

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

Йож Артур Анатолійович

**Алгоритми оцінювання якості навчання студентів на
основі 3D даних / Algorithms for assessing the quality of
student learning based on 3D data**

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

Виконав студент групи Кім-21
А.А. Йож

Науковий керівник
д.т.н., професор, О. М. Березький

Кваліфікаційну роботу допущено
до захисту:

" ____ " _____ 20____ р.

Завідувач кафедри

_____ О.Л. Дубчак

АНОТАЦІЯ

Йож А.А. Алгоритми оцінювання якості навчання студентів на основі 3D даних. – Рукопис.

Кваліфікаційна робота на здобуття освітнього ступеня «магістр» за спеціальністю 123 «Комп'ютерна інженерія», освітньо-професійна програма. Західноукраїнський національний університет, Тернопіль, 2024.

Робота написана обсягом 91 сторінка і містить 12 ілюстрацій, 10 таблиць, 3 додатки та 43 джерела за переліком посилань. Метою роботи є розроблення алгоритмів аналізу панельних даних для оцінки впливу якості навчання фундаментальних дисциплін на якість навчання спеціальних дисциплін студентів.

Методи досліджень базуються на використанні алгоритмів аналізу панельних даних, методах об'єктно-орієнтованого програмування.

Результати дослідження: розроблено алгоритми оцінювання моделі фіксованих ефектів, з випадковими ефектами, динамічної моделі; програмну систему для аналізу панельних даних. Здійснено програмну реалізацію моделей оцінювання якості навчання.

Результати роботи можуть бути використані в багатьох галузях для аналізу панельних даних.

Орієнтовні напрямки розвитку досліджень: дослідження закономірностей впливу спеціальних дисциплін на якість кваліфікаційних робіт.

КЛЮЧОВІ СЛОВА: МОДЕЛЮВАННЯ, ЯКІСТЬ НАВЧАННЯ, АЛГОРИТМ, ПАНЕЛЬНІ ДАНІ, МОДЕЛЬ ФІКСОВАНИХ ЕФЕКТІВ, МОДЕЛЬ З ВИПАДКОВИМИ ЕФЕКТАМИ, ПРОГРАМНА СИСТЕМА.

ANNOTATION

Yozh A.A. Algorithms for assessing the quality of student learning based on 3D data. – Manuscript.

Qualification work for obtaining the educational degree «Master» in specialty 123 «Computer Engineering», educational and professional program. West Ukrainian National University, Ternopil, 2024.

The work is 91 pages long and contains 12 illustrations, 10 tables, 3 applications and 43 sources for references. The aim of the work is to develop panel data analysis algorithms to assess the impact of the quality of teaching fundamental disciplines on the quality of teaching students special disciplines.

Research methods are based on the use of panel data analysis algorithms, object-oriented programming methods.

Research results: algorithms for evaluating the fixed effects model, with random effects, dynamic model have been developed; a software system for analyzing panel data. Software implementation of models for assessing the quality of teaching has been carried out.

The results of the work can be used in many areas for the analysis of panel data.

Approximate areas of research development: study of the patterns of influence of special disciplines on the quality of qualification works.

KEY WORDS: MODELING, LEARNING QUALITY, ALGORITHM, PANEL DATA, FIXED EFFECTS MODEL, RANDOM EFFECTS MODEL, SOFTWARE SYSTEMS.

ЗМІСТ

Вступ.....	6
1 Аналіз якості навчання в університеті і методів та програмних засобів її оцінювання.....	9
1.1 Аналіз навчальних програм в Західноукраїнському національному університеті.....	9
1.2 Аналіз способів оцінювання якості навчання.....	14
1.3 Характеристика панельних даних та методи їх аналізу.....	16
1.4 Аналіз інструментальних засобів та програмного забезпечення для аналізу панельних даних	21
1.5 Висновки до розділу 1	28
2 Алгоритми побудови моделей на основі 3D даних	29
2.1 Алгоритм оцінювання моделі фіксованих ефектів.....	30
2.2 Алгоритм оцінювання моделі з випадковими ефектами	32
2.3 Алгоритм оцінювання динамічної моделі панельних даних.....	37
2.4 Алгоритм тесту Хаусмана	40
2.5 Алгоритм методу різниць у різницях.....	42
2.6 Висновки до розділу 2	44
3 Програмна реалізація алгоритмів оцінювання якості навчання студентів на основі 3D даних	45
3.1 Апаратне забезпечення для реалізації програмної системи	45
3.2 Програмна система для аналізу панельних даних.....	47
3.3 Підготовка та обробка даних для програмної системи.....	50
3.4 Програмна реалізація моделей оцінювання якості навчання в середовищі R.....	54

3.5 Висновки до розділу 3	63
Висновки	64
Список використаних джерел	65
Додаток А Зразки екзаменаційних відомостей	Помилка! Закладку не визначено.
Додаток Б Лістинг коду програми.....	Помилка! Закладку не визначено.
Додаток В Світлокопії виданих публікацій	Помилка! Закладку не визначено.

ВСТУП

Актуальність дослідження. Оцінка якості навчання є однією з ключових проблем сучасної освіти, оскільки вона впливає на розвиток суспільства, економіки та інновацій. Зростаючі вимоги до компетенцій випускників вимагають удосконалення освітніх процесів і забезпечення їх відповідності потребам ринку праці. Актуальність цього дослідження обумовлена кількома ключовими аспектами: підвищенням ефективності освітнього процесу, задоволенням потреб суспільства та економіки, забезпеченням прозорості та підзвітності, впровадженням міжнародних стандартів і підходів, адаптацією до нових викликів таких як дистанційне навчання, цифровізація освіти та пандемії і війна.

Мета і завдання дослідження. Метою кваліфікаційної роботи є дослідження алгоритмів панельних даних для оцінки впливу якості навчання фундаментальних дисциплін на якість навчання спеціальних дисциплін студентів.

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

- проаналізувати навчальні програми Західноукраїнського національного університету;
- провести аналіз способів оцінювання якості навчання;
- проаналізувати методи аналізу панельних даних;
- розробити алгоритм оцінювання моделі фіксованих ефектів;
- розробити алгоритм оцінювання моделі з випадковими ефектами;
- розробити алгоритм оцінювання динамічної моделі;
- розробити програмну систему для аналізу панельних даних;
- здійснити програмну реалізацію моделей оцінювання якості навчання;
- зробити тестування розробленого продукту.

Об'єкт дослідження – процес аналізу успішності студентів.

Предмет дослідження – алгоритми аналізу панельних даних для оцінювання якості навчання студентів.

Методи досліджень. У кваліфікаційній роботі використано алгоритми аналізу панельних даних, методи об'єктно-орієнтованого програмування.

Наукова новизна одержаних результатів полягає в розробці алгоритмів панельних даних для оцінки впливу якості навчання фундаментальних дисциплін на якість навчання спеціальних дисциплін.

Практичне значення одержаних результатів полягає в підготовці даних якості навчання фундаментальних і спеціальних дисциплін та їх моделюванні в програмному середовищі R.

Публікації результатів досліджень. За результатами досліджень опубліковані двоє тез доповідей Всеукраїнської науково-практичної конференції студентів, аспірантів, молодих вчених «Інтелектуальні комп'ютерні системи та мережі» (5 листопада 2024 р., м. Тернопіль, Західноукраїнський національний університет) [1, 2]:

1. Йож А. А., Рутецький Ю. О. Алгоритми та аналіз даних та для оцінювання якості навчання студентів. *Інтелектуальні комп'ютерні системи та мережі* : тези доп. Всеукр. наук.-практ. конф. студентів, аспірантів, молодих вчених (5 лист. 2024 р.). Тернопіль : ЗУНУ, 2024. С. 101-102.

2. Йож А. А., Рутецький Ю. О. Моделювання алгоритмів оцінювання якості навчання студентів. *Інтелектуальні комп'ютерні системи та мережі* : тези доп. Всеукр. наук.-практ. конф. студентів, аспірантів, молодих вчених (5 лист. 2024 р.). Тернопіль : ЗУНУ, 2024. С. 103-104.

Кваліфікаційна робота складається із трьох розділів, висновків, списку використаної літератури та додатків [3].

В першому розділі проведено аналіз навчальних програм Західноукраїнського національного університету, способів оцінювання якості навчання. Також здійснено характеристику панельних даних та методів їх аналізу, проаналізовані інструментальні засоби та програмне забезпечення.

В другому розділі розроблено алгоритми: оцінювання моделі фіксованих ефектів, моделі з випадковими ефектами, динамічної моделі, тесту Хаусмана, методу різниць у різницях.

В третьому розділі здійснено програмну реалізацію алгоритмів оцінювання якості навчання студентів на основі 3D даних, приведена програмна система для аналізу панельних даних, програмна реалізація моделей оцінювання якості навчання в середовищі R.

У додатку приведено зразки екзаменаційних відомостей, лістинг коду програми, світлокопії виданих публікацій.

1 АНАЛІЗ ЯКОСТІ НАВЧАННЯ В УНІВЕРСИТЕТІ І МЕТОДІВ ТА ПРОГРАМНИХ ЗАСОБІВ ЇЇ ОЦІНЮВАННЯ

1.1 Аналіз навчальних програм Західноукраїнського національного університету

Західноукраїнський національний університет є провідним навчальним закладом західного регіону України, який надає високоякісну освіту у сфері економіки, менеджменту, права, комп'ютерних та інформаційних технологій, міжнародних відносин, гуманітарних наук та інших важливих напрямків сучасної освіти [4]. Університет пропонує широкий спектр освітніх програм, орієнтованих на підготовку висококваліфікованих фахівців, які відповідають вимогам сучасного ринку праці.

Основними освітніми напрямами, які пропонує університет, є:

- економіка та менеджмент: програми, що готують студентів до роботи в сфері управління підприємствами, фінансовими організаціями, банками та іншими економічними структурами;
- право: підготовка висококваліфікованих юристів для роботи у державних установах, приватному секторі та міжнародних організаціях;
- міжнародні відносини: освітні програми, спрямовані на формування фахівців у сфері дипломатії, міжнародної політики та глобальних економічних процесів;
- комп'ютерні та інформаційні технології: університет також пропонує сучасні програми навчання в галузі ІТ, що включають інженерію програмного забезпечення, інформаційні системи та технології, кібербезпеку, комп'ютерні науки та інші напрями, пов'язані з цифровими технологіями.

Західноукраїнський національний університет (ЗУНУ) забезпечує своїх студентів усім необхідним для отримання якісної освіти, включаючи сучасні лабораторії, комп'ютерні класи та інноваційні навчальні програми.

Окремо розглянемо факультет комп'ютерних та інформаційних технологій [5]. Завдяки сучасній інфраструктурі, кваліфікованому викладацькому складу та тісним зв'язкам з промисловими партнерами, університет готує фахівців, здатних відповідати на виклики цифрової епохи.

Факультет комп'ютерних інформаційних технологій є одним із провідних факультетів університету, де студенти отримують глибокі знання та практичні навички в різних галузях інформаційних технологій. Факультет пропонує навчальні програми за наступними спеціальностями:

1. Інженерія програмного забезпечення

Навчання на цій спеціальності фокусується на розробці, тестуванні та підтримці програмних систем; на вивченні методологій розробки, управлінні проектами та забезпеченні якості програмного забезпечення.

2. Комп'ютерна інженерія

Ця спеціальність включає в собі поєднання елементів електроніки та комп'ютерних наук для розробки апаратного та програмного забезпечення, а також проектування комп'ютерних систем, мікропроцесорів та вбудованих систем.

3. Інформаційні системи та технології

Навчаючись на цій спеціальності студенти орієнтуються на проектуванні, впровадженні та управлінні інформаційними системами в організаціях, а також на аналізі бізнес-процесів та інтеграції IT-рішень для підвищення ефективності інформаційних систем.

4. Автоматизація та комп'ютерно-інтегровані технології

Закінчивши цю спеціальність, студент отримає знання з автоматизованих систем управління та комп'ютерної інтеграції виробничих процесів; розробки рішень для оптимізації та автоматизації промислових процесів.

5. Кібербезпека

Навчання на цій спеціальності надасть навичок захисту інформаційних систем від кіберзагроз та забезпечення безпеки даних; вивчення криптографії, мережевої безпеки та методів захисту інформації.

6. Комп'ютерні науки

Студенти засвоюють теоретичні основи інформатики, алгоритмів та структур даних; дослідження в галузі штучного інтелекту, машинного навчання та обробки великих даних.

7. Метрологія та інформаційно-вимірвальна техніка

В ЗУНУ з цієї спеціальності є освітньо-професійна програма «Технології інтернету речей». Студенти вивчають принципи вимірювань, метрології та застосування IoT-технологій у різних галузях промисловості.

8. Системний аналіз

Навчаючись на цій спеціальності студенти здійснюють аналіз складних систем та процесів з метою їх оптимізації; а також використовують математичні та комп'ютерні методи для моделювання та прийняття рішень.

Факультет комп'ютерних інформаційних технологій Західноукраїнського національного університету оснащений сучасними лабораторіями та комп'ютерними класами, що забезпечують високий рівень практичної підготовки студентів.

Наведемо опис основних лабораторій та навчальних приміщень факультету:

1. Лабораторія інженерії програмного забезпечення

Приміщення оснащене потужними серверами, комп'ютерами з високопродуктивними процесорами, програмними середовищами для розробки (IDE), системами контролю версій (Git), засобами для тестування та забезпечення якості програмного забезпечення.

Лабораторія створена для розробки, тестування та підтримки програмних проєктів. Студенти працюють над реальними завданнями, співпрацюють у командах, використовують сучасні методології розробки (Agile, Scrum).

2. Лабораторія комп'ютерної інженерії

Лабораторія оснащена обладнанням для проектування апаратних компонентів, мікроконтролерами, вбудованими системами, 3D-принтерами для створення прототипів.

Лабораторія створена для розробки та тестування апаратного забезпечення. Студенти працюють над проектами з розробки мікропроцесорів, вбудованих систем та інтеграції апаратного і програмного забезпечення.

3. Центр інформаційних систем та технологій

Оснащений центр серверними стійками, базами даних, хмарними платформами (AWS, Azure), інструментами для моделювання та управління інформаційними системами.

Центр функціонує для проектування та впровадження інформаційних систем для різних організацій, навчання управлінню базами даних, аналізу бізнес-процесів та інтеграції IT-рішень.

4. Лабораторія автоматизації та комп'ютерно-інтегрованих технологій

Ця лабораторія оснащена програмованими логічними контролерами (PLC), робототехнічними комплексами, системами SCADA, обладнанням для автоматизації виробничих процесів.

Лабораторія створена для розробки та тестування автоматизованих систем управління. Студенти створюють інтегровані рішення для оптимізації та автоматизації промислових процесів.

5. Центр кібербезпеки та мережевих технологій

Він оснащений засобами для тестування захищеності мереж, системами моніторингу та аналізу кіберзагроз, віртуальними лабораторіями для симуляції атак та захисту.

Центр функціонує для навчання методам захисту інформаційних систем від кіберзагроз, для практичних занять з криптографії, мережевої безпеки та реагування на інциденти.

6. Центр комп'ютерних наук

Центр оснащений високопродуктивними обчислювальними кластерами, серверами для обробки великих даних, інструментами для штучного інтелекту та машинного навчання.

В центрі проводяться дослідження в галузі штучного інтелекту, обробки великих даних та алгоритмів. Студенти виконують наукові проекти, аналізують дані та розробляють інноваційні IT-рішення.

7. Лабораторія метрології та інформаційно-виміральної техніки

Лабораторія оснащена пристроями для точних вимірювань, сенсорами, платформами для розробки Інтернету речей (IoT), системами збору та аналізу даних.

Студенти вивчають принципи метрології та застосування IoT-технологій у різних галузях. Студенти розробляють та тестують системи для автоматизованого збору та аналізу даних.

8. Системний аналітичний центр

Центр оснащений програмним забезпеченням для моделювання систем, інструментами для аналізу даних, платформами для підтримки прийняття рішень.

Центр функціонує для аналізу складних систем та процесів з метою їх оптимізації. Студенти використовують математичні та комп'ютерні методи для моделювання, прогнозування та прийняття рішень у різних сферах діяльності.

На факультеті є ряд комп'ютерних класів, які оснащені сучасними комп'ютерами з високою продуктивністю, інтерактивними дошками, програмним забезпеченням для навчання та досліджень, швидкісним інтернетом.

В комп'ютерних класах проводяться лекції, семінари та практичні заняття. Комп'ютерні класи забезпечують студентів доступом до необхідних ресурсів для навчання та самостійної роботи.

