

МІНІСТЕРСТВО ОСВІТИ І НАУКИ УКРАЇНИ

**ХАРКІВСЬКИЙ НАЦІОНАЛЬНИЙ ЕКОНОМІЧНИЙ УНІВЕРСИТЕТ
ІМЕНІ СЕМЕНА КУЗНЕЦЯ**

ПРОГРАМНІ ПРОДУКТИ ОБРОБКИ ВЕЛИКИХ МАСИВІВ ІНФОРМАЦІЇ

**Методичні рекомендації
до лабораторних та самостійних робіт
для студентів спеціальності
051 "Економіка"
першого (бакалаврського) рівня**

**Харків
ХНЕУ ім. С. Кузнеця
2021**

УДК 004.04(07.034)

П78

Укладач В. І. Дериховська

Затверджено на засіданні кафедри статистики і економічного прогнозування.

Протокол № 1 від 19.08.2020 р.

Самостійне електронне текстове мережеве видання

Програмні продукти обробки великих масивів інформації П78 [Електронний ресурс] : методичні рекомендації до лабораторних та самостійних робіт для студентів спеціальності 051 "Економіка" першого (бакалаврського) рівня / уклад. В. І. Дериховська. – Харків : ХНЕУ ім. С. Кузнеця, 2021. – 170 с.

Подано методичні рекомендації до виконання лабораторних і самостійних робіт, метою яких є розширення та поглиблення знань щодо ефективного використання сучасних аналітичних інструментів, статистичних методів і моделей аналізу складних масових соціально-економічних явищ та процесів задля ухвалення ефективних управлінських рішень.

Рекомендовано для студентів спеціальності 051 "Економіка" першого (бакалаврського) рівня.

УДК 004.04(07.034)

© Харківський національний економічний університет імені Семена Кузнеця, 2021

Вступ

З кожним днем зростає обсяг даних, відповідно зростає й інтерес з боку компаній, урядових структур та міжнародних організацій на спеціалістів-аналітиків, які володіють різним аналітичним програмним забезпеченням для збору, оброблення та аналізу інформації, оцінювання вірогідності даних, їхнього моделювання та прогнозування.

Знання програм і програмних середовищ дозволяє правильно проводити оброблення, моделювання й аналіз даних з використанням статистичних методів, що, своєю чергою, впливає на ухвалення вірних управлінських рішень на основі аналітичного аналізу та прогнозування, а отже, дозволяють компаніям економити ресурси і збільшувати свій прибуток. Отже, в сучасних конкурентних умовах ефективно оброблення великих масивів накопиченої інформації значно полегшує пошук оптимальної стратегії розвитку компанії.

Навчальна дисципліна "Програмні продукти обробки великих масивів інформації" є вибірковою дисципліною першого (бакалаврського) рівня. Її освоєння дозволить ефективно використовувати сучасні аналітичні інструменти для аналізу складних, масових соціально-економічних явищ і процесів. Студенти навчатимуться вирішувати аналітичні завдання в умовах невизначеності, застосовуючи відповідні статистичні методи і моделі.

Метою навчальної дисципліни є розширення та поглиблення теоретичних знань і формування професійних компетентностей щодо володіння програмними продуктами аналізу бізнес-інформації з використанням статистичних методів і моделей задля ухвалення ефективних управлінських рішень.

Вивчення початкової дисципліни "Програмні продукти обробки великих масивів інформації" вимагає, поряд із засвоєнням теоретичних положень, практичної реалізації основних розділів навчальної дисципліни шляхом виконання відповідних лабораторних і самостійних робіт.

Виконання лабораторних і самостійних роботи з цієї навчальної дисципліни дозволить бакалаврам опанувати знання щодо сучасного аналітичного програмного забезпечення та ситуацій, в яких застосовувати ті чи інші програмні засоби для аналізу даних залежно від: типу, обсягу і структури даних; встановлення залежності, тренду або тенденції на основі аналізу досліджуваних масивів даних; встановлення зв'язку між досліджуваними ознаками; формулювання коректних висновків та використання методів візуалізації даних.

У результаті вивчення навчальної дисципліни студенти компетентностей, наведених у табл. 1.

Таблиця 1

Компетентності та результати навчання за дисципліною

Компетентності	Результати навчання
Здатність набувати теоретичні знання щодо програмних засобів аналізу даних та отримувати навички формування інформаційного простору дослідження	Уміння орієнтуватися в інформаційному просторі задля формування якісної інформаційної бази дослідження. Здатність проводити первинний аналіз інформаційного простору дослідження. Здатність формувати і попередньо підготувати масив даних для оброблення сучасними методами бізнес-аналітики
Здатність доцільно використовувати вбудовані програмні середовища оброблення й аналізу даних	Здатність до раціонального використання пакетів прикладних програм (MS Excel, Access; Project Expert) для вирішення конкретних бізнес-завдань. Розуміння сфери використання MS Excel та опанування його можливостей. Здатність до побудови баз даних у середовищі MS Access та опанування його аналітичних можливостей. Здатність до побудови та оптимізації проєктів за допомогою Ms Project
Здатність використовувати спеціальні програмні продукти оброблення великих масивів інформації	Здатність використовувати потужну уніфіковану платформу RapidMiner для створення, передачі й обслуговування наукомістких даних і проведення аналітичних процесів. Здатність вирішувати прикладні задачі за допомогою графічних можливостей EViews. Здатність використовувати мову програмування R (вільне програмне середовище обчислень з відкритим вихідним кодом в рамках проєкту GNU) для статистичного оброблення даних і роботи з графікою
Здатність формувати сценарії управлінських рішень і здійснювати їхній обґрунтований вибір	Здатність до узагальнення результатів аналітичної роботи та представлення їх у вигляді професійно оформлених звітів, аналітичних висновків, рекомендацій

У роботі надано авторський переклад українською мовою панелей інструментів (вікон, команд і функцій) усіх програмних продуктів та додатків, що використовуються.

Змістовий модуль 1

Класифікація програмних продуктів обробки великих масивів даних та їхні види.

Вбудовані програмні продукти

Тема 1. Поняття програмних продуктів обробки великих масивів даних (ПП ОВМД) та їхні види

Лабораторна робота 1.

Оброблення таблиць, редагування графіків та візуалізація в MS Excel

Мета: надбання навичок аналізу великих масивів даних за допомогою зведених таблиць та графічного методу в пакеті *MS Excel*.

Завдання: ознайомитися з графічними можливостями *MS Excel*; побудувати різні типи графіків на основі статистичних даних; навчитися використовувати спарклайни й умовне форматування для візуалізації даних; навчитися обробляти великі масиви даних та створювати зведені таблиці.

Методичні рекомендації

1. Знайомство з графічним редактором MS Excel.

MS Excel – це багатофункціональна програма, яка використовується в будь-якій галузі (сфері) діяльності. Різні інструменти *MS Excel* для розрахунку та аналізу дозволяють обробляти велику кількість даних і візуально представляти результати.

Один із способів візуалізації даних – діаграма. *MS Excel* дозволяє створювати велику кількість різних діаграм. Діаграма – це спосіб візуалізувати інформацію, наведену у вигляді таблиці чисел. Графічний формат діаграми спрощує розуміння великої кількості інформації та зв'язків між різними серіями даних.

На діаграмі також наведено огляд ситуації, що дозволяє представити дані та виявляти важливі тенденції.

Діаграми створюються на основі даних, що містяться в робочих аркушах. Графіки (гістограми чи діаграми) *MS Excel* динамічні, якщо змінювати дані на робочих аркушах, на яких побудована діаграма, вони автоматично оновлюються і на діаграмі.

Діаграма в *MS Excel* містить безліч об'єктів, кожен з яких може бути вибраний та змінений окремо. Коли ви пересуваєте курсор миші над діаграмою, поруч із ним з'являється підказка, вказуючи ім'я об'єкта. На рис. 1 наведено приклад діаграми з усіма основними об'єктами.

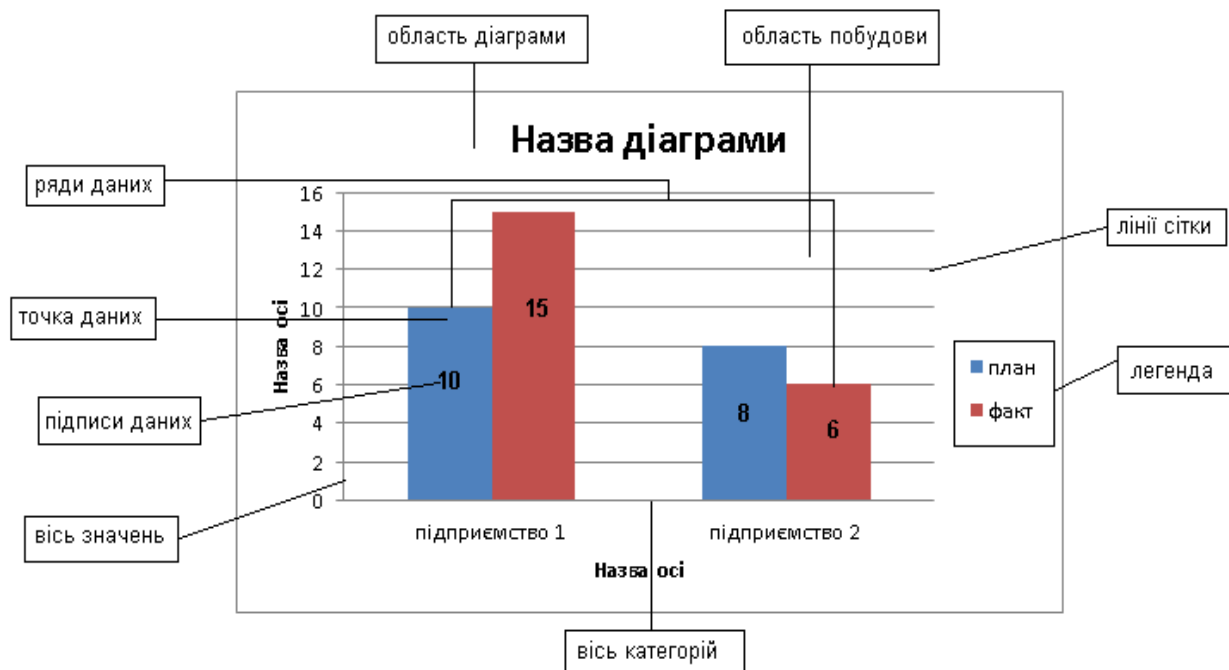


Рис. 1. Приклад діаграми, що містить всі базові елементи

Наявність усіх цих об'єктів є необхідною умовою правильної візуалізації даних за допомогою діаграми.

Є 11 стандартних типів діаграм, що пропонуються *MS Excel*, які дозволяють ефективно представляти дані. Для кожного типу діаграм існує декілька підтипів. Багато діаграм мають об'ємний формат. Більшість діаграм об'ємних форматів не додають нової інформації до презентації даних на плоскій діаграмі, але додають більшої наочності під час презентації підготовці звітів або слайдів для звіту.

На рис. 2 показано, які типи діаграм можуть бути побудовані в *MS Excel*.

1.2. Створення комбінованої діаграми.

Доцільно зазначити, що базові знання щодо побудови стандартних типів і підтипів діаграм *MS Excel*, їхнього призначення та обмеження щодо використання, були одержані під час вивчення навчальної дисципліни "Статистика". Саме тому розглянемо більш детально процеси побудови більш складних елементів візуалізації *MS Excel*, зокрема комбіновану та вторинну кругову діаграми.

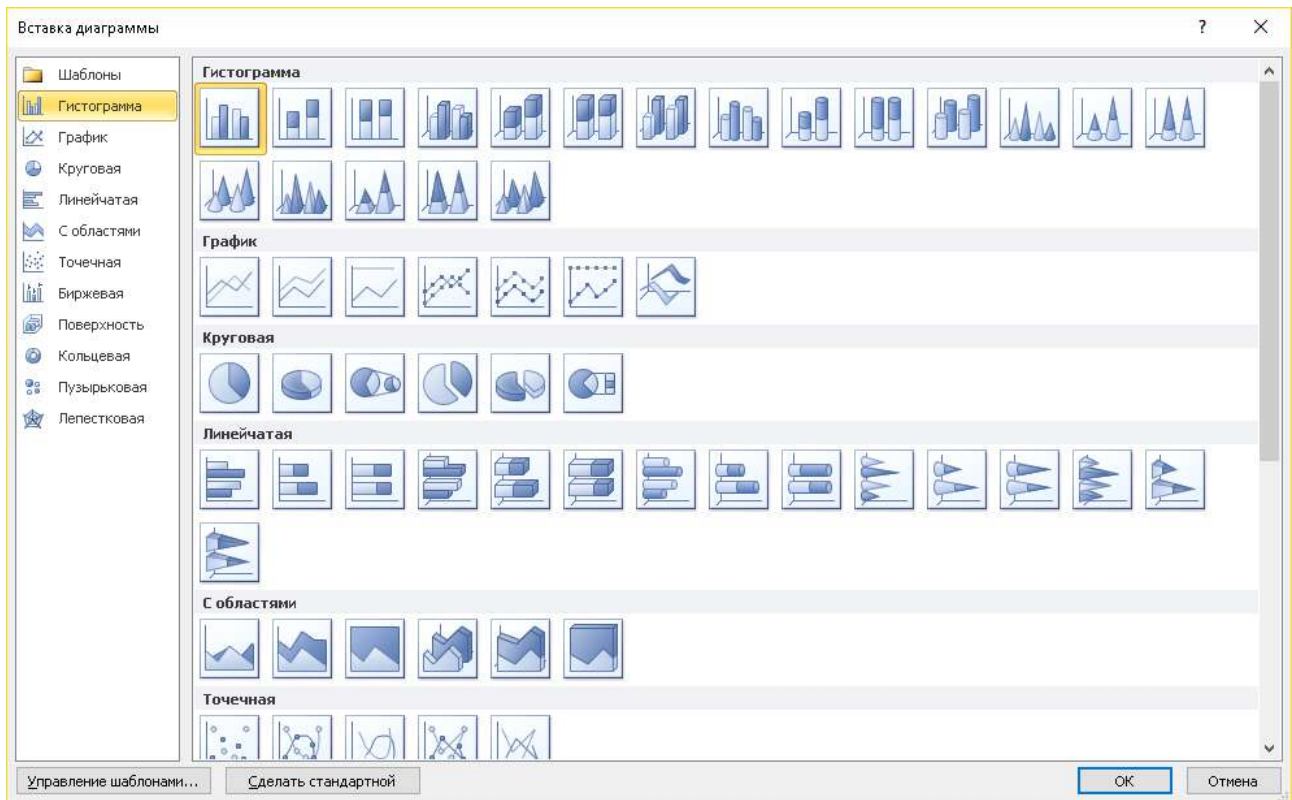


Рис. 2. Вікно для вибору стандартних типів діаграм

Комбінована діаграма складається з декількох серій даних і одночасно використовує різні типи діаграм. У комбінованій діаграмі можна використовувати один тип, наприклад, гістограму, але додати другу вісь значень.

Найпростіший спосіб створити складну (комбіновану) діаграму – це змінити тип діаграми для одного з рядків даних. Щоб змінити тип діаграми для ряду даних, виберіть його (натисніть на нього). На вкладці *Конструктор* у групі *Тип* оберіть команду *Змінити тип діаграми*. У діалоговому вікні *Зміна типу діаграми* змініть тип обраного ряду даних (рис. 3).

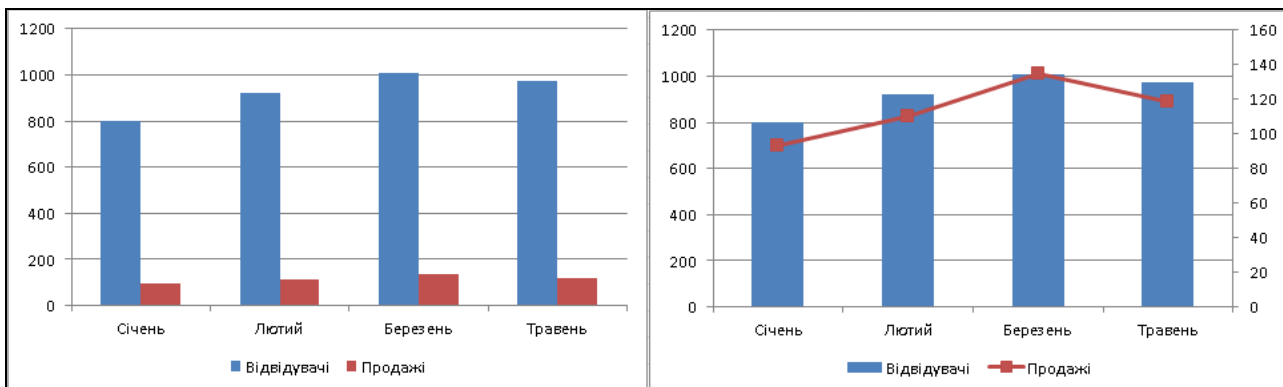


Рис. 3. Зміна типу діаграма для одного з рядів даних

Також використання комбінованої діаграми є раціональним, якщо числа в діаграмі значно відрізняються для кожного ряду даних або за наявності змішаних типів даних (наприклад, ціна та обсяг). В цьому випадку доцільним є використання допоміжної (вертикальної) осі (вісь значень). При цьому масштаб цієї осі відповідає значенням пов'язаного з нею ряду даних. Допоміжна вісь добре підходить для діаграм, в яких поєднуються гістограма і графік. Для створення допоміжної вісі необхідно виділити ряд даних, для якого створюється вісь, у контекстному меню обрати команду *Формат ряду даних*, потім функцію *Параметри ряду* та зробити активною допоміжну вісь (рис. 4).

