

Житомирська політехніка	МІНІСТЕРСТВО ОСВІТИ І НАУКИ УКРАЇНИ ДЕРЖАВНИЙ УНІВЕРСИТЕТ «ЖИТОМИРСЬКА ПОЛІТЕХНІКА» Система управління якістю відповідає ДСТУ ISO 9001:2015	Ф-19.05-05.02/2/072.00.1/М/ ОК6-2023
	Екземпляр № 1	Арк 38 / 1

ЗАТВЕРДЖЕНО

Науково-методичною радою
Державного університету
«Житомирська політехніка»

протокол від 31 серпня 2023 р.
№ 10

МЕТОДИЧНІ РЕКОМЕНДАЦІЇ для проведення практичних занять з навчальної дисципліни «Аналіз великих даних у фінансах / Big Data Analytics in Finance»

для здобувачів вищої освіти освітнього ступеня «магістр»
спеціальності 072 «Фінанси, банківська справа та страхування»
освітньо-професійна програма «Фінанси, банківська справа та страхування»
факультет бізнесу та сфери обслуговування
кафедра фінансів та цифрової економіки

Рекомендовано на засіданні
кафедри фінансів та цифрової
економіки
28 серпня 2023 р.,
протокол № 09

Розробник: к.е.н., доцент кафедри фінансів та цифрової економіки,
ОВАНДЕР Наталія

Житомир
2023

Житомирська політехніка	МІНІСТЕРСТВО ОСВІТИ І НАУКИ УКРАЇНИ ДЕРЖАВНИЙ УНІВЕРСИТЕТ «ЖИТОМИРСЬКА ПОЛІТЕХНІКА» Система управління якістю відповідає ДСТУ ISO 9001:2015	Ф-19.05-05.02/2/072.00.1/М/ ОК6-2023
	Екземпляр № 1	Арк 38 / 2

ЗМІСТ

Вступ	3
Тема 1. Поняття великих даних та їх роль у діяльності економічних суб'єктів	4
Тема 2. Візуалізація великих даних	5
Тема 3. Поняття ринку великих даних. Життєвий цикл аналітики даних. Збір та підготовка даних	8
Тема 4. Методи та інструменти аналізу великих даних	10
Тема 5. Методи моделювання та прогнозування економічного розвитку	15
Тема 6. Аналіз великих даних в банківській сфері	35
Тема 7. Аналіз великих даних в державному управлінні та соціальній сфері	36
Тема 8. Аналіз великих даних у маркетингових дослідженнях	37

Житомирська політехніка	МІНІСТЕРСТВО ОСВІТИ І НАУКИ УКРАЇНИ ДЕРЖАВНИЙ УНІВЕРСИТЕТ «ЖИТОМИРСЬКА ПОЛІТЕХНІКА» Система управління якістю відповідає ДСТУ ISO 9001:2015	Ф-19.05-05.02/2/072.00.1/М/ ОК6-2023
	Екземпляр № 1	Арк 38 / 3

ВСТУП

Конкурентні переваги та інновації в цифровому світі базуються на точному аналізі та розумінні великого масиву даних, які постійно створюються. Для того, щоб використовувати знання, надані великими даними, і в режимі реального часу діяти на їх основі, важливою є бізнес аналітика, яка в епоху великих даних стає найбільш затребуваною професією у наступному десятилітті ХХІ століття від стартапів до великих корпорацій. Високопрофесійні менеджери по роботі з великими даними поєднують в собі бізнес-знання, знання технології Big Data та аналітичні навички для більш швидкого прийняття рішень, що забезпечують новаторське зростання та інновації в будь-якій організації.

Навчальна дисципліна призначена для підготовки нового покоління професіоналів в області менеджменту/бізнес-аналітики, які володіють методами обробки великих даних, сучасними аналітичними інструментами, методами та моделями прийняття управлінських рішень, які інформаційно та інноваційно зорієнтовані на створення нових цінностей для клієнтів.

За допомогою найновіших цифрових технологій (хмарні обчислення, мобільні технології, соціальні технології, машинне навчання, інтернет речей) менеджери/бізнес-аналітики здатні ідентифікувати, збирати, аналізувати великі масиви даних, інтерпретувати та трансформувати їх для глибшого розуміння бізнесу в реальному часі з метою більш швидкого прийняття кращих рішень, підвищення ефективності роботи компанії, оптимізації її конкурентоспроможності, пом'якшення ризиків.

Метою викладання навчальної дисципліни є – підготувати фахівців зі знаннями у галузі великих даних; надання фахівцям навичок у галузі діяльності з удосконалення організації праці, виробництва та управління даними; вивчити принципи, методи та форми організації управління великими даними.

Завданнями вивчення навчальної дисципліни є:

- оволодіти теоретичними основами і набути практичних навичок щодо аналізу економічної інформації;
- вміти видобувати знання шляхом інтеграції та аналізу великих даних, отриманих з різноманітних та різнорідних джерел інформації;
- оволодіти теоретичними основами щодо методів оцінювання достовірності моделі та її параметрів, прогнозних характеристик моделі, побудованих на основі великих даних;
- демонструвати знання сучасного рівня технологій інформаційних систем, практичні навички використання прикладних і спеціалізованих комп'ютерних систем та середовищ з метою їх запровадження у професійній діяльності.

Житомирська політехніка	МІНІСТЕРСТВО ОСВІТИ І НАУКИ УКРАЇНИ ДЕРЖАВНИЙ УНІВЕРСИТЕТ «ЖИТОМИРСЬКА ПОЛІТЕХНІКА» Система управління якістю відповідає ДСТУ ISO 9001:2015	Ф-19.05-05.02/2/072.00.1/М/ ОК6-2023
	Екземпляр № 1	Арк 38 / 4

Тема 1. Поняття великих даних та їх роль у діяльності економічних суб'єктів

Питання для обговорення та дискусії

1. BigData – реалії сучасності
2. Історія становлення BigData
3. Характерні особливості BigData та їх відмінності від звичайних даних
4. Концепції BigData
5. Перспективи BigData в сучасному світі та майбутньому
6. Загрози та виклики які несуть технології BigData.

Питання для контролю:

1. Дайте визначення BigData.
2. Як BigData дозволяють залучати потенційних клієнтів?
3. В яких галузях застосовують технології BigData? Наведіть приклади.
4. З яких джерел можуть надходити великі дані?
5. Коли вперше з'явилося поняття BigData?
6. Чому технології BigData так важливі для сучасного суспільства?
7. Які завдання практично неможливо вирішити без технологій BigData?
8. Сфера BigData характеризується рядом ознак. В чому суть ознаки «об'єм»?
9. Сфера BigData характеризується рядом ознак. В чому суть ознаки «швидкість»?
10. Сфера BigData характеризується рядом ознак. В чому суть ознаки «різноманітність»?
11. Сфера BigData характеризується рядом ознак. В чому суть ознаки «достовірність даних»?
12. Сфера BigData характеризується рядом ознак. В чому суть ознаки «цінність накопиченої інформації»?
13. Що таке принцип «3 V», «5 V», «7 V», «14 V»?
14. Дайте визначення структурованих даних.
15. Дайте визначення неструктурованих даних.
16. Дайте визначення напівструктурованих даних.

Завдання до виконання

1. Опишіть переваги та недоліки (загрози) технологій BigData. Результати оформіть у вигляді порівняльної таблиці.
2. Створіть хронологічну таблицю розвитку технологій BigData.
3. Охарактеризуйте нормативну України, яка регулює розвиток технологій BigData в Україні. Визначте відсутність яких нормативних актів стримує процес запровадження технологій BigData в нашій державі.

Житомирська політехніка	МІНІСТЕРСТВО ОСВІТИ І НАУКИ УКРАЇНИ ДЕРЖАВНИЙ УНІВЕРСИТЕТ «ЖИТОМИРСЬКА ПОЛІТЕХНІКА» Система управління якістю відповідає ДСТУ ISO 9001:2015	Ф-19.05-05.02/2/072.00.1/М/ ОК6-2023
	Екземпляр № 1	Арк 38 / 5

Тема 2. Візуалізація великих даних

Питання для обговорення та дискусії

1. Візуалізація даних необхідність чи просте ілюстрування
2. Види візуалізації, їх позитивні сторони та недоліки
3. Інструменти візуалізації їх порівняльна характеристика
4. Роль та приклади візуалізації у фінансовій сфері
5. Види фінансових графіків та правила їх побудови

Питання для контролю:

1. Що таке візуалізація?
2. Для чого використовують візуалізацію?
3. Що є елементами мови візуалізації?
4. Які ви знаєте галузі використання візуалізації?
5. Типи візуалізації.
6. Завдання візуалізації BigData
7. Вимоги до візуалізації BigData
8. Традиційні види візуалізації
9. Характеристика інструментів для візуалізації

Завдання до виконання

1. Проведіть порівняльний аналіз різних інструментів візуалізації. Результати оформіть у вигляді таблиці.

2. За допомогою одного із інструментів (розглянутого при виконанні першого завдання) здійсніть візуалізацію результатів біржових торгів на одній із світових бірж скориставшись загальною доступною інформацією.

3. За допомогою одного із інструментів (розглянутого при виконанні першого завдання) здійсніть візуалізацію структури населення України за віковими групами.

4. За допомогою одного із інструментів (розглянутого при виконанні першого завдання) здійсніть візуалізацію динаміки курсу долара.

5. За допомогою одного із інструментів (розглянутого при виконанні першого завдання) здійсніть візуалізацію податкових надходжень по регіонах України.

6. Завдання 2.3. Передивитись відео за посиланням [Data Visualization: Then and Now https://www.youtube.com/watch?v=HB9-lbnsHaA](https://www.youtube.com/watch?v=HB9-lbnsHaA) та скласти на ним 15 тестів українською мовою з 5-ма варіантами відповідей. Надіслати файл з тестами. Обов'язково вказати правильні відповіді до тестів. Виконане завдання завантажити на Освітній портал.

7. Публікація в Tableau Public

Безкоштовна версія: Tableau Public

Розвиток: Облако, SaaS, Інтернет; Установлено - Mac / Windows;

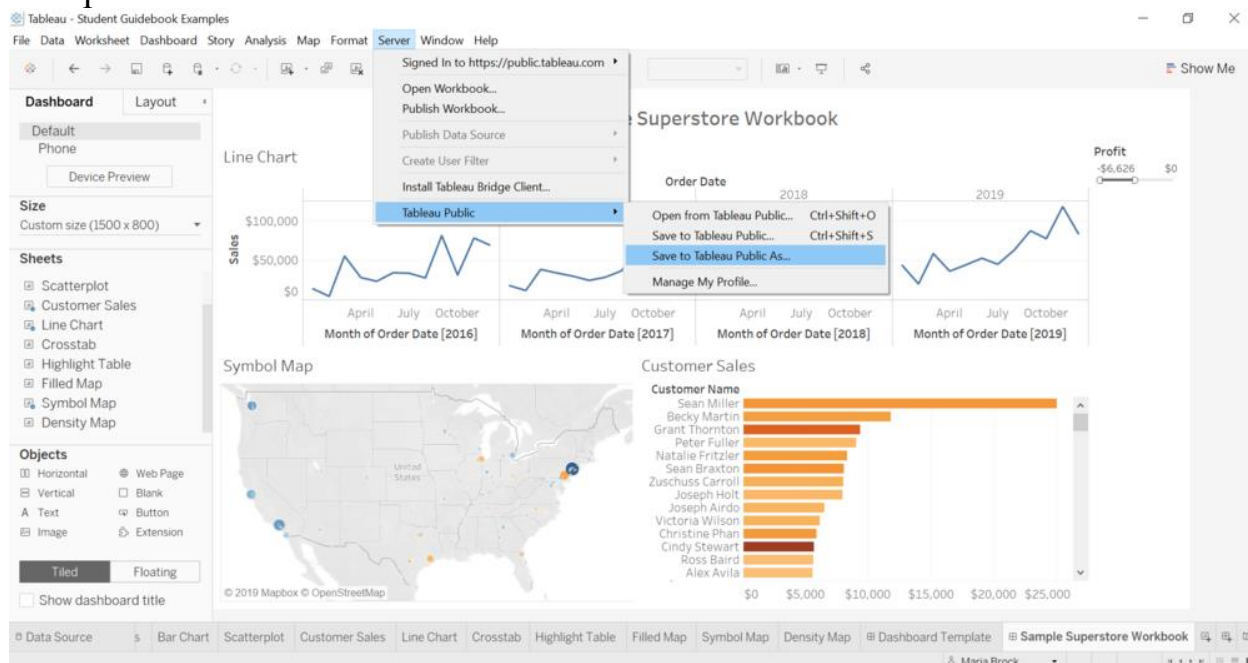
Житомирська політехніка	МІНІСТЕРСТВО ОСВІТИ І НАУКИ УКРАЇНИ ДЕРЖАВНИЙ УНІВЕРСИТЕТ «ЖИТОМИРСЬКА ПОЛІТЕХНІКА» Система управління якістю відповідає ДСТУ ISO 9001:2015	Ф-19.05-05.02/072.00.1/М/ ОК6-2023
	Екземпляр № 1	Арк 38 / 6

Мобільний - Android Native / iOS Native; Особливості ВІ: спеціальний аналіз; Спеціальний запит; Спеціальні звіти; Порівняльний аналіз; Бюджетування і прогнозування; Приборна дошка; Аналіз даних; Візуалізація даних; Ключові показники ефективності; OLAP; Показники ефективності; Прогнозна аналітика; Аналіз рентабельності; Стратегічне планування; Індикатори тренду / проблеми.

Як опублікувати вашу візуалізацію з Tableau Desktop у Tableau Public.