9. Лабораторія віртуальної та доповненої реальності

Лабораторія оснащена VR/AR гарнітурами, потужними комп'ютерами з графічними картами високого класу, спеціалізованим програмним забезпеченням для створення та тестування віртуальних середовищ.

В лабораторії розробляються та впроваджуються віртуальні та доповнені реальності у різні сфери, включаючи освіту, медицину, інженерію та розваги.

Студенти створюють інтерактивні проекти, проводять експерименти та дослідження.

Всі лабораторії та комп'ютерні класи є сучасними, мають інноваційне обладнання: Використання найсучасніших технологій та обладнання забезпечує студентам можливість працювати з актуальними інструментами, які використовуються в промисловості.

Лабораторії є практично орієнтовані. Лабораторії спрямовані на надання практичних навичок, що дозволяє студентам ефективно застосовувати теоретичні знання в реальних проектах.

Студенти мають доступ до великого обсягу ресурсів, включаючи програмне забезпечення, бази даних та обчислювальні потужності, що сприяє успішному навчальному процесу.

Лабораторії сприяють роботі над міждисциплінарними проектами, що дозволяє студентам розширювати свої знання та навички в різних напрямках ІТ.

1.2 Аналіз способів оцінювання якості навчання

Розглянемо яким способом можна оцінювати якість навчання. Оцінювання може проходити за різними підходами й інструментами, які зазвичай спрямовані на вимірювання та аналіз як освітніх результатів учнів/студентів, так і процесів навчання [6]. Основні методи оцінювання якості навчання включають такі компоненти:

1. Формативне оцінювання – це оцінювання під час навчання для навчання [7-9].

Оцінювання в процесі навчання: проводиться під час навчального процесу, щоб допомогти учням зрозуміти свій рівень, виявити сильні та слабкі сторони.

Зворотний зв'язок: викладач надає зворотний зв'язок після кожного завдання чи модуля, що допомагає студентам швидко коригувати свої підходи.

Тестування: короткі тести допомагають оцінити, наскільки студенти засвоїли матеріал, і визначити прогалини в знаннях.

2. Підсумкове оцінювання включає в себе [8]:

- екзамени та контрольні роботи: використовуються для підсумкового оцінювання знань за певний період навчання (семестр, рік);
- стандартизовані тести: тести, наприклад, як зовнішнє незалежне оцінювання (ЗНО) чи єдиний вступний іспит (ЄВІ), дозволяють порівнювати результати студентів та загальний рівень знань.

3. Оцінювання результатів навчання виділяє наступні [10]:

- когнітивні результати: оцінка рівня знань, умінь та навичок, які були здобуті в процесі навчання;
- соціальні та поведінкові результати: вивчення, як навчання вплинуло на соціальні навички, ставлення до навчання, здатність до самостійної роботи;
- досягнення цілей навчальної програми: перевірка, чи досягли студенти цілей, встановлених програмою, та наскільки ефективно були засвоєні основні компетентності.

4. Методологія 360-градусного оцінювання полягає в:

- анкетуванні учасників: анкети можуть бути розроблені для студентів, батьків, викладачів, які можуть висловити свою думку про процес і результати навчання;
- самооцінюванні студентів: студенти можуть оцінювати власний прогрес та ефективність навчання;
- оцінюванні викладачами та адміністрацією: викладачі оцінюють ефективність уроків, а адміністрація – результати, щоб визначити можливі покращення.

5. Аналітика та звітування включає в себе:

- статистичний аналіз: аналіз оцінок та показників успішності студентів за допомогою статистичних методів;
- відстеження прогресу: використання програм для відстеження прогресу кожного студента протягом навчального року;

- звіти для батьків та студентів: надання детальних звітів із результатами навчання, що показують, де студент досяг успіху, а де потрібно більше зусиль.

6. Оцінювання за компетентностями [11] можна поділити на:

- практичні завдання та проекти: оцінювання вміння застосовувати знання на практиці;
- інтерв'ю та презентації: оцінювання за допомогою інтерв'ю та публічних виступів, де студент може продемонструвати свої навички комунікації та критичного мислення.

Кожен з цих методів забезпечує комплексне уявлення про якість навчання та допомагає викладачам і адміністрації освітніх закладів робити обґрунтовані висновки про те, наскільки ефективно працює навчальний процес.

В підрозділі розглянуто Західноукраїнський національний університет, зосереджено увагу на факультет комп'ютерних інформаційних технологій, що надає всебічну підготовку фахівців, готових до роботи в сучасному динамічному середовищі ІТ. Також проведено аналіз способів оцінювання якості навчання.

1.3 Характеристика панельних даних та методи їх аналізу

Існує низка підходів для оцінювання, які базуються на використанні панельних даних (3D), просторових моделей і часових рядів, таких як ARIMA [12-17] та ін.

Панельні дані (також відомі як дані панелі або довгострокові дані) представляють собою комбінацію часових та міжоб'єктних даних [12, 18-23]. Іншими словами, панельні дані включають спостереження за кількома об'єктами (це можуть бути особи, підприємства, країни) протягом декількох періодів часу. Така структура дозволяє одночасно аналізувати як зміни з часом, так і відмінності між об'єктами.

В таблиці 1.1 приведені дані, у яких спостереження реєструються за кількома суб'єктами, в таблиці 1.2 – спостереження реєструються за часом, в таблиці 1.3 – приклад структури панельних даних, де спостереження реєструються за суб'єктами і за часом.

Таблиця 1.1 – Перехресний зріз (Cross-sectional data)

Індивідуальний ID	Показник 1	Показник 2
1	10	5
2	12	6
3	14	8

Таблиця 1.2 – Часовий ряд (Time series data)

Період	Показник 1	Показник 2
2021	10	5
2022	12	6
2023	14	8

Таблиця 1.3 – Панельні дані (Panel data)

Індивідуальний ID	Період	Показник 1	Показник 2
1	2021	10	5
1	2022	12	6
1	2023	14	8
2	2021	8	4
2	2022	9	5
2	2023	10	6

Панельні дані характеризуються двовимірністю. Найчастіше по осі X (вертикальній) представляють різні об'єкти дослідження (різні компанії, країни, особи), а по осі Y (горизонтальній) представляють часові періоди (наприклад, роки, квартали, місяці).

Таким чином, для кожного об'єкта доступні дані за кілька часових періодів, що дозволяє відстежувати зміни в характеристиках об'єкта з часом.

Також існує можливість контролю за незмінними характеристиками, а саме, панельні дані дозволяють контролювати ефекти, які не змінюються з часом, наприклад, культурні особливості, географічне розташування тощо.

В таблиці 1.4 приведено порівняння з іншими типами даних: крос-секційними та часовими рядами. Як видно, панельні дані є поєднанням цих даних.

Таблиця 1.4 – Порівняння різних типів даних

Порівняльні характеристики	Тип даних		
	Крос-секційні дані	Часові ряди	Панельні дані
Вид спостережень	Містять спостереження за багатьма об'єктами, але лише за один часовий період	Містять спостереження за одним об'єктом протягом багатьох часових періодів	Містять спостереження за багатьма об'єктами протягом багатьох часових періодів
Вид аналізу	Не дозволяють аналізувати динаміку змін з часом	Не дозволяють порівнювати між різними об'єктами	Одночасно аналізують різницю між об'єктами та зміни з часом

Використання панельних даних має свої переваги і недоліки.

До переваг можна віднести:

- контроль за незмінними характеристиками (можливість враховувати індивідуальні ефекти, які не змінюються з часом, що знижує спотворення результатів);
- більша ефективність оцінок. Завдяки збільшенню обсягу даних панельні дані можуть забезпечити більш точні та надійні оцінки параметрів моделей;
- аналіз динаміки. Дозволяють вивчати динамічні процеси, такі як вплив попередніх значень змінних на поточні результати;
- виявлення причинно-наслідкових відносин допомагає краще ідентифікувати причинно-наслідкові зв'язки завдяки можливості контролю за незмінними факторами.

До недоліків можна віднести:

- складність збору даних. Збір панельних даних може бути витратним і часовим процесом, особливо при великій кількості об'єктів і періодів;
- проблеми з відсутніми даними. Існують ситуації, коли, за певних причин, деякі спостереження відсутні, що може ускладнити аналіз;
- складність моделювання. Панельні дані вимагають використання складніших статистичних методів та моделей для коректного аналізу.

Таблиця 1.5 приводить методи та моделі аналізу панельних даних.

Таблиця 1.5 – Методи та моделі аналізу панельних даних

№п/п	Метод	Опис	Формула
1	2	3	4
1	Модель фіксованих ефектів (Fixed effects model)	– використовується для контролю за незмінними характеристиками об'єктів – підходить, коли індивідуальні ефекти корельовані з незалежними змінними	$Y_{it} = \alpha_i + \beta X_{it} + \varepsilon_{it},$ де Y_{it} – залежна змінна для об'єкта i в час t , α_i – фіксований ефект для об'єкта i , X_{it} – незалежні змінні, ε_{it} – похибка
2	Модель випадкових ефектів (Random effects model)	– припускає, що індивідуальні ефекти не корельовані з незалежними змінними; – дозволяє використовувати як міжчасові, так і міжоб'єктні варіації	$Y_{it} = \alpha + \beta X_{it} + u_i + \varepsilon_{it},$ де u_i – випадковий ефект для об'єкта i , який не корельований з X_{it}
3	Тест Хаусмана (Hausman Test)	– використовується для визначення, яка модель (фіксованих чи випадкових ефектів) є більш відповідною для даних	
4	Моделі з динамічними ефектами (Dynamic panel models)	– ці моделі враховують залежність поточного значення залежної змінної від її попередніх значень.	$Y_{it} = \gamma Y_{it-1} + \beta X_{it} + \alpha_i + \varepsilon_{it}$

Продовження таблиці 1.5

1	2	3	4
		Часто використовуються методи, такі як система GMM (Generalized method of moments)	
5	Метод різниць у різницях (Difference-in-Differences, DiD)	– метод DiD використовується для оцінки причинно-наслідкових ефектів шляхом порівняння змін у результатах між об'єктами, які зазнали втручання, і тими, що його не зазнали	$Y_{it} = \alpha + \beta \text{Post}_t + \gamma \text{Treatment}_i + \delta (\text{Post}_t \times \text{Treatment}_i) + \varepsilon_{it}$, де δ – оцінка ефекту втручання
6	Машинне навчання для панельних даних	– застосування методів машинного навчання до панельних даних стає все більш популярним. Це включає використання алгоритмів, таких як випадкові ліси (Random Forests), градієнтний бустінг (Gradient Boosting), нейронні мережі та інші для передбачення та класифікації	Random Forests, Gradient Boosting

Панельні дані застосовують в економіці (наприклад, при аналізі впливу макроекономічних політик на економічне зростання різних країн протягом декількох років); соціології (наприклад, при вивченні змін у соціальних поведінках окремих груп населення з часом); фінансах (наприклад, при оцінці ризиків та доходності акцій різних компаній протягом кількох років); медицині (наприклад, при дослідженні ефективності медичних втручань у різних регіонах протягом часу); маркетингу (наприклад, при аналізі впливу рекламних кампаній

на продажі різних продуктів у різних регіонах протягом декількох періодів) тощо.

В підрозділі описано характеристику панельних даних та методи аналізу панельних даних.

1.4 Аналіз інструментальних засобів та програмного забезпечення для аналізу панельних даних

Аналіз панельних даних є важливим інструментом у сфері економіки, соціальних наук, фінансів та інших галузей. Для виконання такого аналізу доступні різні програмні засоби та інструменти, які варіюються за функціональністю, гнучкістю та спеціалізацією [24-26]. Наведемо розширений огляд основних інструментів і програмного забезпечення, які використовуються для аналізу панельних даних. Огляд проведемо врахувавши особливості, переваги та недоліки програмних засобів.

1. Stata є одним з найпопулярніших програмних продуктів для економетричного аналізу панельних даних [27]. Він пропонує широкий набір команд і функцій для оцінки моделей фіксованих і випадкових ефектів, динамічних моделей, а також для проведення тестів на гетероскедастичність і автокореляцію.

Особливостями цього програмного продукту є:

- вбудовані команди для панельних даних, такі як ``xtreg'` (регресія для панельних даних), ``xtabond'` (GMM моделі для динамічних панельних даних), ``hausman'` (тест Хауса-Мангела);
- підтримка різних економетричних тестів для вибору між моделями фіксованих і випадкових ефектів;
- зручний графічний інтерфейс та можливість автоматизації аналізу за допомогою скриптів.

Перевагами цього програмного продукту є:

- широкий вибір методів для економетричного аналізу;
- висока продуктивність при роботі з великими наборами даних;
- інтуїтивно зрозумілий інтерфейс для початківців і потужні можливості для професіоналів.

Недоліками цього програмного продукту є:

- платне програмне забезпечення з ліцензійною вартістю, що може бути високою для окремих користувачів або малих організацій.

2. R – це потужне програмне середовище з відкритим вихідним кодом, яке пропонує широкий набір пакетів для аналізу панельних даних [28]. Пакети ``plm`` (Panel Linear Models) та ``lme4`` забезпечують широкий функціонал для оцінки моделей фіксованих і випадкових ефектів.

Особливостями цього програмного середовища є можливості вбудованих пакетів:

- пакет ``plm`` містить функції для оцінки моделей фіксованих і випадкових ефектів, тестів на гетероскедастичність, автокореляцію та інші важливі економетричні інструменти;
- пакет ``lme4`` забезпечує можливості для моделювання змішаних ефектів (random effects);
- додаткові пакети для динамічних моделей панельних даних (``panelAR``, ``pht``), візуалізації та підготовки звітів.

Перевагами цього програмного середовища є:

- безкоштовне програмне забезпечення з відкритим кодом;
- гнучкість та можливість адаптації під специфічні завдання за допомогою написання власних функцій;
- велика спільнота користувачів, що постійно оновлює та підтримує пакети.

Недоліками цього програмного середовища є:

- потребує певних навичок програмування для використання;

- можливо, менше готових інструментів у порівнянні з комерційними рішеннями, що вимагає додаткових налаштувань.

3. EViews – це програмне забезпечення, орієнтоване на економетричний та статистичний аналіз, зокрема на аналіз часових рядів і панельних даних [29-30]. Його інтуїтивно зрозумілий інтерфейс робить його популярним серед економістів і фінансистів.

Особливостями цього програмного забезпечення є:

- підтримка панельних регресій, динамічних моделей та тестів для панельних даних;
- інструменти для виконання GMM (Generalized Method of Moments) оцінок та інших спеціалізованих економетричних моделей;
- можливість легкої інтеграції з Excel та іншими форматами даних.

Перевагами цього програмного забезпечення є:

- інтуїтивно зрозумілий графічний інтерфейс для тих, хто не володіє програмуванням;
- потужні інструменти для візуалізації та побудови графіків;
- підтримка широкого спектру економетричних методів.

Недоліками цього програмного забезпечення є:

- платне програмне забезпечення з обмеженими можливостями в безкоштовних версіях;
- менш гнучке для користувацьких налаштувань у порівнянні з R або Python.

4. Python, як універсальна мова програмування, також пропонує багато інструментів для аналізу панельних даних через бібліотеки `'pandas'`, `'statsmodels'`, `'linearmodels'` та інші [31]. Він використовується для побудови складних моделей, в тому числі динамічних панельних моделей та моделей машинного навчання.

Особливостями цієї мови програмування полягає в її інструментах:

- бібліотека `'statsmodels'` дозволяє проводити регресійний аналіз панельних даних та виконувати тести для моделей фіксованих і випадкових ефектів;
- бібліотека `'linearmodels'` містить широкий набір функцій для аналізу панельних даних, включаючи GMM;
- інструменти для машинного навчання, доступні через `'scikit-learn'`, дозволяють комбінувати економетричні моделі з новітніми технологіями аналізу даних.

Перевагами цієї мови програмування є:

- безкоштовне і відкрите програмне забезпечення;
- висока гнучкість та потужні інструменти для обробки та візуалізації даних;
- легкість інтеграції з іншими системами та бібліотеками для виконання складних аналізів.

Недоліками цієї мови програмування є:

- необхідність програмування та знання мов програмування для роботи;
- більш стрімка крива навчання для новачків.

5. SAS є одним з найстаріших і найвідоміших програмних продуктів для аналізу даних, включаючи панельні дані. Він широко використовується у великих корпораціях для статистичного аналізу та обробки великих наборів даних.

