аналітиків [24]. Prognoz Platform випускається з 1992 року.

Вмонтованими в програму є два модулі: «Моделювання та прогнозування» й «Аналіз часових рядів», на базі яких можлива побудова моделей часових рядів.

Тільки на місяць можна ознайомитися з можливостями програмного продукту можна на безкоштовному web-доступі до ресурсу. Це породжує певні незручності.

Prognoz Platform має певні недоліки. У Prognoz Platform вмонтована статистична база російських даних. При цьому розмір вибірки в програмі змінити не можливо. Значить, досліджуваний ряд даних доповнюється старими даними. Досліджувані тенденції ряду при цьому не будуть відображати дійсність. Окрім цього, не всі заявлені в програмі економетричні процедури та тести програми реалізовані в онлайн-версії. Недоступною є двокрокова процедура Енгла - Гренджера тестування на коінтеграції. У програмі Prognoz Platform висновок результатів деяких тестів відображений тільки розрахунковими та критичними значеннями. Результатів регресії при цьому немає.

MATLAB – комп'ютерна оболонка для інтерактивних, командних обчислень і візуалізації, в якій для числового аналізу вмонтований пакет прикладних програм, та мова програмування [25]. В MATLAB можна здійснити чисельний аналіз, операції з матрицями, сигнальні процеси, побудувати графіки в



зручному для використання середовищі. Задачі та розв'язки представлені в математичному вигляді. В MATLAB зручно проводити обчислення з матрицями, представляти математичний запис функції, працювати з алгоритмами, створювати робочі оболонки (user interfaces) з програмами в інших мовах програмування. Спеціалізується на чисельному обчисленні, при цьому необхідні інструментальні засоби працюють з програмним забезпеченням Maple. Це дає йому можливість бути повноцінною системою при роботі з алгеброю.

У MATLAB є велика кількість функцій для аналізу даних: матриці та лінійна алгебра – алгебра матриць, лінійні рівняння, власні значення і вектори, сингулярність, факторизація матриць; многочлени та інтерполяція – корені многочленів, операції над многочленами та їх диференціювання, інтерполяція та екстраполяція кривих; математична статистика та аналіз даних – статистичні функції, статистична регресія, цифрова фільтрація, швидке перетворення Фур'є; обробка даних – набір спеціальних функцій, включаючи побудову графіків, оптимізацію, пошук нулів, чисельне інтегрування; диференціальні рівняння – вирішення диференціальних і диференційно-алгебраїчних рівнянь, диференціальних рівнянь із запізненням, рівнянь з обмеженнями, рівнянь в часткових похідних та інше; розріджені матриці – спеціальний клас даних пакету MATLAB, що використовується у спеціалізованих додатках; цілочисельна арифметика – виконання операцій цілочисельної арифметики в середовищі MATLAB. Це професійний пакет розв'язування математичних задач різної складності, моделювання, рішення рівнянь, побудова графіків тощо.

В таблиці 1.3 зроблено порівняння наявних економетричних методів і моделей в пакетах програм.

Професійні пакети використовуються користувачами, які мають справу із надзвичайно великими обсягами даних або вузькоспеціалізованими методами аналізу.

Опишемо `gwrCorMapper`: дослідницький інтерактивний інструмент відображення географічно зваженої кореляції та часткової статистики кореляції. Цей веб-додаток `gwrCorMapper` працює з геометричними геопросторовими

даними. Він може бути застосований в багатьох областях, особливо в містобудуванні.

Таблиця 1.3 – Порівняльна таблиця наявних економетричних методів і моделей в пакетах програм

| Економетричні методи   | Statistica | SPSS | Eviews | Stadia | Prognoz Platform | Gretl |
|--|------------|------|--------|--------|------------------|-------|
| Методи багатомірного статистичного аналізу                   | +          | +    | -      | +      | -                | +     |
| Багатофакторна регресія                                      | +          | +    | +      | +      | +                | +     |
| Моделі з обмеженими залежними змінними (логіт, пробіт і ін.) | +          | +    | +      | -      | +                | +     |
| Моделі декомпозиції одновимірних часових рядів               | +          | +    | +      | +      | +                | +     |
| Моделі Бокса-Дженкінса (ARMA, ARIMA)                         | +          | +    | +      | +      | +                | +     |
| Динамічні моделі з розподіленням лагом                       | +          | +    | +      | +      | +                | +     |
| Моделі багатомірних часових рядів                            | -          | -    | +      | +      | +                | +     |
| Моделі умовної гетероскедастичності (ARCH, GARCH)            | -          | -    | +      | -      | +                | +     |
| Лінійні моделі панельних даних                               | -          | -    | +      | -      | -                | +     |
| Динамічні моделі панельних даних                             | -          | -    | +      | -      | -                | +     |

gwrcorMapper – це веб-додаток, що побудований на R із використанням Shiny framework [3, 26, 27]. R – одне з найпопулярніших та активно використовуваних програмних середовищ для геопросторового статистичного аналізу, і R shiny забезпечує засоби для перетворення програм R на інтерактивні веб-додатки. Будується gwrcorMapper з великих геопросторових статистичних бібліотек, які R пропонує як відкрите джерело на таких популярних сховищах, як GitHub та CRAN.

Основним алгоритмом gwrcorMapper є GWrcor [28]. GWrcor розміщений на CRAN і є розширенням функції зведеної статистики GW, gwss (), що знаходиться у популярній бібліотеці R GWmodel [29, 30], і забезпечує кореляцію GW та частковий аналіз кореляції GW із статистичним *t*-тестом для перевірки, чи спостерігається кореляція зі статистичною значущістю. Як і кореляція GW у бібліотеці GWmodel, часткова кореляція GW обчислює зважену часткову кореляцію між двома змінними, утримуючи одну або кілька змінних постійними. Типи кореляції Пірсона та Спірмена для кореляції GW або часткової кореляції GW доступні в GWrcor. GWrcor також включає статистичний *t*-тест і повертає *p*-значення коефіцієнтів кореляції та часткової кореляції.

Розрахунки як для кореляції GW, так і для часткової кореляції GW починаються з розрахунку географічно зваженої коваріаційної матриці ( $\Sigma$ ) у кожній *i*-й точці даних, розташування якої може бути описано координатами  $(u_i, v_i)$ .

#### 1.4 Постановка задач дослідження

Метою кваліфікаційної роботи є програмна реалізація побудови економетричних моделей оцінки розташування українських ІТ-компаній.

Для досягнення поставленої мети необхідно:

- провести аналіз розвитку ІТ-індустрії в Україні та м. Тернополі;

- проаналізувати економетричні моделі аналізу даних;
- провести аналіз програмних засобів реалізації економетричних моделей;
- розробити алгоритми побудови матриці просторових ваг для областей України;
- розробити алгоритми обчислення показників просторової кореляції;
- розробити алгоритми кластеризації областей України за кількістю ІТ-компаній;
- здійснити комп'ютерне моделювання алгоритмів просторової економетрики;
- розробити модулі: побудови матриці просторових ваг, отримання показників просторової автокореляції, тестування просторових моделей.
- провести тестування реалізованих алгоритмів.

## 1.5 Висновки до розділу 1

Результатами першого розділу є:

- проведений аналіз розвитку ІТ-індустрії в Україні та м. Тернополі;
- проведений аналіз економетричних моделей аналізу даних;
- проведений аналіз програмних засобів реалізації економетричних моделей.

Таким чином, на основі проведеного аналізу показано актуальність проведення наступних досліджень та здійснено постановку задач на наступні розділи.

## 2 АЛГОРИТМИ ПОБУДОВИ МОДЕЛЕЙ ОЦІНКИ РОЗТАШУВАННЯ УКРАЇНСЬКИХ ІТ-КОМПАНІЙ

### 2.1 Алгоритми побудови матриці просторових ваг

В підрозділі 1.2 описані моделі аналізу даних: кореляційний аналіз, дисперсійний аналіз, регресійний аналіз, аналіз часових рядів, які входять в економетрику.

В основі регресійного аналізу є ряд гіпотез і припущень, однією з яких виступає ідея незалежності досліджуваних об'єктів [31]. Фактично це припущення означає, що незалежні змінні в одному з аналізованих об'єктів можуть впливати на залежну змінну виключно в тому ж самому об'єкті, ніяким чином не впливаючи на залежну змінну в інших об'єктах.

Аналізуючи деякі досить схожі випадки, об'єднані в єдину групу з територіальної, культурної чи релігійної приналежності, допускається взаємозалежність між досліджуваними об'єктами. Можна припустити, що аналізовані процеси обумовлені не тільки набором факторів всередині окремо взятого об'єкта, а й своєрідними зовнішніми факторами, тобто процесами, що протікають в суміжних об'єктах (сусідніх областях, наприклад). Іншими словами, ми говоримо про ті випадки, коли припущення про незалежність досліджуваних об'єктів буде порушуватися. Описом і подальшим вивченням подібного взаємозв'язку і займається просторова економетрика – відносно новий напрямок в економетрики та сфері аналізу даних.

Просторові виміри настільки ж важливі, як і часові виміри в аналізі глобальної та локальної нерівності. Географічно зважені моделі включають узагальнену статистику (включаючи середнє значення, дисперсію та кореляцію), регресію та аналіз основних компонентів.

Моделювання просторових взаємозалежностей складається з побудови матриці просторових ваг, перевірки змінних моделі і залишків на можливу просторову кореляцію, побудови просторових моделей.

Опишемо визначення і обрання матриці просторових ваг. Матриця ваг допомагає припустити, що регіон пов'язаний з деякою сукупністю сусідніх регіонів. З допомогою матриці проходить опис ступеня взаємозалежності між досліджуваними об'єктами і враховується можлива взаємодія між досліджуваними об'єктами в просторі. Матриця ваг ( $W$ ) покликана визначити, які з спостережень в якій мірі просторово залежать один від одного.

Для кожного місця розташування в системі вона визначає, які з інших місцерозташувань у системі впливають на значення в цьому розташуванні.

Просторові ваги суттєво залежать від визначення набору околу для кожного спостереження. Це досягається шляхом вибору для кожного розташування  $i$  (як рядка) сусідів як стовпців, що відповідають ненульовим елементам  $w_{ij}$  у фіксованій (нестохастичній) та позитивній матриці просторових ваг  $W$ .

Існує кілька видів таких матриць [32-40]:

- матриця граничних сусідів;
- матриця  $k$  найближчих сусідів;
- матриця відстаней;
- матриця відстаней, що враховує розмір (потужність) регіону.

Всі матриці ваг мають розмір  $N \times N$ , тобто є квадратними. На головній діагоналі є нулі. З допомогою цього враховується вплив тільки сусідніх регіонів (в заданому розумінні сусідства) та виключається вплив регіону самого на себе (це можна зробити безпосередньо в моделі). Кожен рядок матриці (наприклад,  $i$ -ий) містить ваги, з якими всі інші регіони впливають на заданий регіон ( $i$ -ий регіон). В основному, матриця ваг стандартизується по рядках (сума ваг рядка дорівнює одиниці). Стандартизація проводиться з метою досягнення хороших властивостей матриці (щоб уникнути проблем зі обертанням матриць в економетричних пакетах), і, у випадку матриці відстаней, щоб взяти до уваги не абсолютні, а відносні відстані між регіонами.

Приведемо матриці ваг:

1) бінарна матриця граничних сусідів (contiguity matrix). Це найпростіша матриця, що враховує випадок врахування просторових зв'язків:

$$w_{ij} = \begin{cases} 1, & \text{якщо регіон } i \text{ та } j \text{ має спільний кордон,} \\ 0, & \text{якщо } i = j, \\ 0, & \text{якщо регіон } i \text{ та } j \text{ не має спільного кордону.} \end{cases} \quad (2.1)$$

В матриці граничних сусідів припускається, що на економіку регіону впливають лише регіони, які мають спільні кордони, а вплив сусідів 2-го, 3-го кругу є несуттєвим. Цей підхід не завжди є реалістичним економічно і географічно.

Важливо зазначити, що елементи матриці ваг є нестохастичними та екзогенними для моделі. Як правило, вони засновані на географічному розміщенні спостережень або суміжності. Вага відмінна від нуля, коли два місця мають спільну межу або знаходяться на певній відстані одне від одного. Однак це поняття є абсолютно загальним, і альтернативні специфікації просторових ваг (наприклад, економічна відстань) також можуть бути розглянуті. Більш комплексне розуміння сусідства реалізовано в матриці

2)  $k$  найближчих сусідів:

$$w_{ij}(k) = \begin{cases} 1, & \text{якщо } d_{ij} \leq d_i(k), \\ 0, & \text{якщо } i = j, \\ 0, & \text{якщо } d_{ij} > d_i(k), \end{cases} \quad (2.2)$$

де  $d_{ij}$  – відстань від регіону  $i$  до регіону  $j$  (обчислюється як відстань між регіональними центрами);  $d_i(k)$  – найбільша з  $k$  найменших відстаней між регіонами, так що кожен регіон має точно  $k$  сусідів. Обчислення проходить таким чином: спочатку для кожного регіону розраховуються відстані до всіх інших регіонів, після чого вибирається найменших  $k$  відстаней, так що  $k$ -а відстань для даного регіону служить кордоном відсікання. Для тих регіонів, до яких відстань більша, ніж  $k$ -а, міжрегіональні взаємодії вважаються несуттєвими. Для тих регіонів, які попали в число сусідів, в комірках відповідних рядків матриці ваг

ставляться одиниці, в іншому випадку – нулі. Прийнято, з метою стійкості отриманих результатів розглядати матриці 10, 15, 20 і 25 найближчих сусідів.