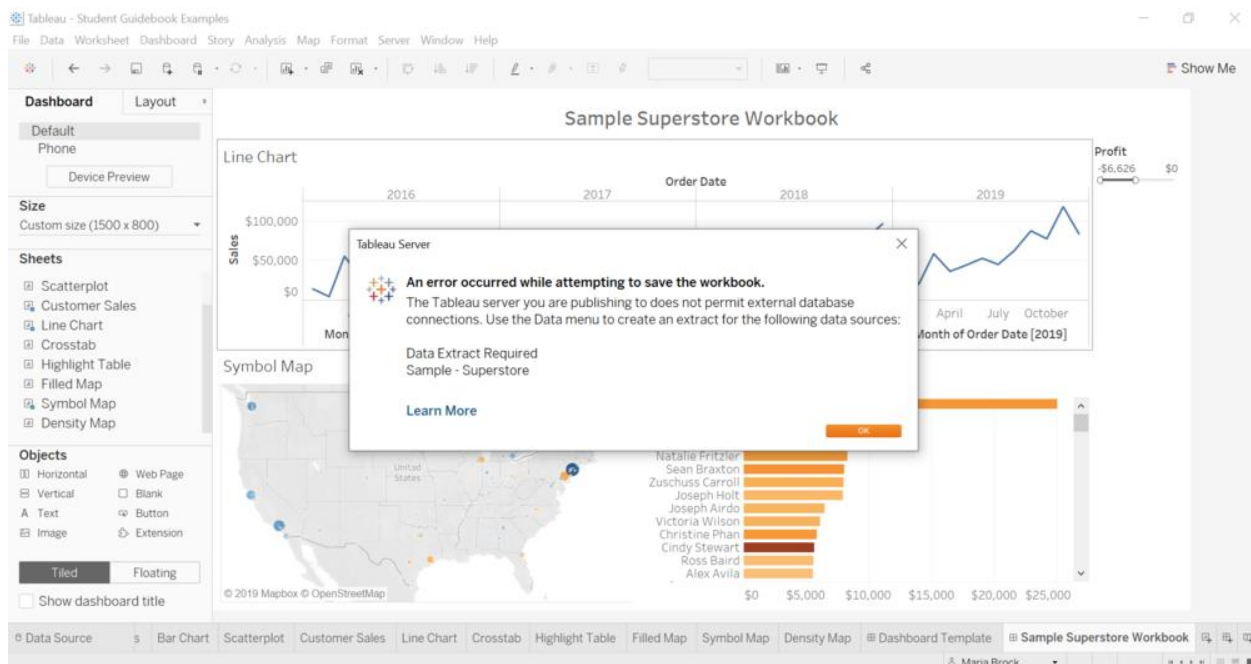
Після цього, як ви створили свою інформаційну панель, виконати наступні дії: - Сервер, Tableau Public, Сохранить в Tableau Public як

* Можливо, спочатку вам запропонують зайти у свій профіль у Tableau Таблица робить аналіз даних інтуїтивно зрозумілим та простим у використанні.



Виберіть назву для своєї візуалізації та натисніть «Зберегти». Далі ви можете переглянути повідомлення у всьому цьому:

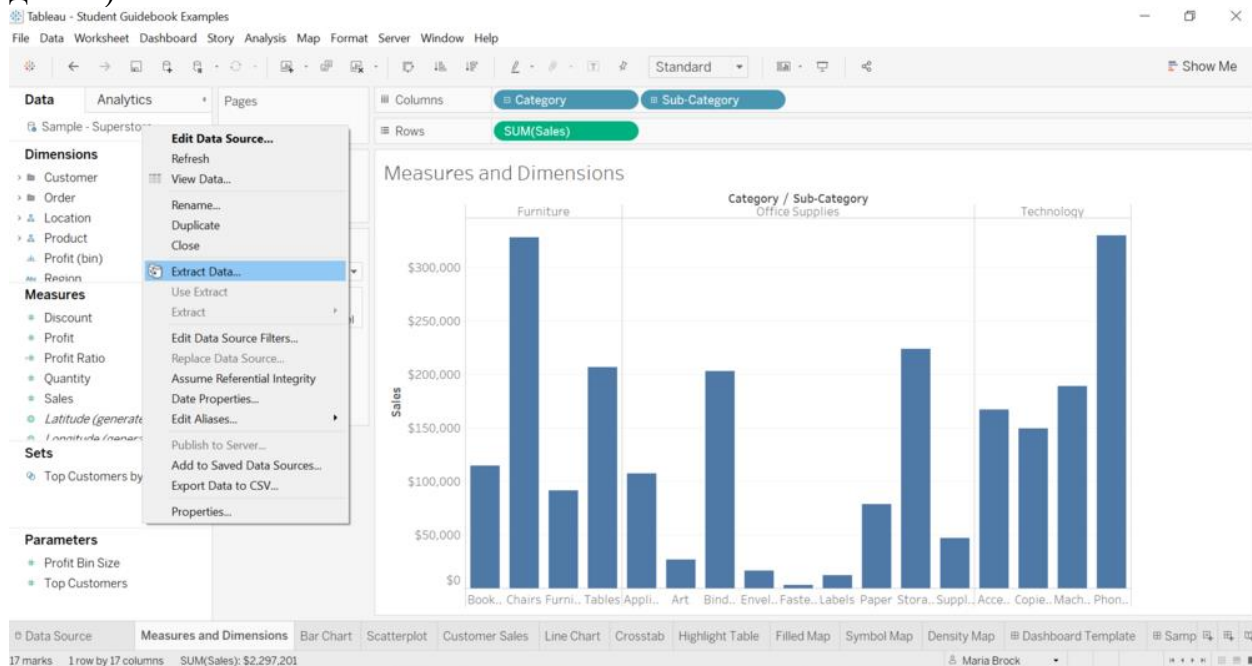
Житомирська політехніка	МІНІСТЕРСТВО ОСВІТИ І НАУКИ УКРАЇНИ ДЕРЖАВНИЙ УНІВЕРСИТЕТ «ЖИТОМИРСЬКА ПОЛІТЕХНІКА» Система управління якістю відповідас ДСТУ ISO 9001:2015	Ф-19.05-05.02/072.00.1/М/ ОК6-2023
	Екземпляр № 1	Арк 38 / 7



Це повідомлення означає, що ваше з'єднання з набором даних Sample-Superstore тепер встановлено. Tableau Public не може розміщувати живі з'єднання, тому вам потрібно змінити ваше з'єднання у вибір (наприклад, скриншот екрану ваших даних).

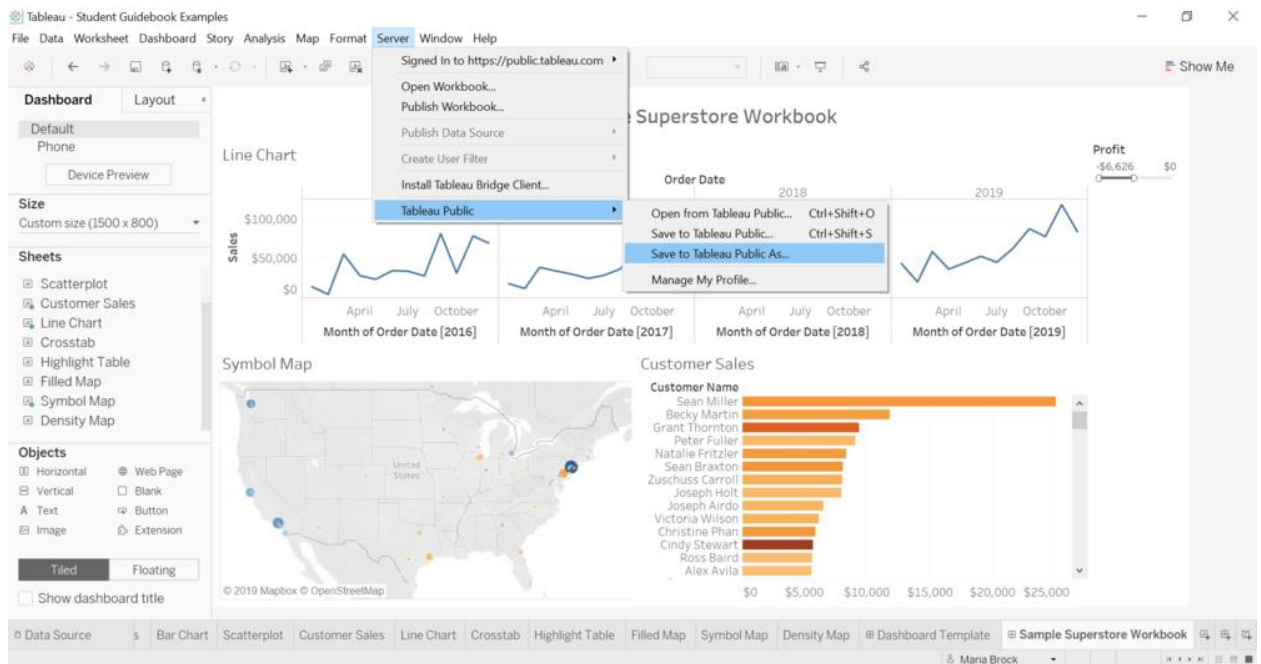
Перейдіть до одного з ваших робочих листів та виконайте наступні дії:

-Ще раз натисніть праву кнопку «Sample-Superstore» (Вибірка-Супермагазин) у верхній частині екрана та виберіть «Extract Data» (Видалення даних).



Натисніть кнопку «Витяг» (Видалення) та перегляньте інформаційну панель, щоб повторити публікацію. Коли закінчите, натисніть «Зберегти», ви потрапите на сторінку Tableau Public.

Житомирська політехніка	МІНІСТЕРСТВО ОСВІТИ І НАУКИ УКРАЇНИ ДЕРЖАВНИЙ УНІВЕРСИТЕТ «ЖИТОМИРСЬКА ПОЛІТЕХНІКА» Система управління якістю відповідає ДСТУ ISO 9001:2015	Ф-19.05-05.02/072.00.1/М/ ОК6-2023
	Екземпляр № 1	Арк 38 / 8



Тут ви бачите, що інформаційна панель вже опублікована. Вона неактивна, тому що автоматично встановлює свої візуалізації у «приховані» (скриті). Ви можете зробити так, або немає, перейти на сторінку настрою. Ми зробили це тому, що якщо візуалізація не буде скритою, вона з'явиться у стрічці діяльності ваших підписників, навіть якщо ви видалите її привітання. Якщо ви налаштували автоматичне вкриття, щоб побачити, як його відредагувати, перегляньте його в браузері, а потім, коли ви будете готові, ви можете показати його, натиснувши на значок у маленькому оці.

Тема 3. Поняття ринку великих даних. Життєвий цикл аналітики даних. Збір та підготовка даних

Питання для обговорення та дискусії

1. Ринок великих даних: переваги, недоліки та ризики
2. Перспективи вітчизняних компаній на ринку великих даних
3. Технології збору великих даних: їх технічна та етична сторона
4. Технології великих даних та приватність інформації – де межа?
5. Джерела великих даних та їх різноманітність.
6. Відкритість даних та державна таємниця
7. Аналіз п'яти сил конкуренції Портера на ринку Big Data.
8. Матриця ризиків суб'єктів ринку великих даних

Питання для контролю:

1. Які країни займають провідне місце на ринку великих даних?
2. Які дані займають найбільшу частку на ринку BigData?

Житомирська політехніка	МІНІСТЕРСТВО ОСВІТИ І НАУКИ УКРАЇНИ ДЕРЖАВНИЙ УНІВЕРСИТЕТ «ЖИТОМИРСЬКА ПОЛІТЕХНІКА» Система управління якістю відповідає ДСТУ ISO 9001:2015	Ф-19.05-05.02/2/072.00.1/М/ ОК6-2023
	Екземпляр № 1	Арк 38 / 9

3. Які елементи життєвого циклу великих даних вам відомі?
4. Охарактеризуйте етап життєвого циклу великих даних «оцінювання бізнес ситуації»
5. Охарактеризуйте етап життєвого циклу великих даних «ідентифікації даних»
6. Охарактеризуйте етап життєвого циклу великих даних «збір та фільтрація даних»
7. Охарактеризуйте етап життєвого циклу великих даних «виокремлювання даних»
8. Охарактеризуйте етап життєвого циклу великих даних «перевірка та очищення даних»
9. Охарактеризуйте етап життєвого циклу великих даних «агрегування та подання даних»
10. Які способи очищення та перевірки даних вам відомі?
11. Чи є абсолютно відкритими для аналізу усі дані фірми?
12. Яких ви знаєте вітчизняних постачальників великих даних?
13. Які великі дані в Україні є у загальному доступі?
14. Як убезпечитись від недостовірних чи спотворених даних?
15. Для чого проводять агрегування даних?
16. Чи доречним є зберігання даних які відсіяні підчас перевірки для виконання конкретного проекту?

Завдання до виконання

1. Проведіть SWOT-аналіз технологій BigData та зробіть відповідні висновки.
2. Доповніть таблицю постачальників відкритих даних.

Сайт з відкритими даними	Опис
Data.gov	Центр відкритих даних уряду США
https://open-data.europa.eu/	Центр відкритих даних Європейської комісії
Data.worldbank.org	Проект відкритих даних всесвітнього банку
Data.gov.ua	Портал відкритих даних Міністерства цифрової трансформації України

3. Користуючись Порталом відкритих даних Міністерства цифрової трансформації України та інших відкритих джерел зберіть максимальну інформацію про членів своєї родини.
4. Яку інформацію про себе ви можете зібрати в мережі з відкритих та конфіденційних джерел. Зберіть дану інформацію та порівняйте її з результатами отриманими при виконанні попереднього завдання.
5. Користуючись доступними сторінками соціальних мереж складіть портрети своїх 2-3 товаришів користуючись виключно інформацією

Житомирська політехніка	МІНІСТЕРСТВО ОСВІТИ І НАУКИ УКРАЇНИ ДЕРЖАВНИЙ УНІВЕРСИТЕТ «ЖИТОМИРСЬКА ПОЛІТЕХНІКА» Система управління якістю відповідає ДСТУ ISO 9001:2015	Ф-19.05-05.02/2/072.00.1/М/ ОК6-2023
	Екземпляр № 1	Арк 38 / 10

отриманою таким шляхом. Зробіть висновок чи доречною та безпечною є інформація поширена ними.

6. Порівняйте відкритість інформації в Україні та інших країнах. Результат оформіть у вигляді таблиці

Тема 4. Методи та інструменти аналізу великих даних

Питання для обговорення та дискусії

1. Типи завдань, пов'язаних з великими даними.
2. В чому полягає сутність методик і методів аналізу великих даних згідно звіту McKinsey? Дайте коротку характеристику кожного з них.
3. Опишіть методи інтелектуального аналізу даних (Data Mining) та які задачі вони розв'язують?
4. Аналіз часових рядів та його основні характеристики.
5. Основні методи технології Text Mining
6. Яким чином проводиться А/В тестування?
7. Технологія MapReduce.
8. Платформа Hadoop.
9. Аналіз програмного забезпечення, що використовується при роботі з Big Data у бізнес середовищі.
10. Розподілена файлова система.
11. Фреймворк Apache Spark.
12. Основні властивості та переваги платформ великих даних.
13. Постачальники та програмне забезпечення великих даних

Питання для контролю:

1. Що означає акронім МАД?
2. Які існують групи методів аналітики Великих даних?
3. Що таке асоціативні правила (*Association Rule Learning*)?
4. В чому полягає метод дерев рішень (*Decision Trees*)?
5. Для чого використовується кластерний аналіз (*Cluster Analysis*)?
6. Що таке еволюційні (генетичні) алгоритми?