Особливостями цього програмного продукту є:

- пропонує розширені інструменти для аналізу панельних даних, у тому числі моделі фіксованих і випадкових ефектів;
- підтримка великих обсягів даних та розширена аналітика для корпоративного рівня;
- вбудовані функції для динамічного моделювання, прогнозування та машинного навчання.

Перевагами цього програмного продукту є:

- висока продуктивність та потужність для роботи з великими обсягами даних;
- підтримка корпоративних систем та інтеграція з іншими програмними продуктами для бізнес-аналітики;
- широка підтримка користувачів і наявність навчальних матеріалів.

Недоліками цього програмного продукту є:

- дуже висока вартість, що робить його недоступним для малих організацій та індивідуальних дослідників;
- потребує спеціалізованих знань для повного використання можливостей.

6. LIMDEP/NLOGIT – це спеціалізоване програмне забезпечення для економетричного аналізу панельних даних, особливо орієнтоване на логістичні моделі та моделі вибору [32]. Це програмне забезпечення часто використовується в транспортній економіці, соціології та інших галузях.

Особливостями цього програмного забезпечення є:

- підтримка моделей фіксованих і випадкових ефектів, логістичних моделей і динамічних моделей;
- вбудовані функції для моделювання рішень та вибору, а також панельних даних;
- моделювання складних систем вибору та аналізу дискретних даних.

Перевагами цього програмного забезпечення є:

- спеціалізоване для аналізу вибору та логістичних моделей;
- підтримка широкого спектра економетричних методів та моделей панельних даних.

Недоліками цього програмного забезпечення є:

- вартість. Це комерційне програмне забезпечення, тому його використання вимагає ліцензії, яка може бути дорогою для малих організацій або індивідуальних користувачів;
- складність моделювання. Хоча програмне забезпечення є потужним, його можливості можуть вимагати значного рівня знань у користувача для

повного використання, особливо при роботі з динамічними та структурними моделями.

Окрім згаданих вами інструментів (Stata, R, EViews, Python, SAS, LIMDEP/NLOGIT, STATISTICA), є ще кілька пакетів, які також мають можливості для оцінювання панельних даних:

7. Gretl вільно поширюваний пакет з відкритим кодом, який підтримує аналіз панельних даних [33].

Він містить інструменти для побудови моделей з фіксованими та випадковими ефектами, а також інші функції для роботи з панельними даними.

Є простим у використанні, підходить для навчання та базового аналізу.

8. MATLAB має пакет Econometrics Toolbox, який включає функції для аналізу панельних даних, зокрема моделі з фіксованими та випадковими ефектами, а також динамічні панельні моделі [34].

Він підтримує налаштування ГММ-оцінювань, корисних для складних панельних даних.

MATLAB є дуже гнучким, що дозволяє створювати власні функції для економетричного аналізу.

9. Пакет SPSS має обмежені можливості для аналізу панельних даних, але за допомогою надбудови SPSS Advanced Statistics Module можна виконувати деякі типи панельного аналізу [35].

Він є не настільки розширений для панельних даних, як інші програми, але підходить для простих моделей.

10. MPlus підтримує аналіз панельних даних, особливо в контексті структурного моделювання (SEM).

Він може бути корисний для роботи з панельними даними, які включають латентні змінні або складні ієрархічні структури.

А також добре підходить для психометричного аналізу, а також для досліджень в галузях соціальних наук.

11. Мова Julia також набирає популярності, і є кілька пакетів, що підтримують панельний аналіз, зокрема FixedEffects.jl і PanelData.jl.

Julia відзначається високою продуктивністю для великих обсягів даних і дозволяє налаштовувати аналіз.

12. OхMetrics – програмне забезпечення для економетричного аналізу, яке включає модулі для роботи з панельними даними, зокрема PcGive і G@RCH.

OхMetrics підходить для оцінювання моделей з фіксованими та випадковими ефектами, а також для більш просунутих економетричних моделей.

13. Gauss являє собою потужне програмне середовище для математичного та статистичного аналізу, яке має можливості для панельних даних [36].

Хоча Gauss не настільки популярний для економетричного аналізу, як Stata чи EViews, він використовується в наукових колах для специфічних задач.

14. STATISTICA теж має можливості роботи з панельними даними. Вона дозволяє виконувати регресійний аналіз, оцінку моделей з фіксованими та випадковими ефектами, а також використовувати інші методи панельного аналізу [37-38].

Моделі з фіксованими та випадковими ефектами. В STATISTICA можна оцінювати моделі з фіксованими і випадковими ефектами, що є важливим для панельного аналізу, особливо коли потрібно враховувати індивідуальні відмінності між суб'єктами (наприклад, регіонами або підприємствами) або часову залежність.

Можливість налаштування ГММ-оцінювань. STATISTICA дозволяє використовувати метод моментів (ГММ), який особливо корисний для динамічних панельних моделей.

Візуалізація панельних даних. STATISTICA має інструменти для побудови графіків, діаграм і статистичних звітів, що допомагає аналізувати структуру панельних даних і зручніше інтерпретувати результати.

Ці програми розширюють можливості для аналізу панельних даних, дозволяючи дослідникам використовувати різноманітні підходи та налаштування для конкретних типів даних і моделей.

Також Excel можна використовувати для роботи з панельними даними, хоча можливості цих програм для такого аналізу відрізняються.

15. Excel має обмежені можливості для панельних даних: Excel не спеціалізується на панельному аналізі, і тому його можливості в цьому контексті обмежені [39-43]. Проте для базового аналізу та обробки даних Excel може бути використаний, особливо якщо потрібно лише підготувати дані перед перенесенням їх до іншого програмного забезпечення.

Excel дозволяє виконувати регресійний аналіз, але не має вбудованих інструментів для моделей з фіксованими або випадковими ефектами. Проте можна створити деякі форми панельного аналізу вручну за допомогою макросів або надбудов, але це потребує значних зусиль.

В Excel можна використати надбудови. Існують надбудови для Excel (наприклад, XLSTAT або Real Statistics), які можуть розширити його функціонал для економетричного аналізу, зокрема панельних даних.

Обробка та візуалізація даних: Excel підходить для початкової обробки панельних даних та візуалізації базових трендів, але для детального панельного аналізу краще підходять спеціалізовані інструменти приведені вище.

Отже, здійснено огляд інструментів та програмного забезпечення для аналізу панельних даних.

1.5 Висновки до розділу 1

Результатами даного розділу є:

- аналіз навчальних програм Західноукраїнського національного університету, зокрема факультету комп'ютерних інформаційних технологій, який готує фахівців ІТ;
- аналіз способів оцінювання якості навчання;
- характеристика панельних даних та методи аналізу панельних даних;
- аналіз інструментальних засобів та програмних продуктів, що використовуються для аналізу панельних даних.

2 АЛГОРИТМИ ПОБУДОВИ МОДЕЛЕЙ НА ОСНОВІ 3D ДАНИХ

Для оцінювання якості навчання використання панельних (3D) даних має ряд переваг і є обґрунтованим підходом. Ось кілька причин, чому панельні дані є корисними для такого аналізу:

1. Врахування індивідуальної неоднорідності: Панельні дані дозволяють досліджувати одну й ту ж вибірку студентів, вузів чи викладачів протягом часу. Це дозволяє враховувати індивідуальні особливості кожного студента чи навчального закладу, що можуть впливати на якість навчання. Наприклад, здібності учнів або рівень підготовки викладачів можуть значно впливати на результати навчання, і використання панельних даних допомагає врахувати ці індивідуальні особливості.

2. Аналіз динаміки змін: Панельні дані дозволяють оцінити, як змінюються показники якості навчання в часі. Це дозволяє виявляти тренди та динаміку змін, такі як покращення чи погіршення результатів навчання впродовж кількох років. Наприклад, можна побачити, як впровадження нових навчальних методик впливає на результати студентів у довгостроковій перспективі.

3. Контроль за зовнішніми факторами: Панельні дані дозволяють краще контролювати зовнішні фактори, які впливають на якість навчання. Наприклад, економічні умови, зміни в політиці освіти, або соціальні чинники можуть впливати на рівень освіти. Панельні дані допомагають виявити, які з цих змін є дійсно значущими, оскільки вони дозволяють розділити часові ефекти та індивідуальні ефекти.

4. Більша статистична потужність: Оскільки панельні дані містять більше спостережень (оскільки дані збираються по кількох точках у часі для одних і тих же індивідів), це підвищує статистичну потужність моделі. Це дозволяє отримувати більш точні результати та робити значущі висновки щодо впливу різних факторів на якість навчання.

5. Можливість використання сучасних економетричних методів: Панельні дані дозволяють застосовувати більш складні економетричні моделі, такі як моделі фіксованих та випадкових ефектів, динамічні панельні моделі та моделі з відкладеним впливом. Це дає можливість глибше аналізувати ефекти різних факторів на якість навчання.

6. Врахування часової кореляції: Панельні дані дозволяють враховувати часову кореляцію, тобто той факт, що результати навчання одного року можуть бути пов'язані з результатами попереднього року. Наприклад, успішність студента в поточному році може значною мірою залежати від його успішності в попередньому році.

2.1 Алгоритм оцінювання моделі фіксованих ефектів

Перейдемо до алгоритму оцінювання моделі фіксованих ефектів (Fixed Effects Model) [12, 18-20]. Основні етапи цього процесу подамо за кроками:

1. Збір даних.

Цей крок полягає в отриманні панельних даних, що містять спостереження за кількома об'єктами протягом певного періоду часу.

2. Попередня обробка даних.

Цей крок можна деталізувати як:

- виявлення пропущених значень і аномалій;
- перетворення змінних у необхідний формат;
- перевірка наявності часових і міжоб'єктних варіацій.

3. Визначення незалежних і залежних змінних.

На даному етапі вибираються змінні для моделювання. Залежна змінна Y_{it} і незалежні змінні X_{it} .

4. Формулювання моделі фіксованих ефектів.

Модель фіксованих ефектів має вигляд:

$$Y_{it} = \alpha_i + \beta X_{it} + \varepsilon_{it},$$

де α_i – індивідуальні ефекти для кожного i -го об'єкта, β – коефіцієнти регресії, а ε_{it} – похибка.

5. Центрування змінних (within transformation).

Цей крок полягає у відніманні середніх значень для кожного i -го об'єкта, що дозволить усунення індивідуальних ефектів:

$$Y_{it} - \bar{Y}_i = \beta (X_{it} - \bar{X}_i) + (\varepsilon_{it} - \bar{\varepsilon}_i)$$

і також надасть можливість виключити індивідуальні незмінні характеристики.

6. Оцінка параметрів моделі.

Використання методу найменших квадратів (МНК) для оцінки коефіцієнтів β після трансформації.

7. Тестування моделі.

Проводиться ряд тестувань, серед яких:

- проведення тестів на значущість коефіцієнтів β ;
- перевірка моделі на гетероскедастичність, автокореляцію та інші проблеми.

8. Інтерпретація результатів.

На цьому кроці необхідно інтерпретувати результати через:

- аналіз отриманих оцінок для залежних та незалежних змінних;
- оцінку впливу кожної незалежної змінної на залежну.

9. Вирішальна перевірка моделі.

Цей етап є остаточним. Він включає в себе:

- перевірку на відповідність даних та теоретичної моделі;
- тестування гіпотез про можливі взаємозв'язки між змінними.

10. Висновки.

Етап висновку полягає в оцінці важливих змінних, прийнятті або відхиленні гіпотез, що впливають з аналізу.

Рисунок 2.1 демонструє алгоритм оцінювання моделі фіксованих ефектів.

2.2 Алгоритм оцінювання моделі з випадковими ефектами

Приведемо для порівняння алгоритм оцінювання моделі з випадковими ефектами (Random Effects Model, REM) [12, 18-20]. Він може бути розбитий на кілька етапів. Це покрокова процедура оцінювання, яка враховує як індивідуальні, так і часові варіації в панельних даних. Основні кроки:

1. Підготовка даних

Нехай маємо панельні дані, де i – індекс об'єкта (індивідуума, фірми тощо), а t – індекс часу. Формат даних:

$$Y_{it} = \alpha + \beta X_{it} + u_i + \varepsilon_{it},$$

де Y_{it} – залежна змінна для i -го об'єкта у часі t ,

X_{it} – незалежні змінні (предиктори),

u_i – індивідуальні випадкові ефекти,

ε_{it} – залишки (помилки моделі).

2. Оцінка загальної регресії МНК

Спочатку оцінюємо модель звичайними методами найменших квадратів (МНК), ігноруючи панельну структуру:

$$Y_{it} = \alpha + \beta X_{it} + \varepsilon_{it}.$$

Це необхідно для отримання початкових оцінок параметрів, які будуть використовуватися для подальшого оцінювання випадкових ефектів.

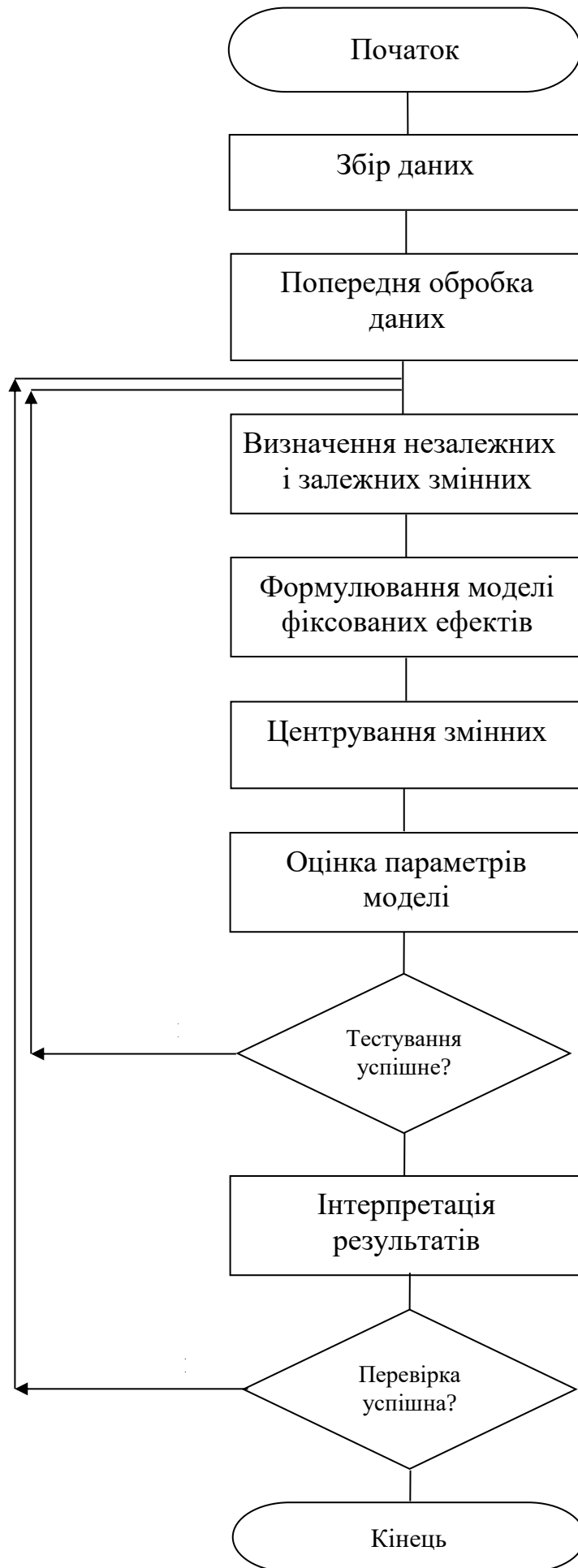


Рисунок 2.1 – Алгоритм оцінювання моделі фіксованих ефектів

3. Оцінка дисперсії компонентів

Для побудови моделі зі випадковими ефектами потрібно оцінити дисперсії випадкових складових. Модель припускає, що загальна помилка складається з двох частин:

u_i – індивідуальні ефекти з дисперсією σ_u^2 ,

ε_{it} – залишки моделі з дисперсією σ_ε^2 .

Для оцінювання дисперсій σ_u^2 і σ_ε^2 використовується метод дисперсійної компоненти (Variance Components Method).

4. Розрахунок параметра Фітса (θ)

Для перетворення моделі в змішану форму використовується параметр θ , який визначається так:

$$\theta = 1 - \sqrt{\frac{\sigma_\varepsilon^2}{\sigma_\varepsilon^2 + T \cdot \sigma_u^2}},$$

де T – кількість спостережень у часі для кожного об'єкта.