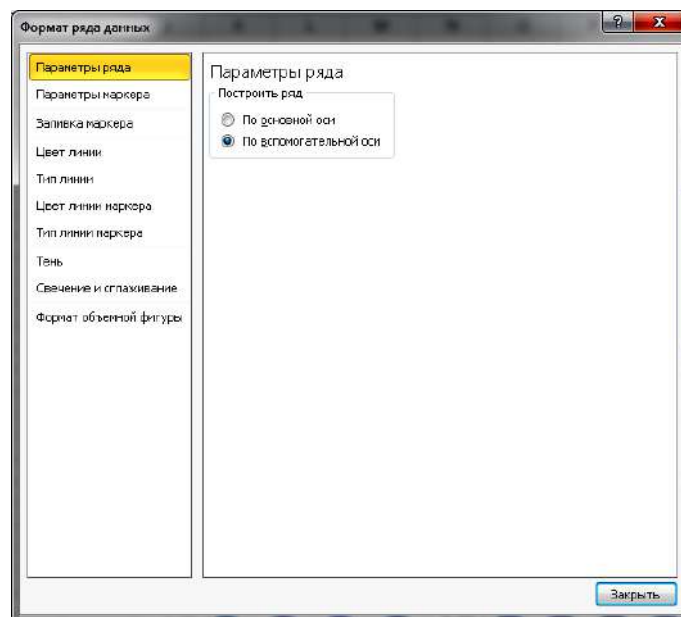


Рис. 4. Додавання допоміжної вісі для одного з рядів даних

Також ви можете змінити діапазон вихідних даних, представлених на діаграмі, використовуючи команди панелі управління або контекстне меню (команда контекстного меню *Вибір даних*) (рис. 5).

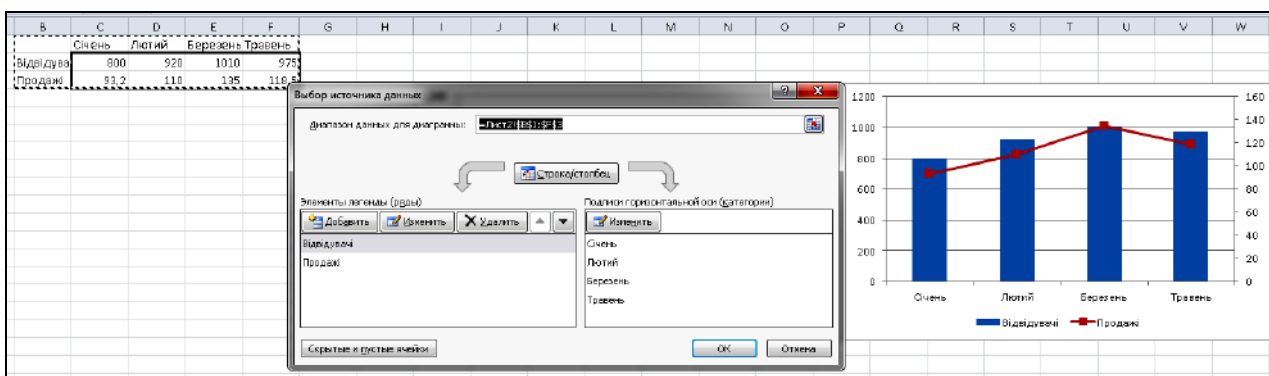


Рис. 5. Редагування вихідних даних

Наведене на рис. 5 діалогове вікно також дозволяє змінити назву рядків та їхню послідовність.

На основі даних, наведених в табл. 2, побудуємо комбіновану діаграму.

Таблиця 2

Середньомісячна вартість витрат на одне домогосподарство

Показники	Роки						
	2012	2013	2014	2015	2016	2017	2018
Середньомісячна вартість витрат на одне домогосподарство, грн, зокрема	3 073,3	3 458	3 592,1	3 820,3	4 048,9	4 952	5 720,4
харчові та безалкогольні напої	1 585,8	1 773,9	1 799,6	1 913,9	2 101,4	2 629,5	2 848,8
алкогольні напої, тютюнові вироби	104,45	117,6	125,7	133,7	137,7	163,4	165,9
непродовольчі товари та послуги	1 072,6	1 224,1	1 336,3	1 398,2	1 469,8	1 807,5	2 316,8
інші витрати	310,4	342,3	330,5	374,4	340,1	351,6	388,9

Для побудови комбінованої діаграми засобами MS Excel, використаємо стандартний алгоритм створення діаграми: на першому етапі виділяємо весь масив вихідних даних та обираємо тип діаграми *Гістограма стовпчаста*, на другому етапі задаємо необхідні параметри (підписи осей, сітка, ряди даних, легенда) діаграми, потім виділяємо кожен ряд даних, що містить інформацію про окремий вид витрат домогосподарства та через контекстне меню змінюємо тип діаграми для ряду даних на *Графік з маркерами* (рис. 6).

На останньому етапі створюємо необхідний дизайн комбінованої діаграми (обираємо фон, заливку для кожного ряду даних). Для редагування та форматування дизайну створеної комбінованої діаграми використовуємо вкладки *Конструктор*, *Макет*, *Формат* у MS Excel.

1.3. Створення вторинної кругової діаграми.

Розглянемо приклад побудови вторинної кругової діаграми на прикладі. Маємо дані про обсяги доходів певного агропромислового підприємства, що спеціалізується на вирощуванні та реалізації сільськогосподарської продукції (табл. 3).

Загальні середньомісячні витрат на одне домогосподарство

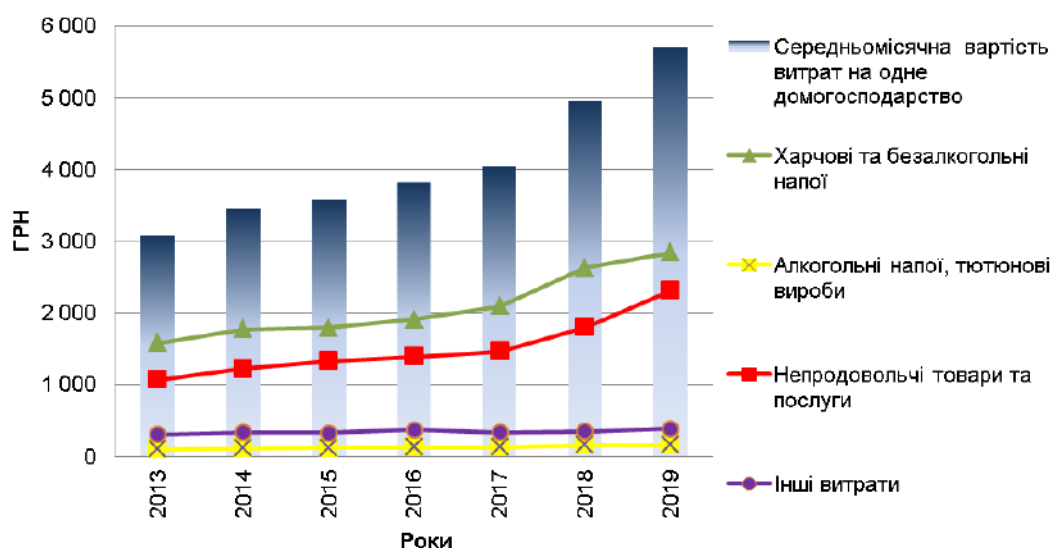


Рис. 6. Комбінована діаграма розподілу середньомісячних витрат на одне домогосподарство за період 2012 – 2018 рр.

Таблиця 3

Обсяги доходу від реалізації сільськогосподарської продукції у 2018 р.

Вид сільськогосподарської продукції	Обсяг отриманого від реалізації доходу, млн дол. США
Соняшникова олія наливом	1 067,5
Соняшникова олія бутельована	140,7
Фермерські продукти	381,3
Зерно, зокрема	877,14
ячмінь	129,26
пшениця	360,09
кукурудза	387,79

Побудуємо вторинну кругову діаграму, що буде відображати структуру доходів агропромислового підприємства.

Кругова діаграма показує як абсолютну величину, так і внесок (питому вагу) кожного елемента ряду даних у загальну суму. Кругова діаграма використовується для графічного зображення структури та пропорційності досліджуваного явища чи процесу. Однак на круговій діаграмі може бути представлений лише один ряд даних.

Вторинна кругова діаграма застосовується для деталізації одного з секторів кругової діаграми. Такий спосіб візуалізації допомагає спростити і деталізувати перегляд невеликих областей (секторів) кругової діаграми.

Під час побудови вторинної кругової діаграми важливо дотримуватися таких правил:

1. Головна і вторинна діаграми – частини одного ряду даних. Їх не можна відформатувати незалежно один від одного.

2. Сектори на вторинному колі теж показують частки, що і на звичайній діаграмі. Однак сума відсотків часток не дорівнює 100, а становить загальну величину значення на секторі головної кругової діаграми (від якої відокремлена вторинна).

3. За замовчуванням на вторинному колі відображається остання третина даних. Якщо, наприклад, у вихідній таблиці 7 рядків (для діаграми – 7 секторів), то останні три ряди даних будуть візуалізовані на вторинній (другорядній) діаграмі (рис. 7).

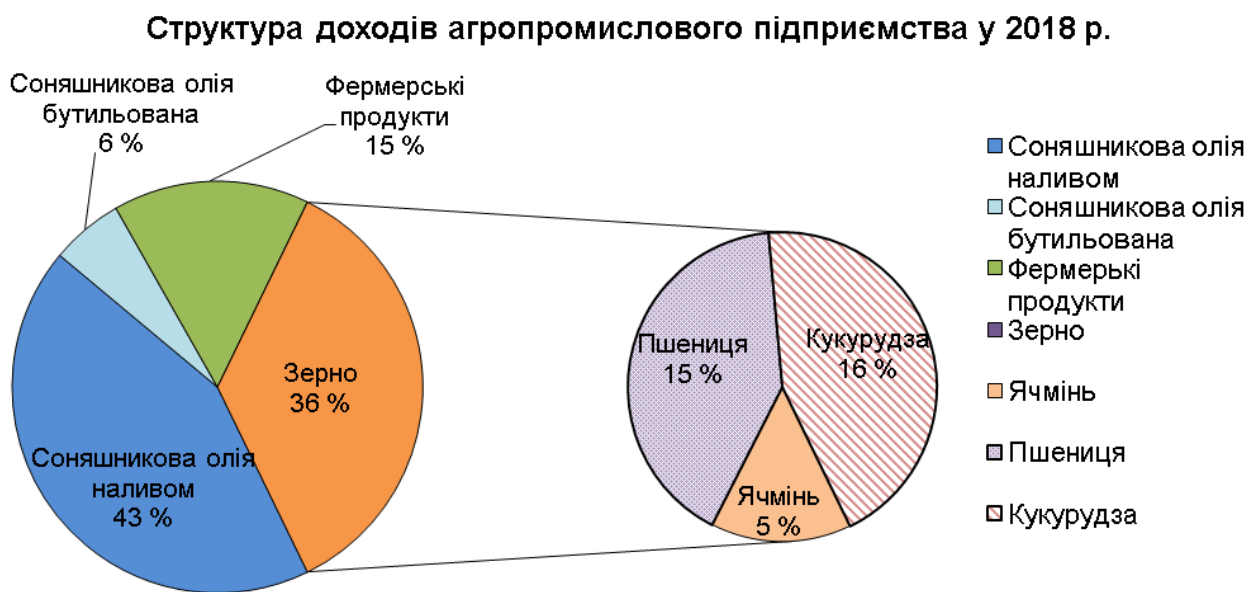


Рис. 7. Вторинна кругова діаграма структури доходів підприємства

1.4. Візуалізація даних з використанням *sparkline MS Excel*.

Часто, працюючи з великою кількістю даних, їх важко зрозуміти. Використовуючи невеликі діаграми, що називаються *sparklines* (*infolines* або *infocells*) та відображаються поряд з даними, ви можете візуально представити значення цих даних, а також спростити виявлення тенденцій у них.

Sparklines (спарклайни) – це невеликі діаграми, які поміщуються всередині окремих клітинок на аркуші *MS Excel*. Завдяки своєму розміру,

спарклайни дозволяють компактно і дуже чітко візуалізувати закономірності у великих наборах даних. Sparklines дозволяють звернути увагу на такі явища, як економічні цикли або сезонне зростання або зниження будь-яких показників, а також висвітлити максимальні та мінімальні значення в різних кольорах. Спарклайни найбільш ефективні, коли вони розміщені близько до даних, які вони представляють.

У MS Excel існують три типи спарклайнів: графік, гістограма (стовпець) та перемога/програв (рис. 8).

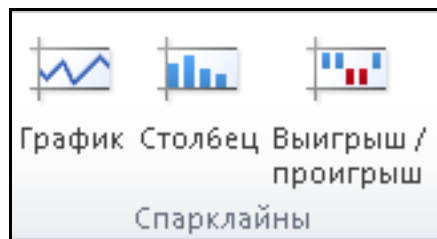


Рис. 8. Типи спарклайнів

Спарклайн "Графік" (Sparkline graphics). На такому типі спарклайну лінія йде зліва направо та відображає зміни у даних у формі тенденцій (рис. 9). Точки з мінімальними значеннями можна представити в іншому кольорі. Спарклайни лише дозволяють швидко визначити, що відбувається, і не дають відповіді на запитання "чому це відбувається?".

	Обсяги реалізації, тис. грн				
Продукція	1 квартал	2 квартал	3 квартал	4 квартал	СПАРКЛАЙН ГРАФІК
Шоколад	4944	3594	3751	2901	
Какао	2895	4996	4879	4670	
Кофе	2730	2157	3921	4494	
Чай	1988	2245	2133	3044	

Рис. 9. Спарклайн "Графік"

Спарклайн "Гістограма" (Sparkline histograms) – це рядки вертикальних стовпчиків, де чим більше значення точки даних, тим вища колонка (рис. 10). Щоб спростити сприйняття, максимальні та мінімальні значення можуть бути виділені іншим кольором.

Спарклайн "Виграш/Програв" (Sparklines of winning/losing). Цей тип спарклайну показує позитивні значення, які відображаються прямокутниками у верхній частині клітинки – це перемоги, негативні значення відповідають прямокутникам знизу – це втрати. Втрати виділені червоним кольором, тому ви можете одразу дізнатись, які значення були збитковими (рис. 11).

	Дохід на початок року, млн. грн			
Філіали	Січень	Лютий	Березень	СПАРКЛАЙН ГІСТОГРАМА
Харків	608	737	574	
Суми	870	667	571	
Чернігів	852	521	944	
Донецьк	730	690	809	

Рис. 10. Спарклайн "Гістограма"

Динаміка вартості акцій компанії						
31.12.2015	31.12.2016	31.12.2017	31.12.2018	31.12.2019	31.12.2020	СПАРКЛАЙН ВИГРАШ/ПРОГРАШ
37%	-9%	24%	12%	-48%	19%	

Рис. 11. Спарклайн "Виграш/Програш"

1.4.1. Створення спарклайнів.

Щоб створити спарклайни, потрібно виділити діапазон даних, який необхідно проаналізувати, потім на вкладці *Вставити* в групі *Спарклайни*, вибрати тип спарклайну, а потім вказати місце розміщення спарклайну (рис. 12).

	Кількість клієнтів станом на 2019 р.				Динаміка зростання клієнтської бази
Філіали	1 кв.	2 кв.	3 кв.	4 кв.	
Київ	560	620	825	1045	
Харків	430	590	720	930	
Дніпро	390	402	605	650	
Одеса	385	430	435	450	

Создание спарклайнов

Выберите нужные данные

Диапазон данных: B22:E25

Выберите место для размещения спарклайнов

Диапазон расположения: \$F\$22:\$F\$25

OK Отмена

Рис. 12. Створення спарклайну

Як правило, спарклайни створюються на основі безперервної серії значень одного ряду даних. Результати побудови спарклайну наведено на рис. 13.