Завдання до виконання

1. Ознайомитись з архітектурою Hadoop

Що нам дає архітектура Hadoop на прикладі референсної архітектури розміщення кластера в ЦОД від Cloudera (рис. 4.1).

У кластері присутні вузли різного призначення:

Name Node - це «Master» вузол кластера, найбільш критичний вузол і, за сумісництвом, єдина точка відмови, зберігає всю мета-інформацію про файлову систему кластера і підтримує її працездатність. Немає Name Node -

Житомирська політехніка	МІНІСТЕРСТВО ОСВІТИ І НАУКИ УКРАЇНИ ДЕРЖАВНИЙ УНІВЕРСИТЕТ «ЖИТОМИРСЬКА ПОЛІТЕХНІКА» Система управління якістю відповідає ДСТУ ISO 9001:2015	Ф-19.05-05.02/2/072.00.1/М/ ОК6-2023
	Екземпляр № 1	Арк 38 / 11

кластер перетворюється в купу непотрібного заліза. На ньому, напевно, можна буде Майні криптовалюта, але для цього є більш ефективні рішення. Зменшити ризики, пов'язані з його втратою, допоможе Standby Name Node. Природно, розташовуватися вони повинні в різних стійках і бути підключені до різних комутаторів.

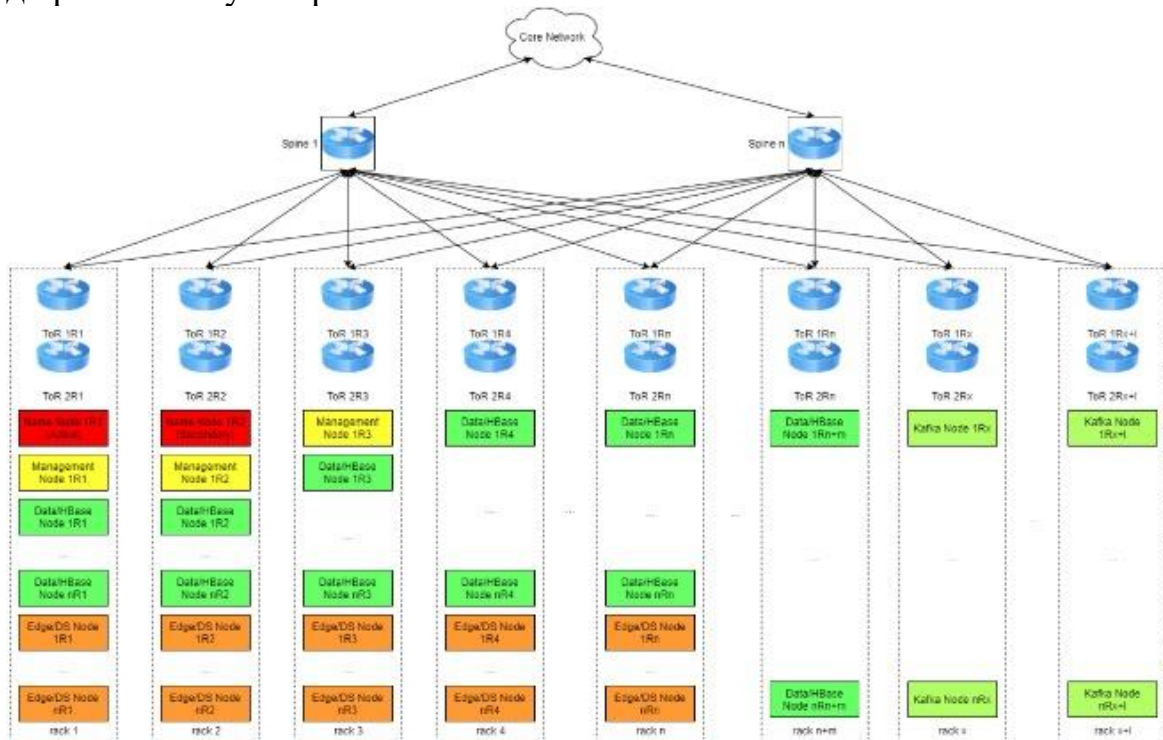


Рис. 4.1. Архітектура розгортання вузлів кластера Hadoop в ЦОД.

Management Node - це вузли, що дозволяють працювати базовим сервісів кластера також в відмовостійкості режимі. Таких вузлів може бути досить багато, відмова одного з них не вплине сильно на роботу кластера.

Data / HBase Node - це вузли, на яких вже безпосередньо зберігаються і обробляються дані. Це найчисленніший тип вузлів кластера. Вони рівномірно «розмазуються» по стійках і генерують високий трафік (особливо в процесах реплікації).

Edge / DS Node - строго кажучи, не члени кластера. Ці вузли потрібні для завантаження даних в кластер / вивантаження з кластера, а також для обробки даних у випадках, коли паралельна обробка на кластері неможлива. У ряді випадків ці вузли містять GPU для виконання завдань навчання моделей Data Science, на них розгортається ETL / ELT інструментарій, сервера додатків, моніторингу, перевірки якості і т.д. На цих вузлах працюють і користувачі.

Kafka Node - вузли розподіленого брокера повідомлень Kafka, використовуються в задачах потокової обробки з архітектурою Pub / Sub. Вузли також рівномірно «розмазуються» по стійках.

Крім того, в кожній стійці є дубльовані ToR-комутатори, а кожен вузол

Житомирська політехніка	МІНІСТЕРСТВО ОСВІТИ І НАУКИ УКРАЇНИ ДЕРЖАВНИЙ УНІВЕРСИТЕТ «ЖИТОМИРСЬКА ПОЛІТЕХНІКА» Система управління якістю відповідає ДСТУ ISO 9001:2015	Ф-19.05-05.02/2/072.00.1/М/ ОК6-2023
	Екземпляр № 1	Арк 38 / 12

кластера по 2x10Gb інтерфейсів з'єднаний з обома комутаторами. У свою чергу, ToR комутатори з'єднуються з комутаторами рівня Spine і найбільш інтенсивний трафік, як видно зі схеми, переміщується у напрямку «Схід-Захід», тобто не покидає межі ЦОД-а. На рівні вище знаходяться комутатори ядра мережі.

Така архітектура допомагає нам виконати вимоги:

- Висока доступність
- Відмовостійкість
- Надійність
- Горизонтальна масштабованість
- Висока продуктивність.

2. Ознайомитись з фреймворком Apache Spark

Apache Spark - це Big Data фреймворк з відкритим вихідним кодом для розподіленої пакетної і потокової обробки неструктурованих і слабоструктурованих даних, що входить в екосистему проектів Hadoop.

Класичний MapReduce, Apache компонент Hadoop для обробки даних, проводить обчислення в два етапи:

1. Map, коли головний вузол кластера (master) розподіляє завдання по робочих вузлів (node).

2. Reduce, коли дані згортаються і передаються назад на головний вузол, формуючи остаточний результат обчислень.

Поки всі процеси етапу Map не закінчаться, процеси Reduce не почнуться. При цьому всі операції проходять по циклу читання-запис з жорсткого диска. Це обумовлює затримки в обробці інформації. Таким чином, технологія MapReduce добре підходить для задач розподілених обчислень в пакетному режимі, але через затримки (latency) не може використовуватися для потокової обробки в режимі реального часу. Для вирішення цієї проблеми було створено Apache Spark і інші Big Data фреймворки розподіленої потокової обробки (Storm, Samza, Flink).

На відміну від класичного обробника ядра Apache Hadoop с дворівневої концепцією MapReduce на базі дискового сховища, Spark використовує спеціалізовані примітиви для рекуррентної обробки в оперативній пам'яті. Завдяки цьому багато обчислювальні завдання реалізуються в Спарк значно швидше. Наприклад, можливість багаторазового доступу до завантажених в пам'ять призначених для користувача даних дозволяє ефективно працювати з алгоритмами машинного навчання (Machine Learning) (рис. 4.2).

Hadoop MapReduce	Apache Spark
Fast	100x faster than MapReduce
Batch Processing	Real-time Processing
Stores Data on Disk	Stores Data in Memory
Written in Java	Written in Scala

Рис. 4.2. Порівняння Apache Hadoop і Spark

Як влаштований Apache Spark: архітектура і принцип роботи Спарк складається з наступних компонентів:

Ядро (Core);

SQL - інструмент для аналітичної обробки даних за допомогою SQL-запитів;

Streaming - надбудова для обробки поточкових даних, про яку докладно ми розповідали тут і тут;

MLlib - набір бібліотек машинного навчання;

GraphX - модуль розподіленої обробки графів.

Spark може працювати як в середовищі кластера Hadoop під керуванням YARN, так і без компонентів ядра хаду, наприклад, на базі системи управління кластером Mesos. Спарк підтримує кілька популярних розподілених систем зберігання даних (HDFS, OpenStack Swift, Cassandra, Amazon S3) і мов програмування (Java, Scala, Python, R), надаючи для них API-інтерфейси (рис. 4.3).

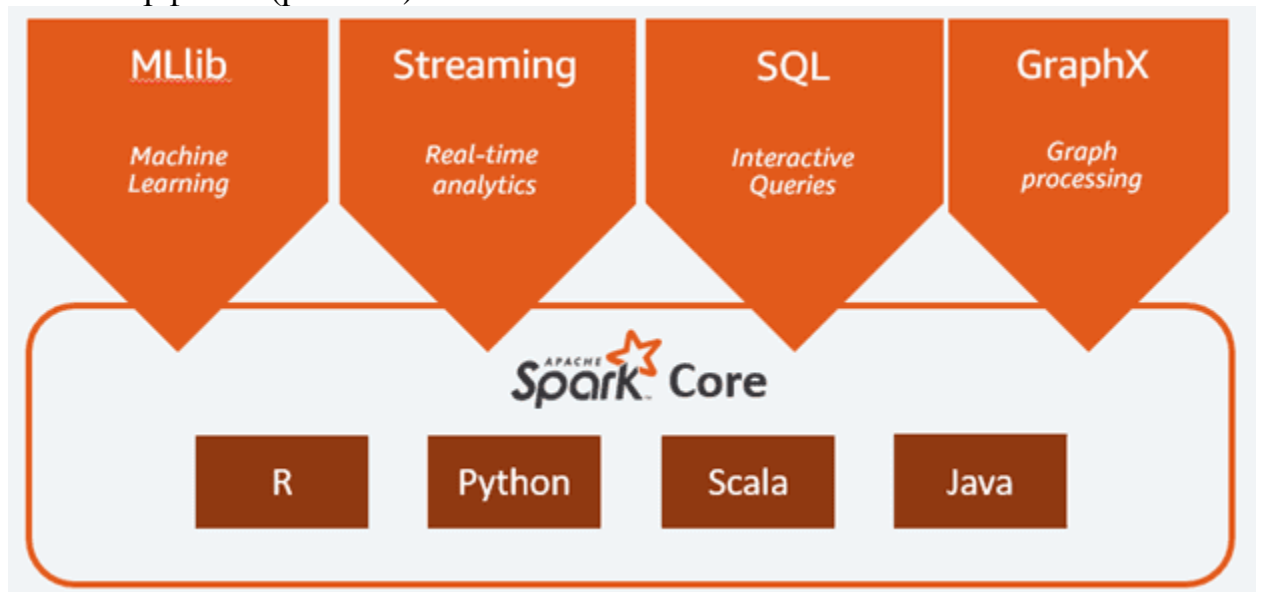


Рис. 4.3. Компоненти Спарк

Справедливості заради варто відзначити, що Spark Streaming, на відміну від, наприклад, Apache Storm, Flink або Samza, і не виконує жодних потоки Big

Житомирська політехніка	МІНІСТЕРСТВО ОСВІТИ І НАУКИ УКРАЇНИ ДЕРЖАВНИЙ УНІВЕРСИТЕТ «ЖИТОМИРСЬКА ПОЛІТЕХНІКА» Система управління якістю відповідає ДСТУ ISO 9001:2015	Ф-19.05-05.02/2/072.00.1/М/ ОК6-2023
	Екземпляр № 1	Арк 38 / 14

Data цілком. Замість цього реалізується мікропакетний підхід (micro-batch), коли потік даних розбивається на невеликі пакети тимчасових інтервалів. Абстракція Spark для потоку називається DStream (discretized stream, Дискретизований потік) і являє собою мікро-пакет, що містить кілька відмовостійких розподілених датасета, RDD (resilient distributed dataset).

Саме RDD є основним обчислювальним примітивом Спарк, над яким можна робити паралельні обчислення і перетворення за допомогою вбудованих і довільних функцій, в тому числі за допомогою тимчасових вікон (window-based operations). Детальніше про тимчасові вікна ми розповідали тут на прикладі Apache Kafka Streams (рис. 4.4.).

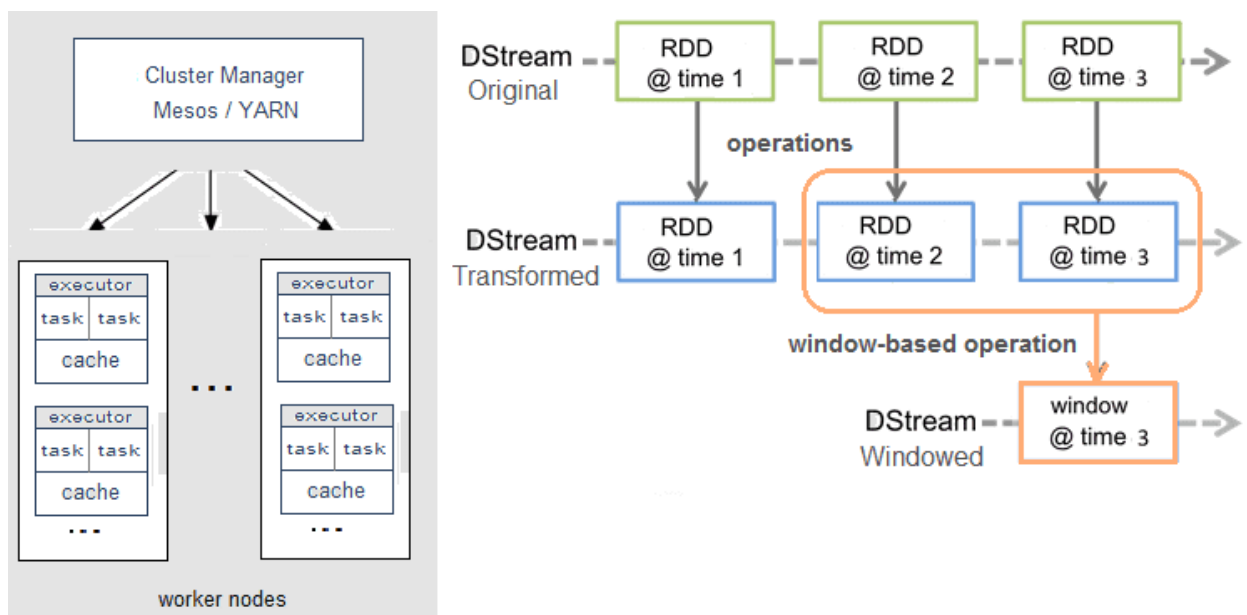


Рис. 4.4.