5. Трансформація змінних

Виконується трансформація залежної та незалежних змінних за допомогою параметра θ :

$$Y_{it}^* = Y_{it} - \theta \cdot \bar{Y}_i,$$

$$X_{it}^* = X_{it} - \theta \cdot \bar{X}_i,$$

де $\bar{Y}_i$ – середнє значення залежної змінної для кожного об'єкта,

$\bar{X}_i$ – середнє значення незалежних змінних для кожного об'єкта.

Ці трансформації дозволяють усунути кореляцію між індивідуальними ефектами u_i та предикторами X_{it} .

6. Оцінка перетвореної моделі

Далі використовуємо трансформовані змінні Y_{it}^* і X_{it}^* для оцінювання моделі за допомогою звичайного методу найменших квадратів (МНК):

$$Y_{it}^* = \alpha + \beta X_{it}^* + \varepsilon_{it}^*.$$

Це дає оцінки параметрів β , які є незміщеними і ефективними для моделі зі випадковими ефектами.

7. Перевірка гіпотез та значущості моделі

Після оцінювання моделі потрібно перевірити значущість отриманих коефіцієнтів β та інших параметрів:

Тест Бройша-Пагана (Breusch-Pagan Test) для перевірки того, чи є випадкові ефекти значущими. Якщо тест показує, що індивідуальні ефекти мають значення, це підтверджує використання REM замість звичайного МНК.

8. Інтерпретація результатів

Отримані оцінки коефіцієнтів β інтерпретуються як вплив змінних X_{it} на залежну змінну Y_{it} , з урахуванням варіацій між об'єктами та у часі.

Додаткові кроки:

- коригування стандартних помилок: Для панельних даних часто потрібне коригування стандартних помилок на гетероскедастичність та автокореляцію;
- перевірка на ендогенність: Якщо є підозра на кореляцію між незалежними змінними і випадковими ефектами, можуть використовуватись інструментальні змінні або методи, такі як Generalized Method of Moments (GMM).

Таким чином, оцінювання моделі зі випадковими ефектами полягає в урахуванні як внутрішньої варіації в об'єктах, так і загальної структури даних через врахування дисперсій компонентів і трансформацію змінних.

Рисунок 2.2 позначає алгоритм оцінювання моделі з випадковими ефектами.

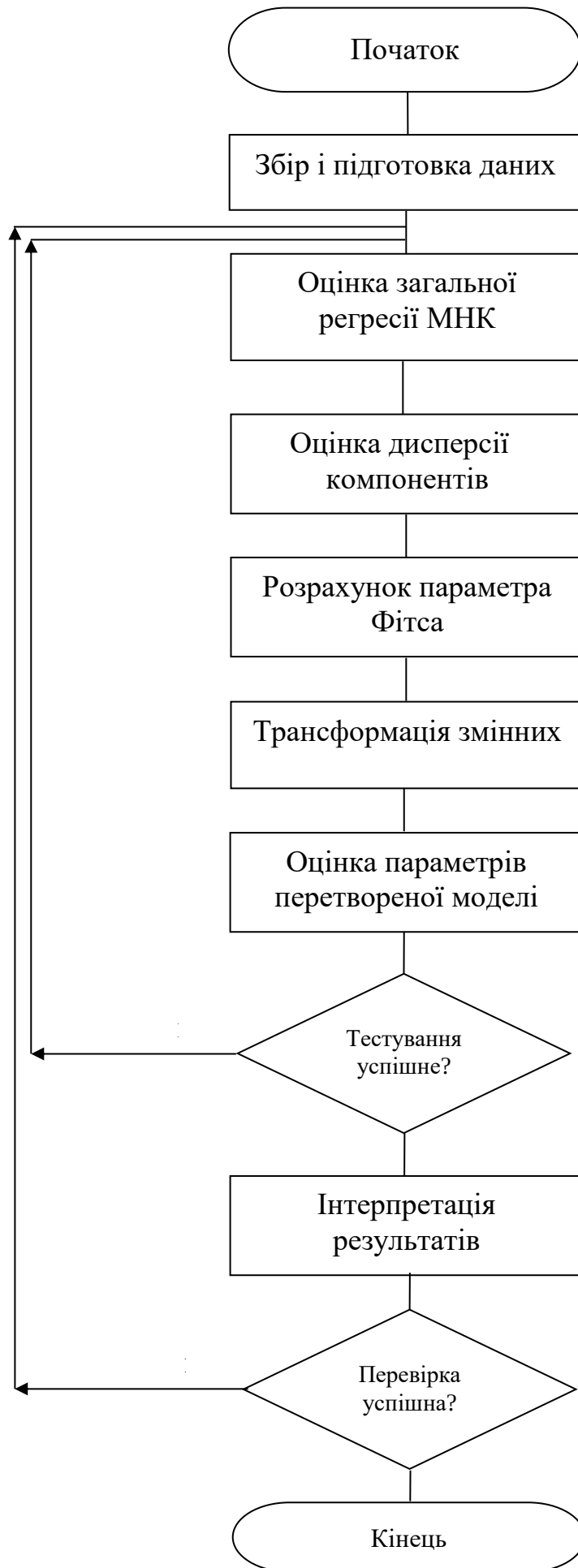


Рисунок 2.2 – Алгоритм оцінювання моделі з випадковими ефектами

2.3 Алгоритм оцінювання динамічної моделі панельних даних

Динамічна модель панельних даних використовується для аналізу панельних даних, коли залежна змінна в попередніх періодах впливає на її значення в поточному періоді. Цей підхід допомагає врахувати динамічні залежності, наприклад, вплив попередніх доходів на поточні або вплив минулих інвестицій на майбутні результати. Для оцінки таких моделей часто використовується генералізований метод моментів (ГММ), який може вирішувати проблему ендогенності та інші виклики, що виникають при роботі з динамічними панельними даними.

Алгоритм оцінювання динамічної моделі панельних даних за допомогою ГММ представимо за кроками [12, 18-20]:

1. Побудова моделі. Динамічна модель панельних даних включає лагове значення залежної змінної як незалежну змінну. Загальна форма динамічної моделі виглядає так:

$$y_{it} = \rho y_{i,t-1} + \beta x_{it} + u_i + \varepsilon_{it},$$

де y_{it} – залежна змінна для i -го суб'єкта в період часу t ,

$y_{i,t-1}$ – лагове значення залежної змінної (значення в попередній момент часу),

x_{it} – набір незалежних змінних для i -го суб'єкта період t ,

u_i – індивідуальний ефект (який може бути випадковим або фіксованим),

ε_{it} – випадкова похибка,

ρ – коефіцієнт лагу залежної змінної.

2. Проблема ендогенності

Лагове значення залежної змінної $y_{i,t-1}$ корелює з індивідуальним ефектом u_i , що викликає ендогенність. Це порушує припущення класичної регресії про некорельованість пояснювальних змінних з похибкою, і тому традиційні методи, як оцінка за допомогою МНК, не можуть бути застосовані без зміщення.

3. Генералізований метод моментів (ГММ)

Метод ГММ допомагає вирішити цю проблему шляхом використання інструментів – змінних, які корелюють з пояснювальними змінними, але не корелюють з похибкою. ГММ дозволяє оцінювати модель без зміщення навіть при наявності ендегенних змінних.

4. Двокрокова оцінка моделі за допомогою ГММ

4.1. Первинна трансформація моделі

Для усунення індивідуальних ефектів u_i , модель трансформують за допомогою першої різниці (differencing). Це дозволяє виключити незмінні індивідуальні ефекти з рівняння:

$$y_{it} - y_{i,t-1} = \rho (y_{i,t-1} - y_{i,t-2}) + \beta (x_{it} - x_{i,t-1}) + (\varepsilon_{it} - \varepsilon_{i,t-1})$$

Така трансформація видаляє індивідуальний ефект u_i , але породжує кореляцію між $y_{i,t-1} - y_{i,t-2}$ і похибкою, що вимагає використання інструментальних змінних.

4.2. Вибір інструментальних змінних

Важливим кроком у ГММ є правильний вибір інструментальних змінних. Інструментами можуть слугувати лаги змінних, які не корелюють з похибкою.

Наприклад, для оцінки моделі з першими різницями використовується другий і старші лаги залежної змінної як інструменти:

$$y_{i,t-2}, y_{i,t-3}, \dots$$

Окрім лагів залежної змінної, можна використовувати лаги незалежних змінних x_{it} .

4.3. Оцінка моделі ГММ

Застосовуються дві версії ГММ:

однокроковий ГММ (One-Step GMM): на першому кроці обчислюється матриця інструментів і оцінюються коефіцієнти.

двокроковий ГММ (Two-Step GMM): після первинної оцінки моделі обчислюється вагова матриця для покращення ефективності оцінки коефіцієнтів.

5. Оцінка похибок

ГММ дозволяє оцінювати не тільки коефіцієнти, але й дисперсію похибок. Це важливо для перевірки статистичної значущості оцінених коефіцієнтів.

6. Тести для перевірки правильності моделі

Після оцінки моделі ГММ необхідно провести кілька тестів для перевірки її коректності та відповідності припущенням.

Тест Хансена (Hansen J-Test). Використовується для перевірки правильності інструментальних змінних. Якщо інструменти коректні, похибки повинні бути некорельованими з інструментами. Нульова гіпотеза цього тесту полягає в тому, що інструменти є правильними, тобто їх коректність не відкидається.

Тест на серійну автокореляцію. Оскільки ГММ працює з панельними даними, важливо перевірити наявність автокореляції в похибках. Перший порядок автокореляції допустимий, але другий порядок свідчить про неправильність моделі.

7. Інтерпретація результатів

Після оцінки моделі та проведення тестів важливо інтерпретувати результати, враховуючи наступне:

- коефіцієнт ρ показує, наскільки сильно минулі значення залежної змінної впливають на поточні;
- значення коефіцієнтів для незалежних змінних β показують вплив кожного фактора x_{it} на залежну змінну;
- важливо перевірити статистичну значущість оцінених коефіцієнтів і правильність використаних інструментів.

Оцінювання динамічних панельних моделей за допомогою ГММ є потужним методом для аналізу динамічних процесів у панельних даних, проте вимагає уважного вибору інструментів і перевірки коректності моделі.

На рисунку 2.3 представлено алгоритм оцінювання динамічних панельних моделей за допомогою ГММ.

2.4 Алгоритм тесту Хаусмана

Як було сказано в підрозділі 1.2 тест Хаусмана використовується для перевірки, яка модель (модель фіксованих ефектів або модель випадкових ефектів) є більше підходить для аналізу панельних даних [12, 18-20]. Тест перевіряє нульову гіпотезу, що оцінки моделі випадкових ефектів є консистентними. Якщо ця гіпотеза відхиляється, модель фіксованих ефектів вважається кращою.

Зробимо позначення:

- β_{FE} – оцінки коефіцієнтів з моделі фіксованих ефектів (Fixed Effects Model);
- β_{RE} – оцінки коефіцієнтів з моделі випадкових ефектів (Random Effects Model);
- $Var(\beta_{FE})$ – дисперсія коефіцієнтів з моделі фіксованих ефектів;
- $Var(\beta_{RE})$ – дисперсія коефіцієнтів з моделі випадкових ефектів.

Приведемо гіпотези тесту:

- нульова гіпотеза (H_0): різниці між оцінками коефіцієнтів у моделях фіксованих і випадкових ефектів немає. Модель випадкових ефектів є кращою (ефективнішою);
- альтернативна гіпотеза (H_1): Оцінки коефіцієнтів між моделями фіксованих і випадкових ефектів є статистично різними. Модель фіксованих ефектів є кращою.

Алгоритм застосування тесту наступний.

Крок 1: Оцінка моделей:

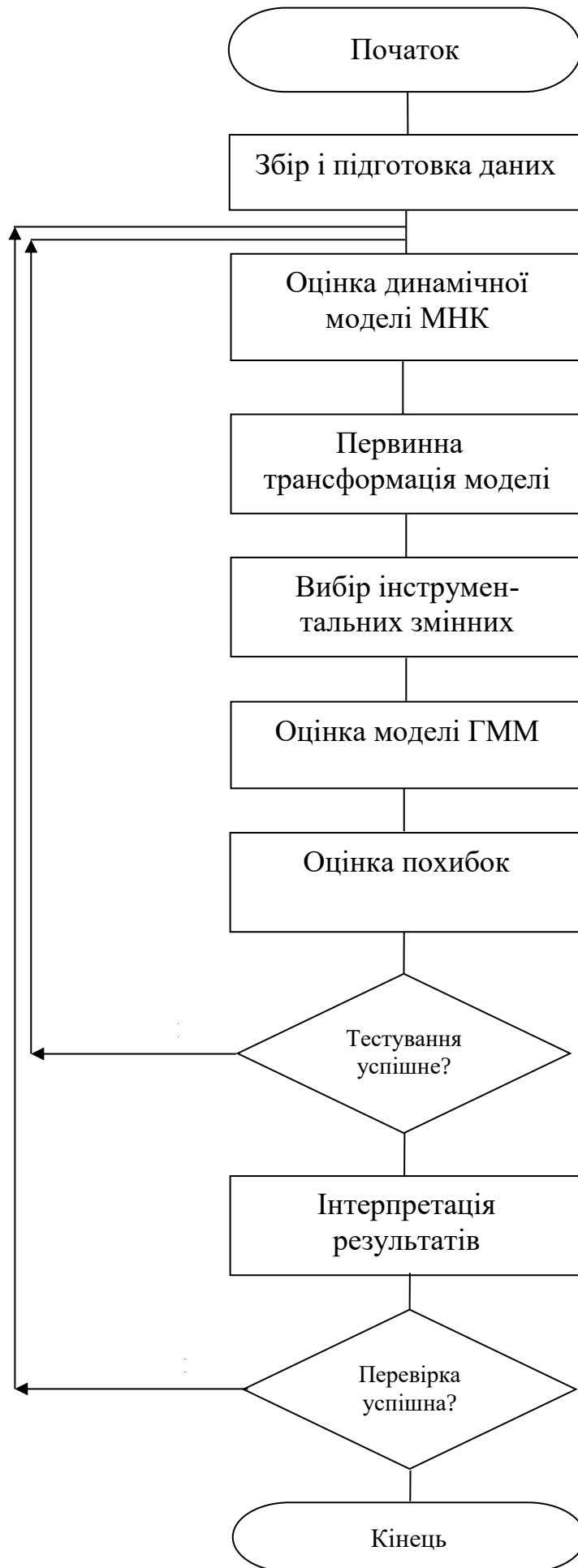


Рисунок 2.3 – Алгоритм оцінювання динамічної панельної моделі

- проходить оцінка моделі фіксованих ефектів і знаходять оцінки коефіцієнтів (β_{FE}) та їх дисперсія ($Var(\beta_{FE})$);
- проходить оцінка моделі випадкових ефектів і знаходять оцінки коефіцієнтів (β_{RE}) та їх дисперсія ($Var(\beta_{RE})$).

Крок 2: Обчислення тестової статистики.

Формула для статистики тесту Хаусмана:

$$H = (\beta_{FE} - \beta_{RE})^T [Var(\beta_{FE}) - Var(\beta_{RE})]^{-1} (\beta_{FE} - \beta_{RE})$$

Це статистика χ^2 з кількістю ступенів свободи, рівною кількості регресорів (без урахування константи).

Крок 3: Перевірка гіпотез

- порівняння значення H з критичним значенням χ^2 для заданого рівня значущості (α , зазвичай 0.05);
- якщо H перевищує критичне значення χ^2 , відхиляємо нульову гіпотезу (H_0) і вибираємо модель фіксованих ефектів;
- якщо H менше критичного значення χ^2 , не відхиляємо нульову гіпотезу і вибираємо модель випадкових ефектів.

2.5 Алгоритм методу різниць у різницях

Метод різниць у різницях (DiD) використовується для оцінки причинного впливу події на результативну змінну [18-20]. Цей метод порівнює зміну у результативній змінній у групі, яка зазнала впливу (treatment group), із зміною у контрольній групі, яка не зазнала впливу (control group).

Умови для використання методу DiD:

1. Наявність двох груп:
 - treatment group (група впливу): зазнала інтервенції;

- control group (контрольна група): не зазнала інтервенції.
- 2. Дані до та після події:
 - дані повинні бути зібрані як до (pre-treatment), так і після (post-treatment) події.
- 3. Гіпотеза паралельних трендів:
 - до інтервенції обидві групи повинні мати подібні тренди у результативній змінній.