Той факт, що утворюється бінарна матриця, призводить до того, що внаслідок стандартизації по рядках для заданого регіону вплив сусідів враховується з однаковими вагами. По тій причині, матрицю  $k$  найближчих сусідів доцільно використовувати в тих випадках, коли регіони даної економічної системи досить однорідні за площею і за відстанню між регіональними центрами. В такому випадку охоплення території, на якій регіон зазнає впливу сусідів, буде досить рівномірним від регіону до регіону. Використання рівних ваг в цьому випадку буде виправданим. Якщо застосувати матрицю (2.2) для регіонів різної площі, то вона працює нестабільно і гірше, ніж матриця відстаней. Особливо це видно за величиною коефіцієнтів просторової кореляції. Ті, що розраховані за матрицями  $k$  найближчих сусідів, якщо і виявляються значущими, то істотно меншими за величиною, ніж ті, що розраховані для матриць відстаней.

Відомо [41], що матриці  $k$  найближчих сусідів, часто використовуються в європейських дослідженнях економічного зростання.

3) матриця відстані. Просторові ваги матриці відстані (distance matrix) розраховуються наступним чином:

$$w_{ij}(q) = \begin{cases} 1/d_{ij}^q, & \text{якщо } d_{ij} \leq D(q), \\ 0, & \text{якщо } i = j, \\ 0, & \text{якщо } d_{ij} > D(q), \end{cases} \quad (2.3)$$

де  $d_{ij}$  – відстань від регіону  $i$  до регіону  $j$  (розраховується як відстань між центрами регіонів),  $D(q)$  – квантілі відстаней,  $q = \overline{1,4}$ .

В випадку (2.3), коефіцієнт просторової матриці ваг аналогічний коефіцієнту гравітації, оскільки тяжіння регіонів є обернено пропорційне квадрату відстані. Чим далі знаходяться регіони географічно, тим менший вплив вони роблять один на другого. У разі  $q = 4$ , матриця (2.3) враховує всі відстані і лише



на головній діагоналі будуть нулі. У випадку якщо  $q < 4$ , критичною відстанню відсікання, далі якої взаємовплив вважається несуттєвим, вважається відповідний кватиль відстані. Найчастіше  $\gamma$  вибирають рівним 2 ( $\gamma = 2$ ).

Часто за  $d_{ij}$  замість відстані між центрами регіонів вибирають мінімальний час між ними в дорозі автомобільними дорогами.

4) матриця відстаней, що враховує розмір регіону. Поряд зі звичайною матрицею відстаней, використовується матриця відстаней, що враховує розмір регіону, ваги якої розраховуються за формулою:

$$w_{ij}(q) = \begin{cases} A_j / d_{ij}^\gamma, & \text{якщо } d_{ij} \leq D(q), \\ 0, & \text{якщо } i = j, \\ 0, & \text{якщо } d_{ij} > D(q), \end{cases} \quad (2.4)$$

де  $A_j$  – розмір або потужність сусіднього регіону  $j$  (ВРП, обсяг промислової продукції і ін.).

В кваліфікаційній роботі використовувалися тільки матриці найближчих сусідів і матриці відстаней, в яких ваги розраховувалися тільки за географічною дистанцією між регіонами – найкоротшими відстанями автомобільними дорогами між регіональними центрами (в кілометрах).

Матриці просторових ваг за граничними сусідами побудовано за наступними кроками:

1. Використовуючи карту України, визначаємо, які регіони мають спільні границі, а які не межують між собою. Присвоюємо кожному регіону порядковий номер.

2. Будуємо матрицю сусідства  $A = [a_{ij}]$ . Вона має квадратний розмір і складається з нулів і одиниць.  $a_{ij} = 0$  у випадку, якщо  $i$ -ий район не межує з  $j$ -им районом і  $a_{ij} = 1$  – в протилежному випадку. При  $i = j$   $a_{ii} = 0$ .

3. Будуємо матрицю просторових ваг  $W = [w_{ij}]$ . Для цього робимо нормування по сумі в рядку. Тобто  $w_{ij} = a_{ij} / \sum_j a_{ij}$ . Тоді  $\forall i - \sum_j w_{ij} = 1$ .

Рисунок 2.1 представляє блок-схему алгоритму побудови матриці просторових ваг за граничними сусідами.

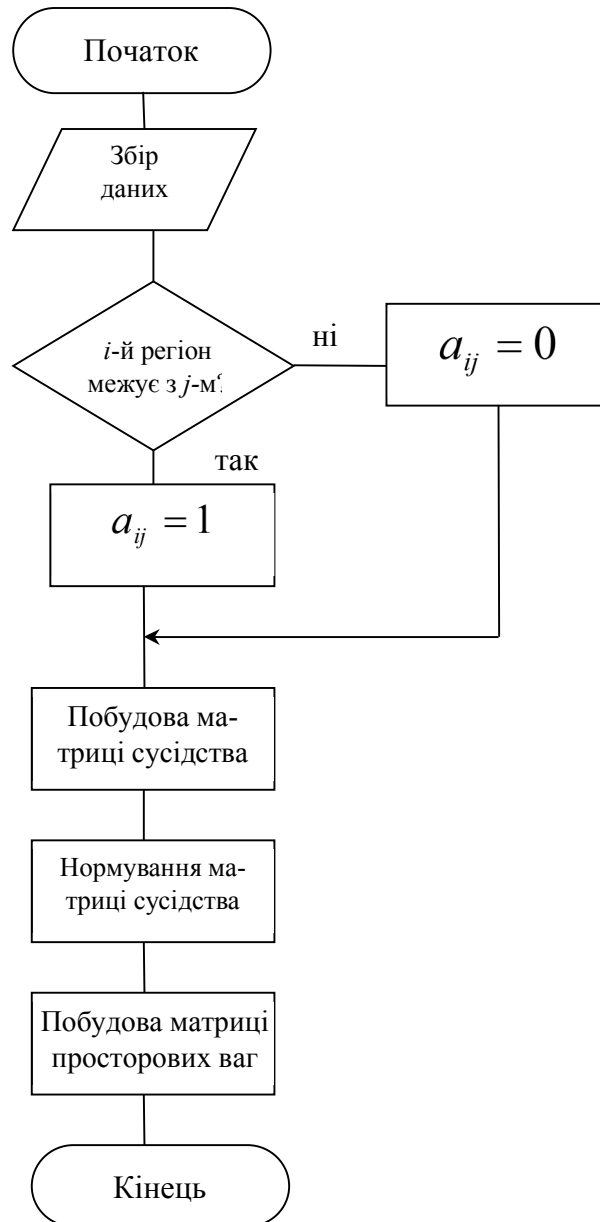


Рисунок 2.1 – Блок-схема алгоритму побудови матриці просторових ваг за граничними сусідами

В роботі реалізовано також матрицю відстані. Алгоритм знаходження матриці просторових ваг за відстанню відрізняється від граничних сусідів лише на кроці 2. Опишемо детальніше крок 2:

2.1) визначаємо відстань між обласними центрами. Будуємо квадратну матрицю відстаней  $A = [a_{ij}]$ ,  $a_{ij}$  дорівнює відстані між  $i$ -м обласним центром і  $j$ -м. При  $i = j$   $a_{ii} = 0$ ;

2.2) для кожного рядка цієї матриці визначаємо мінімальний (min), максимальний елементи (max), кватиль  $(0,25(\max - \min))$  і  $D(q) = 0,25(\max - \min) + \min$ ;

2.3)  $w_{ij}$  обчислюємо за формулою (2.3) при  $\gamma = 2$ .

## 2.2 Алгоритми обчислення показників просторової кореляції

Наступний етап в просторовому аналізі – перевірка змінних моделі і залишків, оцінених методом найменших квадратів, на можливу просторову кореляцію. Таблиця А.2 представляє основні коефіцієнти (статистики), які використовуються при дослідженні просторових ефектів [33, 37, 38, 42].

Найчастіше в дослідженнях за допомогою заданої матриці ваг розраховується показник загальної просторової автокореляції (Global Moran's I):

$$I = \frac{n}{S_0} \cdot \frac{\sum_{i=1}^n \sum_{j=1}^n w_{ij} (Y_i - \bar{Y})(Y_j - \bar{Y})}{\sum_{i=1}^n (Y_i - \bar{Y})^2}, \quad (2.5)$$

де  $Y$  – ознака, що досліджується,  $S_0 = \sum_{i=1}^n \sum_{j=1}^n w_{ij}$  – сума ваг просторової матриці  $W$

$$\bar{Y} = \frac{1}{n} \sum_{i=1}^n Y_i.$$

В матричному вигляді формула (2.5) матиме вигляд:

$$I = \frac{n}{S_0} \cdot \frac{Z'WZ}{Z'Z}, \quad (2.6)$$

де  $Z = Y - \bar{Y}$  – відхилення ознаки від середнього значення  $\bar{Y}$ .

Коефіцієнт просторової автокореляції  $I$  показує ступінь лінійного взаємозв'язку між вектором  $Z$  (нормованих значень ознаки  $Y$ ) і вектором  $WZ$  (просторово зважених середніх ознаки  $Y$  в сусідніх регіонах), який називається просторовим лагом. Якщо значення статистики  $I$  є більшими від математичного сподівання  $M(I) = -1/(n-1)$ , то існує позитивна просторова автокореляція, що означає, що великим значенням зважених середніх (відповідно до заданої матриці) ознаки  $Y$  в сусідніх регіонах відповідають великі значення  $Y$  в кожному з розглянутих регіонів. Якщо ж значення статистики  $I$  є меншими від математичного сподівання, то існує негативна просторова автокореляція [36-38], що означає, що великим значенням зважених середніх (відповідно до заданої матриці) ознаки  $Y$  в сусідніх регіонах відповідають малі значення  $Y$  в кожному з розглянутих регіонів.

Отже, якщо  $I > 0$ , то існує додатна просторова автокореляція, тобто в цілому значення спостережень в сусідніх регіонах є подібними;  $I < 0$  – від'ємна автокореляція, тобто в цілому значення спостережень в сусідніх регіонах відрізняються;  $I = 0$  – значення спостережень в сусідніх регіонах розміщені випадковим чином.

Крім обчислення коефіцієнта  $I$ , ще будують просторову діаграму розсіювання, з допомогою якої робиться візуалізація розкиду значень ознаки відносно просторового лага. По осі абсцис відкладаються значення вектора  $Z$ , по

осі ординат – значення  $WZ$ . На діаграмі зображується лінія регресії  $WZ$  на  $Z$ , тангенс кута нахилу якої дорівнює коефіцієнту загальної просторової автокореляції (у випадку якщо матриця просторових ваг стандартизована по рядках).

Діаграма розсіювання розділена на чотири квадранти (НН, LH, LL, HL) (рисунок 2.2) кожен з яких характеризується певним типом просторової близькості.

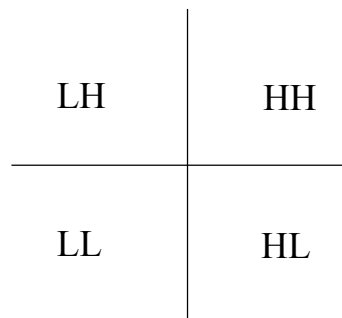


Рисунок 2.2 – Квадранти діаграми розсіювання

В правому верхньому квадранті діаграми (НН) збираються регіони, які з високим значенням ознаки знаходяться в оточенні регіонів так само з високими значеннями досліджуваної ознаки;

В лівому верхньому (LH) квадранті знаходяться регіони з низьким значенням, які оточені регіонами з високим значенням ознаки;

В лівому нижньому (LL) – регіони з низьким значенням, які оточені регіонами так само з низькими значеннями ознаки;

В правому нижньому (HL) квадранті знаходяться регіони з високим значенням, які оточені регіонами з низьким значенням ознаки.

В квадрантах НН і LL відображається групування регіонів зі схожими значеннями ознаки, тому вони характеризуються позитивною просторовою автокореляцією. В квадрантах LH і HL, навпаки, відображається групування регіонів з несхожими значеннями ознаки, тому вони характеризуються негативною просторовою автокореляцією. Таким чином, просторова діаграма

розсіювання може допомогти у виявленні нетипових регіонів, що належать до квадрантів LH або HL і типових, що належать до квадрантів HH і LL.