Принцип роботи спарк де і як використовується apache spark.

Завдяки наявності різнопрофільних інструментів для аналітичної обробки даних «на льоту» (SQL, Streaming, MLLib, GraphX), Спарк активно використовується в системах інтернету речей (Internet of Things, IoT) на стороні IoT-платформ, а також в різних бізнес-додатках, в т.ч. на базі методів Machine Learning. Наприклад, Спарк застосовується для прогнозування відтоку клієнтів (Churn Predict) і оцінки фінансових ризиків. Однак, якщо тимчасова затримка обробки даних (latency) - це критичний чинник, Apache Spark не підійде і варто розглянути альтернативу у вигляді клієнтської бібліотеки Kafka Streams або фреймворків Storm, Flink, Samza.

По набору компонентів і функціональними можливостями Spark можна порівняти з іншим Big Data інструментом розподіленої потокової обробки - Apache Flink.

Житомирська політехніка	МІНІСТЕРСТВО ОСВІТИ І НАУКИ УКРАЇНИ ДЕРЖАВНИЙ УНІВЕРСИТЕТ «ЖИТОМИРСЬКА ПОЛІТЕХНІКА» Система управління якістю відповідає ДСТУ ISO 9001:2015	Ф-19.05-05.02/2/072.00.1/М/ ОК6-2023
	Екземпляр № 1	Арк 38 / 15

Тема 5. Методи моделювання та прогнозування економічного розвитку

Питання для обговорення та дискусії

1. Чим відрізняються гіпотеза, прогноз, план та програма
2. Часовий ряд та його компоненти
3. Основні характеристики ряду динаміки.
4. Прості екстраполяційні методи прогнозування
5. Кореляційно-регресійні методи та моделі.
6. ANOVA-таблиця для регресійної моделі

Питання для контролю:

1. Дайте визначення основним поняттям: модель та метод прогнозування, верифікація прогнозу, прогнозний фон.
2. Що таке кореляційний аналіз?
3. Що таке регресійний аналіз?
4. Що відображає коефіцієнт кореляції Пірсона? Які його властивості?
5. Коефіцієнт детермінації та що він відображає? Які його властивості?

Завдання до виконання

1. Показники аналізу рядів динаміки

⇒ за даними 10 років, наведеними в Excel таблиці «Обсяги реалізованої продукції (товарів, послуг) підприємств за видами економічної діяльності в Україні», що розміщена на освітньому порталі, визначити базисні та ланцюгові характеристики динаміки: абсолютні прирости, темпи зростання і приросту, абсолютні значення 1 % приросту;

⇒ обчислити середньорічні: абсолютний приріст, темп зростання та темп приросту;

⇒ зробити відповідні висновки.

Вимоги до виконання завдання:

- виконати розрахунки завдання та звести у формі таблиці – приклад форми представлення інформації наведений у таблиці 5.1 (назву галузі, кожен студент заміняє на свою):

Таблиця 5.1

Аналіз динаміки обсягів реалізованої продукції підприємств **сільськогосподарського та лісового та рибного господарств** України у 2010–2019 роках, тис. грн.

Роки	Обсяги реалізованої продукції	Абсолютний приріст, тис. грн.		Темп зростання, %		Темп приросту, %		Значення 1 % приросту
		базисний	ланцюговий	базисний	ланцюговий	базисний	ланцюговий	
2010	99 891 436,0							
2011	126 961 245,2							

Житомирська політехніка	МІНІСТЕРСТВО ОСВІТИ І НАУКИ УКРАЇНИ ДЕРЖАВНИЙ УНІВЕРСИТЕТ «ЖИТОМИРСЬКА ПОЛІТЕХНІКА» Система управління якістю відповідає ДСТУ ISO 9001:2015	Ф-19.05-05.02/2/072.00.1/М/ ОК6-2023
	Екземпляр № 1	Арк 38 / 16

2012	162 611 131,2							
2013	161 130 336,0							
2014	213 929 847,2							
2015	362 309 989,7							
2016	403 645 802,4							
2017	454 380 102,6							
2018	525 096 891,4							
2019	556 325 872,1							
Разом:								

- висновки до таблиці є обов'язковими (формули для розрахунків та приклад висновків наведені нижче);

- зміст висновку має включати лише положення, які спираються на цифровий матеріал таблиці (застосування абстрактних або загальних положень, що не підкріплені цифровим матеріалом, не допускається).

Кожен студент виконує завдання за окремою галуззю.

Методичні рекомендації до виконання завдання 5.1.

За даними, наведеними в табл. 5.2:

⇒ визначити базисні та ланцюгові характеристики динаміки: абсолютні прирости, темпи зростання і приросту, абсолютні значення 1 % приросту;

⇒ обчислити середньо кварталні: абсолютний приріст, темп зростання та темп приросту;

⇒ розрахунки представити у вигляді таблиці 3 та зробити відповідні висновки.

Таблиця 5.2

Дані ТОВ «Клен», тис. грн.

2021 рік	Обсяги реалізації
I-й квартал	356 900
II-й квартал	379 800
III-й квартал	450 200
IV-й квартал	356 890

Таблиця 5.3

Аналіз динаміки обсягів реалізації продукції ТОВ «Клен» за 2022 рік

Період	Обсяг реалізації, грн.	Абсолютний приріст, грн.		Темп зростання, %		Темп приросту, %		Значення 1 % приросту
		базисний	ланцюговий	базисний	ланцюговий	базисний	ланцюговий	
I кв-л	356 900							
II кв-л	379 800							
III кв-л	450 200							
IV кв-л	356 890							

Житомирська політехніка	МІНІСТЕРСТВО ОСВІТИ І НАУКИ УКРАЇНИ ДЕРЖАВНИЙ УНІВЕРСИТЕТ «ЖИТОМИРСЬКА ПОЛІТЕХНІКА» Система управління якістю відповідає ДСТУ ISO 9001:2015	Ф-19.05-05.02/2/072.00.1/М/ ОК6-2023
	Екземпляр № 1	Арк 38 / 17

Динамічним рядом (рядом динаміки) називають ряд статистичних показників, що розташовані в хронологічній послідовності і характеризують зміну явища в часі.

Абсолютний приріст (Δy_i) розраховується як різниця між двома рівнями динамічного ряду. Він показує на скільки одиниць в абсолютному вираженні рівень одного періоду більший або менший від певного попереднього рівня, а тому може мати знак “+” (при збільшенні) або “-” (при зменшенні). В аналітичних розрахунках використовують ланцюгові та базисні абсолютні прирости:

⇒ ланцюговий – характеризує приріст значення звітної періоду (y_i) від попереднього (y_{i-1}):

$$\Delta y_i = y_i - y_{i-1} \quad (1)$$

⇒ базисний – характеризує приріст значення звітної періоду від значення, взятого за базу порівняння (y_0).

$$\Delta y_i = y_i - y_0. \quad (2)$$

За базовий рівень (тобто той рівень, з яким проводиться порівняння) залежно від завдання дослідження може прийматися певний постійний для всіх рівень (наприклад, перший рівень ряду) або для кожного наступного рівня попередній йому.

Темп зростання (t_i) – відносний показник, який розраховується як частка від ділення двох рівнів динамічного ряду і показує у скільки разів показник, що аналізується, більший або менший за базисний. Темп зростання (t_i), виражений у відсотках, показує скільки відсотків становить показник, що аналізується, щодо базисного, прийнятого за 100%. Темп зростання розраховується наступним чином:

⇒ ланцюговий:

$$t_i = \frac{y_i}{y_{i-1}} \times 100 \% \quad (3)$$

⇒ базисний:

$$t_i = \frac{y_i}{y_0} \times 100 \% \quad (4)$$

Таким чином, у першому випадку отримують темпи зростання з постійною базою (базисні), у другому випадку – із змінною базою (ланцюгові).

Темп приросту (T) – відносний показник, який показує на скільки відсотків один рівень більший (або менший) базового рівня. Він може бути розрахований трьома способами:

1) відношення абсолютного приросту до попереднього чи базисного рівня.

⇒ ланцюговий:

Житомирська політехніка	МІНІСТЕРСТВО ОСВІТИ І НАУКИ УКРАЇНИ ДЕРЖАВНИЙ УНІВЕРСИТЕТ «ЖИТОМИРСЬКА ПОЛІТЕХНІКА» Система управління якістю відповідає ДСТУ ISO 9001:2015	Ф-19.05-05.02/072.00.1/М/ ОК6-2023
	Екземпляр № 1	Арк 38 / 18

$$T = \frac{\Delta y_i}{y_{i-1}} \times 100\% \quad (5)$$

⇒ базисний:

$$T = \frac{\Delta y_i}{y_0} \times 100\% \quad (6)$$

2) віднімання 100 % від темпу зростання;

$$T = t_i - 100\% \quad (7)$$

3) процентне відношення різниці між двома рівнями ряду до того базового рівня, у порівнянні з яким абсолютний приріст розрахований:

⇒ ланцюговий:

$$T = \frac{y_i - y_{i-1}}{y_{i-1}} \times 100\% \quad (8)$$

⇒ базисний:

$$T = \frac{y_i - y_0}{y_0} \times 100\% \quad (9)$$

Так як кожному відносному показнику відповідають певні абсолютні величини, то при вивченні темпів приросту необхідно враховувати, яка абсолютна величина відповідає кожному відсотку приросту, який його зміст. Для цього розраховують абсолютне значення 1 % приросту (А %) як співвідношення абсолютного приросту за певний період і темпу приросту у відсотках за той самий період. Алгебраїчно це співвідношення дорівнює 0,01 рівня, прийнятого за базу порівняння:

$$A\% = \frac{y_i - y_{i-1}}{100 \left(\frac{y_i - y_{i-1}}{y_{i-1}} \right)} = \frac{y_{i-1}}{100} \quad (10)$$

Середній абсолютний приріст розраховують як середню арифметичну просту з ланцюгових абсолютних приростів:

$$\bar{\Delta} = \frac{\sum_{k=1}^k \Delta_k}{k}, \text{ де } \Delta_k \text{ – ланцюгові прирости} \quad (11)$$

або

$$\bar{\Delta} = \frac{y_n - y_1}{n - 1}$$

Середній темп зростання розраховують за формулою середньої геометричної.

$$\bar{t} = \sqrt[n-1]{t_1 \times t_2 \times \dots \times t_n} \cdot 100 \quad (12)$$

або

Житомирська політехніка	МІНІСТЕРСТВО ОСВІТИ І НАУКИ УКРАЇНИ ДЕРЖАВНИЙ УНІВЕРСИТЕТ «ЖИТОМИРСЬКА ПОЛІТЕХНІКА» Система управління якістю відповідає ДСТУ ISO 9001:2015	Ф-19.05-05.02/2/072.00.1/М/ ОК6-2023
	Екземпляр № 1	Арк 38 / 19

$$\bar{t} = n-1 \sqrt[n]{\frac{y_n}{y_1}} 100$$

Середній темп приросту:

$$t_{np} = \bar{t} - 1 \quad (\text{або } 100\%) \quad (13)$$

Таблиця 5.3

Аналіз динаміки обсягів реалізації продукції ТОВ «Клен» за 2021 рік

Період	Обсяг реалізації, грн.	Абсолютний приріст, грн.		Темп зростання, %		Темп приросту, %		Значення 1 % приросту
		базисний	ланцюговий	базисний	ланцюговий	базисний	ланцюговий	
I кв-л	356 900	–	–	–	–	–	–	–
II кв-л	379 800	22 900	22 900	106,416	106,416	6,416	6,416	3 569
III кв-л	450 200	93 300	70 400	126,142	118,536	26,142	18,536	3 798
IV кв-л	356 890	-10	-93 310	99,997	79,274	-0,003	-20,726	4 502

Середньо кварталний абсолютний приріст

$$\bar{\Delta} = \frac{y_n - y_1}{n - 1} = \frac{356890 - 356900}{4 - 1} = -3,33 \text{ тис.грн.}$$

Середній темп зростання

$$\bar{t} = n-1 \sqrt[n]{\frac{y_n}{y_1}} 100 = 4-1 \sqrt[4]{\frac{356890}{356900}} \times 100 = 99,999066\%$$

Середній темп приросту:

$$t_{np} = \bar{t} - 100 = 99,999066 - 100 = -0,000934\%$$

Висновки.

За результатами дослідження спостерігається тенденція до збільшення обсягів реалізації ТОВ «Клен». Лише в IV-му кварталі в порівнянні з I-м обсяги реалізації зменшилися на 10 грн. або на 0,003 %. Проте в порівнянні з попереднім (III-м) кварталом відбулося суттєве зменшення, яке склало 20,73 %, що в абсолютному виразі – 93310 грн.

Певне зменшення обсягів реалізації в грошовому виразі може бути викликане зменшенням обсягів реалізації натурі, сезонними коливаннями, зменшенням ціни. Зауважимо, що чим більші обсяги реалізації продукції, тим більше доходу отримує підприємство, а за умови, що продукція рентабельна – більше прибутку.

Найбільші зміни відбулися:

Житомирська політехніка	МІНІСТЕРСТВО ОСВІТИ І НАУКИ УКРАЇНИ ДЕРЖАВНИЙ УНІВЕРСИТЕТ «ЖИТОМИРСЬКА ПОЛІТЕХНІКА» Система управління якістю відповідає ДСТУ ISO 9001:2015	Ф-19.05-05.02/2/072.00.1/М/ ОК6-2023
	Екземпляр № 1	Арк 38 / 20

– за базисним способом розрахунку в III-му кварталі в порівнянні з I-м кварталом обсяги реалізації збільшилися на 93300 грн. і вклали 126,14 % обсягу реалізації I-го кварталу;

– за ланцюговим способом розрахунку в III-му кварталі в порівнянні з II-м обсяг реалізації збільшився на 70400 грн., що склало 118,54 % попередніх обсягів реалізації.