Позначення:

- Y_{it} : значення результативної змінної для індивіда i в момент часу t ;
- D_i : індикатор, чи належить індивід i до групи впливу (1, якщо так, інакше 0);
- T_t : індикатор, чи відбувся вплив у момент часу t (1, якщо після інтервенції, інакше 0);
- $D_i \times T_t$: взаємодія, яка показує, чи перебуває індивід i у групі впливу після події;
- α : константа;
- γ : часовий ефект (загальний тренд у часі);
- δ : ефект від інтервенції (оцінюваний параметр);
- ε_{it} : похибка.

Метод DiD реалізується через наступну лінійну модель:

$$Y_{it} = \alpha + \beta_1 D_i + \beta_2 T_t + \delta(D_i \times T_t) + \varepsilon_{it},$$

де δ – це ключовий параметр, який показує вплив інтервенції (різницю в різницях).

Приведемо алгоритм застосування DiD:

1. Підготовка даних:
 - організувати дані у формат панельних даних із двома групами (treatment і control) та двома часовими періодами (до та після події);
 - створити індикаторні змінні:

- D_i : група впливу (1 для групи treatment, 0 для control).
- T_t : часовий індикатор (1 після події, 0 до події).
- $D_i \times T_t$: перемноження D_i і T_t , яке позначає індивідів з групи впливу після події.

2. Оцінка моделі:

- оцінити модель за допомогою лінійної регресії або методу найменших квадратів:

$$Y_{it} = \alpha + \beta_1 D_i + \beta_2 T_t + \delta(D_i \times T_t) + \varepsilon_{it};$$

- оцінити ключовий параметр δ .

3. Перевірка гіпотези паралельних трендів:

- перевірити, чи результати до інтервенції мають схожі тренди у treatment і control групах.

4. Перевірка значущості:

- перевірити p -значення для δ . Якщо $p < 0.05$, вплив інтервенції вважається статистично значущим.

5. Інтерпретація:

- якщо $\delta > 0$, це свідчить про позитивний вплив інтервенції;
- якщо $\delta < 0$, це свідчить про негативний вплив інтервенції.

2.6 Висновки до розділу 2

Отже, в даному розділі приведені алгоритми (і їх блок-схеми) оцінювання моделі фіксованих ефектів, з випадковими ефектами, динамічної моделі, алгоритм тесту Хаусмана, алгоритм методу різниць у різницях.

3 ПРОГРАМНА РЕАЛІЗАЦІЯ АЛГОРИТМІВ ОЦІНЮВАННЯ ЯКОСТІ НАВЧАННЯ СТУДЕНТІВ НА ОСНОВІ 3D ДАНИХ

3.1 Апаратне забезпечення для реалізації програмної системи

В підрозділі 1.3 приведено аналіз інструментальних засобів та програмного забезпечення для аналізу панельних даних. Виділено, що найбільш поширеними у використанні є інструменти R та EViews [28, 29].

Приведемо мінімальні вимоги до апаратного забезпечення для встановлення R.

R – це інструмент, який не вимагає потужного апаратного забезпечення. Мінімальні вимоги для R:

Операційна система:

- Windows: Windows 7 і вище (Windows 10 рекомендовано);
- macOS: macOS 10.13 High Sierra і вище;
- Linux: R доступний для більшості дистрибутивів Linux.

Процесор:

- будь-який сучасний процесор (Intel або AMD) з архітектурою x86 або x86-64.

Оперативна пам'ять (RAM):

- мінімум: 1 ГБ (2 ГБ або більше рекомендовано);
- для великих наборів даних і складних розрахунків треба 4 ГБ або більше.

Місце на диску:

- мінімум: 150 МБ для основної інсталяції R;
- додаткове місце для пакетів R, які ви завантажуються.

Відеокарта:

- достатньо інтегрованої відеокарти.

Мінімальні вимоги до апаратного забезпечення для EViews.

EViews – це комерційне програмне забезпечення для економетричного аналізу, яке також має порівняно низькі вимоги до апаратного забезпечення, але трохи вищі, ніж у R.

Операційна система:

- Windows: Windows 8 або Windows 10 (32-bit або 64-bit);
- macOS і Linux: EViews офіційно не підтримує macOS і Linux. Проте його можна запускати на цих операційних системах через віртуальні машини або емулюючі програми (наприклад, Parallels на macOS).

Процесор:

- мінімум: 1 GHz (бажано багатоядерний процесор для кращої продуктивності).

Оперативна пам'ять (RAM):

- мінімум: 1 ГБ;
- рекомендовано: 2 ГБ або більше, особливо для великих економетричних моделей і обробки великих обсягів даних.

Місце на диску:

- мінімум: 400 МБ для основної інсталяції;
- додатковий простір для збереження файлів даних та результатів аналізу.

Відеокарта:

- достатньо інтегрованої відеокарти.

Рекомендації для оптимальної продуктивності обох програм:

1. Оперативна пам'ять: 8 ГБ або більше RAM, для роботи з великими даними або складними економетричними моделями.
2. Процесор: Багатоядерний процесор (наприклад, Intel i5 або вище) для швидшої обробки даних.
3. Диск SSD: Використання SSD-накопичувача замість жорсткого диска (HDD) може значно прискорити завантаження великих файлів даних і підвищити загальну продуктивність.

3.2 Програмна система для аналізу панельних даних

Програмна система для аналізу панельних даних включає інструменти для оцінювання моделей фіксованих ефектів, випадкових ефектів, динамічних моделей, тесту Хаусмана та методу різниць у різницях, є потужним інструментом для аналізу панельних даних. Наведемо концептуальний опис такої системи.

Основними компонентами програмної системи є:

1. Інтерфейс користувача (UI):

- Графічний інтерфейс (GUI) для інтерактивного використання.
- Командний інтерфейс (CLI) для автоматизованого запуску аналізів.
- Інструменти для завантаження, перевірки та попередньої обробки даних.

2. Модуль попередньої обробки даних:

- Завантаження даних із різних форматів (CSV, Excel, SQL-бази даних).
- Перевірка якості даних: обробка пропусків, дублювання, перевірка форматів.
- Перетворення даних у формат панельних даних.

3. Модуль оцінювання моделей:

- Модель фіксованих ефектів:
 - підтримка оцінки основної моделі фіксованих ефектів;
 - виведення статистичних показників (коефіцієнти, р-значення, R^2).
- Модель випадкових ефектів:
 - Оцінка моделі випадкових ефектів із виведенням статистичних характеристик.
- Динамічна модель:
 - врахування лагових змінних залежної змінної;
 - підтримка методів оцінки, таких як GMM.

4. Модуль тестування:

- Тест Хаусмана:
 - порівняння моделей фіксованих і випадкових ефектів;
 - автоматичний вивід результатів тесту з рекомендацією щодо вибору моделі.
- Тест гіпотези паралельних трендів (для DiD):
 - перевірка, чи тренди в контрольній і групі впливу були схожими до інтервенції.

5. Метод різниць у різницях (DiD):

- Оцінка впливу політики чи події через різницю в різницях.
- Побудова моделей з індикаторними змінними:
 - індикатор часу (до/після);
 - індикатор групи (група впливу/контрольна група);
 - взаємодія часу та групи (основний ефект).
- Інтерпретація ключового параметра (оцінка впливу).

6. Модуль візуалізації:

- Побудова графіків:
 - тренди результативної змінної в контрольній і групі впливу;
 - оцінка залишків і їх розподілу.
- Таблиці з основними результатами для експорту (Excel, PDF).

7. Експорт результатів:

Збереження оцінок моделей, тестів, діагностичних графіків у зручних форматах (CSV, Excel, Word, PDF).

Рисунок 3.1 демонструє програмну систему для аналізу панельних даних.

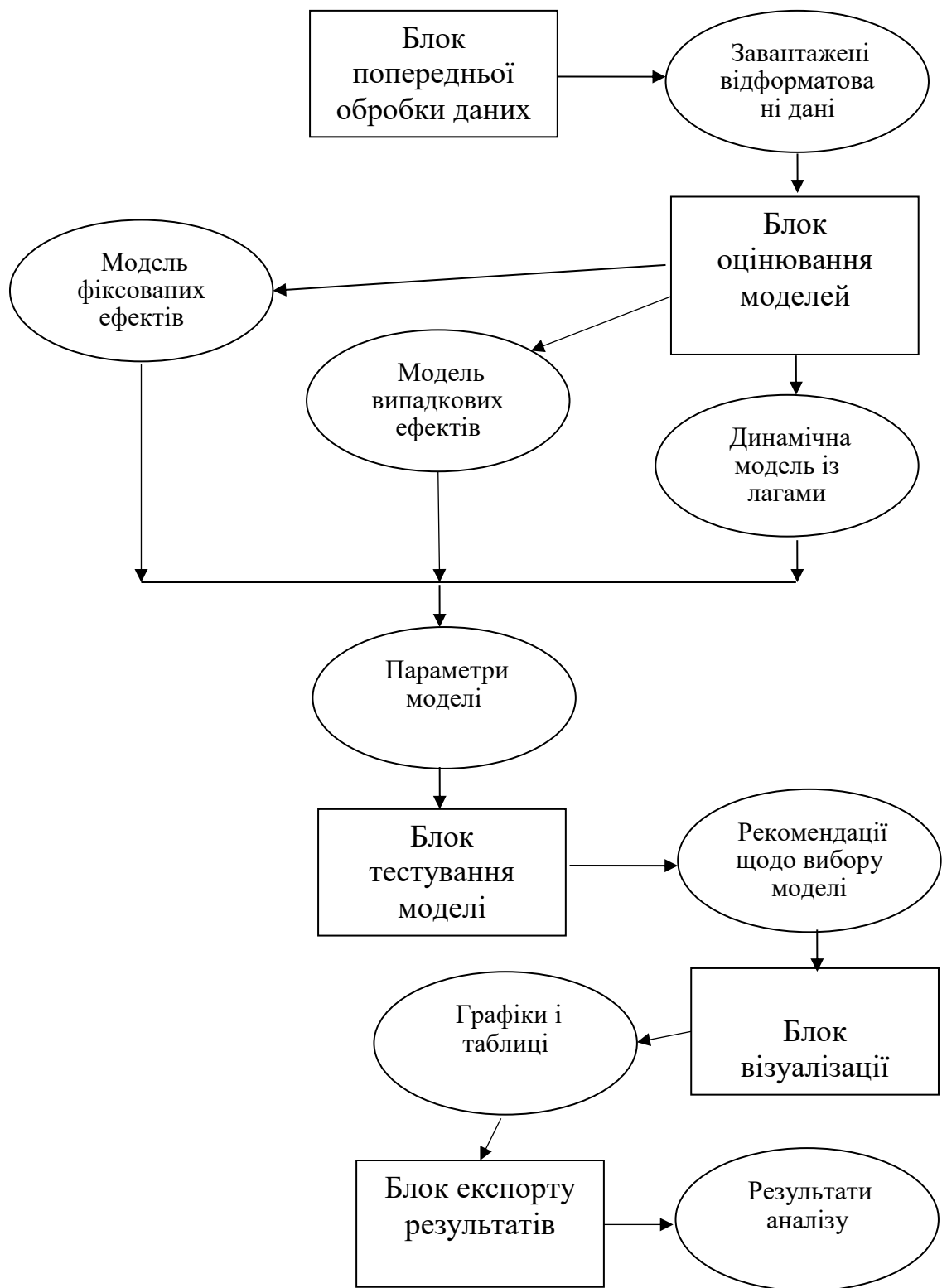


Рисунок 3.1 – Структура програмної системи для аналізу панельних даних

3.3 Підготовка та обробка даних для програмної системи

Оцінювання якості навчання проводилося на основі екзаменаційно-залікових відомостей студентів спеціальностей «Інженерія програмного забезпечення» (ІПЗ), «Комп'ютерна інженерія» (КІ), «Комп'ютерні науки» (КН) Західноукраїнського національного університету. Опрацьовано екзаменаційно-залікові відомості за 2017-2023 рр., за якими підраховано середню успішність студентів з вищої математики, теорії ймовірностей і математичної статистики, дискретної математики, основ програмування, операційних систем, алгоритмів та структур даних. Було розглянуто близько 350 відомостей, за якими підраховано середній бал. Зразок відомостей приведено в додатку А.

В таблиці 3.1 приведені середні успішності з вищої математики, теорії ймовірностей і математичної статистики, дискретної математики. В таблиці 3.2 приведені середні успішності з основ програмування, операційних систем, алгоритмів та структур даних.

На рисунках 3.2 – 3.7 зображені тренди успішності з цих предметів, які реалізовані в R.

Таблиця 3.1 – Середня успішність з вищої математики, теорії ймовірностей і математичної статистики, дискретної математики студентів спеціальностей КІ, КН, ІПЗ

Рік	Вища математика			ТІМС			Дискретна математика		
	КІ	КН	ІПЗ	КІ	КН	ІПЗ	КІ	КН	ІПЗ
2017	76,6	71	72	71,4	71,8	72,1	74	75,2	74,8
2018	61,88	70	71	73,6	73,52	73,01	74,53	74,2	75
2019	73,32	69	72	71,8	71,74	71,6	77,5	75	79
2020	73,87	70	73	71	71,05	72	75,82	76	75
2021	74,18	68	74	73	72,96	72	76,52	76,4	74
2022	67,95	67	71	64,6	65,8	65,9	66,69	67	66
2023	62,45	68	68	63,92	63,94	64,1	63,91	62	64

Таблиця 3.2 – Середня успішність з основ програмування, операційних систем, алгоритмів та структур даних студентів спеціальностей КІ, КН, ІПЗ

Рік	Основи програмування			Операційні системи			Алгоритми та структури даних		
	КІ	КН	ІПЗ	КІ	КН	ІПЗ	КІ	КН	ІПЗ
2017	77,38	78	77,01	75,53	75,2	75,8	75,3	75,22	74,87
2018	78,01	77,9	78,55	73,88	74,2	73,33	72,71	72,91	72,4
2019	76,98	77,02	76,55	74,01	75	76,1	62,08	62,22	62,04
2020	81,61	80,44	82,7	74,8	75,22	74	68,77	68,25	69
2021	78,01	78,4	78,53	76,35	76,4	76,12	67,77	68	68,02
2022	72,61	72,71	72,75	71,15	71,14	71,24	67,45	67	68,4
2023	73,59	73,49	73,9	71,25	72,01	71,15	63,28	63,9	64

Проаналізувавши тренди, можемо зробити висновок, що середня успішність з усіх розглянутих дисциплін (рисунки 3.2-3.7) суттєво зросла в післяковідний період (2020-2021 рр.) і знизилася на всіх трьох спеціальностях в воєнний період (2022-2023 рр.).