Приведемо алгоритм обчислення статистики Морана (показника загальної просторової автокореляції):

1. Збираємо статистичні дані кількості ІТ-компаній, по обласних центрах України. Для цього використовуємо статистичні дані з інтернету. З цих даних формуємо вектор-рядок  $Y$ .

2. Обчислюємо вектор-рядок

$$Z' = Y - \bar{Y}, \text{ де } \bar{Y} = \frac{1}{n} \sum_{i=1}^n Y_i. \quad (2.7)$$

3. Множимо вектор-рядок  $Z'$  на попередньо побудовану матрицю просторових ваг  $W$ . Отримуємо вектор-рядок  $Z'W$ .

4. Будуємо вектор-стовпець  $Z$  транспонуванням вектора-рядка  $Z'$ . Виконуємо множення  $Z'WZ$ .

5. Виконуємо множення  $Z'Z$ .

6. Знаходимо суму всіх елементів матриці  $W$ :  $S_0 = \sum_{i=1}^n \sum_{j=1}^n w_{ij}$ .

7. Обчислюємо показник загальної просторової автокореляції (статистику Морана) за формулою 2.6.

Блок-схема алгоритму обчислення показника загальної просторової автокореляції приведена на рисунку 2.3.

Показник загальної просторової автокореляції  $I$  показує тільки, що значення досліджуваної змінної  $Y$  просторово кластеризовані більшою мірою, ніж при випадковому розподілі, однак не пояснює, чому це відбувається. Для глибшого аналізу треба здійснити перевірку гіпотез про характер просторових взаємодій [43-45]. Оскільки індекс Морана має нормальний розподіл, то гіпотеза про його значущість перевіряється традиційним способом з допомогою  $z$ -статистики.

Перевірка гіпотез про просторову автокореляцію:

- нульова гіпотеза  $H_0$ : просторова автокореляція дорівнює нулю;

- альтернативна гіпотеза  $H_1$ : просторова автокореляція не дорівнює нулю.

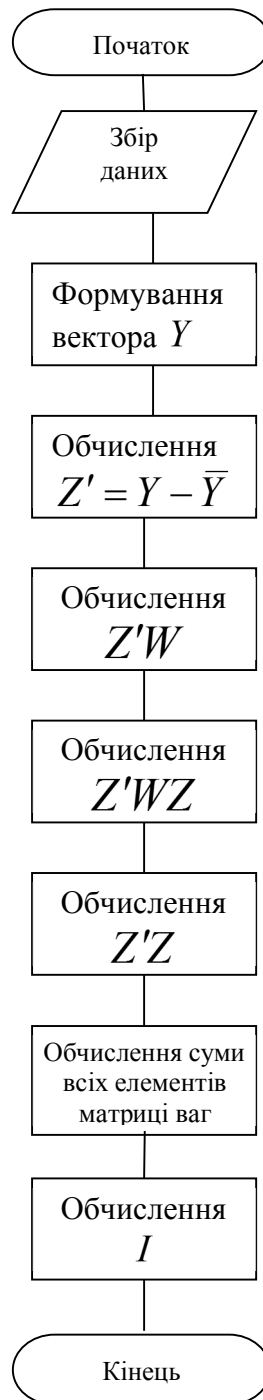


Рисунок 2.3 – Блок-схема алгоритму обчислення показника загальної просторової автокореляції

$z$ -статистика розраховується за формулою:

$$z_I = \frac{(I - E(I))}{\sigma_I}, \quad (2.8)$$

де  $E(I) = \frac{-1}{N-1}$ ,  $\sigma_I$  – стандартна помилка коефіцієнту Морана,

$$\sigma_I^2 = \frac{NS_4 - S_3S_5}{(N-1)(N-2)(N-3)(\sum_i \sum_j w_{ij})^2},$$

$$\text{де } S_1 = \frac{1}{2} \sum_i \sum_j (w_{ij} + w_{ji})^2, S_2 = \frac{\sum_i (\sum_j w_{ij} + \sum_j w_{ji})^2}{1}, S_3 = \frac{N^{-1} \sum_i (y_i - \bar{y})^4}{(N^{-1} \sum_i (y_i - \bar{y})^2)^2},$$

$$S_4 = \frac{(N^2 - 3N + 3)S_1 - NS_2 + 3(\sum_i \sum_j w_{ij})^2}{1}, S_5 = S_1 - 2NS_1 + \frac{6(\sum_i \sum_j w_{ij})^2}{1}.$$

З  $z$ -таблиць знаходиться  $p$ -значення ( $p$ -value) – величина, що використовується при тестуванні статистичних гіпотез і дорівнює ймовірності того, що випадкова величина з даним розподілом (розподілом тестової статистики при нульовій гіпотезі) прийме значення, не менше, ніж фактичне значення тестової статистики [46, 47].

- Якщо  $p$ -value  $< 0,05$ , то  $H_0$  відхиляється з рівнем значущості  $0,05$ , а приймається  $H_1$ .

- Якщо  $p$ -value  $\geq 0,05$ , то  $H_0$  приймається з надійністю  $0,95$ , а відхиляється  $H_1$ .

- Якщо  $H_0$  відхиляється, то  $z > 0$  говорить про додатну просторову автокореляцію,  $z < 0$  – про від’ємну.

Блок-схема алгоритму перевірки гіпотез на просторову автокореляцію приведена на рисунку 2.4.

Якщо перевірка просторової автокореляції показала її наявність, то можна будувати моделі. Зазвичай в просторовому аналізі розглядають три типи моделей: модель просторового лагу, модель лагу тільки на екзогенні змінні і модель просторової помилки [48-50].



## 2.3 Побудова просторових моделей

Просторові моделі використовуються для опису зв'язків між просторовими даними, які визначаються набором: {значення, координати}.

### 2.3.1 Модель просторового лагу

Найпростішою авторегресійною моделлю, призначеною для оцінювання зв'язків між групами близько розташованих об'єктів за одним фактором близькості (просторова близькість або близькість з точки зору часу в дорозі між об'єктами), є авторегресійна модель першого порядку, в якій в якості регресорів виступає просторовий лаг залежної змінної:

$$y = \alpha + \rho W y + \xi, \quad \xi \sim N(0, \sigma^2 I) \quad (2.9)$$

де  $y$  – вектор залежної змінної розмірності  $n \times 1$ ,  $W$  – стандартна просторова матриця ваг розмірності  $n \times n$ ,  $\rho$  – параметр, який відповідає за силу просторової залежності,  $\xi$  – вектор нормально розподіленого шуму.

Для оцінки параметра  $\rho$  застосовується метод максимальної правдоподібності, так як метод найменших квадратів дає зміщену і неспроможну оцінку.

Розширенням просторової авторегресійної моделі першого порядку є змішана просторова авторегресійна модель, що включає додатково вектор пояснюючих змінних аналогічно традиційним моделям регресійного аналізу. У матричній формі модель просторового лагу записується у вигляді:

$$Y = \alpha + X\beta + \rho W Y + \varepsilon \quad \varepsilon \sim N(0, \sigma^2 I) \quad (2.10)$$

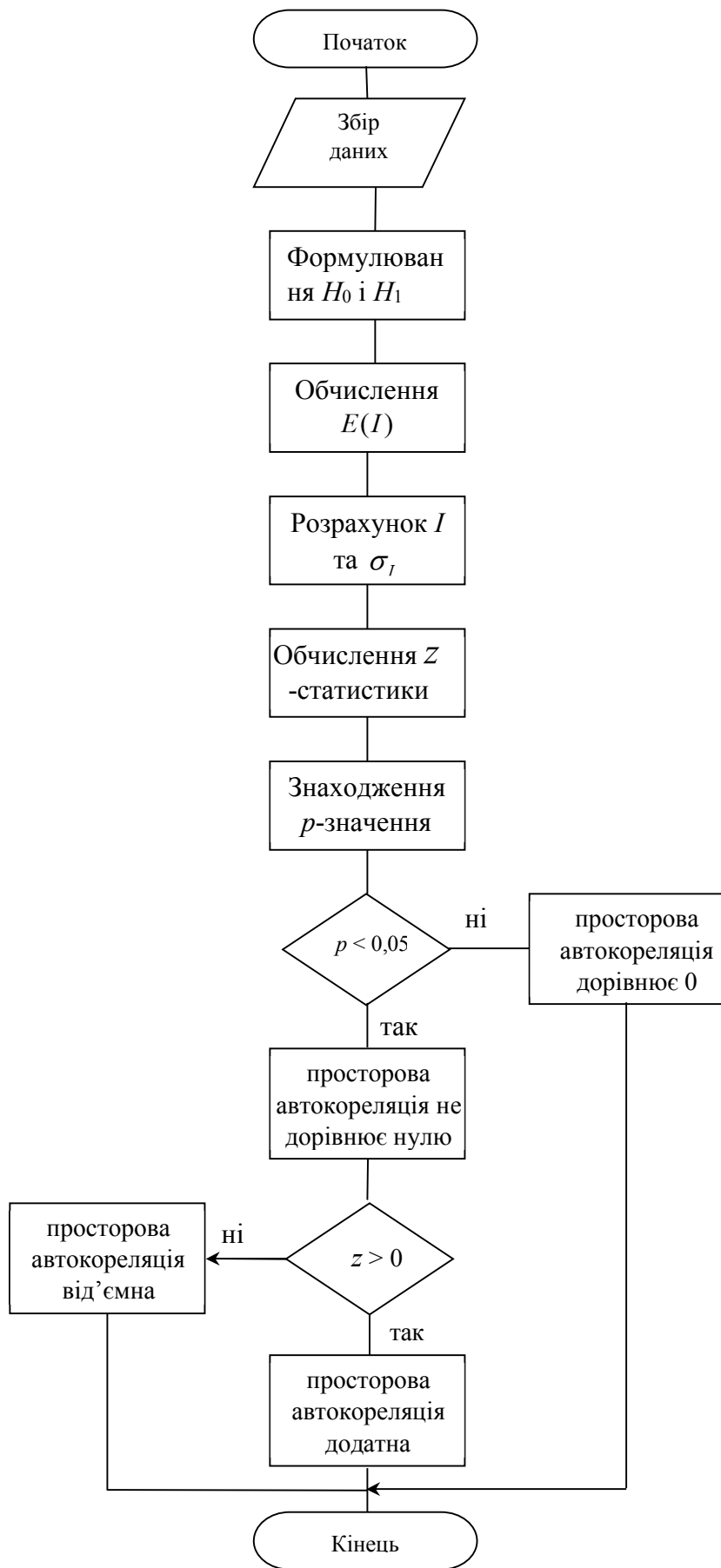


Рисунок 2.4 – Блок-схема алгоритму перевірки гіпотез на просторову автокореляцію

де  $Y$  –  $(n \times 1)$  вектор значень залежної змінної,  $X$  –  $(n \times k)$  матриця пояснюючих змінних,  $\beta$  –  $(k \times 1)$  вектор коефіцієнтів,  $WY$  – просторовий лаг ендогенної змінної,  $\rho$  – коефіцієнт просторової авторегресії. Кожен елемент просторового лага  $WY$  є зважене середнє (відповідно до ваг екзогенно заданої матриці  $W$ ) значень змінної  $Y$  сусідніх регіонів. У присутності стохастичного регресора  $WY$ , який завжди корельований з випадковою компонентою  $\varepsilon$ , метод найменших квадратів буде приводити до неспроможних оцінок [32, 44]. Тому оцінка моделі (2.10) проводиться методом максимальної правдоподібності або методом інструментальних змінних.

Модель конвергенції з просторовим лагом на ендогенну змінну має вигляд:

$$g_T = \alpha + \beta y_0 + \rho W g_T + \varepsilon \quad \varepsilon \sim N(0, \sigma^2 I) \quad (2.11)$$

Таким чином, основна гіпотеза, що перевіряється в такій моделі, полягає в тому, що темпи зростання ознаки, досліджуваної на збіжність, залежать не тільки від значень ознаки в початковий момент часу, але так само і від темпів зростання цього показника в сусідніх регіонах. Іншими словами, регіони в оточенні швидкокорослих регіонів так само будуть рости більш високими темпами. При цьому модель (2.11) може бути інтерпретована як мінімальна модель умовної конвергенції.

Дослідження включає модуль побудови моделей просторового лагу (Spatial Lag Model).

Побудова проходить за кроками:

1. Обчислення зваженого середнього:

$$[Wy]_i = w_{i1}y_1 + w_{i2}y_2 + \dots + w_{iN}y_N = \sum_{j=1}^N w_{ij}y_j, \quad i = \overline{1, N}. \quad (2.12)$$

2. Побудова моделей виду

$$y_i = \alpha + \rho \sum_{j=1}^N w_{ij} y_j + \varepsilon_i, \quad i = \overline{1, N}. \quad (2.13)$$

### 3. Перевірка коефіцієнтів на значущість:

Емпіричне значення відношення  $t_j$  для перевірки нульової гіпотези стосовно параметрів  $w_{ij}$  знаходиться за формулою:

$$t_j = \frac{|\hat{w}_{ij}|}{\sigma_{\hat{w}_{ij}}},$$

де  $\sigma_{\hat{w}_{ij}}$  – стандартна похибка оцінки  $\hat{w}_{ij}$ .