2. Регресійні моделі в прогнозуванні

Завдання. Побудувати економетричну лінійну множинну модель продуктивності праці, що описує залежність між продуктивністю праці (Y), фондомісткістю продукції (X_1) та коефіцієнтом плинності робочої сили (X_2). Оцінити якість, адекватність і статистичну значимість побудованої моделі для рівня значимості $\alpha=0,05$. Розрахувати точковий та інтервальний прогнози продуктивності праці на наступний місяць з рівнем надійності $P=0,95$, якщо очікувані значення чинників, впливають на неї дорівнюють:

- фондомісткість продукції – 45 тис. грн. / тис. грн.;
- коефіцієнт плинності робочої сили – 9%;

Оцінити граничний вплив кожної пояснюючої змінної моделі на продуктивність праці.

Написати висновки за зразком.

Варіант 1

Y	38,9	39,8	40,2	41,4	41,7	43,4	44,3	46,7	47,3	49,0	49,3	51,2	51,3	51,6	51,7
X_1	15	16	17	18	18	19	20	20	20	21	21	21	21	22	23
X_2	5,2	5,3	4,8	5,6	5,6	5,6	4,9	6,0	6,1	6,0	6,1	6,5	6,6	6,9	6,4

Варіант 2

Y	31,1	32,2	32,3	32,4	32,5	32,7	32,8	33,4	33,6	33,6	33,9	34,4	34,9	35,1	36,2
X_1	15	16	17	18	18	19	20	20	20	21	21	21	21	22	23
X_2	4,7	5,7	5,7	5,2	5,0	5,2	5,2	5,3	5,7	5,9	6,0	6,6	7,6	7,2	7,3

Варіант 3

Y	31,7	31,8	31,9	32,1	32,5	32,7	33,0	41,7	41,9	42,0	42,1	52,5	53,6	54,6	55,6
X_1	23	24	24	26	26	27	28	28	29	30	30	30	30	32	32
X_2	4,7	5,7	5,7	5,2	5,0	5,2	5,2	5,3	5,7	5,9	6,0	6,6	7,6	7,2	7,3

Варіант 4

Y	23,8	25,0	25,0	25,2	25,3	25,6	25,6	25,9	26,1	26,1	26,3	26,8	26,8	27,3	27,8
X_1	15	16	17	18	18	19	20	20	20	21	21	21	21	22	23
X_2	5,2	5,3	4,8	5,6	5,6	5,6	4,9	6,0	6,1	6,0	6,1	6,5	6,6	6,9	6,4

Варіант 5

Y	38,9	39,8	40,2	41,4	41,7	43,4	44,3	46,7	47,3	49,0	49,3	51,2	51,3	51,6	51,7
X_1	15	16	17	18	18	19	20	20	20	21	21	21	21	22	23
X_2	4,8	5,6	5,9	4,9	5,1	4,9	5,8	6,5	6,5	5,9	6,3	6,0	6,4	6,7	6,8

Житомирська політехніка	МІНІСТЕРСТВО ОСВІТИ І НАУКИ УКРАЇНИ ДЕРЖАВНИЙ УНІВЕРСИТЕТ «ЖИТОМИРСЬКА ПОЛІТЕХНІКА» Система управління якістю відповідає ДСТУ ISO 9001:2015											Ф-19.05-05.02/072.00.1/М/ ОК6-2023		
	Екземпляр № 1											Арк 38 / 21		

Варіант 6

Y	31,1	32,2	32,3	32,4	32,5	32,7	32,8	33,4	33,6	33,6	33,9	34,4	34,9	35,1	36,2
X ₁	15	15	17	17	18	18	19	19	19	20	20	20	21	21	22
X ₂	4,8	5,6	5,9	4,9	5,1	4,9	5,8	6,5	6,5	5,9	6,3	6,0	6,4	6,7	6,8

Варіант 7

Y	31,7	31,8	31,9	32,1	32,5	32,7	33,0	41,7	41,9	42,0	42,1	52,5	53,6	54,6	55,6
X ₁	23	24	24	26	26	27	28	28	29	30	30	30	30	32	32
X ₂	5,4	6,0	6,3	5,5	5,6	5,5	6,3	6,9	7,0	6,7	7,0	6,8	7,1	7,4	7,5

Варіант 8

Y	23,8	25,0	25,0	25,2	25,3	25,6	25,6	25,9	26,1	26,1	26,3	26,8	26,8	27,3	27,8
X ₁	15	20	24	22	27	31	29	33	29	26	26	27	28	38	40
X ₂	5,4	6,0	6,3	5,5	5,6	5,5	6,3	6,9	7,0	6,7	7,0	6,8	7,1	7,4	7,5

Варіант 9

Y	38,9	39,8	40,2	41,4	41,7	43,4	44,3	46,7	47,3	49,0	49,3	51,2	51,3	51,6	51,7
X ₁	15	15	17	17	18	18	19	19	19	20	20	20	21	21	22
X ₂	5,2	5,3	4,8	5,6	5,6	5,6	4,9	6,0	6,1	6,0	6,1	6,5	6,6	6,9	6,4

Варіант 10

Y	31,7	31,8	31,9	32,1	32,5	32,7	33,0	41,7	41,9	42,0	42,1	52,5	53,6	54,6	55,6
X ₁	15	20	24	22	27	31	29	33	29	26	26	27	28	38	40
X ₂	4,7	5,7	5,7	5,2	5,0	5,2	5,2	5,3	5,7	5,9	6,0	6,6	7,6	7,2	7,3

ПРИКЛАД РОЗВ'ЯЗАННЯ ЗАВДАННЯ

Для деякого підприємства отримано наступні результати вибірових статистичних спостережень за останні 24 місяці (2 роки), що містять дані по продуктивності праці та факторам, що впливають на цей показник (дані наведені в табл. 1):

Місяць	Продуктивність праці, тис.грн. /люд.-год.	Фондомісткість продукції, тис.грн. / тис.грн.	Коефіцієнт плинності робочої сили, %	Рівень втрат робочого часу, %
1	60	30	13,0	15,0
2	61	35	12,5	14,3
3	58	33	12,0	12,0
4	59	34	11,0	12,8
5	62	36	10,0	13,0
6	63	38	9,0	12,5
7	65	40	8,5	11,0
8	60	41	8,2	11,5
9	68	45	8,0	10,0
10	69	45	5,5	9,0
11	70	46	5,0	8,0
12	72	48	4,7	7,5

Житомирська політехніка	МІНІСТЕРСТВО ОСВІТИ І НАУКИ УКРАЇНИ ДЕРЖАВНИЙ УНІВЕРСИТЕТ «ЖИТОМИРСЬКА ПОЛІТЕХНІКА» Система управління якістю відповідає ДСТУ ISO 9001:2015	Ф-19.05-05.02/2/072.00.1/М/ ОК6-2023
	Екземпляр № 1	Арк 38 / 22

13	73	47	4,6	6,5
14	78	50	4,0	6,0
15	75	49	4,1	6,2
16	80	51	4,2	5,8
17	81	50	4,5	5,5
18	83	53	4,0	5,0
19	81	55	4,0	4,5
20	85	56	3,0	4,7
21	87	58	4,0	5,0
22	88	58	5,0	5,1
23	90	59	5,0	4,8
24	92	60	6,0	5,2

У припущенні щодо лінійної залежності між наведеними показниками побудувати економетричну модель продуктивності праці, що описує залежність між продуктивністю праці і наведеними вище факторами. Оцінити якість, адекватність і статистичну значимість побудованої моделі для рівня значимості $\alpha=0,05$. Розрахувати прогноз продуктивності праці на наступний місяць з рівнем надійності $P=0,95$, якщо очікувані значення чинників, що впливають на неї дорівнюють:

- фондомісткість продукції – 60 тис. грн. / тис. грн.;
- коефіцієнт плинності робочої сили – 4%;
- рівень втрат робочого часу – 5%.

Оцінити граничний вплив кожної пояснюючої змінної моделі на продуктивність праці.

РОЗВ'ЯЗАННЯ.

1. Виконуємо специфікацію економетричної моделі.

1.1. На етапі специфікації економетричної моделі виконуємо ідентифікацію змінних моделі та вводимо наступні умовні позначення для залежної (пояснювальної) змінної та незалежних (пояснюючих) змінних моделі: x_1 – фондомісткість продукції; x_2 – коефіцієнт плинності робочої сили; x_3 – рівень втрат робочого часу; y – продуктивність праці.

1.2. Переносимо на робочий аркуш вихідні дані задачі (рис. 1).

1.3. В умові задачі було припущення, що залежність між зазначеними економічними показниками може бути представлена економетричною моделлю багатофакторної лінійної регресії. Тому записуємо в загальному вигляді:

- теоретична модель

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \varepsilon$$

де y – продуктивність праці; x_1 – фондомісткість продукції; x_2 – коефіцієнт плинності робочої сили; x_3 – рівень втрат робочого часу; $\beta_0, \beta_1, \beta_2, \beta_3$ – параметри теоретичної моделі; ε – стохастична складова моделі;

Житомирська політехніка	МІНІСТЕРСТВО ОСВІТИ І НАУКИ УКРАЇНИ ДЕРЖАВНИЙ УНІВЕРСИТЕТ «ЖИТОМИРСЬКА ПОЛІТЕХНІКА» Система управління якістю відповідає ДСТУ ISO 9001:2015	Ф-19.05-05.02/2/072.00.1/М/ ОК6-2023
	Екземпляр № 1	Арк 38 / 23

- вибіркове рівняння регресії

$$\hat{y} = b_0 + b_1 x_1 + b_2 x_2 + b_3 x_3$$

де \hat{y} – оцінка математичного сподівання продуктивності праці; x_1, x_2, x_3 – див. вище;

b_0, b_1, b_2, b_3 – параметри вибіркової регресії;

- вибіркова економетрична модель

$$y = b_0 + b_1 x_1 + b_2 x_2 + b_3 x_3 + e$$

де y – продуктивність праці; x_1, x_2, x_3 – див. вище; b_0, b_1, b_2, b_3 – параметри вибіркової моделі; e – залишки моделі.

	A	B	C	D	E
	Місяць	Продуктивність в праці, тис.грн. /люд.-год.	Фондомісткість продукції, тис.грн. / тис.грн.	Коефіцієнт плинності робочої сили, %	Рівень втрат робочого часу, %
		Y	x1	x2	x3
1					
2					
3	1	60	30	13	15
4	2	61	35	12,5	14,3
5	3	58	33	12	12
6	4	59	34	11	12,8
7	5	62	36	10	13
8	6	63	38	9	12,5
9	7	65	40	8,5	11
10	8	60	41	8,2	11,5
11	9	68	45	8	10
12	10	69	45	5,5	9
13	11	70	46	5	8
14	12	72	48	4,7	7,5
15	13	73	47	4,6	6,5
16	14	78	50	4	6
17	15	75	49	4,1	6,2
18	16	80	51	4,2	5,8
19	17	81	50	4,5	5,5
20	18	83	53	4	5
21	19	81	55	4	4,5
22	20	85	56	3	4,7
23	21	87	58	4	5
24	22	88	58	5	5,1
25	23	90	59	5	4,8
26	24	92	60	6	5,2
27					

Рис. 1. Вихідні дані задачі

1.4. З метою оцінювання параметрів багатofакторної лінійної регресії використовуємо інструмент «Регресія» пакету «Аналіз даних». Пакет «Аналіз даних» активується зі вкладки «ДАНІ» в розділі «Аналіз».

Якщо ця команда відсутня в меню «Аналіз», її потрібно завантажити шляхом активації ряду команд. Детально ця процедура наведена в методичних рекомендаціях до лабораторної роботи №2.

У вікні інструменту «Регресія» вводимо вихідні дані та вказуємо спосіб виводу результатів (рис. 2). Роз'яснення щодо діалогового вікна інструменту

«Регресія» наведені в методичних рекомендаціях до лабораторної роботи №2.

1.5. На окремому аркуші отримаємо результати розрахунків «ВИВІД ІТОГІВ» (рис. 3). Детальна інформація щодо змісту полів наведена в методичних рекомендаціях до лабораторної роботи №2.

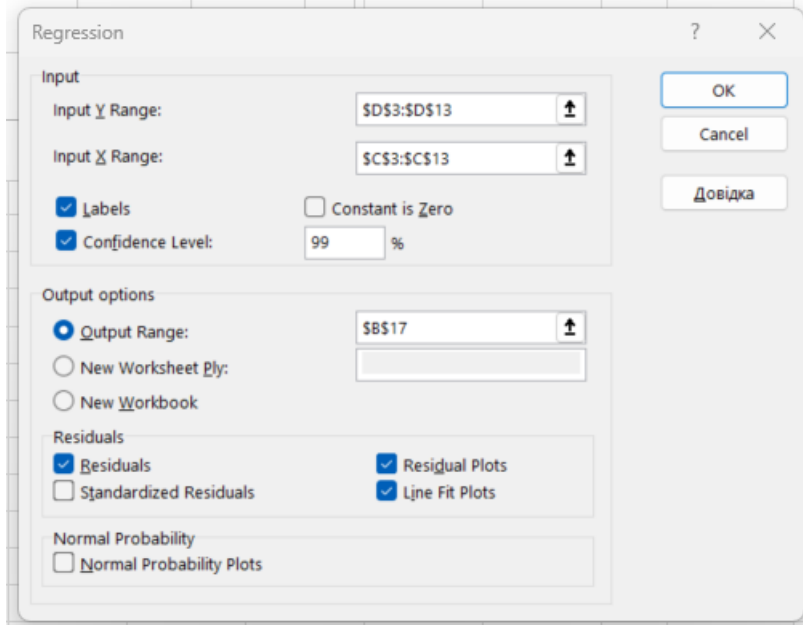


Рис. 2. Меню діалогового вікна інструменту «Регресія / Regression»

Регресія / Regression									
SUMMARY OUTPUT									
<i>Regression Statistics</i>									
Multiple R	0,993								
R Square	0,986								
Adjusted R Square	0,985								
Standard Error	0,179								
Observations	10								
<i>ANOVA</i>									
	<i>df</i>	<i>SS</i>	<i>MS</i>	<i>F</i>	<i>Significance F</i>				
Regression	1	18,478	18,478	573,6707867	0,000000009838				
Residual	8	0,258	0,032						
Total	9	18,736							
	<i>Coefficients</i>	<i>Standard Error</i>	<i>t Stat</i>	<i>P-value</i>	<i>Lower 95%</i>	<i>Upper 95%</i>	<i>Lower 99,0%</i>	<i>Upper 99,0%</i>	
Intercept	-1,8994	0,2155	-8,814	0,0000216043	-2,396	-1,402	-2,623	-1,176	
Xi	1,2115	0,0506	23,951	0,0000000098	1,095	1,328	1,042	1,381	

Рис. 3. Результати розрахунків інструменту «Регресія» для задачі

1.6. З отриманих результатів (рис. 3) перепишемо значення оцінок параметрів багатofакторної лінійної регресії (стовпчик «Коефіцієнти» - комірки B17:B20): $b_0=27,710$; $b_1=1,049$; $b_2=1,514$; $b_3= - 1,584$.