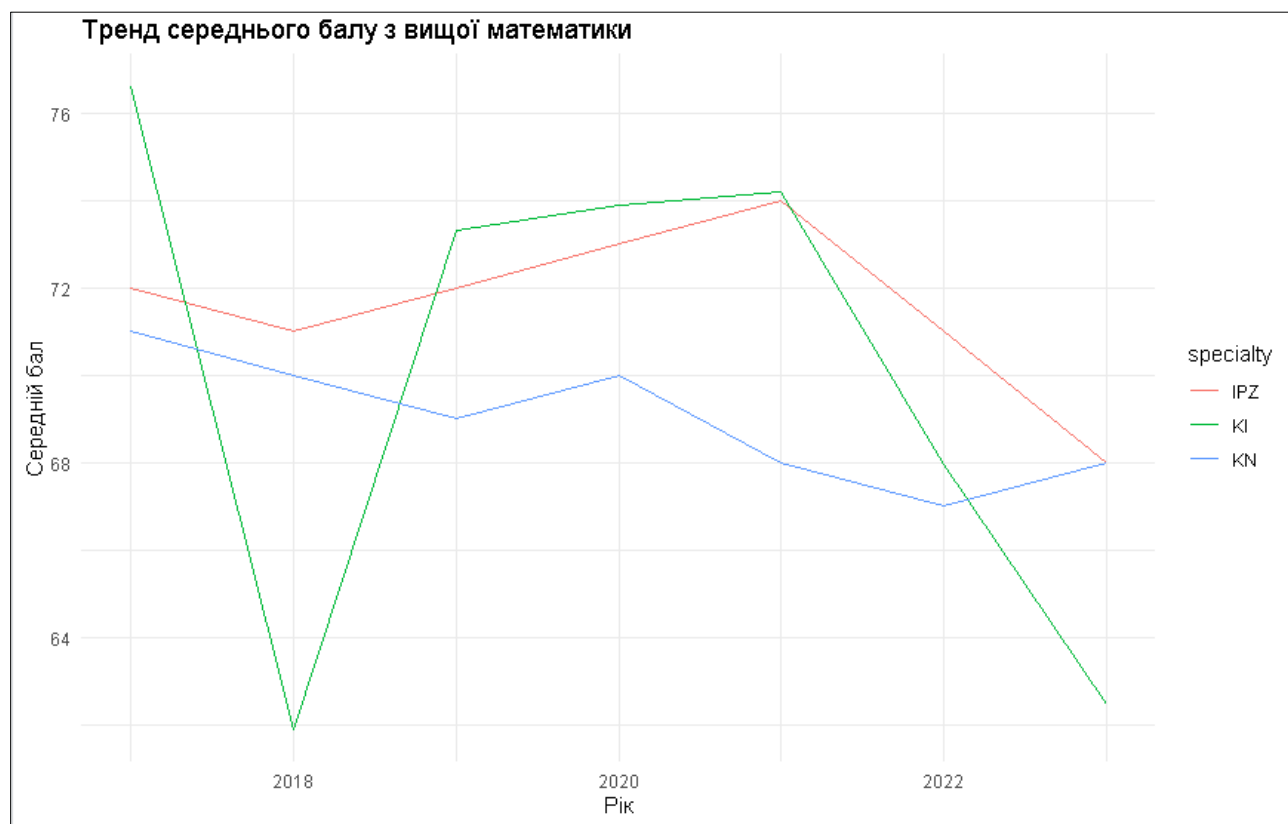


Рисунок 3.2 – Тренд середнього балу з вищої математики

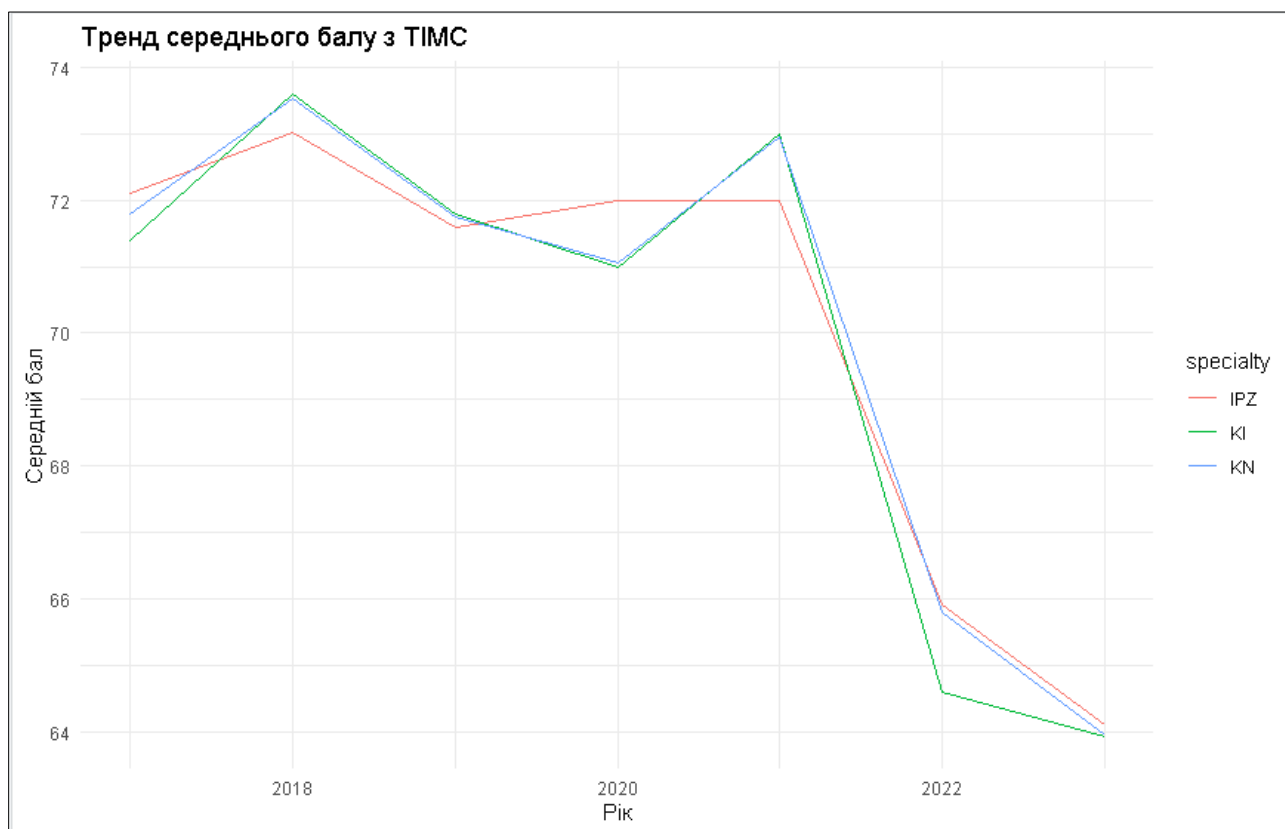


Рисунок 3.3 – Тренд середнього балу з теорії ймовірностей і математичної статистики

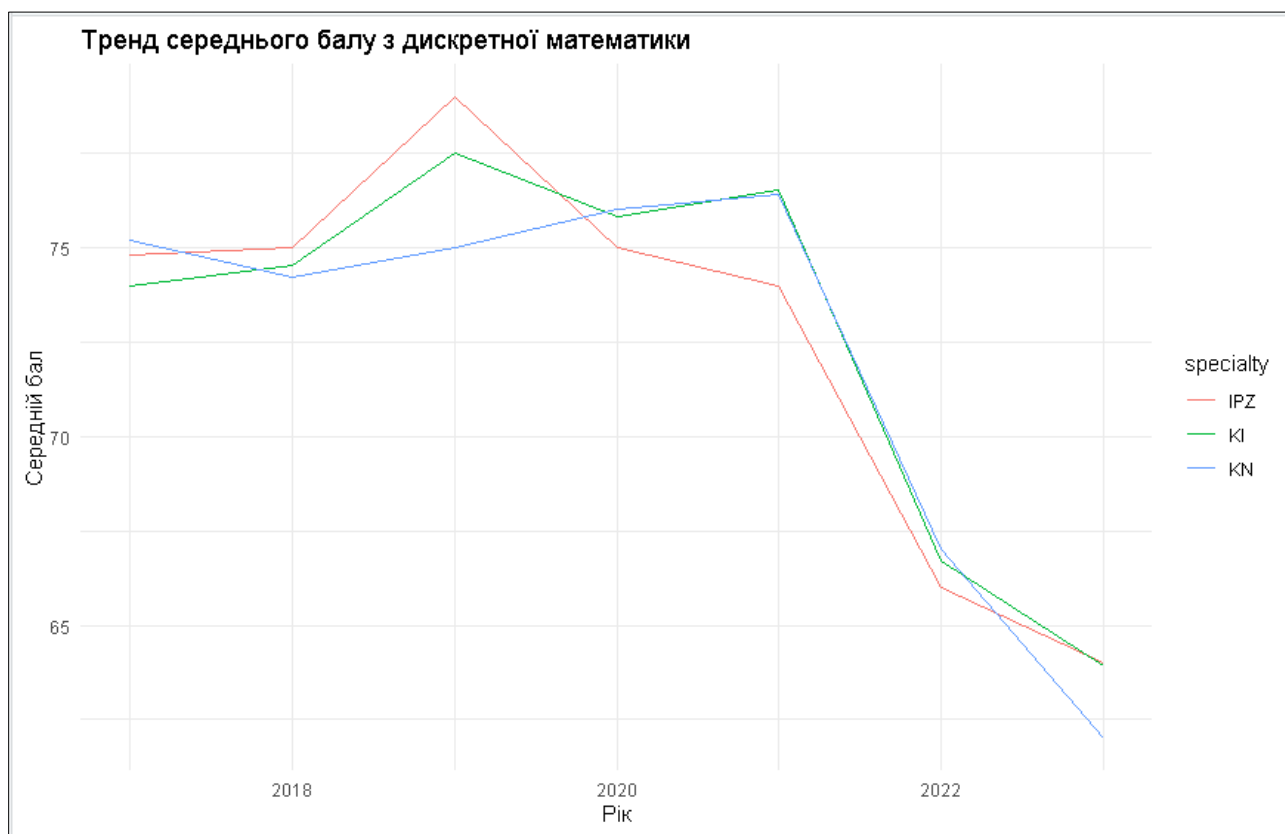


Рисунок 3.4 – Тренд середнього балу з дискретної математики

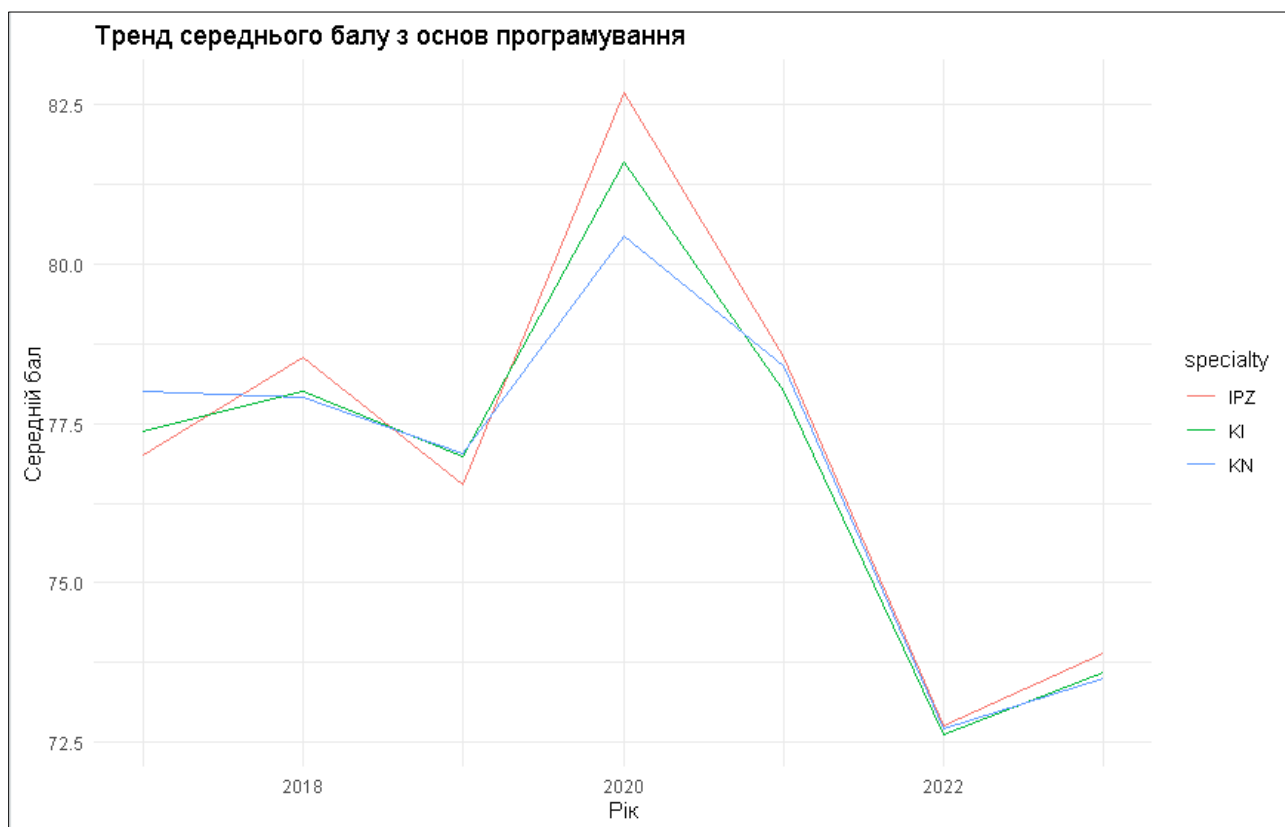


Рисунок 3.5 – Тренд середнього балу з основ програмування

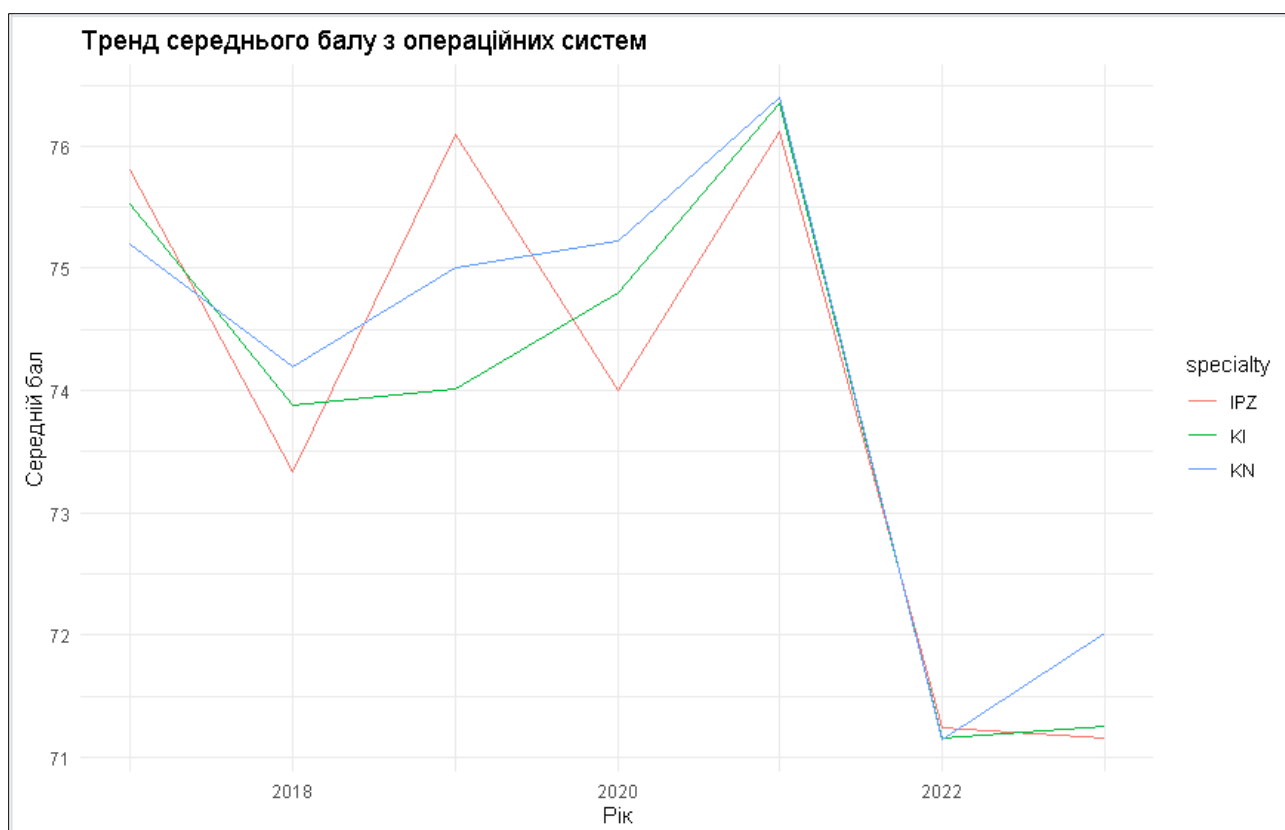


Рисунок 3.6 – Тренд середнього балу з операційних систем

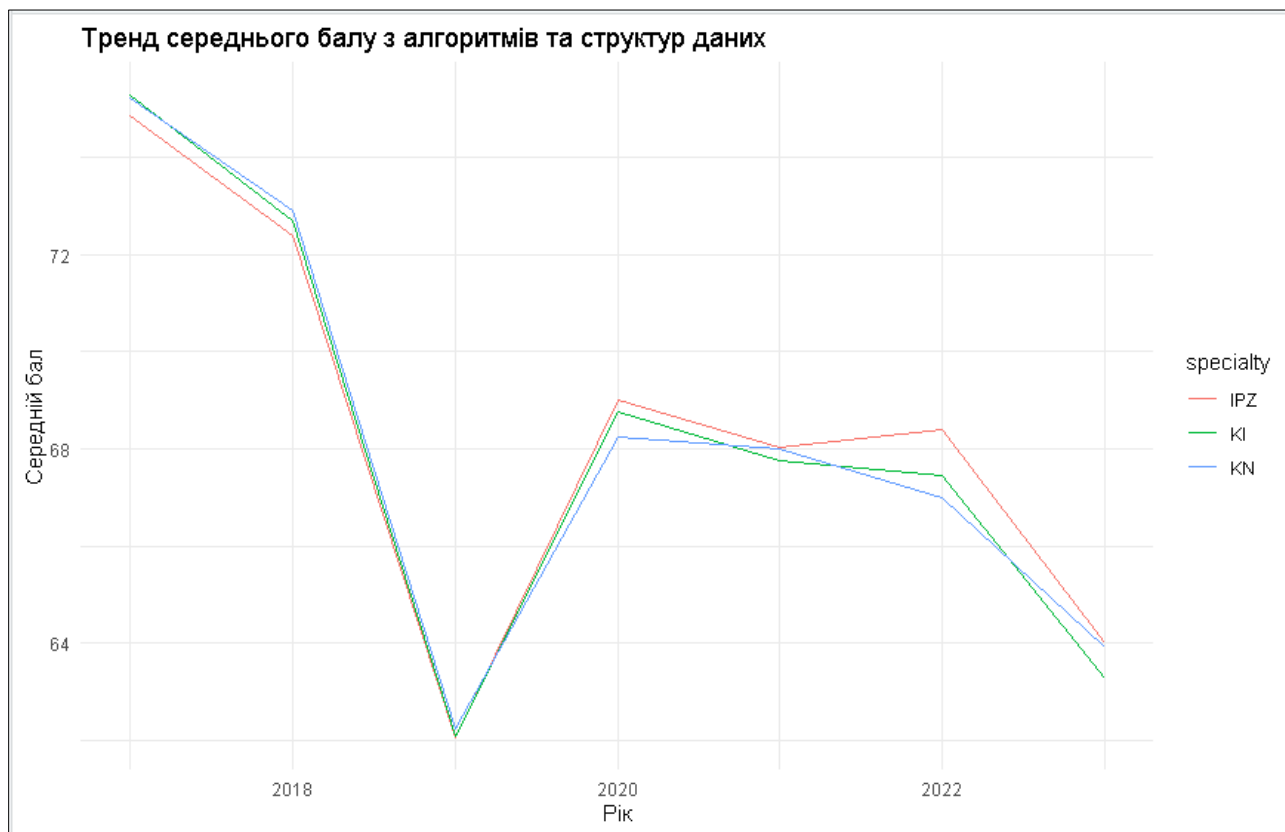


Рисунок 3.7 – Тренд середнього балу з алгоритмів та структур даних

3.4 Програмна реалізація моделей оцінювання якості навчання в середовищі R

В даному підрозділі опишемо результати реалізації в R.