Емпіричне значення оцінки  $t_j$  порівнюють з критичним, знайденим за таблицями Стюдента, для заданого рівня значущості  $\gamma$  і  $k = n - m - 1$  ступенів вільності. Якщо  $|t_j| \leq t_{кр}$  то гіпотеза про рівність нулю параметра  $w_{ij}$  в генеральній сукупності не відхиляється, оцінка  $\hat{w}_{ij}$  є статистично незначущою; якщо  $|t_{емп}| > t_{кр}$  то із заданим рівнем значущості  $\gamma$  гіпотезу  $H_0$  треба відхилити і відповідну оцінку вважають статистично значущою.

### 4. Прогноз за моделями просторового лагу

$$y_{i_{n+k}} = \alpha + \rho \sum_{j=1}^N w_{ij} y_{j_{n+k}} + \varepsilon_i.$$

де  $y_{j_{n+k}}$  – очікувані значення змінних.

На рисунку 2.5 зображена блок-схема алгоритму побудови моделей просторового лагу (Spatial Lag Model).

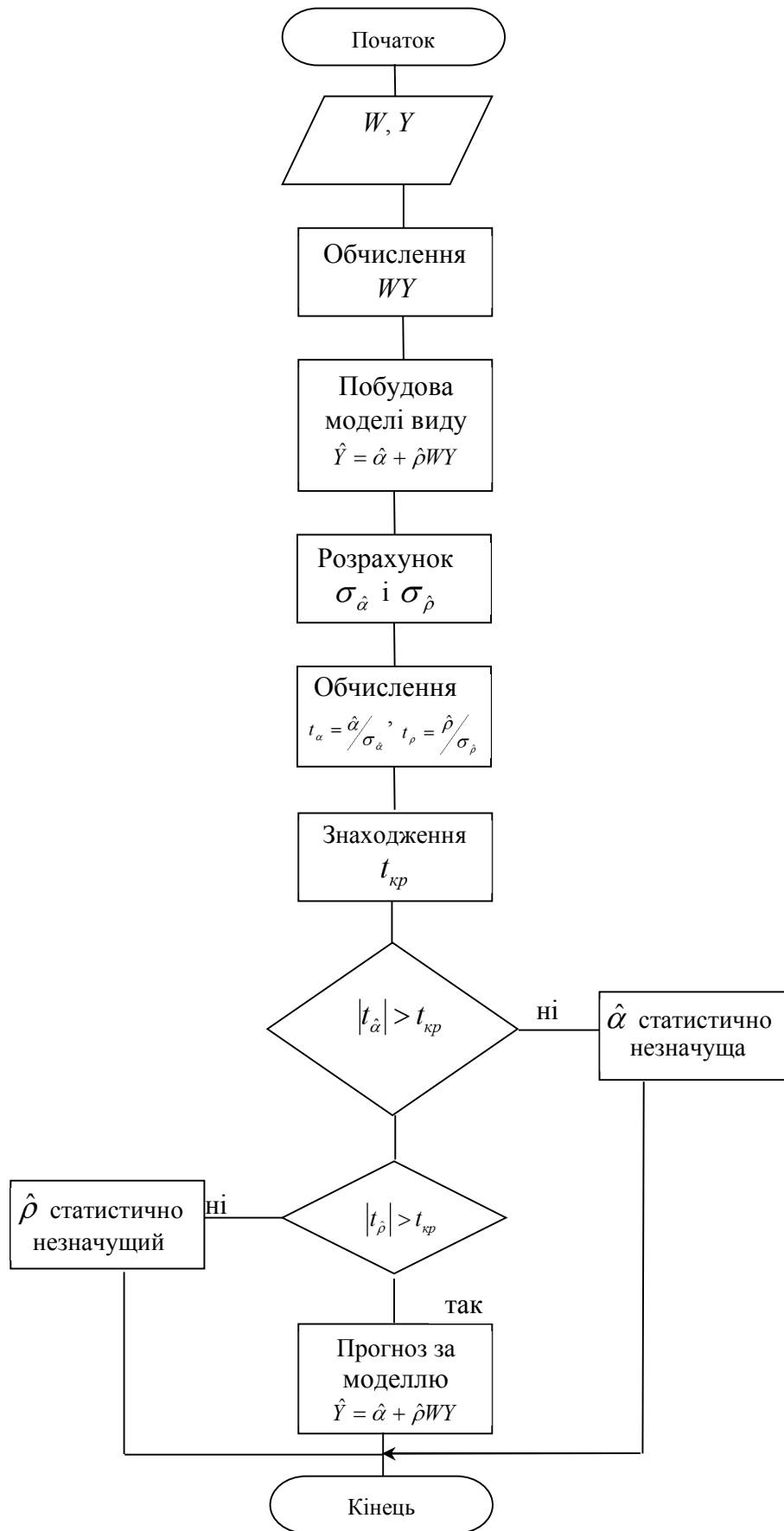


Рисунок 2.5 – Блок-схема алгоритму побудови моделі просторового лагу

### 2.3.2 Модель просторового лагу на екзогенні змінні

Альтернативний спосіб моделювання взаємовпливу регіонів – включити в модель просторовий лаг тільки на пояснючі змінні. У загальному випадку модель (Spatial Cross Regressive Model) записується у вигляді:

$$Y = \alpha + X\beta + WX\gamma + \varepsilon \quad \varepsilon \sim N(0, \sigma^2 I) \quad (2.14)$$

де  $\gamma$  –  $(k \times 1)$  вектор просторових коефіцієнтів. У присутності екзогенного просторового лага модель можна оцінювати безпосередньо методом найменших квадратів. При аналізі збіжності економічних показників, вважаючи єдиним екзогенним фактором початкове значення змінної, отримуємо окремий випадок моделі (2.14):

$$g_T = \alpha + \beta y_0 + \rho W y_0 + \varepsilon \quad \varepsilon \sim N(0, \sigma^2 I) \quad (2.15)$$

Модель (2.15) являє собою мінімальну модель умовної конвергенції з просторовим лагом на початкове значення. Змістовно, темпи зростання регіону залежать, по-перше, від початкового значення в самому регіоні, і, по-друге, від початкових значень змінної в сусідніх регіонах.

### 2.3.3 Модель просторової помилки

У моделі просторової помилки (Spatial Error Model) передбачається, що випадкова компонента слідує просторовому авторегресійному процесу першого порядку:

$$Y = \alpha + X\beta + u \quad u = \lambda W u + \varepsilon \quad \varepsilon \sim N(0, \sigma^2 I) \quad (2.16)$$

де  $\lambda$  – коефіцієнт просторової кореляції залишків. Оцінка моделі методом найменших квадратів приводить до неефективних оцінок. До того ж оцінки

можуть виявитися зміщеними за рахунок пропущених змінних. Тому така модель оцінюється методом максимальної правдоподібності або узагальненим методом моментів. Виражаючи випадкові помилки  $u = (1 - \lambda W)^{-1} \varepsilon$  і підставляючи їх у вихідне рівняння, модель (2.16) можна перетворити до вигляду:

$$Y = (1 - \lambda W)\alpha + X\beta + \lambda WY + WX\gamma + \varepsilon \quad \varepsilon \sim N(0, \sigma^2 I) \quad (2.17)$$

з обмеженням на коефіцієнти  $\gamma + \lambda\beta = 0$ . Модель (2.17) називається просторовою моделлю Дарбіна (Spatial Durbin model) і включає просторовий лаг, як на ендогенну, так і на всі екзогенні змінні. З позицій аналізу конвергенції модель

$$g_T = (1 - \lambda W)\alpha + \beta y_0 + \lambda Wg_T + \gamma W y_0 + \varepsilon \quad \varepsilon \sim N(0, \sigma^2 I) \quad (2.18)$$

має наступну цікаву інтерпретацію. Темпи зростання в регіоні можуть бути пов'язані з темпами зростання в сусідніх регіонах через ендогенний просторовий лаг і з початковими значеннями досліджуваної ознаки в сусідніх регіонах за допомогою екзогенного просторового лага.

## 2.4 Висновки до розділу 2

В даному розділі розроблено алгоритми: побудови матриці просторових ваг, обчислення показників просторової кореляції, перевірки гіпотез про просторову автокореляцію та побудови моделі просторового лагу, що дало змогу розробити алгоритми кластеризації регіонів за кількістю ІТ-компаній.

## 3 КОМП'ЮТЕРНЕ МОДЕЛЮВАННЯ АЛГОРИТМІВ ПРОСТОРОВОЇ ЕКОНОМЕТРИКИ

### 3.1 Вибір програмних засобів та системні вимоги

Для програмної реалізації поставленої задачі, необхідно вибрати програмні засоби, мову програмування.

Виходячи із поставленого завдання нами вибрано програмний додаток `gwrcorMapreg`. Це – програмний додаток для відображення географічно зваженої кореляції та часткової кореляції у великих багатовимірних наборах даних [3, 51]. `gwrcorMapreg`, дає дослідникам можливість взаємодіяти з компонентами карти, що описують локальні кореляційні відносини. `gwrcorMapreg` розроблений на R з використанням фреймворку Shiny. Основний алгоритм розробки був успадкований від `GWrcor` бібліотеки R, для розрахунку географічно виваженої кореляції та статистики часткової кореляції. `gwrcorMapreg` корисний як при виборі змінних, так і при налаштуванні параметрів для географічно зваженої статистики. `gwrcorMapreg` використовує великі геопросторові статистичні бібліотеки R, розташовані бібліотеки на таких сховищах, як GitHub та CRAN.

Інтерфейс `gwrcorMapreg` складається з трьох панелей (рисунок 3.1): I – панелі даних (I), панелі карти (II) та панелі діаграми розсіювання (III). Панель даних містить елементи керування для завантаження даних та налаштування параметрів (вхід), тоді як панелі карти та розсіювання відображають результуючі коефіцієнти кореляції як на карті, так і на графіку розсіювання відповідно (вихід).

Доступні типи ядер в полі `Kernel type` представлені на рисунку 3.2.

Опис панелей інтерфейсу `gwrcorMapreg` приведено в таблиці 3.1.

Функції ядра, використовувані в `gwrcorMapreg` для обчислення матриці ваг приведені в таблиці 3.2.  $w_{ij}$  – це вага  $j$ -го елемента в матриці просторових ваг  $W$ , який розраховується для  $i$ -го спостереження в матриці даних,  $d_{ij}$  – відстань між  $j$ -ю точкою спостереження та  $i$ -ю та  $b = d_i(k)$  – ефективна смуга пропускання (2.2).