Записуємо оцінене вибіркове рівняння регресії

$$\hat{y} = 27,710 + 1,049x_1 + 1,514x_2 - 1,584x_3$$

та вибіркочну економетричну модель

$$y = 27,710 + 1,049x_1 + 1,514x_2 - 1,584x_3 + e.$$

Житомирська політехніка	МІНІСТЕРСТВО ОСВІТИ І НАУКИ УКРАЇНИ ДЕРЖАВНИЙ УНІВЕРСИТЕТ «ЖИТОМИРСЬКА ПОЛІТЕХНІКА» Система управління якістю відповідає ДСТУ ISO 9001:2015	Ф-19.05-05.02/2/072.00.1/М/ ОК6-2023
	Екземпляр № 1	Арк 38 / 25

2. Визначаємо показники якості вибіркової моделі.

2.1. В рядку «Множинний R» з комірки з адресою B4 (рис. 3) вписуємо значення коефіцієнта множинної кореляції $R=0,980$. В рядку «R-квадрат» з комірки B5 вписуємо значення коефіцієнта детермінації $R^2 = 0,961$. В рядку «Стандартна похибка» з комірки B7 вписуємо значення стандартної похибки моделі $\sigma_\varepsilon = 2,332$.

В результаті аналізу отриманих показників, можна зробити наступні висновки: оцінивши силу лінійного взаємозв'язку між змінними за допомогою коефіцієнта множинної кореляції можна сказати, що між продуктивністю праці та факторами, що впливають на цей показник існує прямий та сильний кореляційний зв'язок оскільки коефіцієнт множинної кореляції $R=0,980$, тобто його значення наближується до одиниці. Коефіцієнт детермінації $R^2 = 0,961$, показує, що зміна значення продуктивності праці на 96,1% пояснюється зміною значень фондомісткості продукції, коефіцієнта плинності робочої сили та рівня втрат робочого часу, а на 3,9% - іншими випадковими факторами, що свідчить про високий рівень адекватності моделі в цілому.

2.2. Перевірка адекватності побудованої економетричної моделі. Для оцінки рівня адекватності побудованої економетричної моделі експериментальним даним користуємося F-критерієм Фішера. З комірки E12 вписуємо розрахункове значення F-критерію Фішера $F^*=162,4$.

За статистичними таблицями F-критерію (розподілу) Фішера для рівня значимості $\alpha=0,05$ і ступенів вільності $\nu_1 = m = 3$ (де m – кількість незалежних (пояснюючих) змінних моделі) та $\nu_2 = n - k = 24 - 4 = 20$ (де n – кількість спостережень в статистичній вибірці; k - число параметрів моделі) визначаємо критичне значення F-критерію Фішера. Або з цією ж метою використовуємо вбудовану статистичну функцію «F.ОБР.ПХ» (див. рис. 4) (або «FРАСПОБР» для Excel 2007).

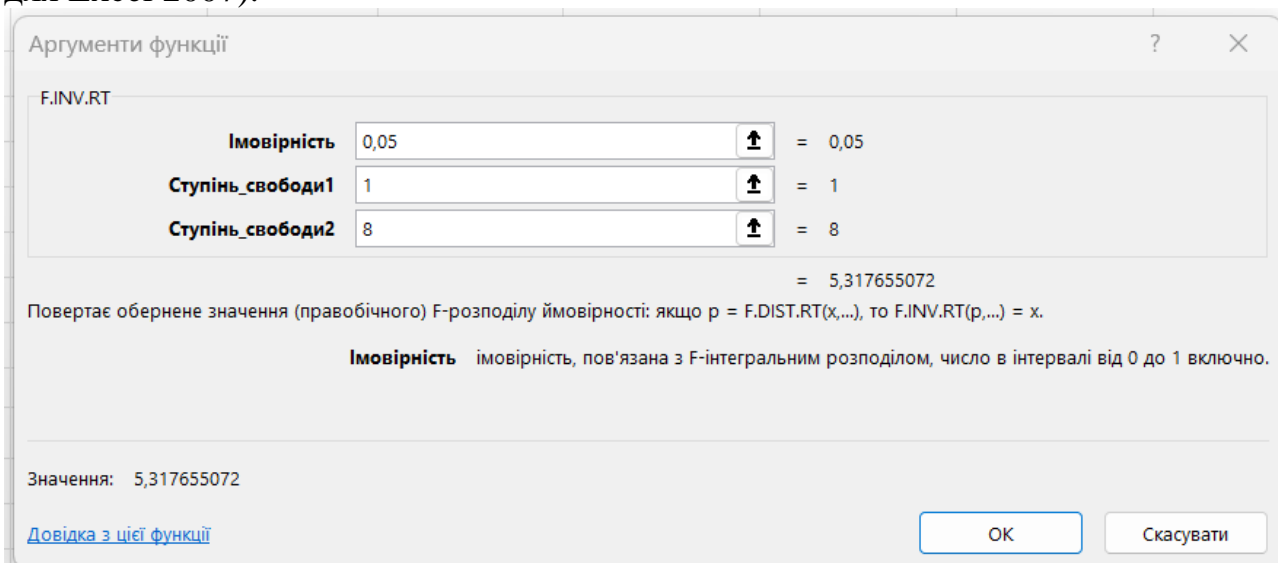


Рис. 4. Діалогове вікно функції «F.INV.RT» з даними моделі

Житомирська політехніка	МІНІСТЕРСТВО ОСВІТИ І НАУКИ УКРАЇНИ ДЕРЖАВНИЙ УНІВЕРСИТЕТ «ЖИТОМИРСЬКА ПОЛІТЕХНІКА» Система управління якістю відповідає ДСТУ ISO 9001:2015	Ф-19.05-05.02/2/072.00.1/М/ ОК6-2023
	Екземпляр № 1	Арк 38 / 26

В результаті проведених розрахунків отримуємо $F_{кр}=3,098$.

Оскільки виконується умова $F^* > F_{кр}$ ($162,4 > 3,098$), то робимо висновок про адекватність побудованої економетричної моделі експериментальним даним.

2.3. Знаходження довірчих інтервалів для оцінок.

З діапазону комірок F17:G20 виписуємо інтервали довіри для параметрів моделі: $1,903 < \beta_0 < 53,517$; $0,666 < \beta_1 < 1,431$; $0,522 < \beta_2 < 2,506$; $-2,985 < \beta_3 < -0,184$.

2.4. Визначаємо показники статистичної значимості вибіркової моделі.

Для цього розрахункові значення t-критерію Стьюдента для параметрів моделі виписуємо зі стовпчика «t-статистика» (комірок D17:D20, рис. 3) і, таким чином, маємо $t_{b_0}^* = 2,240$; $t_{b_1}^* = 5,721$; $t_{b_2}^* = 3,184$ та $t_{b_3}^* = -2,359$.

Для рівня значимості $\alpha=0,05$ та ступеня вільності $\nu = n - k = 24 - 4 = 20$ (де n – кількість спостережень в статистичній вибірці; k - число параметрів моделі) за статистичними таблицями t-критерію (розподілу) Стьюдента визначаємо критичне значення t-критерію Стьюдента. Або з цією ж метою використовуємо вбудовану статистичну функцію «СТЮДЕНТ.ОБР.2Х» (див. рис. 5) (або «СТЮДРАСПОБР» для Excel 2007).

В результаті проведених розрахунків (рис. 5) маємо критичне значення t-критерію Стьюдента $t_{кр} = 2,086$.

Використаємо вбудовану в Ексель статистичну функцію

Функція	СТЮДЕНТ.ОБР.2Х	2010
	T.INV.2T	Англомова версія
	СТЮДЕНТРАСПОБР	2007 та більш ранні версії

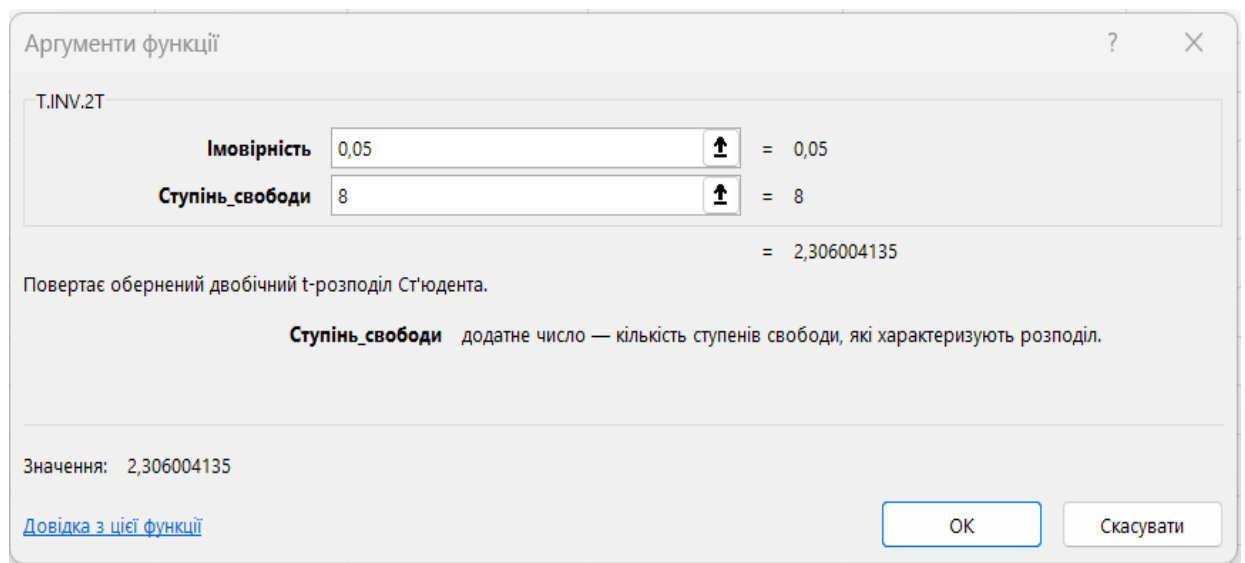


Рис. 5 Діалогове вікно функції T.INV.2T з даними моделі

Житомирська політехніка	МІНІСТЕРСТВО ОСВІТИ І НАУКИ УКРАЇНИ ДЕРЖАВНИЙ УНІВЕРСИТЕТ «ЖИТОМИРСЬКА ПОЛІТЕХНІКА» Система управління якістю відповідає ДСТУ ISO 9001:2015	Ф-19.05-05.02/2/072.00.1/М/ ОК6-2023
	Екземпляр № 1	Арк 38 / 27

Перевірку статистичної значимості кожного оціненого параметру здійснюємо шляхом перевірки виконання умови $\left|t_{b_j}^*\right| > t_{кр}$, ($j = \overline{0, 3}$). Оскільки умова перевірки виконується - $\left|t_{b_0}^*\right| > t_{кр}$, $\left|t_{b_1}^*\right| > t_{кр}$, $\left|t_{b_2}^*\right| > t_{кр}$, $\left|t_{b_3}^*\right| > t_{кр}$ відповідно $2,240 > 2,086$; $5,721 > 2,086$; $3,184 > 2,086$ та $|-2,359| > 2,086$ це свідчить про те, що всі параметри багатофакторної лінійної регресії є статистично значимі.

Виконаємо перевірку статистичної значимості коефіцієнта множинної кореляції R^2 . Для цього розраховуємо значення t-статистики для коефіцієнта парної кореляції за наступною залежністю:

$$t_R^* = \frac{R \sqrt{n-k}}{\sqrt{1-R^2}},$$

де n – розмір статистичної вибірки, k - число параметрів моделі.

В результаті проведених розрахунків отримуємо наступне значення цього показника

$$t_R^* = \frac{R \sqrt{n-k}}{\sqrt{1-R^2}} = \frac{0,980 \sqrt{24-4}}{\sqrt{1-0,961}} = 22,193.$$

Порівнявши розрахункове значення t-статистики для коефіцієнта множинної кореляції з критичним значенням критерію Стюдента $\left|t_R^*\right| > t_{кр}$ ($22,193 > 2,086$) можна зробити висновок, що коефіцієнт множинної кореляції R є статистично значимий.

Таким чином за результатами вищеведених досліджень можна зробити висновок про те, що отримана оцінена багатофакторна лінійна регресійна модель є якісною та статистично значимою. Тому на її основі можна здійснювати адекватний економіко-математичний аналіз і прогнозування.

3. Виконаємо прогнозні розрахунки на основі отриманої економетричної моделі.

З використанням отриманої моделі $\hat{y} = 27,710 + 1,049x_1 + 1,514x_2 - 1,584x_3$ можна отримати прогнозні значення залежної змінної при задаванні перспективного значення пояснювальних змінних. Прогноз на перспективу є двох видів: точковий та інтервальний.

Точковий прогноз продуктивності праці (\hat{Y}_{np}) визначається за формулою:

$$\hat{Y}_{np} = X_{np}^T \cdot B,$$

Житомирська політехніка	МІНІСТЕРСТВО ОСВІТИ І НАУКИ УКРАЇНИ ДЕРЖАВНИЙ УНІВЕРСИТЕТ «ЖИТОМИРСЬКА ПОЛІТЕХНІКА» Система управління якістю відповідає ДСТУ ISO 9001:2015	Ф-19.05-05.02/2/072.00.1/М/ ОК6-2023
	Екземпляр № 1	Арк 38 / 28

де $X_{np}^T = \begin{pmatrix} 1 \\ x_{1 np} \\ x_{2 np} \\ x_{3 np} \end{pmatrix} = (1 \quad x_{1 np} \quad x_{2 np} \quad x_{3 np})$ - транспонований вектор

прогнозних значень пояснюючих змінних; $x_{1 np}$ - прогнозне (очікуване) значення фондомісткості продукції; $x_{2 np}$ - прогнозне (очікуване) значення коефіцієнта плинності робочої сили; $x_{3 np}$ - прогнозне (очікуване) значення рівня витрат робочого часу;

$B = \begin{pmatrix} b_0 \\ b_1 \\ b_2 \\ b_3 \end{pmatrix}$ - вектор оцінок параметрів моделі.