	Кількість клієнтів станом на 2019 р.				Динаміка зростання клієнтської бази
Філіали	1 кв.	2 кв.	3 кв.	4 кв.	
Київ	560	620	825	1045	
Харків	430	590	720	930	
Дніпро	390	402	605	650	
Одеса	385	430	435	450	

Рис. 13. Спарклайн "Графік"

1.4.2. Редагування спарклайнів.

Для роботи з спарклайнами використовуйте контекстну вкладку *Конструктор*, яка з'являється автоматично, коли ви робите активною клітинку, в якій розміщується спарклайн (рис. 14).

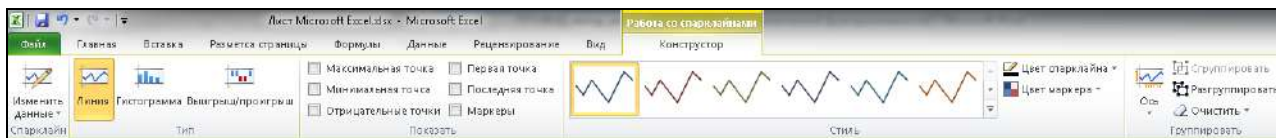


Рис. 14. Майстер редагування спарклайну

Після створення спарклайну ви можете змінити його тип. Для цього в групі *Тип* контекстної вкладки *Конструктор* оберіть кнопку з іншим типом спарклайну.

Змінити стиль або формат спарклайну можливо у вкладці *Конструктор* у групі *Стиль*. Також у цій групі ви можете змінити колір спарклайну, товщину лінії на графіку та управляти кольором маркеру (рис. 14). Важливо знати, що кольори ліній (стовпців) можна змінити незалежно від обраного стилю. Для типів спарклайну "Гістограми" та "Виграш/Програш" неможливо змінити товщину контуру стовпців.

Спарклайни, розташовані в сусідніх осередках, згруповані за замовчуванням. Коли ви вибираєте одну клітинку зі спарклайном, увесь діапазон виділяється. У цьому випадку всі дії конфігурації та дизайн будуть автоматично виконуватися з усіма спарклайнами в діапазоні, навіть якщо обрано лише одну клітинку. Щоб мати можливість працювати з кожним спарклайном індивідуально, використовуйте контекстну вкладку *Конструктор* та оберіть команду *Розгрупувати* (рис. 15). Також щоб вилучити один елемент спарклайну, виберіть клітинку або кілька клітинок зі спарклайнами та в контекстній вкладці групи *Конструктор*, натисніть кнопку *Очистити* (рис. 15).

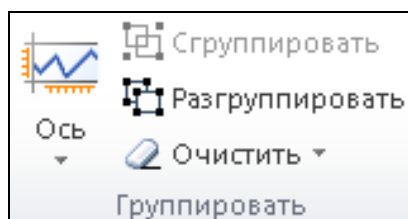


Рис. 15. Розбивка спарклайну на окремі елементи

Для того щоб вибрати конкретні дані на вкладці *Конструктор*, обираємо групу *Показати*. Ця група управляє відображенням точок на спарклайні. Це, своєю чергою, дозволяє вибрати індивідуальні точки даних або всі маркери (значення) даних, вказавши їхню видимість (рис. 16). Для того щоб увімкнути всі значення, установіть прапорець *Маркери*.

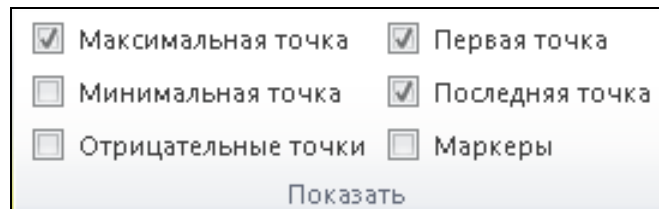


Рис. 16. Відображення точок даних на спарклайні

Стандартні спарклайни всіх типів не мають осей. Якщо є негативні значення, ви можете відобразити горизонтальну вісь. Для цього виділіть клітинку зі спарклайном. На контекстній вкладці *Конструктор* у групі *Групування* натисніть кнопку *Вісь*, а в наступному меню виберіть *Показати вісь* (рис. 17).

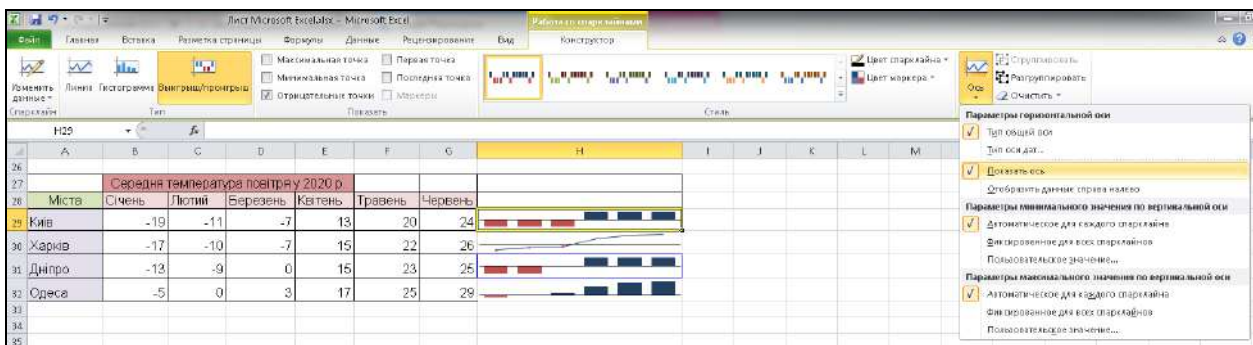


Рис. 17. Додавання осі на спарклайн

1.5. Візуалізація даних з використанням умовного форматування.

Умове форматування забезпечує наочність інформації під час вивчення та аналізу даних, змінюючи формат представлення даних у клітинах залежно від значень вихідних даних. Умове форматування дозволяє підкреслювати незвичайні значення та візуалізувати дані за допомогою гістограм, шкали кольорів і наборів піктограм, які відповідають певним варіаціями даних.

Умове форматування можна застосувати до діапазону клітинок електронної таблиці Microsoft MS Excel або звіту PivotTable. На рис. 18 зображено робочий аркуш зі шістьма блоками, для кожного з яких застосовується окремий тип умовного форматування.

між 10 та 20					Менше середнього					Значення, що повторюються				
23	14	2	22	12	9	18	15	13	12	19	5	12	19	2
20	8	22	21	3	21	13	15	20	11	14	22	15	1	22
3	22	5	1	23	9	10	11	15	14	13	9	10	1	4
24	17	10	18	6	15	1	5	14	21	21	12	4	10	7
12	18	22	13	8	5	1	23	13	17	25	23	24	13	22
9	15	21	18	5	16	4	20	23	11	12	16	12	23	1

Гістограми					Кольорова шкала					Набір значків				
25	2	9	20	11	16	20	13	15	12	✓ 18	🏆 15	✓ 21	✗ 7	✗ 2
6	23	24	5	25	2	8	12	11	20	✗ 15	✗ 3	✓ 25	✓ 21	✗ 7
4	8	1	21	20	1	13	14	22	7	✗ 8	✓ 22	🏆 12	✓ 20	✗ 7
22	16	12	4	21	2	15	18	15	18	🏆 14	11	✗ 4	🏆 16	✗ 9
17	17	21	12	12	22	11	5	14	9	✗ 9	✓ 24	✗ 5	✗ 6	🏆 16
9	17	7	22	7	5	17	6	4	9	✗ 5	✗ 6	✓ 18	🏆 13	✗ 8
6	13	24	22	11	10	19	21	11	13	✓ 25	🏆 11	✓ 19	✓ 20	✗ 8

Рис. 18. Типи умовного форматування

Щоб застосувати умовне форматування до клітинки або діапазону клітинок, ви маєте вибрати одне із правил форматування на вкладці *Головна* в групі *Стилі* у випадковому списку *Умовне форматування* (рис. 19).

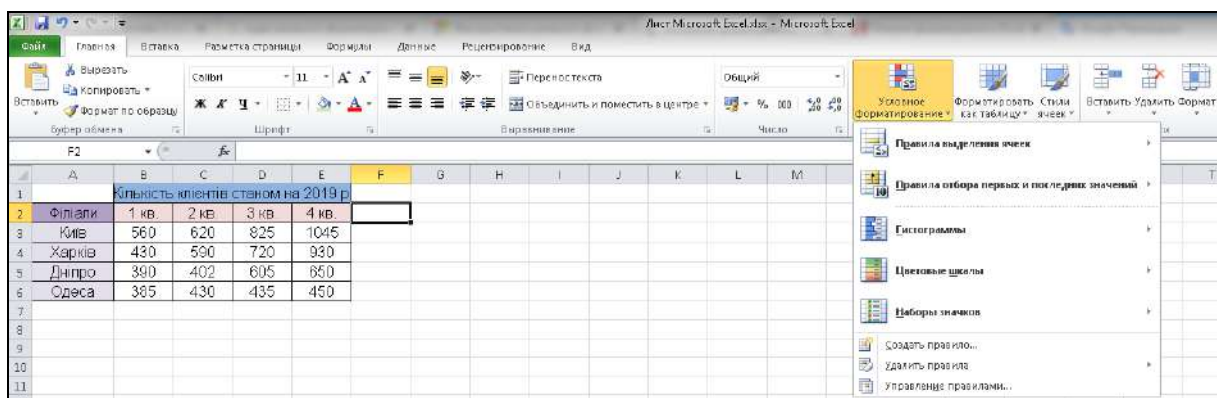


Рис. 19. Створення умовного форматування

Список містить групи правил умовного форматування:

1. *Правило розподілу клітин*. Правила цієї групи ґрунтуються на порівнянні вмісту клітинки з заданим значенням. Серед цих правил існують також правила, які шукають появу певного текстового рядка у вмісті клітинки та перевіряють наявність дублікатів значень.

2. *Правило вибору першого та останнього значень*. У правилах цієї групи виділяються перші або останні 10 значень, перші або останні 10 % значень, а також значення, які знаходяться вище або нижче середнього значення.

3. *Гістограми*. Застосування цих правил призводить до безпосереднього відображення в клітинці горизонтальних стовпчиків, розмір яких

є пропорційним значенню, що міститься в клітинці. Гістограми допомагають вивчити значення в клітинках шляхом їхнього співставлення з іншими значеннями діапазону даних. Гістограми оптимальні для визначення основних показників, особливо у великій кількості даних.

4. *Кольорові шкали.* Відповідно до цих правил, фон клітинок вибраного діапазону буде заповнено градієнтом, а фоновий відтінок кожної клітинки залежить від розміру значення в цій клітинці. Кольорові шкали – це візуальні елементи, які допомагають зрозуміти розподіл даних. Триколірна шкала дозволяє порівнювати діапазон клітин, використовуючи градацію трьох кольорів. Рівень яскравості кольору відповідає високим, середнім або низьким значенням.

5. *Набори значків ("іконок").* Застосування цих правил призводить до відображення в кожній клітині графічного значка, що відповідає значенню у цій клітинці. набір значків використовується для коментування та класифікації даних у трьох-п'яти категоріях, розділених пороговим значенням. Кожна піктограма відповідає діапазону значень.

6. *Створіть правило.* Ця команда дозволяє створювати власні правила умовного форматування, включно з правилами, засновані на логічних формулах.

7. *Видаліть правила.* Ця команда дозволяє видалити правила умовного форматування з вибраного діапазону клітинок чи всього робочого аркуша.

8. *Управління правилами.* Ця команда відкриває діалогове вікно *Умови умовного форматування*, де ви можете створювати нові правила, редагувати існуючі правила або видаляти вибрані правила.

За допомогою умовного форматування ви можете обрати значення, які відповідають будь-якому стану (більше, менше, між, рівні тощо). Для цього оберіть клітинку або діапазон клітинок. На вкладці *Головна* в групі *Стилі* у випадному списку умовного форматування виберіть команду *Правила вибору клітинок*, а потім виберіть умову. Відкриється діалогове вікно для налаштування параметрів умовного форматування (рис. 20).

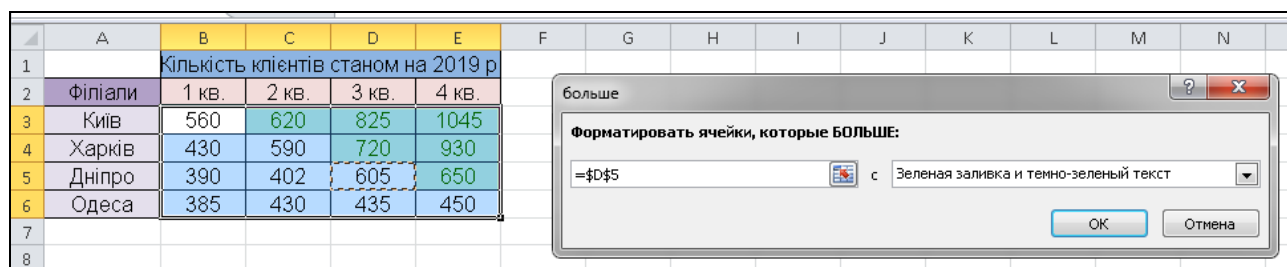


Рис. 20. Налаштування параметрів умовного форматування

Назва і вміст вікна налаштувань параметрів умовного форматування залежать від обраного правила та заданої умови.

Правила умовного форматування можуть бути створені, змінені та видалені. Виберіть клітинку або діапазон клітинок. На вкладці *Головна* в групі *Стилі* у випадному списку умовного форматування натисніть кнопку *Управління правилами*. У діалоговому вікні *Умови умовного форматування* натисніть кнопку *Створити правило* (рис. 21).

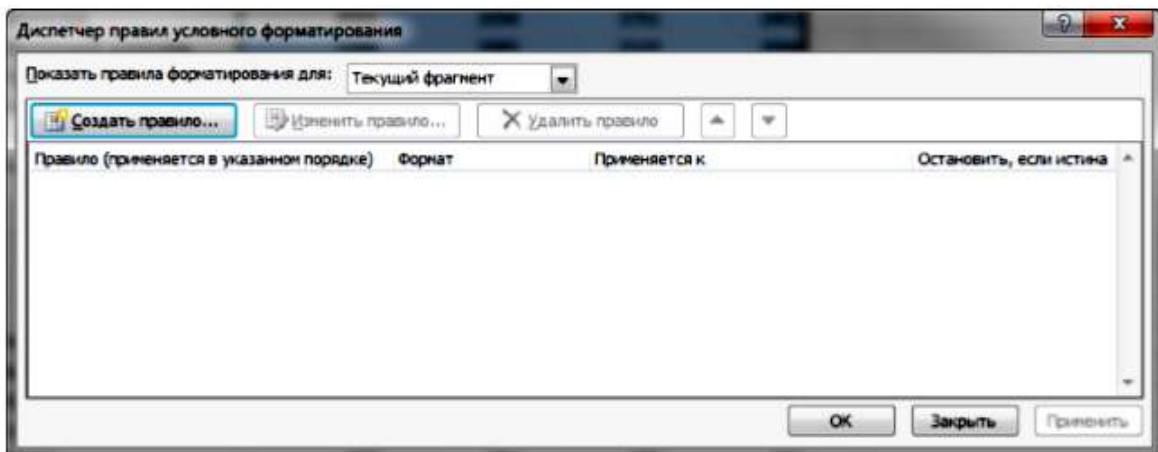


Рис. 21. Розширене форматування

З'явиться діалогове вікно *Створення правила форматування* (рис. 22), в якому ви задаєте нові параметри форматування.