На першому кроці встановлюємо і завантажуюмо пакет `plm`

```
# Встановлення та завантаження пакету plm
install.packages("plm")
library(plm)
```

Дальше завантажуюмо панельні дані з листа Excel

```
# Завантаження панельних даних
install.packages("readxl ")
```

```
library(readxl)
file_path <- file.choose() # Виберіть файл вручну через вікно
my_data <- read_excel(file_path)
```

Дальше проходить побудова графіків для усіх дисциплін, що ми розглянули (рисунки 3.2-3.7).

```
install.packages("ggplot2")
library(ggplot2)
# Побудова графіків трендів для кожного предмета
ggplot(data, aes(x = year, y = high_math, color = specialty))
+
  geom_line() +
  labs(title = "Тренд середнього балу з вищої математики",
        x = "Рік",
        y = "Середній бал") +
  theme_minimal()
```

Щоб визначити який вплив мають дисципліни на середню успішність з операційних систем, будуємо модель фіксованих ефектів.

```
# Модель фіксованих ефектів
fixed_model <- plm(oper_sys ~ discr_math + TIMS + fund_progr +
alg_data_str + high_math, data = pdata, model = "within")
summary(fixed_model)
```

На рисунку 3.8 продемонстровано результат моделювання в середовищі R.

Отже, результатом програми є:

```
Oneway (individual) effect Within Model
Call:
plm(formula = oper_sys ~ discr_math + TIMS + fund_progr +
alg_data_str +
      high_math, data = pdata, model = "within")
Balanced Panel: n = 3, T = 7, N = 21
Residuals:
      Min.      1st Qu.      Median      3rd Qu.      Max.
-1.493888 -0.754242  0.089277  0.755016  1.241229

Coefficients:
              Estimate Std. Error t-value Pr(>|t|)
discr_math      0.1513832  0.1728450  0.8758   0.3970
```

TIMS	0.2187679	0.2600361	0.8413	0.4154
fund_progr	-0.0012533	0.1404259	-0.0089	0.9930
alg_data_str	-0.0231578	0.0733157	-0.3159	0.7571
high_math	0.1124744	0.0841541	1.3365	0.2043

Total Sum of Squares: 74.048
 Residual Sum of Squares: 14.865
 R-Squared: 0.79926
 Adj. R-Squared: 0.69117
 F-statistic: 10.3519 on 5 and 13 DF, p-value: 0.00035688

```

RStudio
File Edit Code View Plots Session Build Debug Profile Tools Help
Go to file/function Addins
Source
Console Terminal Background Jobs
R 4.1.2 ~ /
> fixed_model <- plm(oper_sys ~ discr_math + TIMS + fund_progr + alg_data_str + high_math, data = pdata, model = "within")
> summary(fixed_model)
oneway (individual) effect within Model

Call:
plm(formula = oper_sys ~ discr_math + TIMS + fund_progr + alg_data_str +
    high_math, data = pdata, model = "within")

Balanced Panel: n = 3, T = 7, N = 21

Residuals:
    Min.    1st Qu.    Median     3rd Qu.     Max.
-1.493888 -0.754242  0.089277  0.755016  1.241229

Coefficients:
              Estimate Std. Error t-value Pr(>|t|)
discr_math    0.1513832   0.1728450   0.8758   0.3970
TIMS          0.2187679   0.2600361   0.8413   0.4154
fund_progr   -0.0012533   0.1404259  -0.0089   0.9930
alg_data_str -0.0231578   0.0733157  -0.3159   0.7571
high_math     0.1124744   0.0841541   1.3365   0.2043

Total Sum of Squares: 74.048
Residual Sum of Squares: 14.865
R-Squared: 0.79926
Adj. R-Squared: 0.69117
F-statistic: 10.3519 on 5 and 13 DF, p-value: 0.00035688
  
```

Рисунок 3.8 – Модель фіксованих ефектів

Отримано наступну модель

$$\begin{aligned}
 oper_sys_{it} - \overline{oper_sys}_i = & 0.151 \cdot (discr_math_{it} - \overline{discr_math}_i) + 0.219 \cdot (TIMS_{it} - \\
 & - \overline{TIMS}_i) - 0.001 \cdot (fund_progr_{it} - \overline{fund_progr}_i) - 0.023 \cdot (alg_data_str_{it} - \\
 & - \overline{alg_data_str}_i) + 0.112 \cdot (high_math_{it} - \overline{high_math}_i) + \varepsilon_{it},
 \end{aligned}$$

де $oper_sys_{it}$ – значення залежної змінної $oper_sys$ для індивіда i у момент часу t ;

$\overline{oper_sys}_i$ – середнє значення $oper_sys$ для індивіда i за весь період часу;

$discr_math_{it}$, $TIMS_{it}$, $fund_progr_{it}$, $alg_data_str_{it}$, $high_math_{it}$ – значення незалежних змінних у момент часу t ;

$\overline{discr_math}_i$, $\overline{TIMS}_i$, $\overline{fund_progr}_i$, $\overline{alg_data_str}_i$, $\overline{high_math}_i$ – середні значення відповідних змінних для індивіда i за весь період часу;

ε_{it} – залишок (випадкова помилка).

Зробимо інтерпретацію результатів моделі фіксованих ефектів (Within Model).

Модель фіксованих ефектів, оцінила залежність змінної `oper\_sys` від незалежних змінних `discr\_math`, `TIMS`, `fund\_progr`, `high\_math`, `alg\_data\_str`. Розглянемо кожну частину результату детальніше.

Для кожної незалежної змінної наведено оцінений коефіцієнт, стандартну помилку, t-статистику та р-значення (таблиця 3.3).

Таблиця 3.3 – Результати оцінки і тестування моделі фіксованих ефектів

Змінна	Оцінений коефіцієнт	Стандартна помилка	t-статистика	р-значення
`discr_math`	0.1514	0.1728	0.8758	0.3970
`TIMS`	0.2188	0.2600	0.8413	0.4154
`fund_progr`	-0.0013	0.1404	-0.0089	0.9930
`high_math`	0.1125	0.0842	1.3365	0.2043
`alg_data_str`	-0.0232	0.0733	-0.3159	0.7571

- Коефіцієнти показують, на скільки змінюється залежна змінна середня успішність з операційних систем при зміні відповідної середньої успішності на одиницю, за умови фіксації інших змінних.

- р-значення для всіх змінних більше за 0.05, що означає, що жодна зі змінних не є статистично значущою на рівні значущості 5%. Зокрема:

- середня успішність з основ програмування має дуже високе р-значення (0.9930), що вказує на те, що вона практично не впливає на середню успішність з операційних систем у даній моделі.

- середня успішність з вищої математики має найнижче р-значення серед усіх, але воно все ще вище 0.05, тобто не є статистично значущим.

Але загальні статистичні показники моделі вийшли непогані.

Загальна сума квадратів, яка показує загальну дисперсію в залежній змінній – 74.048. Сума квадратів залишків, що вказує на невідповідність моделі до фактичних даних – 14.865, що є меншим від TSS. $R^2 = 0.79926$ – це значення показує, що модель пояснює приблизно 79.9% варіації в залежній змінній `oper\_sys`. Це доволі високий показник, що свідчить про те, що модель добре пояснює варіацію у даних. Скориговане значення $R^2 = 0.69117$, що враховує кількість змінних і кількість спостережень, і є трохи нижчим, що вказує на помірну кількість пояснюючої здатності моделі. Статистика F-тесту для перевірки значущості всієї моделі – 10.3519. р-значення для F-тесту – 0.00035688, що є менше 0.05, що означає, що модель є статистично значущою на рівні значущості 5%. Тобто, всі незалежні змінні разом мають значущий вплив на `oper\_sys`, навіть якщо жодна з них окремо не є значущою.

Отже, модель добре пояснює варіацію в залежній змінній `oper\_sys` і є загалом значущою. Однак жодна з окремих незалежних змінних не є статистично значущою. Це може свідчити про те, що вплив змінних є слабким індивідуально, але значущим у сукупності.

Розглянемо модель випадкових ефектів за допомогою трансформації Wallace-Hussain.

```
# Модель випадкових ефектів
random_model <- plm(oper_sys ~ discr_math + TIMS + fund_progr
+ alg_data_str + high_math,
                    data = pdata,
                    model = "random",
                    random.method = "walhus") # Wallace-
Hussain method
summary(random_model)
```

Отримуються наступні результати:

```
Oneway (individual) effect Random Effect Model
(Wallace-Hussain's transformation)
Call:
```

```
plm(formula = oper_sys ~ discr_math + TIMS + fund_progr +
alg_data_str +
      high_math, data = pdata, model = "random", random.method =
"walhus")
```

Balanced Panel: n = 3, T = 7, N = 21

Effects:

	var	std.dev	share
idiosyncratic	0.8307	0.9114	1
individual	0.0000	0.0000	0
theta:	0		

Residuals:

Min.	1st Qu.	Median	3rd Qu.	Max.
-1.70094	-0.73173	0.19859	0.57725	1.34609

Coefficients:

	Estimate	Std. Error	z-value	Pr(> z)
(Intercept)	42.1274455	7.8992685	5.3331	9.656e-08 ***
discr_math	0.1450086	0.1625356	0.8922	0.3723
TIMS	0.2381832	0.2456050	0.9698	0.3322
fund_progr	-0.0018646	0.1352426	-0.0138	0.9890
alg_data_str	-0.0251682	0.0701307	-0.3589	0.7197
high_math	0.0930526	0.0766465	1.2140	0.2247

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Total Sum of Squares: 74.404

Residual Sum of Squares: 15.913

R-Squared: 0.78613

Adj. R-Squared: 0.71484

Chisq: 55.1368 on 5 DF, p-value: 1.2235e-10

Отримано наступну модель

$$\begin{aligned} \text{oper_sys} = & 42.13 + 0.15 \cdot \text{discr_math} + 0.24 \cdot \text{TIMS} - 0.002 \cdot \text{fund_progr} - \\ & - 0.03 \cdot \text{alg_data_str} + 0.09 \cdot \text{high_math} + u_i + \varepsilon_{it}. \end{aligned}$$

Для кожної незалежної змінної наведено оцінений коефіцієнт, стандартну помилку, z-статистику та p-значення (таблиця 3.4).

TSS = 74.404, RSS = 15.913, $R^2 = 0.78613$, R^2 скоригований = 0.71484.

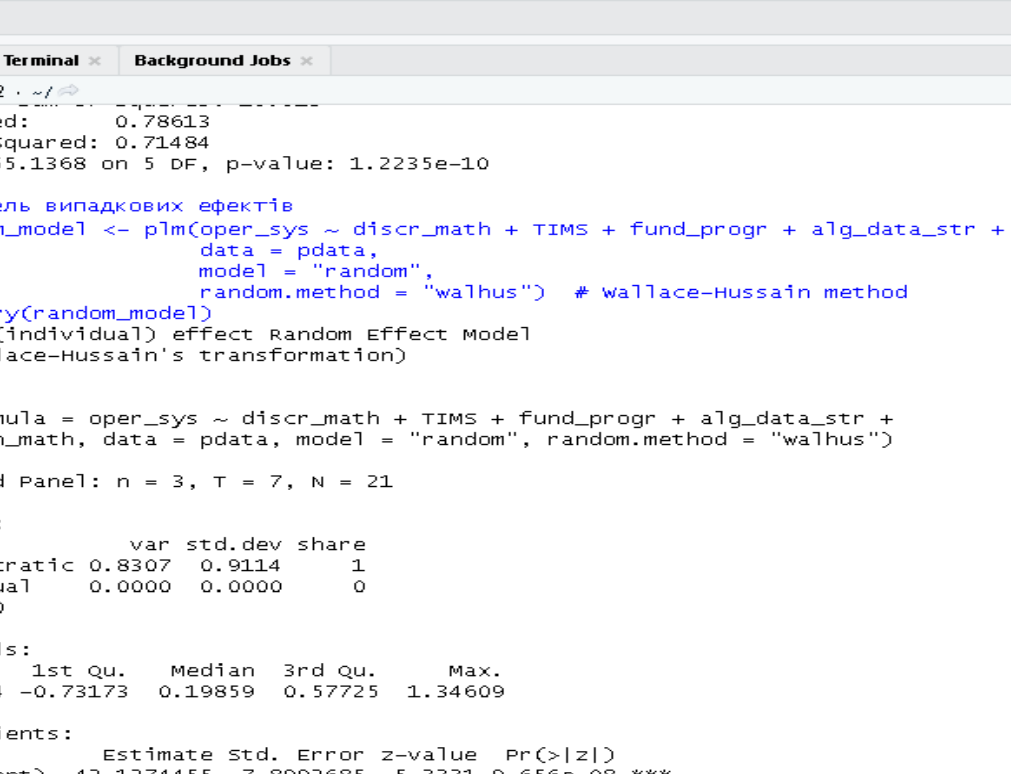
Статистика $\chi^2 = 55.1368$ – показує значущість усієї моделі, p-value = 1.2235e-10 –

значення p для χ^2 -тесту дуже низьке, що вказує на значущість моделі на рівні 0.001.

Таблиця 3.4 – Результати оцінки і тестування моделі випадкових ефектів

Змінна	Оцінений коефіцієнт	Стандартна помилка	z-статистика	p-значення
Intercept	42.1274	7.8993	5.3331	9.656e-08
`discr_math`	0.1450	0.1625	0.8922	0.3723
`TIMS`	0.2382	0.2456	0.9698	0.3322
`fund_progr`	-0.0019	0.1352	-0.0138	0.9890
`high_math`	0.0931	0.0766	1.2140	0.2247
`alg_data_str`	-0.0252	0.0701	-0.3589	0.7197

Результат моделювання продемонстровано на рисунку 3.9.