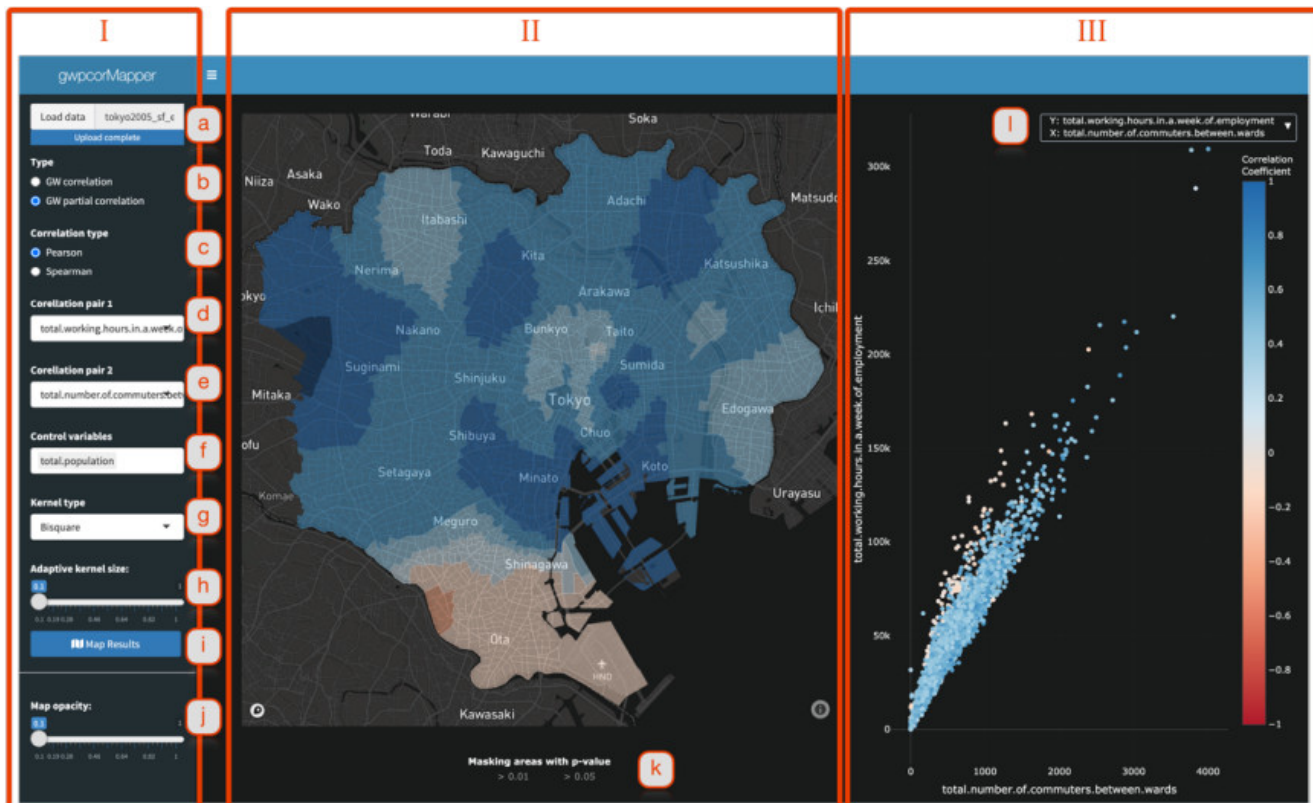


Рисунок 3.1 – Інтерфейс gwpcorMapper

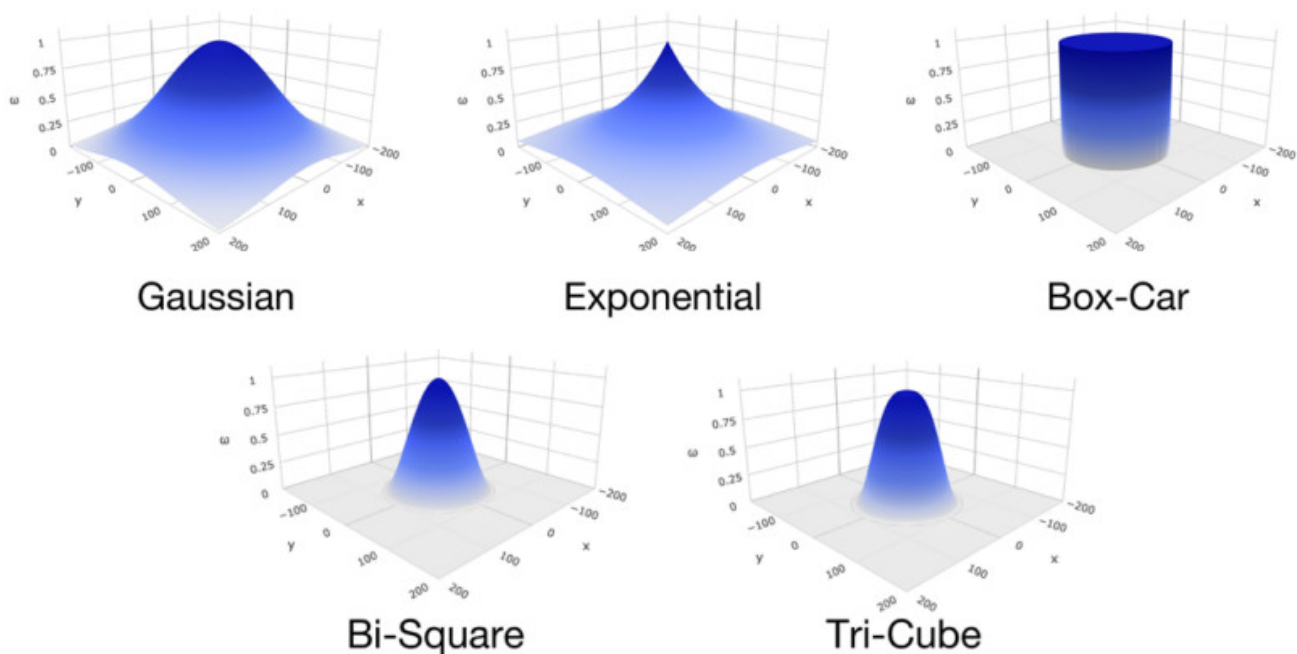


Рисунок 3.2 – Доступні типи ядер в полі Kernel type

Таблиця 3.1 – Опис панелей інтерфейсу gwrcongMapper

| Поле введення        | Опис   | Мітка |
|----------------------|--|-------|
| Load data            | Призначене для завантаження даних. Дані заносяться в файлах форматів .geojson, .json, .csv.                | a     |
| Type                 | Вибір методу кореляції. Користувач може вибирати кореляцію або частинну кореляцію                          | b     |
| Correlation type     | Вибір коефіцієнта кореляції. Користувач може вибирати коефіцієнт кореляції Пірсона або Спірмена            | c     |
| Correlation pair 1   | Випадаючий список переліку змінних для вибору першої змінної для аналізу                                   | d     |
| Correlation pair 2   | Випадаючий список переліку змінних для вибору другої змінної для аналізу                                   | e     |
| Control variables    | Контрольні змінні. Випадаючий список для вибору будь-якої кількості контрольних змінних для аналізу        | f     |
| Kernel type          | Випадаючий список для вибору типу ядра для аналізу   | g     |
| Adaptive kernel size | Адаптивний селектор пропускнуої здатності розміру ядра   | h     |
| Map Results          | Кнопка відображення результатів. Для запуску аналізу та відображення результатів                           | i     |
| Map Opacity          | Непрозорість карти. Контроль непрозорості рівня результатів на карті                                       | j     |
| p-value Mask         | p-значення для запуску маскуванню статистичного t-тесту з використанням значень p на порогах 0,01 або 0,05 | k     |
| Variable Selector    | Випадаючий список для вибору пари змінних для побудови графіку в частковому кореляційному аналізі          | l     |

Таблиця 3.2 – Функції ядра для обчислення матриці просторових ваг

| Тип ядра    | Формула   |
|-------------|---|
| Gaussian    | $w_{ij} = \exp\left(-\frac{1}{2}\left(\frac{d_{ij}}{b}\right)^2\right)$   |
| Exponential | $w_{ij} = \exp\left(-\frac{ d_{ij} }{b}\right)$   |
| Box-Car     | $w_{ij} = \begin{cases} 1 & \text{якщо }  d_{ij}  < b, \\ 0 & \text{в іншому випадку} \end{cases}$  |
| Bi-Square   | $w_{ij} = \begin{cases} \left(1 - \left(\frac{d_{ij}}{b}\right)^2\right)^2 & \text{якщо }  d_{ij}  < b, \\ 0 & \text{в іншому випадку} \end{cases}$   |
| Tri-Cube    | $w_{ij} = \begin{cases} \left(1 - \left(\frac{ d_{ij} }{b}\right)^3\right)^3 & \text{якщо }  d_{ij}  < b, \\ 0 & \text{в іншому випадку} \end{cases}$ |

gwrcorMapper забезпечує: географічно зважену кореляцію; часткову географічно зважену кореляцію із статистичним t-тестом для перевірки, чи спостерігається кореляція зі статистичною значущістю; обчислює діагональну матрицю ваг. Просторові ваги визначаються з допомогою функцій ядер, які доступні в gwrcorMapper: Gaussian, Exponential, Box-Car, Bi-Square та Tri-Cube (рисунок 3.3). Також в програмному додатку можна змоделювати різноманітні просторові моделі: модель просторового лагу (2.10), модель просторового лагу на екзогенні змінні (2.14), модель просторової помилки (2.16), коефіцієнт Морана (2.6) тощо.

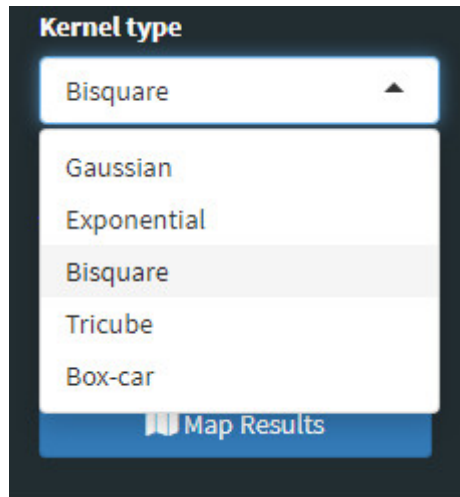


Рисунок 3.3 – Функції ядер

Для правильного функціонування програмного додатка персональний комп'ютер повинен відповідати мінімальним системним вимогам:

- процесор з тактовою частотою  $\geq 2,2$ ГГц;
- оперативна пам'ять обсягом 8Гб та поколінням пам'яті DDR4 або DDR3;
- відеокарта обсягом пам'яті не менш ніж 4Гб та з поколінням пам'яті не нижче GDDR5;
- жорсткий диск розміром 50Гб.

Програмне забезпечення:

- операційна система Windows 10 x64 або Ubuntu x64 або Linux;
- середовище R;
- остання версія драйверів для відеокарти.

### 3.2 Структура програмної системи

В даному підрозділі на базі алгоритмів побудови матриці просторових ваг, обчислення показників просторової кореляції, побудови моделей просторового лагу розроблена програмна система просторової економетрики (рисунок 3.4).

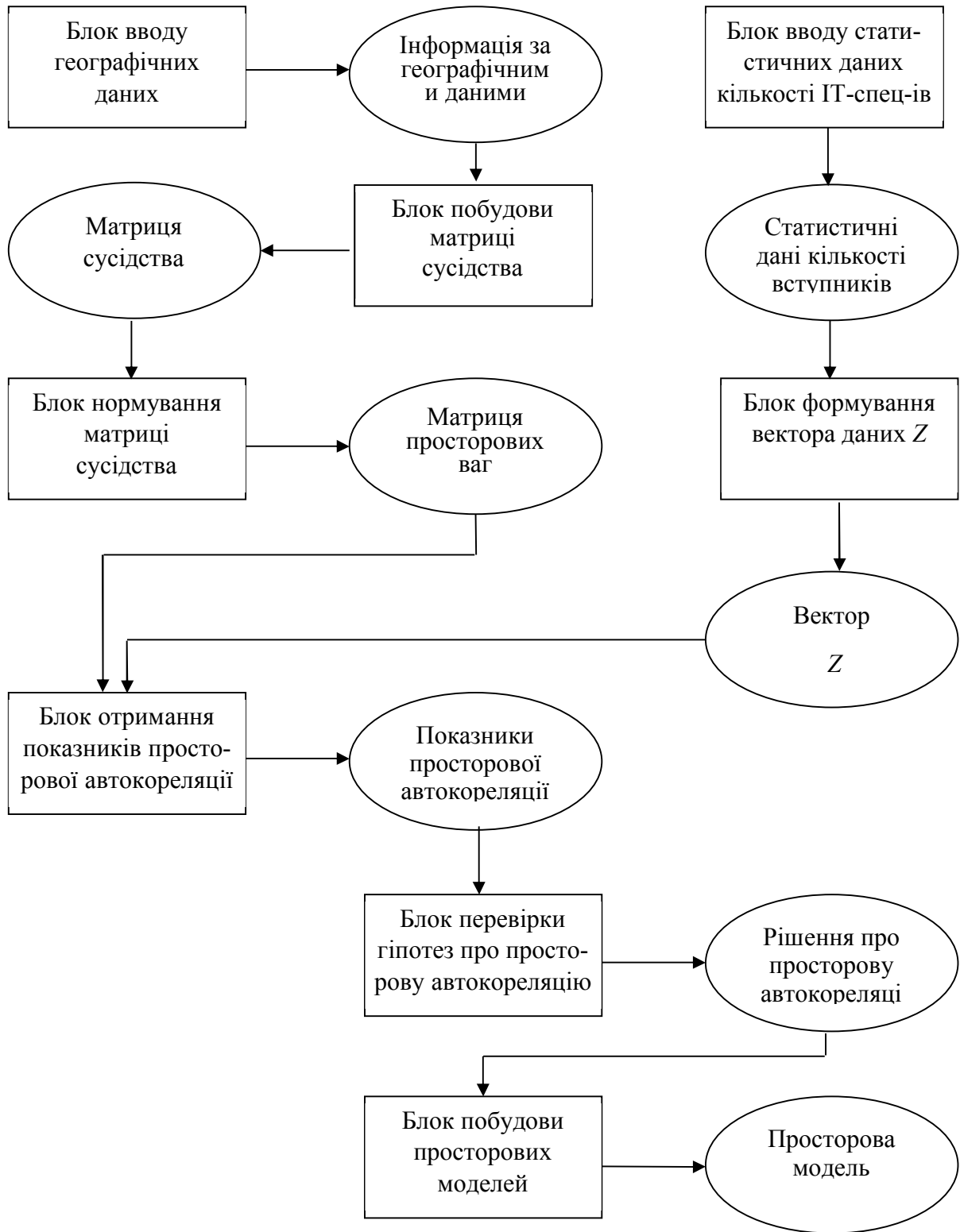


Рисунок 3.4 – Програмна система просторової економетрики

Відповідно до цієї схеми географічні дані подаються для отримання інформації по сусідству областей (довгота і широта місцезнаходження обласних центрів) таблиця А.3, в результаті чого на виході блоку побудови матриці сусідства отримуємо матрицю сусідства і після її нормування – матрицю просторових ваг. Далі вибірка статистичних даних кількості ІТ-спеціалістів за областями подається на вхід блоку формування вектора даних  $Z$  (формула 2.7) і цей сформований вектор разом з матрицею просторових ваг використовуються в блоці обчислення показників загальної просторової автокореляції. В результаті, на виході цього блоку отримуємо статистику Морана, діаграму розсіювання, кластеризацію областей України на 4 квадранти.