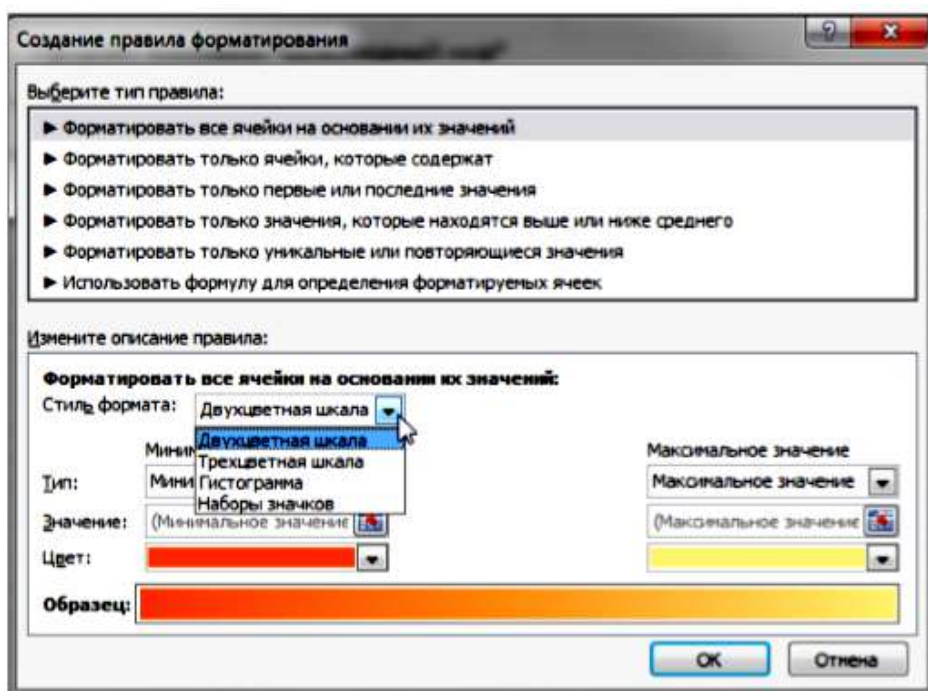


Рис. 22. Створення нового правила форматування

Щоб змінити правило умовного форматування, виберіть правило, а потім натисніть *Редагувати правило*. У діалоговому вікні *Редагування форматування правил* можна вибрати тип іншого правила, змінити умови та параметри форматування (рис. 23).

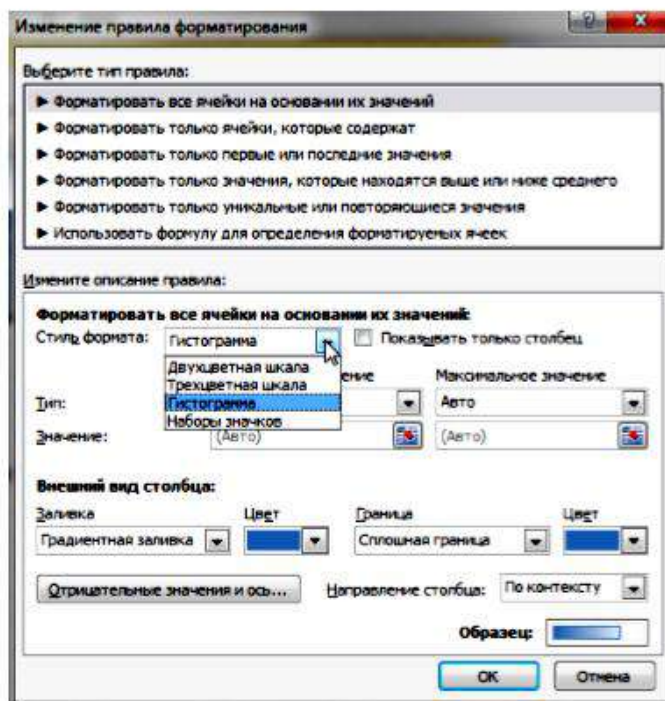


Рис. 23. Редагування правил умовного форматування

Якщо існує декілька правил для одного діапазону, ви можете змінити порядок їхнього застосування, використовуючи кнопки *Вгору* та *Вниз* у діалоговому вікні *Умови умовного форматування*.

2. Створення зведеної таблиці.

Зведена таблиця – це потужний інтерактивний інструмент аналізу великих масивів даних у *MS Excel*, що дозволяє проводити глибокий аналіз наявної інформації за допомогою вибірки, групування, узагальнення та підсумовування великих обсягів даних.

До переваг використання зведених таблиць відносять:

- 1) можливість детально аналізувати числові дані;
- 2) виконання запитів до великих обсягів даних зручними способами;
- 3) отримання проміжних підсумків та агрегації числових даних, зведення даних за категоріями й підкатегоріями та створення настроюваних обчислень і формул;
- 4) розгортання та згортання рівнів даних для виокремлення результатів і деталізації потрібних зведених даних;

5) перетворення рядків на стовпці або стовпців на рядки, завдяки чому можна переглядати різні форми зведення вихідних даних;

6) фільтрування, сортування, групування й умовного форматування найкорисніших і найцікавіших підмножин даних для виділення винятково потрібної інформації;

7) представлення лаконічних і наочних інтерактивних або друкованих звітів із зауваженнями;

8) виконання автоматичного добору інформації та її інтерактивне оновлення (значення зведених таблиць оновлюються автоматично під час зміни значення будь-якої пов'язаної з ними таблиці);

9) на основі зведених таблиць будуються діаграми, що динамічно перебудовуються разом зі зміною зведеної таблиці.

Існує два способи створення зведеної таблиці:

1) використання надбудови *Майстер зведених таблиць та діаграм*;

2) використання стандартної вбудованої функції *Зведена таблиця* на вкладці *Вставка* панелі інструментів.

Розглянемо на прикладі етапи побудови зведеної таблиці.

На основі даних про витрат родини створимо зведену таблицю (табл. 4).

Таблиця 4

Витрати родини

Місяць	Тип витрат	Сума, грн
Січень	Транспорт	740
Січень	Продукти харчування	2 350
Січень	Побутові товари	1 750
Січень	Розваги	1 000
Лютий	Транспорт	1 150
Лютий	Продукти харчування	2 400
Лютий	Побутові товари	2 250
Лютий	Розваги	1 250
Березень	Транспорт	900
Березень	Продукти харчування	2 600
Березень	Побутові товари	2 000
Березень	Розваги	1 200

Виділяємо діапазон клітинок, на основі яких потрібно створити зведену таблицю. При цьому, діапазон даних немає містити пустих рядків або стовпців. Заголовки мають складатися лише з одного рядка.

На панелі інструментів обираємо вкладку *Вставка*, потім натискаємо кнопку *Зведена таблиця* (рис. 24).

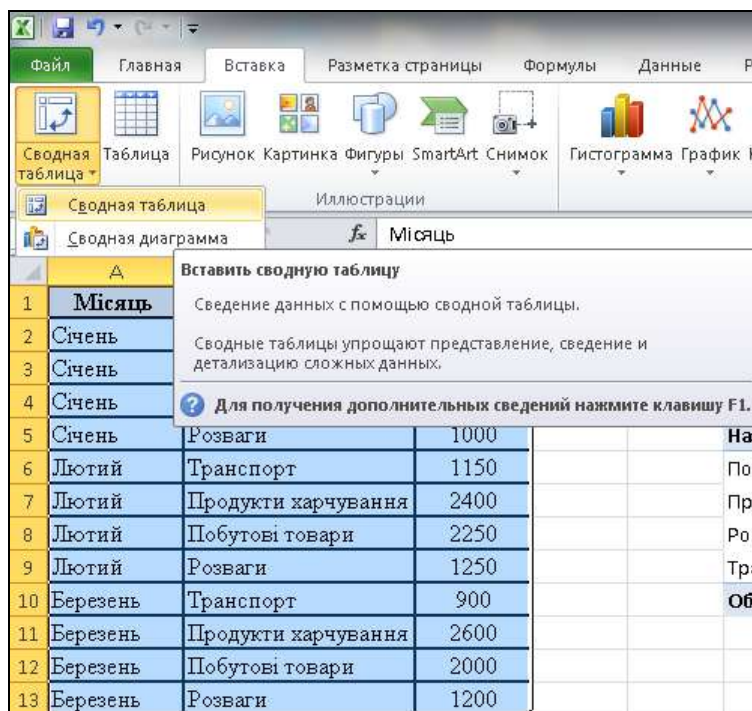


Рис. 24. Створення зведеної таблиці

На наступному етапі задаємо параметри створення зведеної таблиці (рис. 25). У розділі *Виберіть дані для аналізу* встановіть перемикач *Вибрати таблицю або діапазон*. У полі *Таблиця/Діапазон* перевірте обраний діапазон клітинок. У розділі *Виберіть розташування звіту зведеної таблиці* виберіть пункт *Новий аркуш*, щоб розташувати зведену таблицю на новому аркуші або *Наявний аркуш*, а потім задайте потрібне місце розташування готової таблиці. Натисніть кнопку *ОК*.

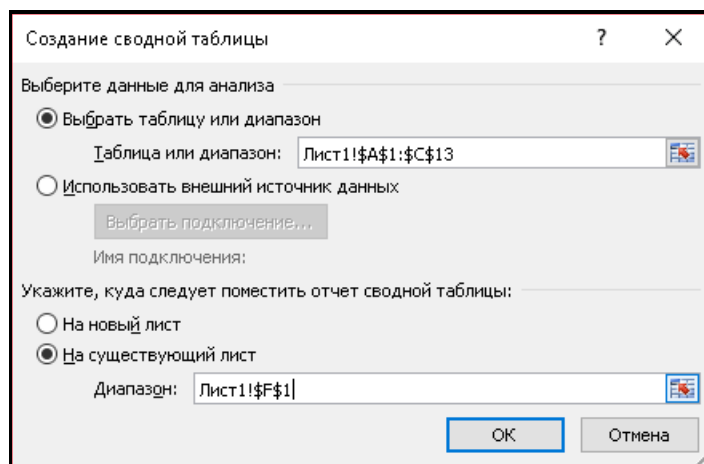


Рис. 25. Визначення параметрів зведеної таблиці

У результаті виконання зазначених операцій відкривається форма створення зведеної таблиці (рис. 26).

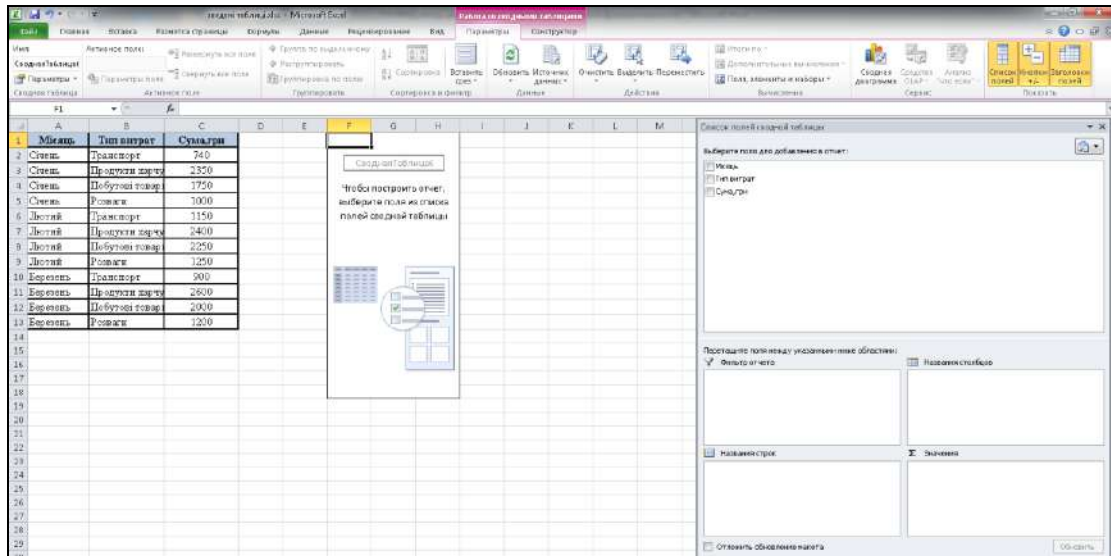


Рис. 26. Форма створення зведеної таблиці

У правій частині вікна розташований список полів таблиці, а нижче чотири області: фільтр звіту, назви рядків, назви стовпців та значення.

Для того щоб заповнити зведену таблицю перетягуємо мишкою потрібні поля таблиці до областей, що відповідають нашим запитам (завданням). Отже, створимо зведену таблицю витрат родини, у якій буде встановлено 2 фільтри (за типами витрат та за місяцями), а також буде визначено загальний підсумок витрат (рис. 27).

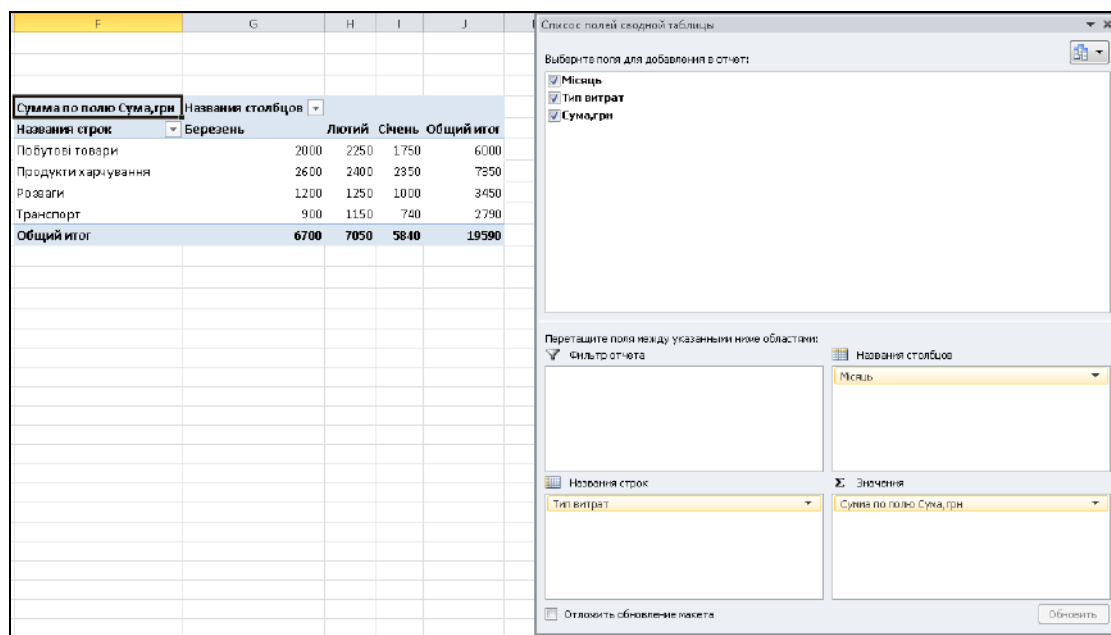


Рис. 27. Заповнення списку полів зведеної таблиці

Потрібно відзначити, що всі арифметичні розрахунки даних підтягнутих з іншої таблиці можливі тільки в останній області – *Значення*. Під час того, як ми робили ці маніпуляції з перенесенням полів в області, відповідно змінювалася і сама таблиця в лівій частині вікна.

Існує також другий спосіб створення зведеної таблиці – *Майстер зведених таблиць та діаграм*.

2.1. Робота зі зведеною таблицею.

Створена зведена таблиця дозволяє аналізувати витрати родини окремо по місяцям, по типам витрат та знаходити загальний підсумок витрат (рис. 28).

Сумма по полю Суммагрн	Названия столбцов			
Названия строк	Березень	Лютий	Січень	Общий итог
Сортировка от А до Я	2000	2250	1750	6000
Сортировка от В до А	2600	2400	2350	7350
Дополнительные параметры сортировки...	1200	1250	1000	3450
Удалить фильтр с "Тип витрат"	900	1150	740	2790
Фильтры по подписи	6700	7050	5840	19590
Фильтры по значению				

Сумма по полю Суммагрн	Названия столбцов		
Названия строк	Лютий	Січень	Общий итог
Сортировка от А до Я	2250	1750	6000
Сортировка от В до А	2400	2350	7350
Дополнительные параметры сортировки...	1250	1000	3450
Удалить фильтр с "Месяць"	1250	1000	3450
Фильтры по подписи	1150	740	2790
Фильтры по значению	7050	5840	19590

Рис. 28. Використання фільтрів зведеної таблиці

Управляти даними в таблиці можна як завгодно. Для цього необхідно увійти у режим форми списку полів (вкладка *Параметри*, кнопка *Список полів*) та змінювати місце полів таблиці в областях зведеної таблиці. Така зміна в автоматичному порядку змінить зовнішній вигляд готової зведеної таблиці.

Редагування зведеної таблиці можливе за допомогою вкладок *Параметри* та *Конструктор* панелі інструментів після того, як зробили активною саму зведену таблицю. Наприклад, використовуючи кнопку *Зведена діаграма* групи *Сервіс* вкладки *Параметри*, користувач може автоматично візуалізувати результати побудови зведеної таблиці (рис. 29). Доцільно зазначити, що зведена діаграма також є динамічною та дозволяє використовувати фільтри під час демонстрації відповідних задач.

На вкладці *Конструктор* ви можете обрати (змінити) макет звіту, задати параметри стилю зведеної таблиці (рис. 30).