Інтервальний прогноз для математичного сподівання продуктивності праці визначається за формулою:

$$\hat{Y}_{np} - t_{\alpha/2} \cdot \hat{\sigma}_\varepsilon \sqrt{X_{np}^T (X^T X)^{-1} X_{np}} \leq M(Y_{np}) \leq \hat{Y}_{np} + t_{\alpha/2} \cdot \hat{\sigma}_\varepsilon \sqrt{X_{np}^T (X^T X)^{-1} X_{np}}.$$

Інтервальний прогноз для індивідуального значення продуктивності праці визначається за формулою:

$$\hat{Y}_{np} - t_{\alpha/2} \hat{\sigma}_\varepsilon \sqrt{1 + X_{np}^T (X^T X)^{-1} X_{np}} \leq Y_{np} \leq \hat{Y}_{np} + t_{\alpha/2} \hat{\sigma}_\varepsilon \sqrt{1 + X_{np}^T (X^T X)^{-1} X_{np}},$$

де $X_{np} = \begin{pmatrix} 1 \\ x_{1 np} \\ x_{2 np} \\ x_{3 np} \end{pmatrix}$ - вектор прогнозних значень пояснюючих змінних; $x_{1 np}$ -

прогнозне (очікуване) значення фондомісткості продукції; $x_{2 np}$ - прогнозне (очікуване) значення коефіцієнта плинності робочої сили; $x_{3 np}$ - прогнозне (очікуване) значення рівня витрат робочого часу.

Матриця $(X^T X)$ – матриця, яка має наступний вигляд:

$$X^T X = \begin{bmatrix} n & \sum_{i=1}^n x_{1i} & \sum_{i=1}^n x_{2i} & \sum_{i=1}^n x_{3i} \\ \sum_{i=1}^n x_{1i} & \sum_{i=1}^n x_{1i}^2 & \sum_{i=1}^n x_{1i} \cdot x_{2i} & \sum_{i=1}^n x_{1i} \cdot x_{3i} \\ \sum_{i=1}^n x_{2i} & \sum_{i=1}^n x_{2i} \cdot x_{1i} & \sum_{i=1}^n x_{2i}^2 & \sum_{i=1}^n x_{2i} \cdot x_{3i} \\ \sum_{i=1}^n x_{3i} & \sum_{i=1}^n x_{3i} \cdot x_{1i} & \sum_{i=1}^n x_{3i} \cdot x_{2i} & \sum_{i=1}^n x_{3i}^2 \end{bmatrix};$$

$(X^T X)^{-1}$ - обернена матриця до матриці $(X^T X)$; $\hat{\sigma}_\varepsilon$ - стандартна похибка моделі; $t_{\alpha/2}$ - критичне значення t-критерію Стьюдента.

3.1. Точковий прогноз продуктивності праці визначаємо в наступній послідовності.

3.1.1. На робочому аркуші, на якому вводили вихідні дані задачі (див. рис. 1), в комірках B28:B31 розміщуємо компоненти вектору прогнозних

значень пояснюючих змінних $X_{np} = \begin{pmatrix} 1 \\ x_{1np} \\ x_{2np} \\ x_{3np} \end{pmatrix} = \begin{pmatrix} 1 \\ 60 \\ 4 \\ 5 \end{pmatrix}$ (рис. 6). Тобто в ці комірки

вводимо прогнозні (очікувані) значення фондомісткості продукції $x_{1np} = 60$ тис.грн./ тис.грн.; коефіцієнта плинності робочої сили $x_{2np} = 4\%$ та рівня втрат робочого часу $x_{3np} = 5\%$.

B35	:	✕	✓	f_x	=МУМНОЖ(B33:E33;Лист3!B17:B20)
	A	B	C	D	E
27					
28		1			
29	<i>X_{np}</i> =	60			
30		4			
31		5			
32					
33	<i>X_{np} транс</i> =	1	60	4	5
34					
35	<i>Y_{np теор}</i> =	88,76187			

Рис. 6. Результати визначення точкового прогнозу продуктивності праці

Житомирська політехніка	МІНІСТЕРСТВО ОСВІТИ І НАУКИ УКРАЇНИ ДЕРЖАВНИЙ УНІВЕРСИТЕТ «ЖИТОМИРСЬКА ПОЛІТЕХНІКА» Система управління якістю відповідає ДСТУ ISO 9001:2015	Ф-19.05-05.02/2/072.00.1/М/ ОК6-2023
	Екземпляр № 1	Арк 38 / 30

3.1.2. За допомогою вбудованої функції **ТРАНСП** виконуємо транспонування вектору X_{np} .

Функція **ТРАНСП** – перетворює горизонтально орієнтований діапазон комірок у вертикально орієнтований та навпаки. Для активізації даної функції використовують «Майстер функцій» та обрають категорію «Посилання і масиви». Транспонування масиву полягає в тому, що перший рядочок масиву 1 стає першим стовпчиком нового масиву 2, другий рядочок – другим стовпчиком і т.д.

Виділяємо діапазон комірок, в яких мають бути розташовані компоненти транспонованого масиву X_{np}^T (в нашому прикладі - В33:Е33).

Викликати функцію **ТРАНСП**. В поле «Масив» ввести аргументи, які відповідають компонентам масиву X_{np} , що розташовані в комірках з адресою В28:В31. Для завершення вводу натиснути комбінацію клавіш Ctrl + Shift + Enter, в результаті чого отримуємо масив $X_{np \text{ транс}}$, наведений на рис. 6. Якщо діалогове вікно функції вже закрили, то для завершення вводу поставити курсор в будь-яку з комірок, де розташований транспонований вектор (в нашому прикладі В33:Е33) та натиснути кнопку «F2», а потім натиснути комбінацію клавіш Ctrl + Shift + Enter.

3.1.3. Використовуючи функцію «**МУМНОЖ**» знайти добуток векторів $X_{np}^T \cdot B$.

Функція «**МУМНОЖ**» – дозволяє отримати добуток двох масивів. Для активації даної функції використовують «Майстер функцій» та обрають категорію «Математичні». Результатом добутку є масив з кількістю рядочків, як в масиві 1 та кількістю стовпців, як в масиві 2.

Стаavimo курсор в комірку, де будемо обраховувати прогнозне значення \hat{Y}_{np} (В35). Викликаємо функцію «**МУМНОЖ**». В діалоговому вікні функції заповнюємо поля. В поле «Массив1» вводимо аргументи, які відповідають вектору X_{np}^T (В33:Е33), в поле «Массив2» вводимо аргументи, які відповідають вектору B (Лист3!В17:В20). Оскільки знайдені нами коефіцієнти вектору B знаходяться на іншому аркуші, вказуємо комірки з назвою аркуша, на якому отримали «**ВИВІД ІТОГІВ**».

Для завершення вводу і отримання результату функції натиснути комбінацію клавіш «Ctrl + Shift + Enter».

В клітинці з адресою В35 отримаємо значення точкового прогнозу продуктивності праці $\hat{Y}_{np} = 88,762$ (рис. 6).

За результатами проведених розрахунків можна зробити висновок про те, що при заданих значеннях фондомісткості продукції 60 тис. грн. / тис. грн., коефіцієнта плинності робочої сили 4% та рівня втрат робочого часу 5%

Житомирська політехніка	МІНІСТЕРСТВО ОСВІТИ І НАУКИ УКРАЇНИ ДЕРЖАВНИЙ УНІВЕРСИТЕТ «ЖИТОМИРСЬКА ПОЛІТЕХНІКА» Система управління якістю відповідає ДСТУ ISO 9001:2015	Ф-19.05-05.02/2/072.00.1/М/ ОК6-2023
	Екземпляр № 1	Арк 38 / 31

точкове прогнозне значення продуктивності праці буде на рівні 88,762 тис.грн./людино-год.

3.2. З метою визначення інтервального прогнозу для математичного сподівання продуктивності праці проведемо розрахунки в наступній послідовності.

3.2.1. В комірках В37:Е40 розміщуємо компоненти матриці ($X^T X$), розраховані згідно формули наведеної вище (див. стор 9). Для цього скористаємось вбудованими функціями, наведеними на рис. 8.

	A	B	C	D	E
37		24	=СУММ(C3:C26)	=СУММ(D3:D36)	=СУММ(E3:E26)
38	$X_{транс}X=$	=СУММ(C3:C26)	=СУММКВ(D3:D26)	=СУММПРОИЗВ(C3:C26;D3:D26)	=СУММПРОИЗВ(C3:C26;E3:E26)
39		=СУММ(D3:D26)	=СУММПРОИЗВ(D3:D26;C3:C26)	=СУММКВ(D3:D26)	=СУММПРОИЗВ(D3:D26;E3:E26)
40		=СУММ(E3:E26)	=СУММПРОИЗВ(E3:E26;C3:C26)	=СУММПРОИЗВ(E3:E26;D3:D26)	=СУММКВ(E3:E26)
41					

Рис. 8. Вбудовані функції MS EXCEL для розрахунку матриці ($X^T X$)

Після розрахунку функцій наведених на рис. 8, отримуємо їх результати в в цих комірках (В37:Е40) як наведено на рис. 9.

	A	B	C	D	E
28		1			
29	$X_{пр} =$	60			
30		4			
31		5			
32					
33	$X_{пр транс} =$	1	60	4	5
34					
35	$Y_{пр теор} =$	88,76187007			
36					
37		24	1117	163,8	200,9
38	$X_{транс}X =$	1117	1286,94	6872,1	8656
39		159,8	6872,1	1286,94	1574,83
40		200,9	8656	1574,83	1966,09
41					
42	$(X_{транс}X)обр =$	0,197	0,001	-0,019	-0,010
43		0,001	0,000	0,000	0,000
44		0,012	0,000	0,038	-0,032
45		-0,035	0,000	-0,028	0,026
46					
47	$X_{прог т} * (X_{т} * X) обр$	0,143	0,000	-0,017	0,002
48					
49	$X_{прог т} * (X_{т} * X) обр * X_{прог}$	0,079			
50					

Рис. 9. Результати виконання розрахунків для прогнозних значень

Житомирська політехніка	МІНІСТЕРСТВО ОСВІТИ І НАУКИ УКРАЇНИ ДЕРЖАВНИЙ УНІВЕРСИТЕТ «ЖИТОМИРСЬКА ПОЛІТЕХНІКА» Система управління якістю відповідає ДСТУ ISO 9001:2015	Ф-19.05-05.02/2/072.00.1/М/ ОК6-2023
	Екземпляр № 1	Арк 38 / 32

3.2.2. З метою визначення компонентів матриці $(X^T X)^{-1}$ (рис. 9) в комірках В42:Е45 використовуємо вбудовану функцію МОБР(В37:Е40).

Функція **МОБР** – дозволяє отримати обернену матрицю. Для активзації даної функції використовують «Майстер функцій» та обрають категорію «Математичні». Вона застосовується до масивів, у яких кількість рядочків дорівнює кількості стовпчиків. Деякі квадратні матриці не можуть бути оберненими, в таких випадках функція МОБР видає повідомлення про помилку «#ЧИСЛО!». Визначник такої матриці дорівнює нулю.

Виділяємо діапазон комірок, де будемо обраховувати обернену матрицю $(X^T X)^{-1}$ - В42:Е45. Викликаємо функцію МОБР. В поле «Масив» вводимо аргументи, які відповідають компонентам масиву $(X^T X)$ – В37:Е40. Для завершення вводу і отримання результату функції натиснути комбінацію клавіш «Ctrl + Shift + Enter».

3.2.3. Для розрахунку добутку матриць $X_{np}^T (X^T X)^{-1} X_{np}$ двічі використовуємо функцію МУМНОЖ. Спочатку множимо вектор X_{np}^T (знаходиться в комірках В33:Е33) на $(X^T X)^{-1}$ (знаходиться в комірках В42:Е45) і отримаємо результат функції МУМНОЖ(В33:Е33;В42:Е45) в комірках В47:Е47 (рис. 9). Потім множимо отриманий вектор $X_{np}^T (X^T X)^{-1}$ (знаходиться в комірках В47:Е47) на X_{np} (знаходиться в комірках В28:В31) і отримаємо результат функції МУМНОЖ(В47:Е47;В28:В31) в комірці В49 (рис. 9).

3.2.4. Визначаємо мінімальне значення інтервального прогнозу для математичного сподівання продуктивності праці – ліву частину нерівності

$$\hat{Y}_{np} - t_{\alpha/2} \cdot \hat{\sigma}_\varepsilon \sqrt{X_{np}^T (X^T X)^{-1} X_{np}} \leq M(Y_{np}) \leq \hat{Y}_{np} + t_{\alpha/2} \cdot \hat{\sigma}_\varepsilon \sqrt{X_{np}^T (X^T X)^{-1} X_{np}},$$

де

$\hat{Y}_{np} = 88,762$ тис.грн./людино-год. (точковий прогноз продуктивності праці); $t_{\alpha/2} = t_{кр} = 2,086$ - критичне значення t-критерію Стьюдента, яке було обраховано за допомогою вбудованої функції «СТЬЮДЕНТ.ОБР.2Х»; $\hat{\sigma}_\varepsilon$ - стандартна похибка моделі – знаходиться на аркуші «ВИВІД ІТОГІВ», комірка В7, в рядку «Стандартна похибка» (див. рис.3).

В результаті отримуємо таке значення лівої частини нерівності:

$$\hat{Y}_{np} - t_{\alpha/2} \cdot \hat{\sigma}_\varepsilon \sqrt{X_{np}^T (X^T X)^{-1} X_{np}} = 88,762 - 2,086 \cdot 2,332 \cdot \sqrt{0,079} = 87,395$$

Визначаємо максимальне значення інтервального прогнозу для математичного сподівання продуктивності праці – праву частину нерівності

$$\hat{Y}_{np} + t_{\alpha/2} \cdot \hat{\sigma}_\varepsilon \sqrt{X_{np}^T (X^T X)^{-1} X_{np}} = 88,762 + 2,086 \cdot 2,332 \cdot \sqrt{0,079} = 90,129$$

Тоді маємо

Житомирська політехніка	МІНІСТЕРСТВО ОСВІТИ І НАУКИ УКРАЇНИ ДЕРЖАВНИЙ УНІВЕРСИТЕТ «ЖИТОМИРСЬКА ПОЛІТЕХНІКА» Система управління якістю відповідає ДСТУ ISO 9001:2015	Ф-19.05-05.02/2/072.00.1/М/ ОК6-2023
	Екземпляр № 1	Арк 38 / 33

$$87,395 \leq M(Y_{np}) \leq 90,129.$$

За результатами проведених розрахунків, з ймовірністю 95%, можна зробити висновок про те, що при заданих значеннях фондомісткості продукції 60 тис. грн. / тис. грн.; коефіцієнта плинності робочої сили 4% та рівня втрат робочого часу 5% середнє значення залежної змінної у генеральній сукупності буде знаходитися в межах 87,395 тис.грн./людино-год. до 90,129 тис.грн./людино-год.