Блок перевірки гіпотез про просторову автокореляцію використовується для аналізу. В результаті цього блоку отримуємо інформацію, чи варто даліше будувати просторові моделі і здійснювати за ними прогнозування. На виході блоку перевірки гіпотез отримуємо рішення про просторову автокореляцію. Якщо нульова гіпотеза відхиляється, а приймається альтернативна гіпотеза з певним рівнем значущості, то переходимо до блоку побудови просторових моделей. В результаті побудови – отримуємо просторові моделі.

Особливості розробленої ІС алгоритмів просторової економетрики наступні:

1) блок отримання показників просторової автокореляції включає запропонований в підрозділі 2.2 алгоритм обчислення показника загальної просторової автокореляції (статистики Морана), побудову діаграми розсіювання і кластеризацію районів на чотири квадранти (НН, ЛН, LL, НЛ) кожен з яких якісно характеризується певним типом просторової близькості;

2) в блоці перевірки гіпотез про просторову автокореляцію застосовується запропонований в підрозділі 2.2 алгоритм перевірки гіпотез про просторову автокореляцію і блок-схема з рисунку 2.3;

3) в блоці побудови просторових моделей використовується алгоритм побудови моделей просторового лагу, запропонований в підрозділі 2.3.1, блок-схема якого приведена на рисунку 2.4.

Розроблена ПС була застосована для аналізу і прогнозу кількості ІТ-спеціалістів і компаній з областей України.

Розроблену ПС алгоритмів просторової економетрики можна розбити на три модулі: модуль побудови матриці просторових ваг, модуль отримання показників просторової автокореляції, модуль тестування просторових моделей. Розглянемо реалізацію кожного з цих модулів.

### 3.3 Модуль побудови матриці просторових ваг

Для побудови матриць просторових ваг використовувалися статистичні дані таблиць А.3 і А.4 додатку А. Таблиця А.3 містить географічні координати обласних центрів і міст України та кількість ІТ-спеціалістів. Інформація про географічні координати основних населених пунктів світу є у мережі Інтернет. Маючи географічні координати – розраховуємо відстані між обраними містами. Таблиця А.4 є таблицею відстаней між обласними центрами і містами України. В таблиці 3.3 приведені обласні центри і окремі міста України. В таблиці 3.4 приведені поля і значення цих полів з бази даних таблиці А.3.

Таблиця 3.3 – Список обласних центрів і міст України

|   |                  |    |               |    |                |
|---|------------------|----|---------------|----|----------------|
| 1 | Сімферополь      | 10 | Київська обл. | 19 | Тернопіль      |
| 2 | Вінниця          | 11 | Кропивницький | 20 | Харків         |
| 3 | Луцьк            | 12 | Луганськ      | 21 | Херсон         |
| 4 | Дніпро           | 13 | Львів         | 22 | Хмельницький   |
| 5 | Донецьк          | 14 | Миколаїв      | 23 | Черкаси        |
| 6 | Житомир          | 15 | Одеса         | 24 | Чернігів       |
| 7 | Ужгород          | 16 | Полтава       | 25 | Чернівці       |
| 8 | Запоріжжя        | 17 | Рівне         | 26 | м. Київ        |
| 9 | Івано-Франківськ | 18 | Суми          | 27 | м. Севастополь |

Таблиця 3.4 – Поля бази даних географічного положення міст України

| Поле таблиці        | Значення  |
|---------------------|---|
| city                | місто   |
| lat                 | географічна широта  |
| lng                 | географічна довгота   |
| country             | країна  |
| iso2                | міжнародне скорочення назви країни                                  |
| admin_name          | приналежність до адміністративної одиниці (до області)              |
| capital             | тип міста: primary – столиця, admin – обласний центр, minor - місто |
| IT specialists      | офіційна кількість ІТ-спеціалістів                                  |
| fact IT specialists | фактична кількість ІТ-спеціалістів за даними «Портрета»             |

В додатку Б приведені матриця просторових ваг за граничними сусідами та матриця просторових ваг за відстанями (відсікання за квантилями відстаней). Обидві є побудовані з допомогою `gwpcorMapper`. На рисунку 3.5 показано фрагмент матриці просторових ваг за граничними сусідами у вікні `gwpcorMapper`.

### 3.4 Модуль отримання показників просторової автокореляції

Наступний модуль – перевірка змінних моделі і залишків, оцінених методом найменших квадратів, на можливу просторову кореляцію. В `gwpcorMapper` за допомогою матриць просторових ваг розраховується показник загальної просторової автокореляції (Global Moran's I).

Гіпотеза про значущість індексу Морана перевіряється за допомогою z-статистики.





Рисунок 3.5 – Фрагмент матриці просторових ваг за граничними сусідами у вікні gwpcorMapper

З таблиці 3.5 видно, що статистика Морана  $I$ , яка показує наявність просторової взаємозалежності між об'єктами виявляється статистично незначущою на 10%-ому рівні за матрицею граничних сусідів і за матрицею відстаней, що приймає нульову гіпотезу про відсутність просторового взаємозв'язку і свідчить про відсутність взаємовпливу між нашими спостереженнями, тобто між областями України.

Діаграма розсіювання Морана для розподілу кількості офіційних ІТ-спеціалістів з використанням матриці граничних сусідів приведена на рисунку 3.6.

Таблиця 3.5 – Індeksi Морана

| Показник            | Матриця граничних сусідів | Матриця відстаней |
|---------------------|---------------------------|-------------------|
| IT specialists      | -0,0735                   | -0,0953           |
| fact IT specialists | -0,061                    | -0,0508           |

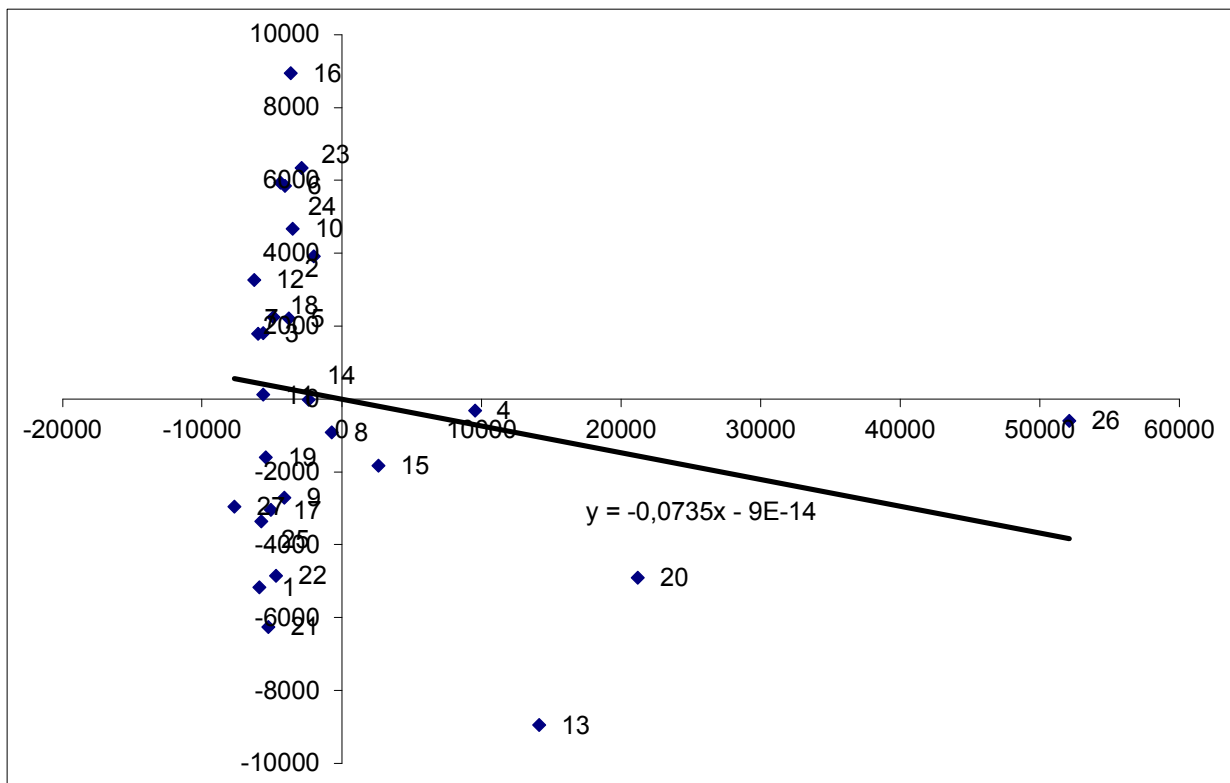


Рисунок 3.6 – Діаграма розсіювання Морана для кількості офіційних ІТ-спеціалістів за матрицею граничних сусідів

В I-ий кластер – регіонів з великою кількістю ІТ-компаній, які знаходяться в оточенні регіонів теж з великою кількістю ІТ-компаній – за матрицею граничних сусідів не попала жодна область.

В II-ий кластер – регіонів з невеликою кількістю ІТ-компаній, які знаходяться в оточенні регіонів з великою кількістю ІТ-компаній – за матрицею граничних сусідів попали: Вінницька обл., Волинська обл., Донецька обл.,

Житомирська обл., Закарпатська обл., Київська обл., Кіровоградська обл., Луганська обл., Миколаївська обл., Полтавська обл., Сумська обл., Черкаська обл., Чернігівська обл.

В III-ій кластер – регіонів з невеликою кількістю ІТ-компаній, які знаходяться в оточенні регіонів теж з невеликою кількістю ІТ-компаній ввійшли: АР Крим, Запорізька обл., Івано-Франківська обл., Рівненська обл., Тернопільська обл., Херсонська обл., Хмельницька обл., Чернівецька обл, м. Севастополь.

В IV-ий кластер – регіонів з великою кількістю ІТ-компаній, які знаходяться в оточенні регіонів з невеликою кількістю ІТ-компаній – за матрицею граничних сусідів попали: Дніпропетровська обл., Одеська обл., Львівська обл., Харківська обл., м. Київ.

Інші діаграми розсіювання Морана показані на рисунках 3.7 – 3.9.

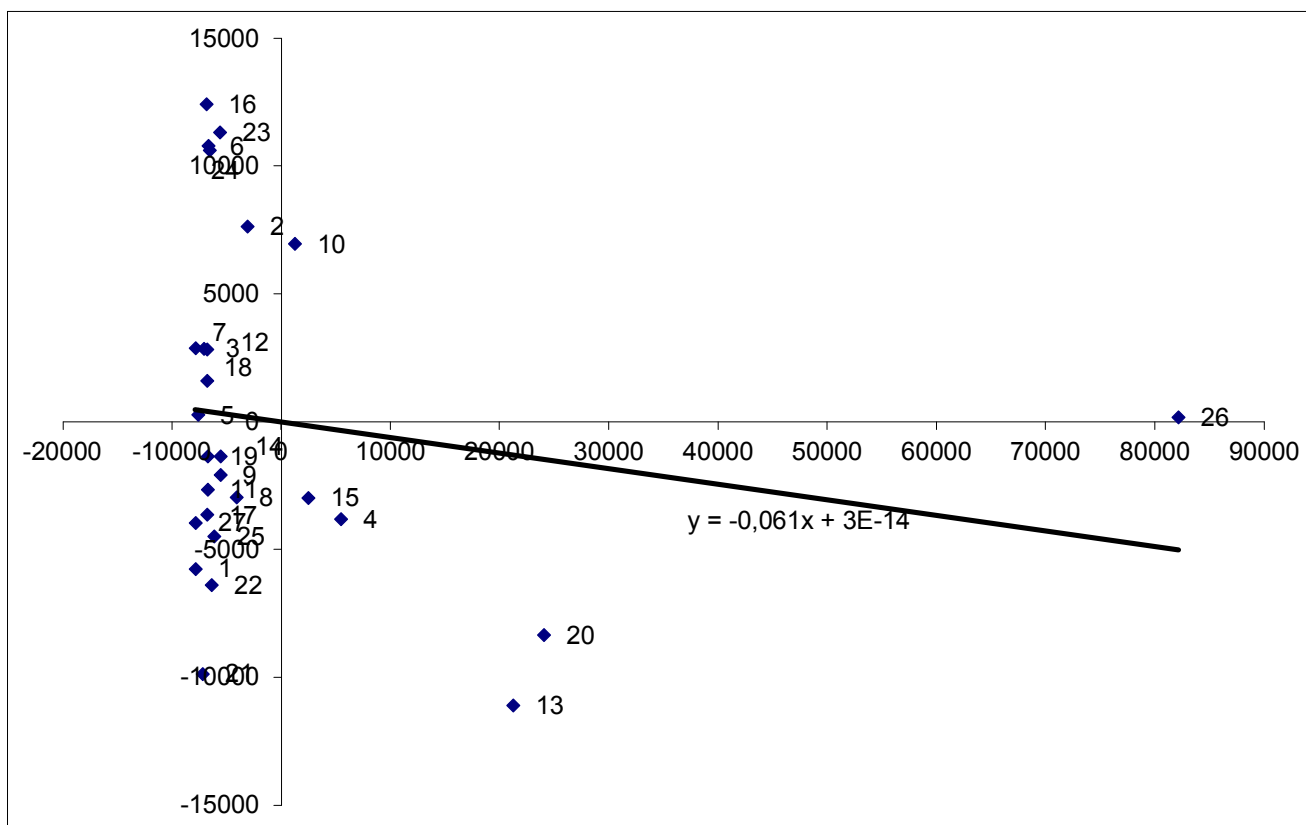


Рисунок 3.7 – Діаграма розсіювання Морана для кількості фактичних ІТ-спеціалістів за матрицею граничних сусідів