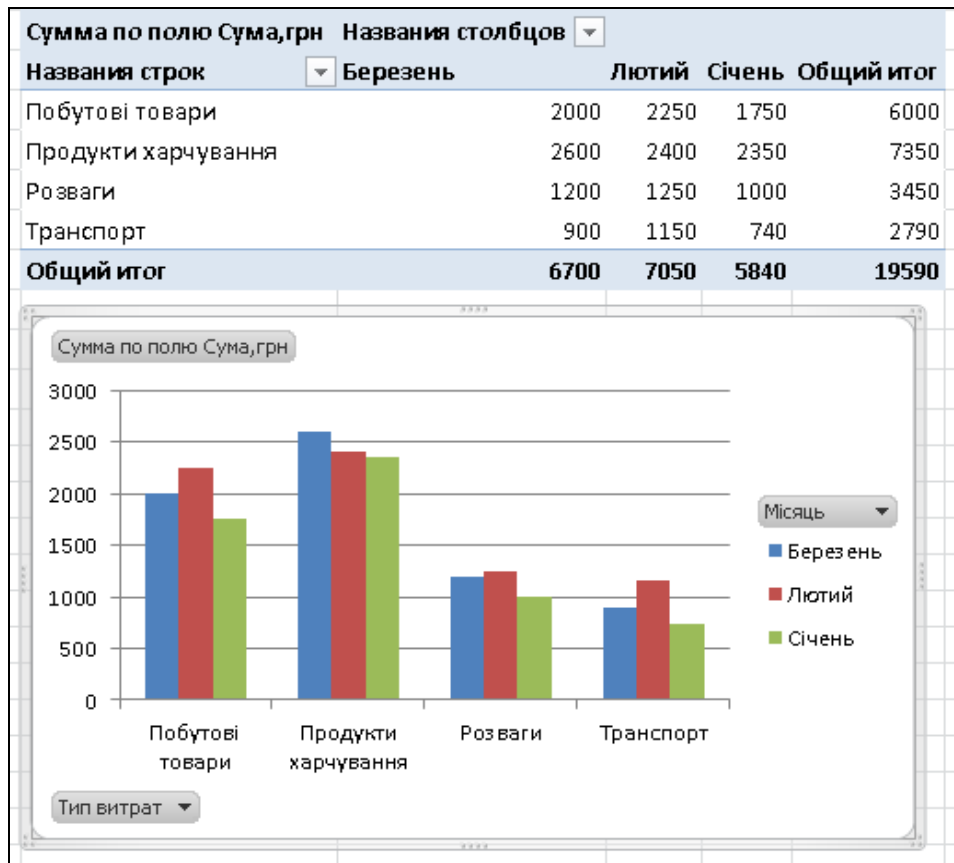


Рис. 29. Побудова зведеної діаграми

Місяць	Тип витрат	Сума,грн
Січень	Транспорт	740
Січень	Продукти харчування	2350
Січень	Побутові товари	1750
Січень	Розваги	1000
Лютий	Транспорт	1150
Лютий	Продукти харчування	2400
Лютий	Побутові товари	2250
Лютий	Розваги	1250
Березень	Транспорт	900
Березень	Продукти харчування	2600
Березень	Побутові товари	2000
Березень	Розваги	1200

Рис. 30. Редагування зведеної таблиці

Доцільно зазначити, що під час роботи зі зведеною таблицею користувач може використовувати стандартні елементи візуалізації даних *MS Excel*, зокрема спарклайни та умовне форматування, що забезпечить презентабельність і дозволить покращити наочність сприйняття аналізованих даних.

Завдання:

1. Використовуючи власний інформаційний простір (статистичні дані за обраною тематикою), необхідно побудувати:

1.1. Комбіновані діаграму.

1.2. Вторинну кругову діаграму.

2. До статистичних даних завдання 1 застосувати будь-які два типи спарклайнів і два типи умовного форматування.

3. Використовуючи дані сайту Державної служби статистики, побудуйте зведену таблицю, що буде відображати соціально-економічне становище регіонів України.

Результати роботи оформіть у вигляді звіту з лабораторної роботи.

Завдання для самостійного опрацювання

Запитання, що виносяться на самостійну роботу студентів:

1. Які сучасні види спеціалізованих програмних продуктів оброблення та аналізу великих масивів даних (ВМД) вам відомі?

2. Які вбудовані програмні продукти *MS Office* для аналізу ВМД використовують у бізнес-аналітиці?

3. Який спектр функцій дозволяють виконувати вбудовані програмні продукти *MS Office* для оброблення ВМД?

4. Здійсніть порівняльний аналіз вбудованих і спеціалізованих програмних продуктів оброблення ВМД щодо типів виконуваних завдань.

5. Наведіть приклади використання спеціалізованих програмних продуктів для вирішення конкретних бізнес-завдань.

Завдання 1. Використовуючи дані табл. 5, побудуйте зведену таблицю, що буде відображати загальну кількість підприємств галузі, сумарний обсяг виробленої продукції, а також встановіть фільтри, що дозволять групувати підприємства за кількістю робітників і середнім розміром основних фондів підприємства.

Відредагуйте стиль стандартної зведеної таблиці відповідно до власних уподобань.

За результатами створеної зведеної таблиці побудуйте зведену діаграму.

Групування підприємств за кількістю робітників

Групи підприємств за кількістю робітників	Кількість підприємств, од.	Кількість робітників, осіб	Обсяг виробленої продукції, млн грн	Середню фактичну продуктивність праці одного робітника, грн	Обсяг основних фондів, тис. грн	Середній розмір основних фондів одного підприємства	Середня величина виробленої продукції одним підприємством, млн грн
43 – 78,5	5	303	850	2,805	725	145	170
78,5 – 114	12	1 170	4 160	3,556	3 301	375	346,667
114 – 149,5	5	646	2 500	3,870	2 128	426	500
149,5 – 185	4	703	2 830	4,026	2 520	630	707,5
185 – 220,5	2	407	1 610	3,956	1 990	995	805
220,5 - 256	2	506	1 910	3,775	2 050	1 025	955

Завдання 2. Використовуючи дані табл. 6, побудуйте два типи спарклайнів і три типи умовного форматування. Вибір типів візуалізації обґрунтуйте.

Таблиця 6

Динаміка продажу товару А у регіональній мережі супермаркетів, тис. шт.

Роки	1 квартал	2 квартал	3 квартал	4 квартал
2016	1 020	1 040	1 000	950
2017	1 450	1 500	1 520	1 510
2018	1 690	1 640	1 690	1 770
2019	2 100	2 160	2 200	2 040

На першому спарклайні позначте мінімальне та максимальне значення. Для другого – застосуйте одночасно різні типи спарклайнів. Для одного з типів умовного форматування створіть власне правило побудови.

Тема 2. Обробка таблиць, графіків і побудова моделей за допомогою MS Excel

Лабораторна робота 2 Аналіз вибірки та регресійний аналіз у MS Excel

Мета: опанувати навички аналізу статистичних даних за допомогою описової статистики, розв'язання оптимізаційних задач і проведення регресійного аналізу даних у пакеті *MS Excel*.

Завдання: за допомогою *MS Excel* та надбудови *Аналіз даних* провести аналіз вибірки та регресійний аналіз даних.

Методичні рекомендації

1. Знайомство з надбудовою Аналіз даних MS Excel.

Для проведення комплексного статистичного або інженерного аналізу даних використовують надбудову *Аналіз даних*, що містить розширений перелік аналітичних макروفункцій. Деякі інструменти надбудови дозволяють представити результати аналізу в графічному вигляді. Повний перелік вбудованих функцій (засобів чи інструментів аналізу) наведено на рис. 31.

- Однофакторный дисперсионный анализ (ANOVA: single factor);
- Двухфакторный дисперсионный анализ с повторениями (ANOVA: two factor with replication);
- Двухфакторный дисперсионный анализ без повторений (ANOVA: two factor without replication);
- Корреляция (Correlation);
- Ковариация (Covariance);
- Описательная статистика (Descriptive Statistics);
- Экспоненциальное сглаживание (Exponential Smoothing);
- Двухвыборочный F-тест для дисперсии (F-test Two Sample for Variances);
- Анализ Фурье (Fourier Analysis);
- Гистограмма (Histogram);
- Скользящее среднее (Moving average);
- Генерация случайных чисел (Random Number Generation);
- Ранг и Перцентиль (Rank and Percentile);
- Регрессия (Regression) - простая регрессия; для множественной регрессии см. [здесь](#);
- Выборка (Sampling);
- Парный двухвыборочный t-тест для средних (t-Test: Paired Two Sample for Means);
- Двухвыборочный t-тест с одинаковыми дисперсиями (t-Test: Two-Sample Assuming Equal Variances);
- Двухвыборочный t-тест с различными дисперсиями (t-Test: Two-Sample Assuming Unequal Variances);
- Двухвыборочный z-тест для средних (z-Test: Two Sample for Means).

Рис. 31. Перелік функцій надбудови Аналіз даних MS Excel

Щоб виконати статистичний аналіз даних за допомогою цієї надбудови, необхідно перейти на вкладку *Дані*, в меню групи *Аналіз* обрати кнопку *Аналіз даних* (рис. 32).

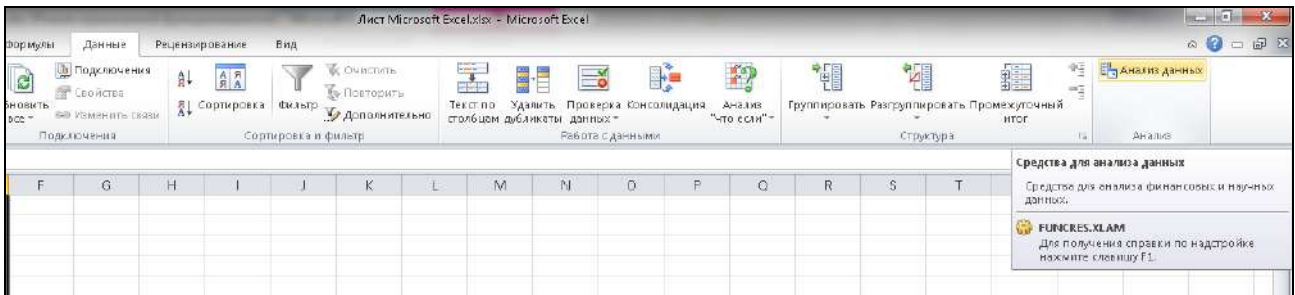


Рис. 32. Виклик надбудови *Аналіз даних*

Якщо надбудова *Аналіз даних* відсутня, її можна встановити (підключити) за допомогою команд *Файл / Параметри / Надбудови / Пакет аналізу VBA* (Analysis ToolPak VBA).

З метою опанування навичок аналізу статистичних даних за допомогою описової статистики та проведення регресійного аналізу даних у пакеті *MS Excel* розглянемо такий приклад.

Необхідно перевірити наявність лінійного множинного зв'язку між прибутком підприємств та показниками, що характеризують їхню діяльність (у табл. 7 наведено значення показників станом на 2019 р.).

Таблиця 7

Вихідні данні для побудови багатofакторної економетричної моделі

Номер підприємства	Прибуток підприємства, млн грн (Y)	Випуск продукції, тис. грн (X1)	Обсяг роздрібного товарообороту підприємств, млн грн (X2)
1	2	3	4
1	6	40	30
2	4,6	20	33
3	4,4	31	20
4	4,5	32	25
5	5,5	34	29
6	4,8	35	20
7	5,1	37	21
8	5,2	32	20

1			
9	7	39	35
10	5,3	35	30
11	7,5	50	35
12	7,7	37	30
13	7,3	50	40
14	7	38	42
15	6,7	50	39
16	5,7	35	35
17	6	46	36
18	6,4	49	38
19	7,1	54	41
20	6,3	45	34

MS Excel має інструмент, за допомогою якого здійснюється оброблення даних з цілого ряду основних статистичних показників. Це *Описова статистика (Descriptive Statistics)*. З його допомогою можна швидко обробити масиви даних та отримати інформацію щодо значимості основних статистичних критеріїв.

Під описової статистикою розуміють систематизацію емпіричних даних з цілого ряду основних статистичних критеріїв. Причому на основі отриманого результату з цих підсумкових показників можна сформулювати загальні висновки про досліджуваний масив даних.

Отже, створюємо вихідний інформаційний простір на аркуші *MS Excel* і викликаємо функцію *Описова статистика (Descriptive Statistics)* (рис. 33).

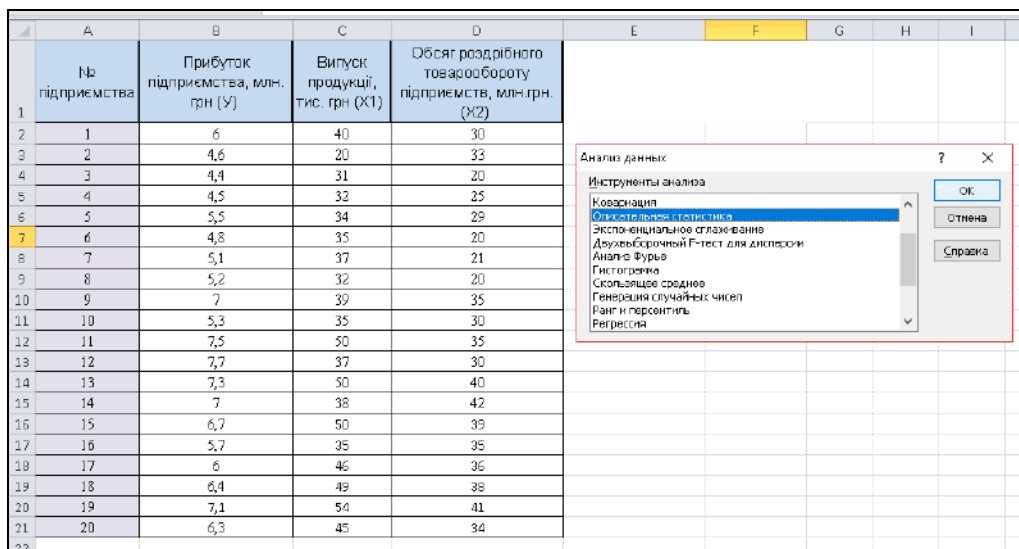


Рис. 33. Виклик функції *Описова статистика*

Інструмент *Описова статистика* пропонує таблицю основних статистичних характеристик для однієї або декількох множин вихідних значень (рис. 34).

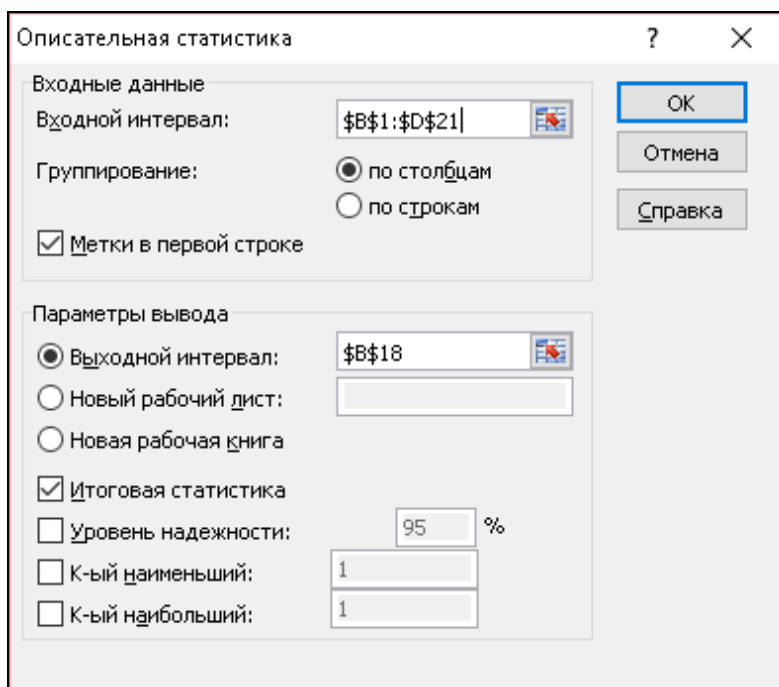


Рис. 34. Встановлення параметрів для функції *Описова статистика*

Вихідний інтервал цього інструменту містить такі статистичні характеристики: середнє значення, стандартна помилка, медіана, мода, стандартне відхилення, дисперсія, коефіцієнт ексцесу, коефіцієнт асиметрії, інтервал (розмах), мінімальне та максимальне значення, сума, кількість значень, k-е найбільше і найменше значення (для будь-якого заданого значення k) та рівень значущості для середнього.

Встановіть прапорець *Підсумкова статистика*, якщо потрібен повний список характеристик, в іншому випадку виберіть конкретні характеристики, які мають бути присутніми у вихідній таблиці. Більшість з отриманих характеристик також можна отримати за допомогою вбудованих статистичних формул *MS Excel*.

Отже, можна розпочинати безпосередній аналіз отриманих результатів. За допомогою інструменту описової статистики були розраховані показники.

Результати використання функції *Описова статистика* наведено на рис. 35.