The screenshot shows the RStudio environment with the following components:

- Menu Bar:** File, Edit, Code, View, Plots, Session, Build, Debug, Profile, Tools, Help.
- Toolbar:** Includes icons for file operations (new, open, save, etc.) and a search bar labeled "Go to file/function".
- Source Panel:** Currently empty.
- Console Panel:** Displays the output of an R session.


```
R R4.1.2 ~ /
R-Squared:      0.78613
Adj. R-Squared: 0.71484
Chisq: 55.1368 on 5 DF, p-value: 1.2235e-10
>
> # Модель випадкових ефектів
> random_model <- plm(oper_sys ~ discr_math + TIMS + fund_progr + alg_data_str + high_math,
+                      data = pdata,
+                      model = "random",
+                      random.method = "wallus") # wallace-Hussain method
> summary(random_model)
Oneway (individual) effect Random Effect Model
(wallace-Hussain's transformation)

Call:
plm(formula = oper_sys ~ discr_math + TIMS + fund_progr + alg_data_str +
    high_math, data = pdata, model = "random", random.method = "wallus")

Balanced Panel: n = 3, T = 7, N = 21

Effects:
              var std.dev share
idiosyncratic 0.8307  0.9114     1
individual    0.0000  0.0000     0
theta: 0

Residuals:
    Min.   1st Qu.   Median   3rd Qu.    Max.
-1.70094 -0.73173  0.19859  0.57725  1.34609

Coefficients:
              Estimate Std. Error z-value Pr(>|z|)
(Intercept)  42.1274455   7.8992685   5.3331 9.656e-08 ***
discr_math    0.1450086   0.1625356   0.8922  0.3723
TIMS          0.2381832   0.2456050   0.9698  0.3322
fund_progr   -0.0018646   0.1352426  -0.0138  0.9890
alg_data_str -0.0251682   0.0701307  -0.3589  0.7197
high_math     0.0930526   0.0766465   1.2140  0.2247
```

Рисунок 3.9 – Модель випадкових ефектів

Отже, модель загалом є значущою, хоча усі незалежні змінні незначущі крім константи. Її значущість вказує на високу середню базову оцінку `oper\_sys` незалежно від впливу змінних.

Розглянемо аналіз динамічної моделі (моделі фіксованих ефектів (Within Model) з використанням лагової змінної). Програмний код:

```
# Динамічна панельна модель
dynamic_model <- plm(oper_sys ~ lag(oper_sys, 1) + discr_math
+ TIMS + fund_progr + alg_data_str + high_math, data = pdata, model
= "within")
summary(dynamic_model)
```

Отримаємо результати:

```
Oneway (individual) effect Within Model
Call:
plm(formula = oper_sys ~ lag(oper_sys, 1) + discr_math + TIMS
+
      fund_progr + alg_data_str + high_math, data = pdata, model
= "within")
Balanced Panel: n = 3, T = 6, N = 18

Residuals:
      Min.   1st Qu.   Median   3rd Qu.    Max.
-1.49653 -0.36030  0.10077  0.25869  1.32628

Coefficients:
              Estimate Std. Error t-value Pr(>|t|)
lag(oper_sys, 1) -0.30980519  0.24379753 -1.2707  0.2357
discr_math       0.20727244  0.21941931  0.9446  0.3695
TIMS             0.18850662  0.29595826  0.6369  0.5400
fund_progr      -0.00063062  0.15931508 -0.0040  0.9969
alg_data_str    -0.00143670  0.15734230 -0.0091  0.9929
high_math       0.08180773  0.09278726  0.8817  0.4009

Total Sum of Squares:    65.584
Residual Sum of Squares: 9.1259
R-Squared:              0.86085
Adj. R-Squared:         0.73716
F-statistic: 9.27981 on 6 and 9 DF, p-value: 0.0019641
```

Отримано наступну модель

$$\begin{aligned} \text{oper_sys}_{it} - \overline{\text{oper_sys}_i} = & -0.310 \cdot (\text{oper_sys}_{it-1} - \overline{\text{oper_sys}_i}) + 0.207 \cdot (\text{discr_math}_{it} - \\ & - \overline{\text{discr_math}_i}) + 0.189 \cdot (\text{TIMS}_{it} - \overline{\text{TIMS}_i}) - 0.001 \cdot (\text{fund_progr}_{it} - \overline{\text{fund_progr}_i}) - \\ & - 0.001 \cdot (\text{alg_data_str}_{it} - \overline{\text{alg_data_str}_i}) + 0.082 \cdot (\text{high_math}_{it} - \overline{\text{high_math}_i}) + \varepsilon_{it}, \end{aligned}$$

де oper_sys_{it-1} – значення залежної змінної для індивіда i , але з лагом у один період часу.

Модель включає лагову змінну $\text{'lag(oper_sys, 1)'}$, яка представляє значення залежної змінної 'oper_sys' за попередній період.

Для кожної незалежної змінної наведено оцінений коефіцієнт, стандартну помилку, t-статистику та p-значення (таблиця 3.5).

Таблиця 3.5 – Результати оцінки і тестування динамічної моделі

Змінна	Оцінений коефіцієнт	Стандартна помилка	t-статистика	p-значення
$\text{'lag(oper_sys, 1)'}$	-0.3098	0.2438	-1.2707	0.2357
'discr_math'	0.2073	0.2194	0.9446	0.3695
'TIMS'	0.1885	0.2960	0.6369	0.5400
'fund_progr'	-0.0006	0.1593	-0.0040	0.9969
'high_math'	0.0818	0.0928	0.8817	0.4009
'alg_data_str'	-0.0014	0.1573	-0.0091	0.9929

Коефіцієнт $\text{'lag(oper_sys, 1)'}$ = -0.3098 є від'ємним, що свідчить про те, що попереднє значення 'oper_sys' має негативний вплив на поточне значення, але цей ефект не є статистично значущим ($p = 0.2357$).

$TSS = 65.584$, $RSS = 9.1259$, $R^2 = 0.86085$, R^2 скоригований = 0.73716.

Статистика $F = 9.27981$ – показує значущість усієї моделі, $p\text{-value} = 0.0019641$ – значення p для F-тесту дуже низьке, що вказує на значущість моделі на рівні 0.05.

Отже, модель є загалом значущою. Тобто всі змінні разом впливають на 'oper_sys' , а індивідуальний вплив кожної змінної, в тому числі лагової, на залежну змінну слабкий.

3.5 Висновки до розділу 3

Отже, в даному розділі приведені мінімальні вимоги до апаратного забезпечення побудованої програмної системи, описана програмна система інструментів для оцінювання моделей для аналізу панельних даних, приведені опис вхідних даних, експериментальні дослідження та результати роботи програмної системи.

ВИСНОВКИ

1. Проведено аналіз навчальних програм Західноукраїнського національного університету, зокрема факультету комп'ютерних інформаційних технологій з метою підготовки даних для оцінки якості навчання студентів.

2. Проаналізовано способи оцінювання якості навчання, які включають формативне та підсумкове оцінювання, оцінювання за компетентностями, методологію 360-градусного оцінювання, що дало змогу підготувати вхідні дані для аналізу.

3. Проведено аналіз моделей панельних (3D) даних; проаналізовано інструментальні засоби та програмні продукти, що використовуються для аналізу панельних даних.

4. Розроблено алгоритми оцінювання моделей фіксованих ефектів, з випадковими ефектами, динамічної моделі, тесту Хаусмана, методу різниць у різницях для їх програмної реалізації в програмному середовищі R.

5. Проведено програмну реалізацію розроблених алгоритмів оцінювання якості навчання студентів, приведено опис вхідних даних, та здійснено експериментальні дослідження на прикладі даних фундаментальних дисциплін навчальних програм факультету комп'ютерних інформаційних технологій.

6. Експериментальні дослідження показали, що вплив фундаментальних дисциплін (вища математика, теорія ймовірностей і математична статистика, дискретна математика, основи програмування, алгоритми та структури даних) на вивчення спеціальної дисципліни операційні системи є значущим з коефіцієнтом детермінації $R^2 = 0.79926$, та р-значенням для F-тесту – 0.00035688, що є менше 0.05, що означає, що модель є статистично значущою на рівні значущості 5%.

СПИСОК ВИКОРИСТАНИХ ДЖЕРЕЛ

1. Йож А. А., Рутецький Ю. О. Алгоритми та аналіз даних та для оцінювання якості навчання студентів. *Інтелектуальні комп'ютерні системи та мережі* : тези доп. Всеукр. наук.-практ. конф. студентів, аспірантів, молодих вчених (5 лист. 2024 р.). Тернопіль : ЗУНУ, 2024. С. 101-102.
2. Йож А. А., Рутецький Ю. О. Моделювання алгоритмів оцінювання якості навчання студентів. *Інтелектуальні комп'ютерні системи та мережі* : тези доп. Всеукр. наук.-практ. конф. студентів, аспірантів, молодих вчених (5 лист. 2024 р.). Тернопіль : ЗУНУ, 2024. С. 103-104.
3. 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.
4. Західноукраїнський національний університет [Електронний ресурс]. – Режим доступу: <https://www.wunu.edu.ua/>.
5. Факультет комп'ютерних інформаційних технологій [Електронний ресурс]. – Режим доступу: <https://www.wunu.edu.ua/fkit/>
6. Оцінювання результатів навчання: рекомендації МОН [Електронний ресурс]. – Режим доступу: <https://osvita.ua/school/87164/>.
7. Формативне оцінювання [Електронний ресурс]. – Режим доступу: <https://osvitoria.media/tag/formatyvne-otsinyuvannya/>.
8. Гривко А., Ващенко Л. Поточне та формувальне оцінювання в базовій та старшій профільній школі. Український педагогічний журнал. 2021. № 2. С. 72-83.
9. Бажміна, Е.А.А. Формувальне оцінювання: цілі, умови, принципи та структура. Вісник Черкаського національного університету ім. Богдана Хмельницького. Серія: Педагогічні науки. № 4, 2020. С. 130-137.
10. Vashchenko, L. Assessment of academic achievements in general secondary education: the view of participants in the educational process. Continuing

Professional Education: Theory and Practice, №2. 2021. P. 55–63.
<https://doi.org/10.28925/1609-8595.2021.2.7>.

11. Смолінчук Л. Проблема метакомпетенцій в сучасних дослідженнях. Теорія та методика управління освітою. Вип. 9. [Електронний ресурс] Режим доступу до журн.: <http://edu.ua/katalog/692-elektronne-naukove-fahove-vydannja-qteorija-ta-metodyka-upravlinnja-osvitojuq-vypusk-9-2012>.

12. Шевченко Н.Г., Лупан І.В. Моделювання з використанням панельних даних. Наукові записки КДПУ. Серія: математичні науки. 2014. Вип. 73. С. 66–79.

13. Березький О., Березька К., Неміш В. Використання методів просторової економетрики для аналізу результатів вступу абітурієнтів. Геометричне моделювання та інформаційні технології. 2016. №1. С. 5–9.

14. Березький О.М., Березька К.М., Ляшук С.В. Аналіз результатів вступу абітурієнтів на основі методів просторової економетрики. Матеріали II всеукраїнської науково-практичної конференції «Прикладна геометрія та інформаційні технології в моделюванні об'єктів, явищ і процесів» (AGIT-2017), м. Миколаїв, 18–20 жовтня 2017 р. Миколаїв: МНУ імені В.О. Сухомлинського, 2017. С. 93–94.

15. Березька К.М., Неміш В.М., Тимчук Ю.С. Прогнозування прибутків комп'ютерної фірми на основі ARIMA-моделей. III Всеукраїнська науково-практична конференція «Прикладна геометрія та інформаційні технології в моделюванні об'єктів, явищ і процесів» (AGIT-2018), м. Миколаїв, 17–19 жовтня 2018 р. С. 120–121.

16. Березька К. М., Маслій В.В. Модифіковані ARIMA-моделі обсягів прямих іноземних інвестицій. Матеріали IX-ої міжнародної конференції «Актуальні проблеми економіки 2015–2016», м. Київ, 19 лютого 2016 р. К.: Національна академія управління, 2016. С. 15 – 20.

17. Маслій В. В., Березька К.М. Вибір та оцінка ARIMA-моделі для прогнозування обсягів прямих іноземних інвестицій. Науковий вісник Міжнародного гуманітарного університету. 2017. Вип. 24, Ч. 2. С. 115–119.

18. Greene W. H. Econometric Analysis, 7th ed., Prentice Hall, 2011. 1054 p.

19. Murray M. P. Econometrics: A Modern Introduction. Prentice Hall. 2005.
20. Stock James H., Mark W. Watson. Introduction to Econometrics (3rd Edition) 2010.
21. Прикладна економетрика : навч. посіб. : у двох частинах. Частина 2 : [Електронне видання] / Л. С. Гур'янова, Т. С. Клебанова, С. В. Прокопович та ін. Харків : ХНЕУ ім. С. Кузнеця, 2016. 252 с.
22. Неміш В. М., Процик А. І., Березька К. М. Практикум з вищої математики: Навчальний посібник. Тернопіль: Економічна думка, 2007. 302 с.
23. Єрмоменко В., Алілуйко А., Березька К., Мартинюк О. Економетрика : навчальний посібник. Тернопіль: Підручники і посібники, 2023. 168 с.
24. Гевлич І.Г. Використання прикладних програм та навичок програмування при вирішенні економетричних задач. Економіка і організація управління. № 2(54). 2024. С.69-79. DOI 10.31558/2307-2318.2024.2.6.
25. Матвійчук Р. Д., Половенко Л. П. Огляд програмного забезпечення для моделювання економічних процесів. Прикладні аспекти сучасних міждисциплінарних досліджень: матеріали II Міжнародної науково-практичної конференції (м. Вінниця, 24 листопада 2023 р.). Вінниця: ДонНУ імені Василя Стуса, 2023. С. 226-228.
26. Манаков С. Ю. Огляд програмних інструментів для дисципліни «Моделювання та аналіз програмного забезпечення». Європейський вибір України, розвиток науки та національна безпека в реаліях масштабної військової агресії та глобальних викликів XXI століття»: матеріали Міжнар.наук.-практ. конф. (м. Одеса, 17 червня 2022 р.). Одеса: Видавничий дім «Гельветика», 2022. Т. 1. С. 692-695.
27. Stata: офіційний веб-сайт. URL: <https://www.stata.com/>.
28. R-project: офіційний веб-сайт. URL: <https://www.r-project.org/>.
29. EViews: офіційний веб-сайт. URL: <https://www.eviews.com/home.html>.
30. Красько Б. В. Переваги використання EVIEWS для аналізу та моделювання панельних даних. Студентський вісник національного університету водного господарства та природокористування. Випуск 2 (14). 2020. С. 115-118.

31. Python: офіційний веб-сайт. URL: <https://www.python.org/>.
32. LIMDEP: Superior Statistical Analysis Software: офіційний веб-сайт. URL: <https://www.limdep.com/products/limdep/>.
33. Gretl: офіційний веб-сайт. URL: <https://gretl.sourceforge.net/>.
34. Matlab: офіційний веб-сайт. URL: <https://www.mathworks.com/products/matlab.html>.
35. IBM SPSS Software: офіційний веб-сайт. URL: <https://www.ibm.com/spss>.
36. Gauss: офіційний веб-сайт. URL: <https://www.aptech.com/>.
37. Озерчук Д. А., Романюк О. Н., Ціхановська О. М. Аналіз найпоширеніших пакетів прикладних програм для економістів. Електронні інформаційні ресурси: створення, використання, доступ: Збірник матеріалів Міжнародної науково-практичної Інтернет конференції 9-10 листопада 2020 р. Суми/Вінниця: НІКО/ВНТУ, 2020. С. 200-202.
38. Роїк М. В., Присяжнюк О. І., Денисюк В. О. Огляд програмних засобів статистичного аналізу даних. Ефективна економіка. 2017. №7. [Електронний ресурс]. – Режим доступу: <http://www.economy.nayka.com.ua/?op=1&z=5676>
39. Статистичні методи оцінки регіонального розвитку: навчальний посібник / під ред. д-ра екон. наук, проф. Раєвської О. В. – Х. : ФОП Александрова К. М., 2016. 328 с.
40. Бахрушин В.Є. Методи аналізу даних : навч. посіб. Запоріжжя : КПУ, 2011. 268 с.
41. Грисенко М.В., Шворак Л.О. Економіко-математичне моделювання світогосподарських процесів : навч. посіб. Ч. І. Теоретичні основи. К.: ВПЦ "Київський університет", 2016. 271 с.
42. Грисенко М.В., Шворак Л.О. Економіко-математичне моделювання світогосподарських процесів : навч. посіб. Ч. II. Прикладні моделі. К.: ВПЦ "Київський університет", 2016. 223 с.
43. Березька К.М. Економетрика: основи теорії та комп'ютерний практикум. Тернопіль: ЗУНУ, 2022. 152 с.