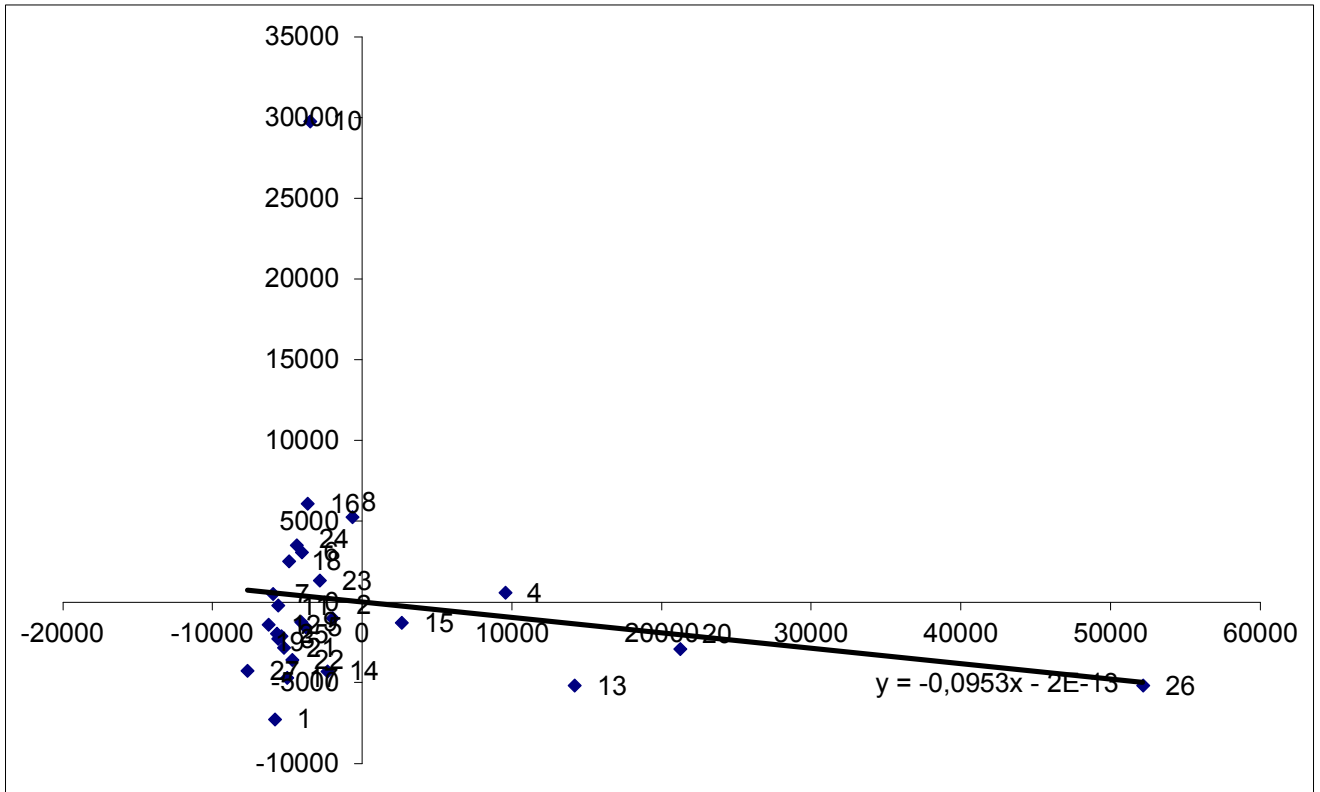


Рисунок 3.8 – Діаграма розсіювання Морана для кількості офіційних ІТ-спеціалістів за матрицею відстані

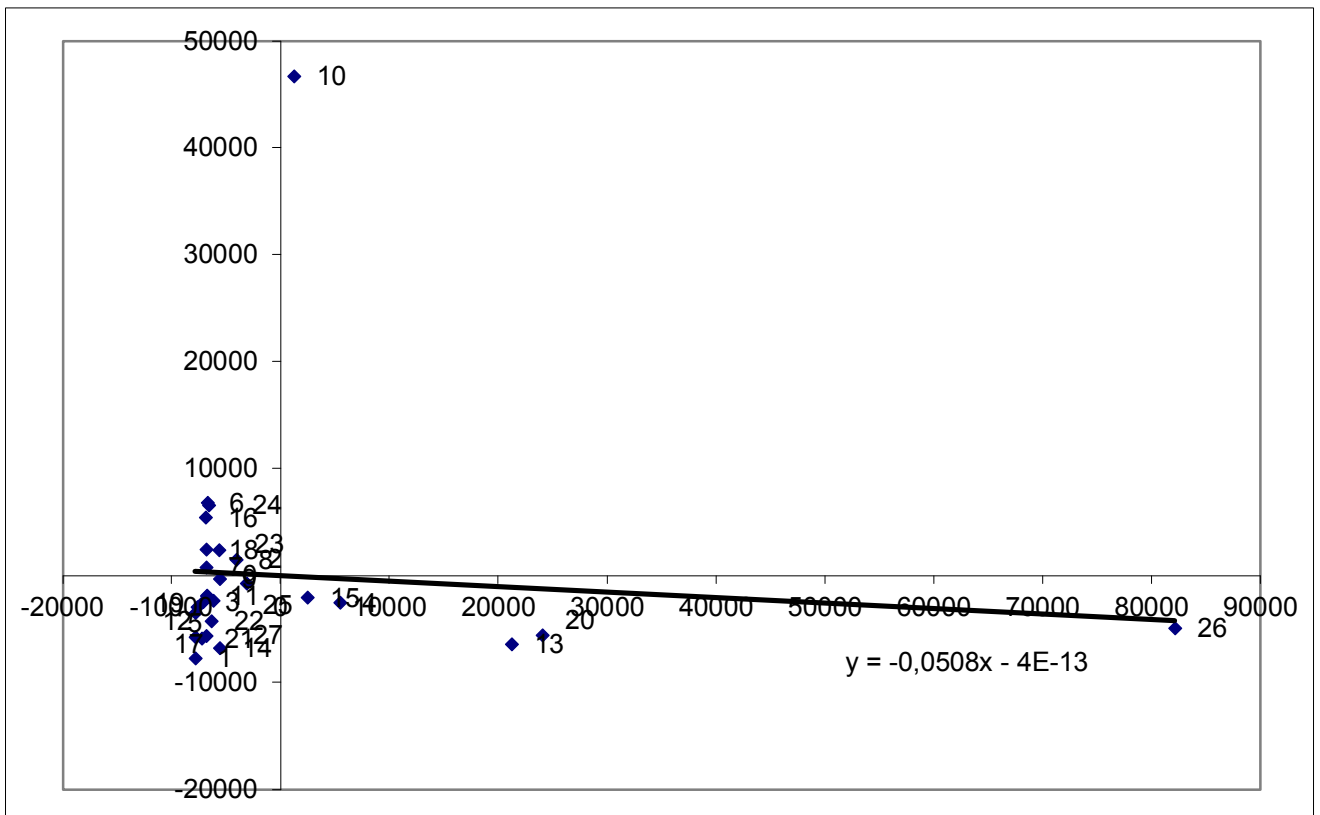


Рисунок 3.8 – Діаграма розсіювання Морана для кількості фактичних ІТ-спеціалістів за матрицею відстані

На основі статистики Морана показано, що спостерігається від'ємна автокореляція, тобто в цілому кількість ІТ-спеціалістів в сусідніх регіонах є різною. Але  $I \approx 0$ , тому кількість ІТ-спеціалістів в сусідніх регіонах розташовані випадково. Через те відпадає необхідність блоці побудови просторових моделей.

В таблиці 3.6 приведена кластеризація регіонів за матрицею граничних сусідів для фактичних ІТ-спеціалістів за даними «Портрета», за матрицею просторових ваг за відстанями (відсікання за квантилями відстаней) для офіційних ІТ-спеціалістів та фактичних ІТ-спеціалістів за даними «Портрета». Нумерація областей і міст співпадає з приведеною в таблиці 3.1.

Таблиця 3.6 – Кластеризація областей

| Вид матриці ваг                             | I-й кластер | II-й кластер                                  | III-й кластер                                     | IV-кластер              |
|---|-------------|---|---|-------------------------|
| граничних сусідів (офіційні ІТ-спеціалісти) | –           | 2, 3, 5, 6, 7, 10, 11, 12, 14, 16, 18, 23, 24 | 1, 8, 9, 17, 19, 21, 22, 25, 27                   | 4, 13, 15, 20, 26       |
| граничних сусідів (фактичні ІТ-спеціалісти) | 10, 26      | 2, 3, 5, 6, 7, 12, 16, 18, 23, 24             | 1, 8, 9, 11, 14, 17, 19, 21, 22, 25, 27           | 4, 13, 15, 20           |
| за відстанями (офіційні ІТ-спеціалісти)     | 4           | 6, 7, 8, 10, 16, 18, 23, 24                   | 1, 2, 3, 5, 9, 11, 12, 14, 17, 19, 21, 22, 25, 27 | 13, 15, 20, 26          |
| за відстанями (фактичні ІТ-спеціалісти)     | 10          | 2, 6, 7, 8, 16, 18, 23, 24                    | 1, 9, 11, 12, 14, 17, 19, 21, 22, 25, 27          | 3, 4, 5, 13, 15, 20, 26 |

Отримані результати кластеризації регіонів відображені на картограмах (рисунки 3.3-3.5).



Рисунок 3.3 – Кластеризація регіонів на діаграмі Морана (за матрицею граничних сусідів – офіційні дані)



Рисунок 3.3 – Кластеризація регіонів на діаграмі Морана (за матрицею граничних сусідів – фактичні дані)





Рисунок 3.3 – Кластеризація регіонів на діаграмі Морана (за матрицею відстаней – офіційні дані)

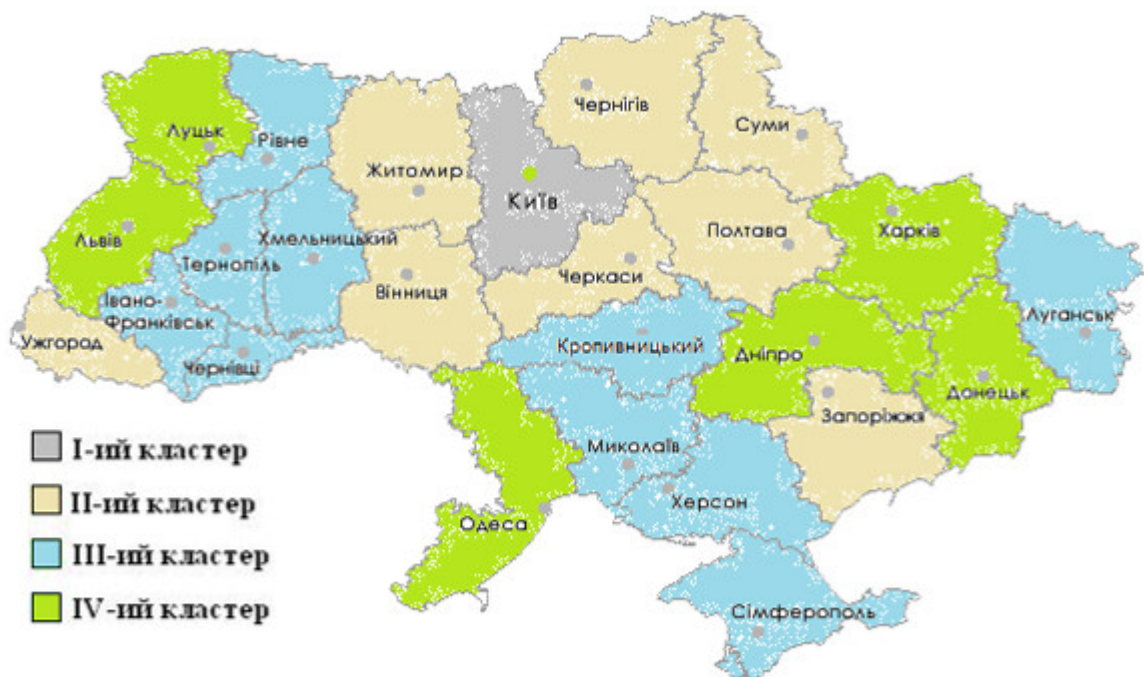


Рисунок 3.3 – Кластеризація регіонів на діаграмі Морана (за матрицею відстаней – фактичні дані)

Для покращення розвитку ІТ-галузі треба збільшувати кількість ІТ-фірм (кількість ІТ-спеціалістів) в тих регіонах, що попали в II-й і III-й кластери.