Прибуток підприємства, млн. грн (У)		Випуск продукції, тис. грн (Х1)		Обсяг роздрібного товарообороту підприємств, млн.грн. (Х2)	
Среднее	6,005	Среднее	39,45	Среднее	31,65
Стандартная оц	0,236584802	Стандартная оц	1,90218	Стандартная ош	1,621038
Медиана	6	Медиана	37,5	Медиана	33,5
Мода	6	Мода	35	Мода	30
Стандартное о	1,058039398	Стандартное о1	8,506808	Стандартное отк	7,249501
Дисперсия выб	1,119447368	Дисперсия выб	72,36579	Дисперсия выбо	52,55526
Эксцесс	-1,2963328	Эксцесс	-0,09613	Эксцесс	-0,93826
Асимметрично	0,012440028	Асимметрично	-0,12549	Асимметричнос	-0,44599
Интервал	3,3	Интервал	34	Интервал	22
Минимум	4,4	Минимум	20	Минимум	20
Максимум	7,7	Максимум	54	Максимум	42
Сумма	120,1	Сумма	789	Сумма	633
Счет	20	Счет	20	Счет	20

Рис. 35. Основні статистичні характеристики для досліджуваного масиву даних

Асиметричність (міра відхилення розподілу частоти від симетричного розподілу). Якщо спостереження підкоряються нормальному розподілу, то асиметрія дорівнює нулю. Для перевірки на нормальний розподіл можна застосовувати таке правило: якщо асиметрія значно відрізняється від нуля, то гіпотезу про те, що дані взяті з нормально розподіленої генеральної сукупності, слід відкинути. Якщо коефіцієнт асиметрії більше нуля, то розподіл ймовірностей є асиметричним і має правобічний нахил. Якщо коефіцієнт асиметрії менше нуля, асиметрія від'ємна, ряд розподілу на графіку буде мати лівобічний нахил.

Інтервал – це розмах варіації у вибірці.

Мінімум – це мінімальне значення ознаки у вибірці.

Максимум – це максимальне значення ознаки у вибірці.

Стандартне відхилення – це міра розкиду виміряних величин. Воно дорівнює квадратному кореню з дисперсії.

Дисперсія вибірки – це квадрат стандартного відхилення. Ця характеристика також є мірою розкиду виміряних величин.

Сума – це сумарне значення елементів вибірки.

Ексцес – це показник гостровершиності розподілу. Ексцес нормального розподілу дорівнює нулю. Якщо ексцес деякого розподілу відмінний від нуля, то крива щільності цього розподілу відрізняється від кривої щільності

нормального розподілу: якщо ексцес додатній, то крива теоретичного має вищу та "гострішу" вершину ніж крива нормального; якщо ексцес від'ємний, то крива теоретичного має нижчу та "плоскішу" вершину, ніж крива нормального.

Середнє значення (арифметичне середнє значень, що аналізуються) визначається як сума значень, поділена на їхню кількість.

Стандартна помилка визначається як стандартне відхилення, поділене на квадратний корінь з обсягу вибірки.

Медіана – це значення ознаки в статистичній одиниці, що стоїть у середині ранжованого ряду та ділить сукупність на дві рівні за чисельністю частини.

Мода – це значення, що найчастіше зустрічається в сукупності.

Рахунок – це обсяг вибірки.

Рівень надійності (95,4 %) – це гранична помилка вибірки, оцінена з заданим рівнем надійності.

Проаналізувавши отримані результати, користувач має загальне уявлення щодо статистичних характеристик досліджуваних показників, закономірності розподілу, тобто може оцінити ступінь однорідності сукупності, що є передумовою використання інших статистичних методів.

2. Регресійний аналіз у MS Excel.

Проведемо регресійний аналіз, тобто застосуємо метод визначення відокремленого і спільного впливу факторів на результативну ознаку та кількісної оцінки цього впливу шляхом використання відповідних критеріїв. Регресійний аналіз проводиться на основі побудованого рівняння регресії та визначає внесок кожної незалежної змінної у варіацію досліджуваної (прогнозованої) залежної змінної величини.

Основним завданням регресійного аналізу є визначення впливу факторів на результативний показник (в абсолютних показниках). Передусім для цього необхідно підібрати та обґрунтувати рівняння зв'язку, що відповідає характеру аналітичної стохастичної залежності між досліджуваними ознаками. Рівняння регресії показує, як в середньому змінюється результативна ознака (Y) під впливом зміни факторних ознак (X_i).

У загальному вигляді рівняння регресії можна подати так:

$$Y_x = f(x_1, x_2, \dots, x_n),$$

де Y_x – залежна змінна величина;

x – незалежні змінні величини (фактори).

Залежно від кількості змінних величин розрізняють два види регресійного аналізу (однофакторний та багатофакторний). У нашому випадку досліджується вплив декількох незалежних (екзогенних) змінних X на одну залежну (ендогенну) змінну Y . Отже, виконаємо багатофакторний регресійний аналіз в *MS Excel*. Для цього викликаємо надбудову *Аналіз даних*, обираємо режим (функцію) *Регресія* та задаємо основні параметри регресії (рис. 36).

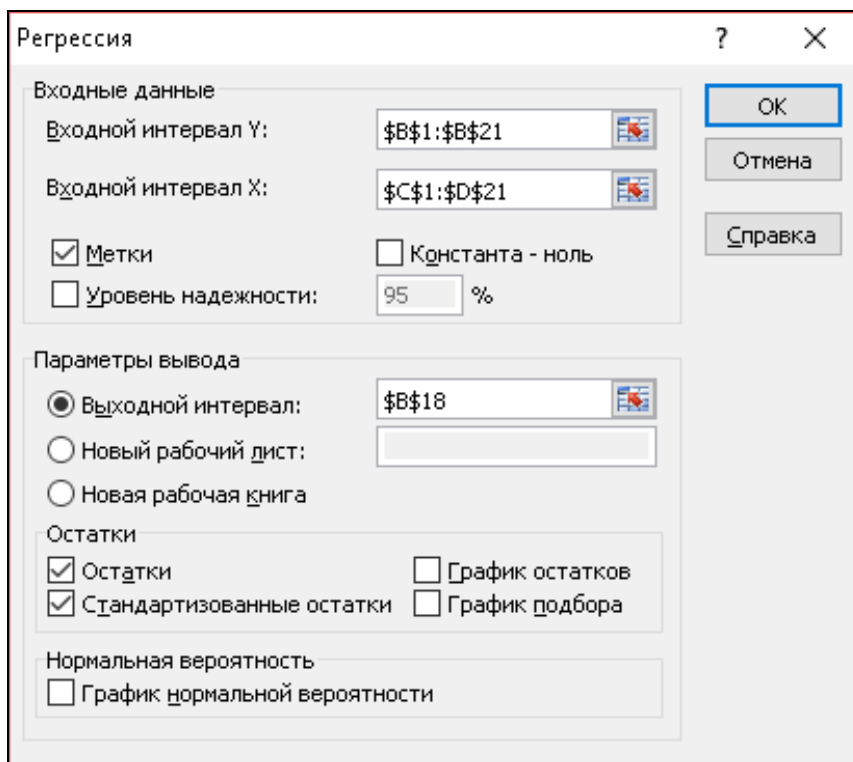


Рис. 36. Значення параметрів режиму *Регресія*

У діалоговому вікні режиму *Регресія* задаються такі параметри:

1. Вхідний інтервал Y – вводяться посилання на клітинки, що містять дані за результативною ознакою. Діапазон має складатися з одного стовпця.
2. Вхідний інтервал X – вводиться посилання на клітинки, що містять факторні ознаки. Максимальна кількість вхідних діапазонів (стовпців) дорівнює 16.
3. Мітки – команда активізується, якщо перший рядок (стовпець) у вхідному діапазоні містить заголовки. Якщо заголовки відсутні, необхідно її деактивувати. В цьому випадку будуть автоматично створені стандартні назви для даних вихідного діапазону.

4. Рівень надійності – активізується, якщо в полі, яке розташовано навпроти, необхідно ввести рівень надійності відмінний від рівня 95 %. Встановлений рівень надійності використовується для перевірки значущості коефіцієнта детермінації R^2 і коефіцієнтів регресії a_i .

5. Константа-нуль – активізується, якщо потрібно, щоб лінія регресії йшла через початок координат.

6. Вихідний інтервал / Новий робочий лист / Нова робоча книга.

У полі *Вихідний інтервал* активізується поле, в яке необхідно ввести посилання на ліву верхню клітинку вихідного діапазону. Розмір вихідного діапазону буде визначений автоматично, й на екрані з'явиться повідомлення у випадку можливого накладення вихідного діапазону на вхідні дані.

У полі *Новий робочий лист* відкривається новий лист, на якому починаючи з клітинки A1, демонструються результати аналізу.

У полі *Нова робоча книга* відкривається нова книга, на першому листі якої, починаючи з клітинки A1, вставляються результати аналізу.

7. Залишки – активізується, якщо потрібно включити в вихідний діапазон стовпець із залишками.

8. Стандартизовані залишки – активізується, якщо потрібно включити в вихідний діапазон стовпець із стандартизованими залишками.

9. Графік залишків – активізується, якщо потрібно вивести на робочий лист графіки залежності залишків від факторних ознак (x_i).

10. Графік підбору – активізується, якщо потрібно вивести на робочий лист графіки залежності теоретичних результативних значень (\hat{y}) від факторних ознак (x_i).

11. Графік нормальної ймовірності – активізується, якщо потрібно вивести на робочий лист графік залежності значень, що спостерігаються (y) від інтервалів персентилей, які автоматично формуються.

Розраховані показники, отримані в режимі *Регресія* наведено на рис. 37.

Отримані результати з регресійної статистики відповідають таким статистичним показникам:

множинний R – коефіцієнт кореляції (R);

R-квадрат – коефіцієнт детермінації (R^2);

стандартна помилка – залишкове стандартне відхилення;

спостереження – кількість спостережень (n).

	A	B	C	D	E	F	G	H	I	J	K	L	M	N
2	1	6	40	30		Вивід итогов								
3	2	4,6	20	33										
4	3	4,4	31	20		<i>Регрессионная статистика</i>								
5	4	4,5	32	25		Множественный R	0,81368							
6	5	5,5	34	29		R-квадрат	0,661994							
7	6	4,8	35	20		Нормированный R-квадрат	0,622223							
8	7	5,1	37	21		Стандартная ошибка	0,650304							
9	8	5,2	32	20		Наблюдения	20							
10	9	7	39	35										
11	10	5,3	35	30		<i>Дисперсионный анализ</i>								
12	11	7,5	50	35										
13	12	7,7	37	30										
14	13	7,3	50	40		Регрессия	2	14,08028967	7,040145	16,64751	9,90487E-05			
15	14	7	38	42		Остаток	17	7,189210327	0,422895					
16	15	6,7	50	39		Итого	19	21,2695						
17	16	5,7	35	35										
18	17	6	46	36										
19	18	6,4	49	38										
20	19	7,1	54	41										
21	20	6,3	45	34										
22						Вывод остатка								
23														
24														
25						<i>Наблюдение</i>	<i>ток либр</i>	<i>статки</i>	<i>залишки</i>					
26						1	5,926786	0,073213664	0,119022					
27						2	4,999764	-0,399764157	-0,64989					
28						3	4,758703	-0,398703304	-0,58314					
29						4	5,145782	-0,645781612	-1,04984					
30						5	5,522975	-0,022974502	-0,03735					
31						6	4,989804	-0,189804113	-0,29881					

Рис. 37. Результати регресійного аналізу

Результати дисперсійного аналізу використовуються для перевірки значущості коефіцієнта детермінації R^2 . Стівпці мають таку інтерпретацію:

стівпець df – кількість ступенів свободи. Для рядка *Регресія* кількість ступенів свободи визначається кількістю факторних ознак m у рівнянні регресії $k_{\Phi} = m$. Для рядка *Залишки* кількість ступенів свободи визначається кількістю спостережень i і кількістю змінних у рівнянні регресії $m + 1$: $k_0 = n - (m + 1)$. Для рядка *Разом* кількість ступенів свободи визначається сумою $k_Y = k_{\Phi} + k_0$;

стівпець SS – сума квадратів відхилень. Для рядка *Регресія* – це сума квадратів відхилень теоретичних даних від середнього: $SS_{\Phi}^2 = \sum(\hat{y}_i - \bar{y})^2$. Для рядка *Залишки* – це сума квадратів відхилень емпіричних даних від теоретичних: $SS_3^2 = \sum(y_i - \hat{y}_i)^2$. Для рядка *Разом* – це сума квадратів відхилень емпіричних даних від середнього: $SS_Y^2 = \sum(y_i - \bar{y})^2$ або $SS_Y^2 = SS_{\Phi}^2 + SS_3^2$;

стівпець MS – дисперсії, що розраховані за формулою: $MS = \frac{SS}{df}$.

Для рядка *Регресія* – це факторна дисперсія σ_{Φ}^2 . Для рядка *Залишки* – це залишкова дисперсія σ_3^2 ;

стівпець F – розраховане значення F-критерію Фішера

$$F_p = \frac{MS (\text{Регресія})}{MS (\text{Залишки})};$$

стовпець *Значущість F* – значення рівня значущості, відповідає розрахованому значенню F_p . Визначається за допомогою функції $=FPASP(F_p; df \text{ (регресія)}; df \text{ (залишок)})$.

Отримані результати аналізу містять значення коефіцієнтів регресії a_i та їхні статистичні оцінки.

Інтерпретація стовпців:

1. Коефіцієнти – значення коефіцієнтів a_i .
2. Стандартна похибка – стандартні похибки коефіцієнтів a_i .
3. $t_{\text{стат}}$ – розраховані значення t-критерію:

$$t_{\text{стат}} = \frac{\text{Коефіцієнти}}{\text{Стандартна похибка}}.$$

4. P-значення – значення рівнів значущості, відповідні розрахованим значенням t_p . Визначається за допомогою функції $=СТЬЮДРАСП(t_p; n - m - 1)$.

5. Нижні 95 % і верхні 95 % – відповідно нижні та верхні межі довірчих інтервалів для коефіцієнтів регресії a_i . Для знаходження меж довірчих інтервалів з допомогою функції $=СТЬЮДРАСПОБР(\alpha; n-m-1)$ розраховується критичне значення t-критерію ($t_{\text{кр}}$), а потім розраховуються відповідно нижні й верхні межі довірчих інтервалів за формулами:

$$\text{Нижні 95 \%} = \text{Коефіцієнт} - \text{Стандартна похибка} \times t_{\text{кр}};$$

$$\text{Верхні 95 \%} = \text{Коефіцієнт} + \text{Стандартна похибка} \times t_{\text{кр}}.$$

Звіт розрахованих залишків містить генеровані теоретичні значення \hat{y} результативної ознаки y й значення залишків. Остання розраховується як різниця між емпіричним y та теоретичним \hat{y} значеннями результативної ознаки y .

Для нашого прикладу рівняння регресії буде мати вигляд:

$$\hat{y} = a_0 + a_1x_1 + a_2x_2.$$

Розраховані коефіцієнти регресії a_i дозволяють побудувати рівняння, що виражає залежність прибутку підприємств (Y) від випуску продукції (x_1) та обсягу роздрібного товарообігу (x_2), і отримати таке рівняння регресії:

$$\hat{y} = 1,69 + 0,056x_1 + 0,066x_2.$$

Коефіцієнт множинної кореляції дорівнює 0,813 (R). Вимірюється коефіцієнт від -1 до $+1$. Так як значення коефіцієнта сильно наближене до 1, то можемо говорити про адекватність моделі.

Значення множинного коефіцієнту детермінації $R^2 = 0,661$ вказує, що 66,1 % загальної варіації результативної ознаки пояснюється впливом факторних ознак x_1, x_2 . Тобто обрані фактори істотно впливають на прибуток підприємств, що підтверджує правильність їхнього включення в побудовану модель.

Розрахований рівень значущості $\alpha_p = 0,000\ 099\ 049\ 713\ 734\ 498 < 0,05$ підтверджує значущість R^2 .

Для перевірки значущості коефіцієнтів регресії a_0, a_1, a_2 порівнюють попарно елементи масивів G18:G20 і H18:H20, згідно з чим абсолютне значення a_0, a_1, a_2 є більшими, ніж їхня стандартна похибка, тому їх слід включити до рівняння регресії. Стандартні похибки коефіцієнтів a_0, a_1, a_2 менше своїх стандартних похибок.