3.3. Для визначення інтервального прогнозу для індивідуального значення продуктивності праці, виконуємо наступні розрахунки.

Визначаємо мінімальне значення інтервального прогнозу для індивідуального значення продуктивності праці, ліву частину нерівності:

$$\hat{Y}_{np} - t_{\alpha/2} \hat{\sigma}_{\varepsilon} \sqrt{1 + X_{np}^T (X^T X)^{-1} X_{np}} \leq Y_{np} \leq \hat{Y}_{np} + t_{\alpha/2} \hat{\sigma}_{\varepsilon} \sqrt{1 + X_{np}^T (X^T X)^{-1} X_{np}}.$$

$$\hat{Y}_{np} - t_{\alpha/2} \hat{\sigma}_{\varepsilon} \sqrt{1 + X_{np}^T (X^T X)^{-1} X_{np}} = 88,762 - 2,086 \cdot 2,332 \cdot \sqrt{1 + 0,079} = 83,709$$

Обчислюємо максимальне значення інтервального прогнозу для індивідуального значення продуктивності праці.

$$\hat{Y}_{np} + t_{\alpha/2} \hat{\sigma}_{\varepsilon} \sqrt{1 + X_{np}^T (X^T X)^{-1} X_{np}} = 88,762 + 2,086 \cdot 2,332 \cdot \sqrt{1 + 0,079} = 93,815$$

Отже маємо

$$83,709 \leq Y_{np} \leq 93,815$$

За результатами проведених розрахунків, з ймовірністю 95%, можна стверджувати, що при заданих значеннях фондомісткості продукції 60 тис. грн. / тис. грн.; коефіцієнта плинності робочої сили 4% та рівня втрат робочого часу 5% індивідуальне (окреме) значення залежної змінної у генеральній сукупності буде знаходитись в межах від 83,709 тис.грн./людино-год. до 93,815 тис.грн./людино-год.

4. Висновки.

Виконуємо економіко-математичний аналіз моделі продуктивності праці в наступній послідовності.

За результатами розрахунків було отримане таке оцінене вибіркове рівняння регресії

$$\hat{y} = 27,710 + 1,049x_1 + 1,514x_2 - 1,584x_3$$

та вибіркова економетрична модель

$$y = 27,710 + 1,049x_1 + 1,514x_2 - 1,584x_3 + e.$$

Отриманий K=коефіцієнт множинної кореляції, R=0,980, вказує на те, що між продуктивністю праці та факторами, що впливають на цей показник існує лінійний, прямий та сильний кореляційний зв'язок оскільки, його значення наближується до одиниці.

Коефіцієнт детермінації $R^2 = 0,961$, показує, що зміна значення продуктивності праці на 96,1% пояснюється зміною значень фондомісткості продукції, коефіцієнта плинності робочої сили та рівня втрат робочого часу, а

Житомирська політехніка	МІНІСТЕРСТВО ОСВІТИ І НАУКИ УКРАЇНИ ДЕРЖАВНИЙ УНІВЕРСИТЕТ «ЖИТОМИРСЬКА ПОЛІТЕХНІКА» Система управління якістю відповідає ДСТУ ISO 9001:2015	Ф-19.05-05.02/2/072.00.1/М/ ОК6-2023
	Екземпляр № 1	Арк 38 / 34

на 3,9% - іншими випадковими факторами, що свідчить про високий рівень адекватності моделі в цілому.

Оскільки виконується умова $F^* > F_{кр}$ ($162,4 > 3,098$), то робимо висновок про адекватність побудованої економетричної моделі експериментальним даним.

Порівнявши розрахункове значення t-статистики для коефіцієнта множинної кореляції з критичним значенням критерію Стюдента $|t_R^*| > t_{кр}$ ($22,193 > 2,086$) можна зробити висновок, що коефіцієнт множинної кореляції R є статистично значимий.

Умова перевірки t-статистики Стюдента стосовно параметрів економетричної моделі b_0, b_1, b_2, b_3 виконується: $|t_{b_0}^*| > t_{кр}$, $|t_{b_1}^*| > t_{кр}$, $|t_{b_2}^*| > t_{кр}$, $|t_{b_3}^*| > t_{кр}$ відповідно $2,240 > 2,086$; $5,721 > 2,086$; $3,184 > 2,086$ та $|-2,359| > 2,086$. Це свідчить про те, що всі параметри багатофакторної лінійної регресії є статистично значимі.

Таким чином за результатами вищенаведених досліджень можна зробити висновок про те, що отримана оцінена багатофакторна лінійна регресійна модель є якісною та статистично значимою. Тому на її основі можна здійснювати адекватний економіко-математичний аналіз і прогнозування.

Точкове прогнозне значення продуктивності праці при заданих значеннях фондомісткості продукції 60 тис.грн./тис.грн., коефіцієнта плинності робочої сили 4% та рівня втрат робочого часу 5% буде на рівні 88,762 тис.грн./людино-год.

Інтервальний прогноз для математичного сподівання продуктивності праці

$$87,395 \leq M(Y_{np}) \leq 90,129.$$

Тобто, з ймовірністю 95%, можна зробити висновок про те, що при заданих значеннях фондомісткості продукції 60 тис.грн./тис.грн.; коефіцієнта плинності робочої сили 4% та рівня втрат робочого часу 5% середнє значення залежної змінної у генеральній сукупності буде знаходитися в межах 87,395 тис.грн./людино-год. до 90,129 тис.грн./людино-год.

Інтервальний прогноз для індивідуального значення продуктивності праці $83,709 \leq Y_{np} \leq 93,815$.

Тобто, з ймовірністю 95%, можна стверджувати, що при заданих значеннях фондомісткості продукції 60 тис. грн. / тис. грн.; коефіцієнта плинності робочої сили 4% та рівня втрат робочого часу 5% індивідуальне (окреме) значення залежної змінної у генеральній сукупності буде знаходитись в межах від 83,709 тис.грн./людино-год. до 93,815 тис.грн./людино-год.

Житомирська політехніка	МІНІСТЕРСТВО ОСВІТИ І НАУКИ УКРАЇНИ ДЕРЖАВНИЙ УНІВЕРСИТЕТ «ЖИТОМИРСЬКА ПОЛІТЕХНІКА» Система управління якістю відповідає ДСТУ ISO 9001:2015	Ф-19.05-05.02/2/072.00.1/М/ ОК6-2023
	Екземпляр № 1	Арк 38 / 35

На основі обчислених коефіцієнтів регресії b_1 , b_2 , b_3 оцінюємо граничний вплив кожного фактору на продуктивність праці. На підставі обчислених значень можна зробити висновки про те, що у випадку збільшення фондомісткості продукції на 1 тис.грн./тис.грн. продуктивність праці в середньому збільшиться на 1,049 тис.грн./людино-год. за умови, що інші фактори такі як коефіцієнт плинності робочої сили та рівень втрат робочого часу, будуть незмінними. При збільшенні коефіцієнта плинності робочої сили на 1%, за умови, що фондомісткість продукції та рівень втрат робочого часу, будуть незмінними, продуктивність праці в середньому збільшиться на 1,514 тис.грн./людино-год. При збільшенні рівня втрат робочого часу на 1% при незмінних значеннях фондомісткості продукції та коефіцієнта плинності робочої сили, продуктивність праці в середньому зменшиться на 1,548 тис.грн./людино-год.

Тема 6. Аналіз великих даних в банківській сфері

Питання для обговорення та дискусії

1. Особливості застосування BigData при аналізі фінансових ринків
2. Особливості застосування BigData в банківській сфері
3. Особливості застосування BigData страховими компаніями
4. Застосування технологій BigData при просуванні фінансових продуктів

Питання для контролю:

1. Як використовуються великі дані при пошуку нових клієнтів
2. Порядок застосування великих даних при просуванні нових послуг на ринок
3. Чи можливо негайно перейти до технологій BigData в усіх сферах фінансового сектору?
4. Що таке цифрова робоча сила у фінансах?
5. Чи можуть програмні роботи виконувати функції фінансових радників?
6. Основні функції великих даних в банках України
7. Якими законодавчими актами регулюється використання персональних даних в банках України
8. Шляхи зниження кредитних ризиків за допомогою технологій BigData
9. Особливості аналізу великих даних для розширення клієнтської бази
10. Що таке фінансові ризики?
11. Які фінансові послуги використовують великі дані
12. Чи можуть використовувати технології BigData податкові органи
13. Наскільки аналітика великих даних може підтримувати дотримання

Житомирська політехніка	МІНІСТЕРСТВО ОСВІТИ І НАУКИ УКРАЇНИ ДЕРЖАВНИЙ УНІВЕРСИТЕТ «ЖИТОМИРСЬКА ПОЛІТЕХНІКА» Система управління якістю відповідає ДСТУ ISO 9001:2015	Ф-19.05-05.02/2/072.00.1/М/ ОК6-2023
	Екземпляр № 1	Арк 38 / 36

фінансових послуг?

14. Поняття скорингу.
15. Що дає аналіз великих даних у фінансових моделях?
16. Які інструменти використовуються для аналізу великих даних у фінансовій сфері

Завдання до виконання

1. Проведіть порівняльний аналіз різних інструментів які можна застосувати для аналізу даних у фінансовому секторі.
2. Проаналізуйте залежність цін на зерно на світовому ринку від погодних факторів, цін на нафту, за допомогою запропонованих викладачем інструментів. Візуалізуйте отриману інформацію та зробіть відповідні висновки
3. Проаналізуйте як впливає на фінансові показники роботи Українських банків ситуація з пандемією, рівнем безробіття, за допомогою запропонованих викладачем інструментів. Візуалізуйте отриману інформацію та зробіть відповідні висновки.
4. Виявіть фактори, які впливають на наповненість бюджету та їх взаємозв'язок на основі відкритих даних міністерства фінансів.

Тема 7. Аналіз великих даних в державному управлінні та соціальній сфері

Питання для обговорення та дискусії

1. Цифрова соціологія.
2. Соціальні наслідки впровадження великих даних
3. Сутність «Програми створення системи соціального кредиту» Китаю.
4. Застосування великих даних в управлінні розвитком міста та регіону.
5. Процес прийняття управлінських рішень на регіональному (місцевому) рівні з використанням технології Big Data.
6. Джерела відкритих даних в Україні

Питання для контролю:

1. Що таке цифрова соціологія?
2. Чим займається провідна європейська дослідницька група «The Digital Methods Initiative» (DMI)?
3. Який механізм реалізації китайської системи соціального рейтингу?
4. Як застосовуються технології великих даних в управлінні просторово-економічним розвитком міста і регіону? Наведіть приклади.
5. Які існують етапи процесу застосування Big Data для виявлення пріоритетних сфер підприємницької діяльності на територіях перспективного

Житомирська політехніка	МІНІСТЕРСТВО ОСВІТИ І НАУКИ УКРАЇНИ ДЕРЖАВНИЙ УНІВЕРСИТЕТ «ЖИТОМИРСЬКА ПОЛІТЕХНІКА» Система управління якістю відповідає ДСТУ ISO 9001:2015	Ф-19.05-05.02/2/072.00.1/М/ ОК6-2023
	Екземпляр № 1	Арк 38 / 37

розвитку?

6. Які ви знаєте джерела відкритих даних в Україні?

Завдання до виконання

1. Ознайомитись з даними, що розміщені на ресурсі «Дія.Відкриті дані» за посиланням <https://diia.data.gov.ua/>. Зібрати дані за темою випускної кваліфікаційної роботи.

2. Зібрати дані за темою випускної кваліфікаційної роботи за допомогою ресурсу «YouControl» <https://youcontrol.com.ua/>.

3. Зібрати дані за темою випускної кваліфікаційної роботи за допомогою ресурсу Opendatabot <https://opendatabot.ua/> та Пошуково-аналітична система .007- Web-ресурс <https://www.007.org.ua/>.

Тема 8. Аналіз великих даних у маркетингових дослідженнях

Питання для обговорення та дискусії

1. Важливість великих даних у сфері торгівлі та цифрового маркетингу
2. Основні методи аналізу Big Data
3. Проблеми та наслідки використання великих даних в маркетингу
4. Аналіз ефективності поштових розсилок за допомогою метрик.
5. Психологічні прийоми в поштових розсилках
6. Інструменти для роботи з email розсилками.
7. Веб-аналітика як важливий інструмент цифрового маркетингу
8. Система метрик в цифровому маркетингу та методика їх визначення
9. «Піратська» воронка AARRR
10. Методика аналізу маркетингових воронок.

Питання для контролю:

1. Які є найпоширеніші ролі великих даних у цифровому маркетингу?
2. Що таке описова аналітика?
3. Що таке прогнозна або предикативна аналітика?
4. Що розуміють під поняттям приписна аналітика?
5. На яке питання відповідає діагностична аналітика?
6. Які виникають проблеми при використанні великих даних в маркетингу?
7. Які метрики треба аналізувати для аналізу ефективності поштових розсилок?
8. Які види електронних листів Ви знаєте?
9. Яка структура електронних листів?
10. Ознаки показників веб-аналітики
11. Що таке маркетингова воронка та як її побудувати?

Житомирська політехніка	МІНІСТЕРСТВО ОСВІТИ І НАУКИ УКРАЇНИ ДЕРЖАВНИЙ УНІВЕРСИТЕТ «ЖИТОМИРСЬКА ПОЛІТЕХНІКА» Система управління якістю відповідає ДСТУ ISO 9001:2015	Ф-19.05-05.02/2/072.00.1/М/ ОК6-2023
	Екземпляр № 1	Арк 38 / 38

12. Що таке спліт-тестування (A/B Split Testing)?

Завдання до виконання

1. Напишіть тригерного електронного листа.
2. Напишіть приклад комерційного електронного листа.
3. Напишіть приклад електронного листа - анонса.
4. Напишіть приклад інформаційного електронного листа.