Виявлення просторової кореляції (кластеризації регіонів) за допомогою коефіцієнта загальної просторової автокореляції та діаграми розсіювання продемонструвало, що ІТ-галузь працює зосереджена в п'яти містах України і практично не має впливу на інші міста.

### 3.6 Висновки до розділу 3

В даному розділі на основі запропонованих алгоритмів розроблено модуль побудови матриці просторових ваг, модуль отримання показників просторової автокореляції, модуль тестування просторових моделей, що дало можливість провести просторову кластеризацію ІТ-компаній в Україні.



## ВИСНОВКИ

1. Здійснено аналіз розвитку ІТ-індустрії в Україні.
2. Проаналізовано методи і задачі регресійного та кореляційного аналізу, що дозволило переконатися у необхідності використовувати просторовий регресійно-кореляційний аналіз у випадку, коли об'єкти характеризуються значенням і координатою.
3. Розроблено алгоритми побудови матриці просторових ваг, алгоритми обчислення показників просторової кореляції, що дало змогу розробити алгоритми кластеризації областей України за кількістю ІТ-компаній.
4. На основі розроблених алгоритмів просторової економетрики здійснено комп'ютерне моделювання, що дало можливість здійснити кластеризацію областей України за кількістю ІТ-компаній.
5. На основі запропонованих алгоритмів розроблено модуль побудови матриці просторових ваг, модуль отримання показників просторової автокореляції, модуль тестування просторових моделей в програмному додатку `gwpcorMapper`.

## СПИСОК ВИКОРИСТАНИХ ДЖЕРЕЛ

1. Скільки IT-спеціалістів в Україні: +29 тисяч за рік згідно з Мін'юстом [Електронний ресурс]. – Режим доступу: <https://dou.ua/lenta/articles/how-many-devs-in-ukraine-2020/>.
2. Березька К. М., Цимбалюк Д. В., Іванов Ю. М. Застосування програмного додатку gwrsofMapper в просторовій економетриці. *Інтелектуальні комп'ютерні системи та мережі* : тези доп. V Наук.-практ. конф. молодих вчених і студентів (2 груд. 2021 р.). Тернопіль : ЗУНУ, 2021. С.
3. Цимбалюк Д. В., Іванов Ю. М. Програмна система оцінки локалізації українських IT-компаній. *Інтелектуальні комп'ютерні системи та мережі* : тези доп. V Наук.-практ. конф. молодих вчених і студентів (2 груд. 2021 р.). Тернопіль : ЗУНУ, 2021. С.
4. Березький О. М., Дубчак Л. О., Мельник Г. М. Методичні рекомендації до виконання кваліфікаційної роботи з освітнього ступеня “Магістр”. Спеціальність: 123 - Комп'ютерна інженерія. Магістерська програма – «Комп'ютерна інженерія». Тернопіль : ЗУНУ, 2021. 32 с.
5. Методичні вказівки до оформлення курсових проектів, звітів про проходження практики, випускних кваліфікаційних робіт для студентів спеціальності «Комп'ютерна інженерія» / І.В. Гураль, Л.О. Дубчак / Під ред. О.М. Березького. Тернопіль: ТНЕУ, 2019. 33 с.
6. Software development in Ukraine: 2019-2020 IT market report [Електронний ресурс]. – Режим доступу: <https://www.n-ix.com/software-development-in-ukraine-2019-2020-market-report/>.
7. Декрет І. Драйвери розвитку IT-компаній в Україні [Електронний ресурс] / Інноваційне підприємництво: стан та перспективи розвитку : зб. матеріалів V Всеукр. наук.-практ. конф., 29–30 берез. 2020 р. Київ : КНЕУ, 2020. С. 86–89.

8. Офіційний сайт НБУ [Електронний ресурс]. – Режим доступу: <https://bank.gov.ua/statistic>.
9. Software development professionals in Ukraine: Overview 2020 [Електронний ресурс]. – Режим доступу: <https://www.ssa.group/blog/software-development-professionals-in-ukraine-overview-2020/>.
10. Офіційний сайт IT Ukraine Association [Електронний ресурс]. – Режим доступу: <https://itukraine.org.ua/en/it-industry-in-ukraine-count-not-neglect.html>
11. Офіційний сайт N-iX [Електронний ресурс]. – Режим доступу: [https://www.n-ix.com/softwaredevelopment-in-ukraine-2019-2020-marketreport/?utm\\_source=dou.ua&utm\\_medium=referral&utm\\_campaign=Dou%20\(Ukraine%20IT%20Report\)%20updated](https://www.n-ix.com/softwaredevelopment-in-ukraine-2019-2020-marketreport/?utm_source=dou.ua&utm_medium=referral&utm_campaign=Dou%20(Ukraine%20IT%20Report)%20updated).
12. Ukrainian IT Industry 2019-2020: \$5 billion and 200,000 specialists [Електронний ресурс]. – Режим доступу: <https://ain.ua/en/2019/08/30/ukrainian-it-industry-2019-2020/>.
13. Винничук Р. О. Особливості розвитку IT-ринку в Україні: стан та тенденції / Р. О. Винничук, Т. В. Склярчук // Вісник Національного університету "Львівська політехніка". Логістика. – 2015. – № 833. – С. 3-8.
14. Юрій Віннічук Рейтинг найбільших IT-компаній України [Електронний ресурс]. – Режим доступу: [https://biz.censor.net/resonance/3273242/rejting\\_nayiblshih\\_tkompanyi\\_ukrani](https://biz.censor.net/resonance/3273242/rejting_nayiblshih_tkompanyi_ukrani).
15. Огляд IT-ринку праці: Тернопіль [Електронний ресурс]. – Режим доступу: <https://dou.ua/lenta/articles/it-market-ternopil/>.
16. IT-компанія JSSolutions повідомляє про розвиток компанії під новим брендом Apiko [Електронний ресурс]. – Режим доступу: <http://fcit.tneu.edu.ua/web-rozrobka/mysql/43-novynty/novynty-fakultetu/913-apiko>.
17. Аналіз даних [Електронний ресурс]. – Режим доступу: [https://uk.wikipedia.org/wiki/Аналіз\\_даних](https://uk.wikipedia.org/wiki/Аналіз_даних).
18. Вох, G., Jenkins G., Reinsel, G., Ljung G. (2016) *Time Series Analysis: Forecasting and Control*, 5th Edition: Wiley.

19. Функції Excel (за категоріями) [Електронний ресурс]. – Режим доступу: <https://support.office.com/uk-ua/article>.
20. Универсальная статистическая диалоговая система STADIA [Електронний ресурс]. – Режим доступу: <http://statsoft.msu.ru/products.htm>.
21. EViews Econometric Modeling Software [Електронний ресурс]. – Режим доступу: <https://www.ihs.com/products/eviews-econometric-modeling-analysis-software.html>.
22. StatSoft's Electronic Statistics Textbook [Електронний ресурс]. – Режим доступу: <http://www.statsoft.com/textbook/stathome.html>.
23. IBM SPSS Statistics [Електронний ресурс]. – Режим доступу: <http://www-03.ibm.com/software/products/ru/spss-statistics>.
24. Prognoz Platform [Електронний ресурс]. – Режим доступу: <https://krit.pro/platforma/>.
25. MATLAB [Електронний ресурс]. – Режим доступу: <https://uk.wikipedia.org/wiki/MATLAB>.
26. R Core Team R. A Language and Environment for Statistical Computing; R Foundation for Statistical Computing: Vienna, Austria, 2020; Available online: <https://www.r-project.org/>.
27. Chang, W.; Cheng, J.; Allaire, JJ.; Xie, Y.; McPherson, J. Shiny: Web Application Framework for R; R package version 1.5.0; 2020; Available online: <https://CRAN.R-project.org/package=shiny>.
28. Percival, J., Tsutsumida, N. Geographically Weighted Partial Correlation for Spatial Analysis. *Gi\_forum* 2017, 1, 36–43, doi:10.1553/giscience2017\_01\_s36.
29. Lu, B.; Harris, P.; Charlton, M.; Brunson, C. The GWmodel R package: further topics for exploring spatial heterogeneity using geographically weighted models. *Geo-spatial Information Sci* 2014, 17, 85–101, doi:10.1080/10095020.2014.917453.
30. Gollini, I.; Lu, B.; Charlton, M.; Brunson, C.; Harris, P. GWmodel: An R Package for Exploring Spatial Heterogeneity Using Geographically Weighted Models. *Journal of Statistical Software* 2015, 63.

31. Rubin D. Bayesian inference for causal effects: The role of randomization / D. Rubin // *The annals of statistics*. - 1978. - Vol. 6.
32. Anselin L. Local indicators of spatial association / L. Anselin // *Geographical analysis*. – 1995. – Vol. 27, № 2. – P. 93-115.
33. Дубровина Н.А. Применение методов пространственной эконометрики в региональных исследованиях / Н.А. Дубровина // *Бізнес Інформ* – 2010. – №5(2). – С. 12-16.
34. Березька К.М., Березький О. М., Маслій В.В. Оцінка диспропорцій регіонального розподілу іноземних інвестицій в Україні. Актуальні проблеми економіки. 2013. № 12(150). С. 106 – 114.
35. Cliff A. *Spatial Autocorrelation* / A. Cliff, J. Ord . – London: Pion, 1973.
36. Fisher W. *Econometric estimation with spatial dependence* / W. Fisher // *Regional and Urban Econometrics*. – 1971. – №1. P. 19 – 40.
37. Kapoor M. *Panel data models with spatially correlated error components* / M. Kapoor, H. Kelejian, I. Prucha // *Journal of Econometrics*. – 1998. – №140. – P. 97 – 130.
38. Koczczevska K. *Ekonometria i statystyka przestrzenna* / K. Koczczevska. – Warszawa: Wydanie I, 2006.
39. Berezka K. The analysis of students admission results to the educational qualification level “bachelor’s degree” on the basis of methods of spatial econometrics / Kateryna Berezka, Oleh Berezsky // *Quantitative methods in accounting and finance*. – Ternopil, 2016. P. 36-43.
40. Березький О. Використання методів просторової економіки для аналізу результатів вступу абітурієнтів / Олег Березький, Катерина Березька, Василь Неміш // *Геометричне моделювання та інформаційні технології*. – 2016. – №1. – С. 5-9.
41. Fingleton B. Externalities, economic geography and spatial econometrics: conceptual and modelling developments / B. Fingleton // *International Regional Science Review*. – 2003. №26(2). – P. 197-207.

42. Moran P. The interpretation of statistical maps / P. Moran // *Journal of the Royal Statistical Society*. – 1948. – Series B, №10. – P. 243 – 251.
43. Anselin L. Spatial dependence in linear regression models with an introduction to spatial econometrics / Luc Anselin, Anil K Bera // *Statistics Textbooks and Monographs*. – 1998. – Vol. 155. – P. 237-290.
44. Anselin L. *Spatial econometrics: methods and models* / Luc Anselin. – Springer Science & Business Media, 2013. – 215 p.
45. Anselin L. Local geographic spillovers between university research and high technology innovations / Luc Anselin, Attila Varga, Zoltan Acs // *Journal of urban economics*. – 1997. – Vol.42. Issue 3. – P. 422-448.
46. Іващук О.Т. Економетричні методи та моделі: Навч. посібник. Тернопіль: ТАНГ “Економічна думка”, 2002. 348 с.
47. Гмурман В.Е. Теория вероятностей и математическая статистика. М.: Высшая школа, 1972. 368 с.
48. Casseti E. Generating Models by the Expansion Methods: Application to Geographic Research. *Geographical Analysis*. 1972. Vol. 4. P. 81–91.
49. Hall S.G., Petroulas P. Spatial Interdependencies of FDI Locations: A Lessening of the Tyranny of Distance? University of Leicester. Working Paper. 2008. № 08/28. 36 p.
50. Davidson R., MacKinnon J.G. *Estimation and inference in econometrics*. New York: Oxford University Press, 1993. 205 p.
51. Percival J., Tsutsumida N., Murakami D., Yoshida T., Nakaya T. gwpcorMapper: an interactive mapping tool for exploring geographically weighted correlation and partial correlation in high-dimensional geospatial datasets [Електронний ресурс]. – Режим доступу: <https://arxiv.org/abs/2101.03491>.