Перевірку на адекватність коефіцієнтів регресії рекомендується проводити за наступними еквівалентними методами:

1. Використання t-критерію. Цей коефіцієнт (критерій Стьюдента) вважається значущим, якщо його абсолютна величина перевищує 2,00 (точніше 1,96 – табличне значення), що відповідає рівню значущості 0,05. У нашому прикладі маємо для коефіцієнтів a_0, a_1, a_2 такі показники критерію Стьюдента: $a_0 = 2,216$; $a_1 = 2,533$ і $a_2 = 2,538$. З представленого ряду витікає, що значущими виявляються всі коефіцієнти нашого рівняння.

2. Використання рівня значущості. В цьому випадку оцінювання проводиться шляхом аналізу показника p , тобто рівня значущості α . Коефіцієнт визнається значущим, якщо розраховане для нього p -значення менше (або дорівнює) 0,05 (тобто для 95-процентної довірчої ймовірності).

Бачимо, що показник p становить для коефіцієнтів a_0, a_1, a_2 такі величини: $p_{a_0} = 0,04$; $p_{a_1} = 0,02$ і $p_{a_2} = 0,02$. Ці дані дозволяють також зробити висновок, що всі розглянуті коефіцієнти статистично значущі. Інакше кажучи, можна зробити висновок про не випадковий характер впливу всіх досліджуваних параметрів.

Отже, перевірка обома методами дає цілком узгоджені результати. Тому в остаточному вигляді наше рівняння регресії (для рівня значущості 0,05) слід записати так:

$$\hat{y} = 1,69 + 0,056x_1 + 0,066x_2.$$

Отримане рівняння регресії можливо інтерпретувати: якщо випуск продукції (x_1) збільшиться на одну одиницю, то прибуток підприємств збільшиться в середньому на 0,056 одиниць, якщо обсяг роздрібного товарообігу

підприємств збільшиться на одну одиницю, то прибуток підприємств в середньому збільшиться на 0,066 одиниць. Якщо випуск продукції та обсяг роздрібного товарообігу залишаться незмінними, то прибуток підприємств буде становити 1,69 одиниць. Інакше кажучи, кожна додаткова гривня прибутку підприємств коштує 0,056 і 0,066 гривень випуску продукції та обсягу роздрібного товарообігу відповідно.

Визначення критеріїв адекватності моделі

Перевірку на адекватність рівняння регресії, тобто статистичне оцінювання отриманого рівняння прийнято виконувати за допомогою трьох критеріїв:

- 1) критерію Фішера (F-критерієм);
- 2) рівня значущості α_0 ;
- 3) коефіцієнта детермінації R^2 .

Перевірка за F-критерієм. Рівняння регресії буде мати високу ступінь адекватності за умови, якщо $F_{\text{розрах.}} > F_{\text{крит.}}$. При цьому $F_{\text{крит.}}$ – це табличне значення, що визначається для ступенів свободи $f_1 = n - k - 1 = 20 - 2 - 1 = 17$ та $f_2 = n - m = 20 - 1 = 19$. Тоді для $\alpha_0 = 0,05$, $F_{\text{крит.}} = 2,2$.

У нашому випадку $F_{\text{розрах.}} > F_{\text{крит.}}$ ($16,65 > 2,2$), тобто отримане значення більше табличного, що підтверджує адекватність моделі.

Тепер виконаємо перевірку з використанням рівня значущості α_0 . На рис. 37 знаходимо позицію *Значущість F*. Там вказана величина $9,90497E-05$, тобто це число $+0,000\ 099\ 049\ 713\ 734\ 498$. Фактично можна визнати, що $\alpha_0 = 0,000$. Це говорить про те, що дійсно виявляється стійка залежність цієї функції y (прибуток підприємств) від факторів, що впливають x_1 і x_2 , тобто дохід підприємств не є чисто випадковою величиною.

Перевірка значущості коефіцієнта детермінації R^2 . Нагадаємо, що за нашими розрахунками коефіцієнт детермінації R^2 становить 0,662 або 66 %. Таблиця для тестування на рівні значущості 5 % у разі вибірки $n = 20$ і кількість змінних $k = 2$ дає критичне значення $R^2_{\text{крит.}} = 0,297$. Оскільки виконується співвідношення $R^2_{\text{розрах.}} > R^2_{\text{крит.}}$, то з імовірністю 95 % можна стверджувати про наявність значущості заданого рівняння регресії.

3. Розв'язання оптимізаційних задач у MS Excel.

Теорія оптимальних рішень пов'язана з такими поняттями: варіант, безліч варіантів і критерій. Наявність альтернатив ухвалених рішень дозволяє зробити вибір найкращого рішення – оптимального. Оптимальне

рішення є оптимальним у певному сенсі та визначається шляхом порівняння. Тобто, має існувати критерій, стосовно якого вибирається найкраще рішення.

Математичне ухвалення рішень здійснюється на основі математичної моделі. Математична модель – це адекватний опис розглянутого об'єкта за допомогою математичних функцій і відносин між ними. Математична модель оптимізаційної задачі має містити математичне уявлення варіанта, безлічі варіантів і критерію.

Варіант (позначаємо його через X) – це вектор, що складається з компонент, кожна з яких представляє якийсь об'єкт у кількісному вираженні, тобто $X = \{x_1, x_2, \dots, x_n\}$. Безліч варіантів представляється математичним описом. Зазвичай це сукупність функцій, аргументами яких будуть компоненти вектора-варіанта. Функції задаються у певних співвідношеннях, оскільки вони обмежують можливі значення компонента. $X = \{x_1, x_2, \dots, x_n\}$ можна вважати варіантом поставленого завдання.

Критерій оптимальності визначається як система показників, що використовуються у різних галузях науки, техніки, економіки для обрання найефективнішого варіанта розв'язання певної задачі з кількох можливих варіантів. Принципи знаходження оптимального варіанта зводяться до того, щоб одержати максимальний ефект за заданих витрат або досягти поставленої мети з мінімальними витратами.

Процес оптимізації пов'язаний із визначенням значень економічних показників, за яких досягається оптимум, тобто найкращий стан системи. Найчастіше оптимуму відповідає досягнення найкращого результату при даних витратах ресурсів або досягнення заданого результату за мінімальних ресурсних витратах.

Оптимальні норми дають можливість за даних техніко-економічних умов досягти найбільшої ефективності у використанні ресурсів або найповнішого задоволення економіко-господарських потреб країни (окремого суб'єкта господарювання). Під час розрахунків оптимальних норм використовують економіко-математичні методи, за допомогою яких визначають рівень, що забезпечує за певних обмежень досягнення заданого критерію.

Модель оптимізаційної задачі складається з таких елементів:

- змінні – невідомі величини, значення яких потрібно знайти в результаті розв'язання задачі;
- цільова функція – величина, яка залежить від змінних і значення якої потрібно максимізувати чи мінімізувати;

- критерій – вимоги мінімізації та максимізації цільової функції;
- обмеження – умови, яким мають задовольняти змінні.

Найпростішою оптимізаційною задачею вважається задача пошуку максимального або мінімального значення функції однієї змінної. Розглянемо приклад математичної моделі такої задачі:

$$f(x) = x + \sin x \rightarrow \min;$$

$$0 \leq x \leq 10.$$

Тут змінною величиною є x . Цільова функція – $f(x)$. Критерієм є вимога мінімізації ($\rightarrow \min$). Обмеження – $0 \leq x \leq 10$.

Різноманітністю оптимізаційних задач є задачі лінійного програмування, тобто задачі, в яких цільова функція та обмеження є лінійними.

Обмеження можуть називатися прямими або непрямими.

Загалом алгоритм розв'язання оптимізаційної задачі такий:

1. Введення позначень. Потрібно визначити, яка змінна яку величину позначатиме.

2. Створення цільової функції та критерію. Записати формулу цільової функції. Нею може бути вартість продукції, обсяг прибутку, обсяг витрат на виробництво та перевезення тощо.

Слід визначити, яка величина буде максимізуватися чи мінімізуватися.

3. Складання системи обмежень. Обмеження – це нерівності або рівності, яким мають задовольняти значення змінних.

4. Розв'язання задачі. Деякі оптимізаційні задачі можна розв'язати аналітично, але такий спосіб трудомісткий, простіше використати засоби табличного процесора *MS Excel*. Це будуть задачі на пошук екстремуму функції однієї змінної, так і задачі лінійного програмування.

Результат діяльності будь-якої економічної системи (фірми, підприємства, галузі) визначається ефективністю використання ресурсів. Це тягне за собою необхідність пошуку найкращого варіанта використання ресурсів, що забезпечує знаходження максимальних значень, наприклад, для прибутку, або ж мінімальних значень – для збитків. Такі завдання отримали назву задач оптимізації. Оптимізаційні задачі широко використовуються в економіці, наприклад, транспортна задача, оптимізація асортименту продукції, оптимізація витрат на рекламу, визначення оптимальної кількості видаваних газет або журналів, яке забезпечить максимальну виручку від продажу, і т. п.

Під розв'язанням оптимізаційних задач розуміється процес вибору таких значень змінних x , які забезпечують оптимальне значення деякої функції $f(x)$ у певному діапазоні її існування. Під оптимальним значенням функції розуміється її екстремум, тобто її максимальне або мінімальне значення.

У *MS Excel* задачі оптимізації розв'язуються за допомогою надбудови *Пошук рішення*.

Програма *Пошук рішень* (MS Excel Solver) – додаткова надбудова табличного процесора *MS Excel*, яка призначена для вирішення певних систем рівнянь, лінійних і нелінійних завдань оптимізації, використовується з 1991 року.

Розмір завдання, яке можна вирішити за допомогою базової версії цієї програми, обмежується такими граничними показниками:

- кількість невідомих (decision variable) – 200;
- кількість формульних обмежень (explicit constraint) на невідомих – 100;
- кількість граничних умов (simple constraint) на невідомих – 400.

Висока ефективність їхнього застосування пояснюється інтеграцією програми оптимізації та табличного бізнес-документа. Завдяки світовій популярності табличного процесора *MS Excel* вбудована в його середовище програма Solver є найбільш поширеним інструментом для пошуку оптимальних рішень у сфері сучасного бізнесу.

За замовчанням у *MS Excel* надбудова *Пошук рішення* відключена. Щоб активізувати її, необхідно перейти на вкладку *Файл*, обрати команду *Параметри*, а потім виберіть категорію *Надбудови*, інструмент *Пошук рішення*, у полі *Управління* виберіть значення *Надбудови MS Excel* і натисніть кнопку *Перейти*. У полі *Доступні надбудови* встановите прапорець поряд з пунктом *Пошук рішення* і натисніть кнопку *ОК*. Після цього на вкладці *Дані* у групі *Аналіз* з'явиться надбудова *Пошук рішення*.

Розглянемо приклади розв'язання оптимізаційної задачі за допомогою надбудови *Пошук рішення*.

Задача 1. Пошуку максимуму цільової функції $Z = 3\,000x_1 + 2\,000x_2$ при наступних обмеженнях:

$$\begin{cases} x_1 + 2x_2 \leq 6 \\ 2x_1 + x_2 \leq 8 \\ x_2 - x_1 \leq 1 \\ x_2 \leq 2 \\ x_1, x_2 \geq 0 \end{cases} .$$

У клітинки А3 і В3 відводяться під значення змінних X_1 та X_2 . У клітинку С4 вводиться цільова функція. У діапазоні клітинок А7:А10 вводяться ліві частини обмежень, а в діапазон клітинок В7:В10 – праві частини обмежень (рис. 38).

	А	В	С
1	Переменные		
2	x1	x2	
3			
4	Функция цели		=3000*A3+2000*B3
5			
6	Ограничения		
7	=A3+2*B3	6	
8	=2*A3+B3	8	
9	=B3-A3	1	
10	=B3	2	

Рис. 38. Встановлення параметрів оптимізаційної задачі

Обираємо команду *Дані / Пошук рішення* і заповнюємо діалогове вікно *Пошук рішення*, як показано на рис. 39.

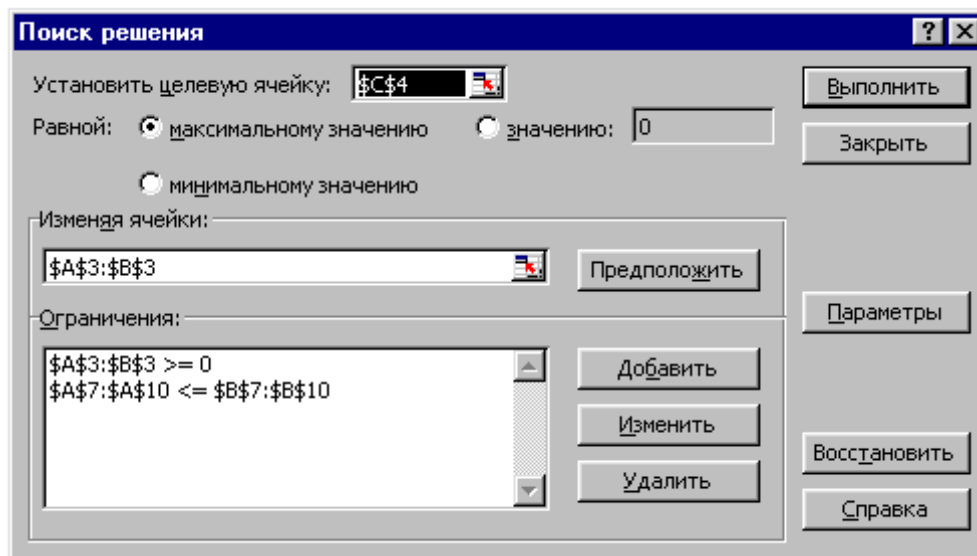


Рис. 39. Діалогове вікно *Пошук рішення*

У діалоговому вікні *Параметри* необхідно встановити прапорці *Лінійна модель* і *Позитивне значення*. Після натискання кнопки *Виконати* відкривається вікно *Результати пошуку рішення*, яке повідомляє, що рішення знайдено. Результати розрахунків показано на рис. 40.

	A	B	C
1	Переменные		
2	x1	x2	
3	3,333333	1,333333	
4	Функция цели		12666,66667
5			
6	Ограничения		
7	6	6	
8	8	8	
9	-2	1	
10	1,333333	2	
11			

Рис. 40. Результати виконання команди *Пошук рішення*

Задача 2. Є дані про нарахування заробітної плати працівників підприємства (табл. 8).

Таблица 8

Вихідні дані

ПІБ працівника	Дата нарахування	Сума заробітної плати, грн	Премія, грн
Бобко С. С	28.01.2018 р.	6 155	
Гламаздіга В. Є.	28.01.2018 р.	5 845	
Виперайленко О. І.	28.01.2018 р.	5 065	
Тмінко А. Л.	28.01.2018 р.	7 524	
Парфйонов Б. Д.	28.01.2018 р.	10 057	
Петрова Г. О.	28.01.2018 р.	7 502	
Коваль Г. Р.	28.01.2018 р.	8 030	

Необхідно розрахувати премію кожного працівника, яка є добутком заробітної плати, зазначеної в окремому стовпці, на певний коефіцієнт. При цьому, загальна сума грошових коштів, що виділяються на премію, дорівнює 30 000 грн. Клітинка, в якій знаходиться задана сума має назву цільової, так як наша мета підібрати розмір премії саме під це число.

Розв'язання

Коефіцієнт, який застосовується для розрахунку суми премії, необхідно обчислити за допомогою функції Пошуку рішень. Клітинка, в якій він розташовується, називається шуканою. Вводимо вихідні дані на лист MS Excel (рис. 41).

	A	B	C	D	E	F	G	H
1								
2								
3	ІПБ	Дата нарахування	Сума заробітної плати, грн	Премія, грн			Коефіцієнт	
4	Бобко С.С	28.01.2018	6155					
5	Гламаздга В.Є.	28.01.2018	5845					
6	Виперайленко О.І.	28.01.2018	5065					
7	Тмінко А.Л.	28.01.2018	7524					
8	Парфйонов Б.Д.	28.01.2018	10057					
9	Петрова Г.О.	28.01.2018	7502					
10	Коваль Г.Р.	28.01.2018	8030					
11	Разом		50178					

Рис. 41. Введення вихідних даних задачі

Цільова та шукана клітинки мають бути пов'язані одна з одною за допомогою формули. У нашому випадку, формула розташовується в цільовій клітинці, і має такий вигляд: $C11*G\$3$, де $G\$3$ – абсолютне посилання на клітинку коефіцієнту, а $C11$ – загальна сума заробітної плати, від якої проводиться розрахунок премії працівникам підприємства (рис. 42).

	A	B	C	D	E	F	G
1							
2							
3	ІПБ	Дата нарахування	Сума заробітної плати, грн	Премія, грн			Коефіцієнт
4	Бобко С.С	43128	6155	=C4*G\$4			
5	Гламаздга В.Є.	43128	5845	=C5*G\$4			
6	Виперайленко О.І.	43128	5065	=C6*G\$4			
7	Тмінко А.Л.	43128	7524	=C7*G\$4			
8	Парфйонов Б.Д.	43128	10057	=C8*G\$4			
9	Петрова Г.О.	43128	7502	=C9*G\$4			
10	Коваль Г.Р.	43128	8030	=C10*G\$4			
11	Разом		=СУММ(C4:C10)	=C11*G\$4			

Рис. 42. Демонстрація формул

Відкриваємо інструмент *Пошук рішення* та заповнюємо вікно параметрів, у яке потрібно внести дані (рис. 43).

У полі *Оптимізувати цільову функцію* потрібно ввести адресу клітинки, де буде розташовуватися загальна сума премії для всіх працівників.

Під вікном з адресою цільової клітинки потрібно встановити параметри значень, які будуть знаходитись у ній. Це може бути максимум, мінімум або конкретне значення. В нашому випадку, це буде останній варіант. Тому, ставимо перемикач позицію *Значення*, і в полі зліва від нього прописуємо число 30 000. Як ми пам'ятаємо, саме це число за умовами задачі становить загальну суму премії для всіх працівників підприємства.

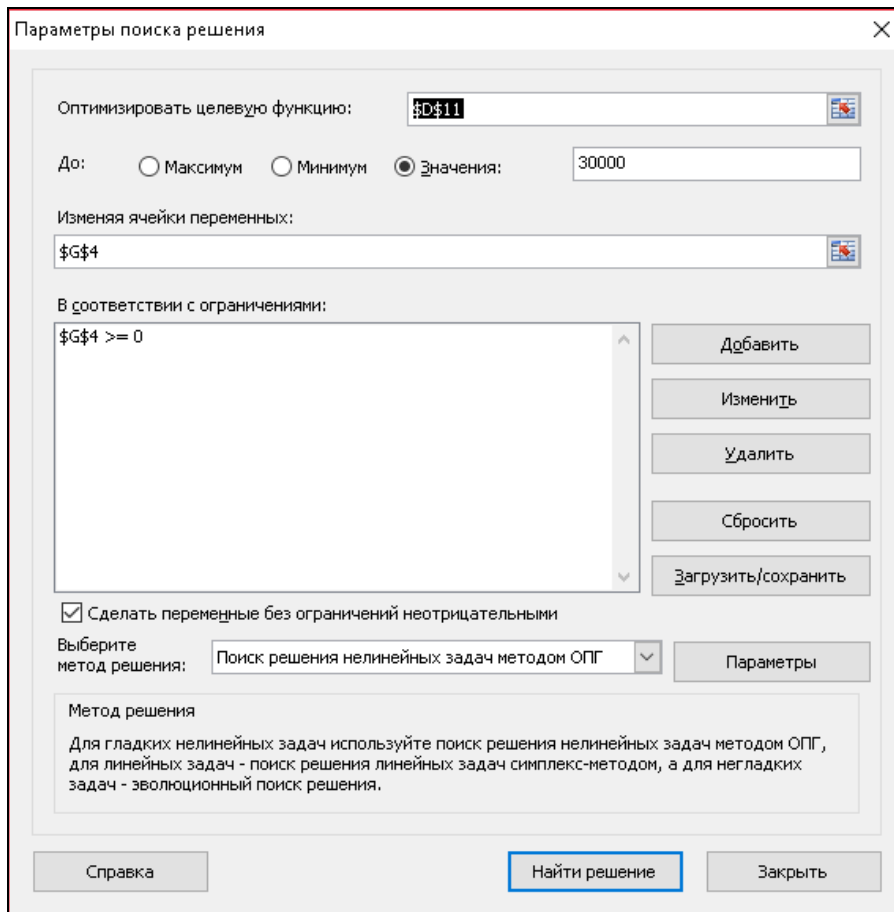


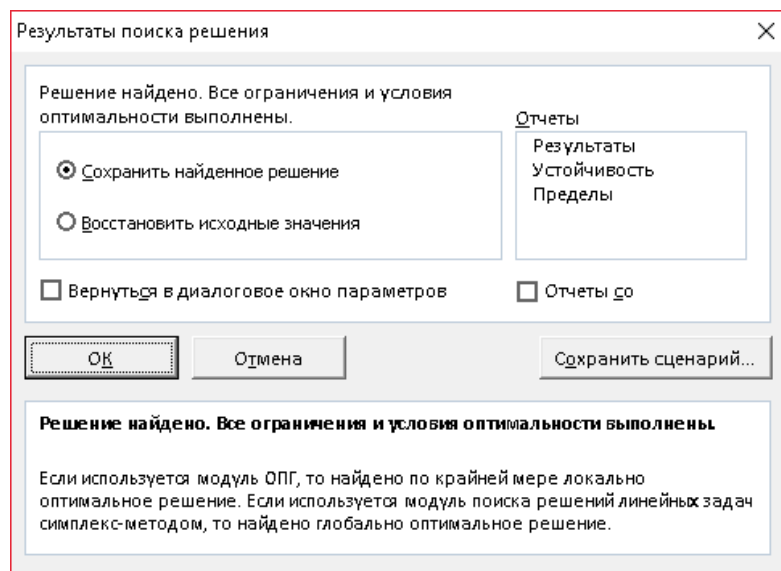
Рис. 43. Встановлення параметрів пошуку рішення

Нижче розташовано поле *Змінюючи значення змінних*. Тут потрібно вказати адресу шуканої клітинки, де знаходиться коефіцієнт, множенням на який розміру основної заробітної плати буде розрахована величина премії. Адресу можна прописати тими самими способами, як ми це робили для цільової клітинки.

У полі *Відповідно до обмежень* можна виставити певні обмеження для даних, наприклад, зробити значення цілими або невід'ємним. Для цього тиснемо на кнопку *Додати*. Після цього відкривається вікно додавання обмеження. У полі *Посилання на клітинки* прописуємо адресу клітинок щодо яких вводиться обмеження. В нашому випадку, це шукана клітинка з коефіцієнтом. Далі ставимо потрібний знак: "менше або дорівнює", "більше або дорівнює", "одне число", "ціле число", "бінарне число" і т. д. У нашому випадку, ми виберемо знак "більше або одне", щоб зробити коефіцієнт додатним числом. Відповідно у полі *Обмеження* вказуємо число 0. Якщо ми хочемо налаштувати ще одне обмеження, то тиснемо на кнопку *Додати*. У зворотному випадку, тиснемо на кнопку *ОК*, щоб зберегти введені обмеження.

Як бачимо, після цього, обмеження з'являється у відповідному полі вікна параметрів пошуку рішення. Також, зробити змінні невід'ємними, можна встановивши галочку біля відповідного параметра трохи нижче. Бажано, щоб встановлений тут параметр не суперечив тим, які ви прописали в обмеженнях, інакше може виникнути конфлікт.

Далі, програма *MS Excel* у клітинках виконує необхідні розрахунки. Одночасно з представленням результатів, відкривається вікно, в якому ви можете зберегти знайдене рішення або відновити вихідні значення, перемістивши перемикач у відповідну позицію. Незалежно від обраного варіанта, встановивши прапорець *Повернутися в діалогове вікно налаштувань*, ви можете знову перейти до налаштувань пошуку рішення. Після того, як виставлені прапорці та перемикачі, тиснемо на кнопку *ОК* (рис. 44).



	А	В	С	Д	Е	F	G
1							
2							
3	ІПБ	Дата нарахування	Сума заробітної плати, грн	Премія, грн			Коефіцієнт
4	Бобко С.С	28.01.2018	6155	3679,9			0,597871577
5	Гламаздига В.Є.	28.01.2018	5845	3494,56			
6	Виперайленко О.І.	28.01.2018	5065	3028,22			
7	Тмінко А.Л.	28.01.2018	7524	4498,39			
8	Парфйонов Б.Д.	28.01.2018	10057	6012,79			
9	Петрова Г.О.	28.01.2018	7502	4485,23			
10	Коваль Г.Р.	28.01.2018	8030	4800,91			
11	Разом		50178	30000			
12							

Рис. 44. Результати виконання команди *Пошук рішення*

Отже, в результаті отримаємо нараховування премій працівникам у межах грошового обмеження 30 000 грн.

Задача 3. Торговельне підприємство випускає два види продукції із сировини чотирьох типів. Показники виробництва наведено в табл. 9. Ціна реалізації одиниці продукції P1 і P2 становить 200 та 300 грн відповідно. Необхідно скласти такий план виробництва продукції, за якого прибуток від реалізації буде максимальним.

Таблиця 9

Вихідні дані

Види сировини	Запаси сировини, од.	Число одиниць сировини, затрачених на виготовлення од. продукції	
		P1	P2
S1	18	1	3
S2	16	2	1
S3	5	–	1
S4	21	3	–

Розв'язання

Складемо оптимізаційну модель задачі. Введемо необхідні позначення: x_1 , x_2 – число одиниць продукції відповідно P1 і P2, запланованих до виробництва.

Для їхнього виготовлення (див. табл. 9) буде потрібно $(1 \times x_1 + 3 \times x_2)$ одиниць ресурсу S1, $(2 \times x_1 + 1 \times x_2)$ одиниць ресурсу S2, $(1 \times x_2)$ одиниць ресурсу S3, $(3 \times x_1)$ одиниць ресурсу S4.

У наслідок того, що споживання ресурсів S1, S2, S3, S4 не має перевищувати їхніх запасів, відповідно 18, 16, 5 і 21 одиниць, то зв'язок між споживанням ресурсів і їхніми запасами виражається системою обмежень:

$$\begin{cases} x_1 + 3x_2 \leq 18 \\ 2x_1 + x_2 \leq 16 \\ x_2 \leq 5 \\ 3x_1 \leq 21 \end{cases}$$

Змінні позитивні $x_1 > 0$, $x_2 > 0$. Сумарний прибуток F від реалізації продукції буде становити: $F = x_1 + x_2$.

Для розв'язання задачі використовуємо інструмент *Пошук рішення*.

Першою дією необхідно ввести вхідні дані на робочий аркуш MS Excel, як це показано на рис. 45.

	А	В	С	Д	Е
1	Види сировини	Запаси сировини, од.	Число одиниць сировини, затрачених на виготовлення од. продукції		Обмеження
2			Р1	Р2	
3	1	18	1	3	=СУММПРОИЗВ(\$C\$8:\$D\$8;C3:D3)
4	2	16	2	1	=СУММПРОИЗВ(\$C\$8:\$D\$8;C4:D4)
5	3	5		1	=СУММПРОИЗВ(\$C\$8:\$D\$8;C5:D5)
6	4	21	3		=СУММПРОИЗВ(\$C\$8:\$D\$8;C6:D6)
7	Ціна проукції		200	300	
8	План виробництва				
9	Прибуток		=СУММПРОИЗВ(C8:D8;C7:D7)		
10					
11					

Рис. 45. Шаблон розв'язання задачі оптимізації використання ресурсів

Для формування обмежень задачі в клітинку Е5 вводимо функцію: СУММПРОИЗВ (С8:Д8;С3:Д3).

Копіюємо функцію у клітинки С6:С8 за допомогою маркера заповнення.

У клітинку С9 вводимо цільову функцію:

СУММПРОИЗВ (С8:Д8;С7:Д7).

Рішення цього завдання економічного планування зводиться до розв'язання систем лінійних рівнянь за допомогою інструменту *Пошук рішення*.

Для розв'язуваної задачі цільовою функцією буде функція у клітинці \$C\$9, що дорівнює максимальному значенню, змінними даними – діапазон \$C\$8:\$D\$8, що містить кількість видів продукції, діапазон \$E\$3:\$E\$6 – використовується для визначення обмежень задачі.

Для ведення обмежень потрібно натиснути на кнопку *Додати*.

1. Обмеження за ресурсами.

Посилання на клітинки: \$E\$3:\$E\$6. Вид обмеження: <=. Обмеження: посилання на клітинки: \$B\$3: \$B\$6. Додати.

2. Обмеження на незаперечність змінних.

Посилання на клітинки: \$C\$8:\$D\$8. Вид обмеження: >=. Обмеження: 0. Додати.

3. Обмеження на цілісність змінних.

Посилання на клітинки: \$C\$8:\$D\$8, вид обмеження: ціле. ОК.

Після активації команди *Пошук рішення*, необхідно заповнити відповідними параметрами діалогове вікно (рис. 46).

Після натискання кнопки *Знайти рішення* надбудова *Пошук рішення* розпочинає ітерацію. Після обчислень відкриває діалогове вікно *Результати пошуку рішення* (рис. 47), у якому виводиться повідомлення про результати розв'язання задачі.

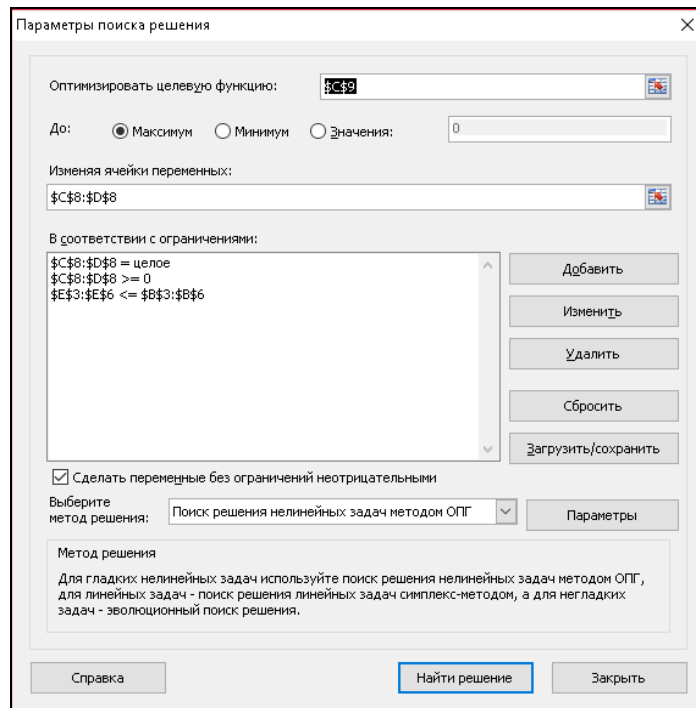
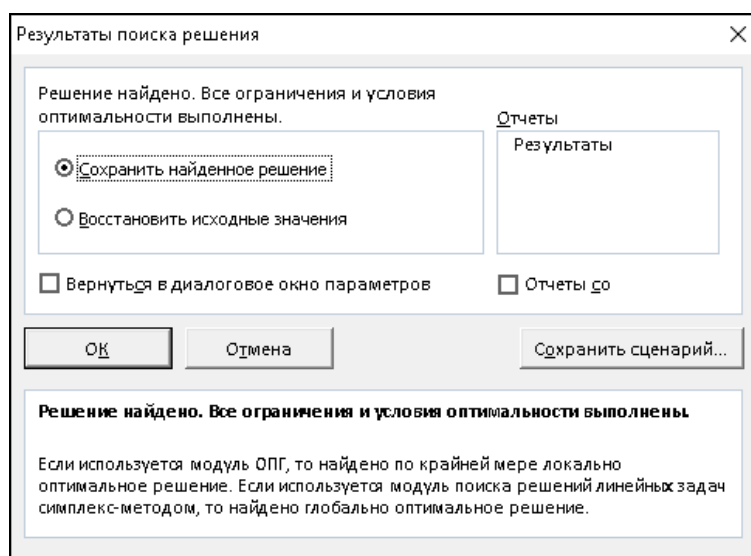


Рис. 46. Внесення параметрів пошуку рішення



	A	B	C	D	E
	Виды сировини	Запаси сировини, од.	Число единиц сировини, затраченных на изготовление од. продукции		Обмеження
			P1	P2	
1	1	18	1	3	18
2	2	16	2	1	16
3	3	5		1	4
4	4	21	3		18
5	Ціна продукції		200	300	
6	План виробництва		6	4	
7	Прибуток		2400		

Рис. 47. Вікно результатів пошуку рішення

Отже, можна зробити висновок, що максимальний прибуток підприємства буде становити 2 400 грн, продукції типу P1 необхідно виробляти 6 од., продукції типу P2 – 4 од.

Завдання:

1. Використовуючи власний інформаційний простір (статистичні дані за обраною тематикою) та аналітичні інструменти MS Excel, необхідно:

1.1. Розрахувати основні описові статистики для показників, що досліджуються. Оцінити ступінь однорідності сукупності та зробити висновок щодо закономірності розподілу досліджуваної сукупності.

1.2. Побудувати багатофакторну регресійну модель та оцінити її адекватність.

2. Використовуючи інструмент *Пошук рішення MS Excel* розв'язати систему рівнянь:

$$-6X - 9Y - 3Z = 42.$$

$$5X - Y - 5Z = 75.$$

$$-3X - 10Y - Z = 44.$$

Результати роботи оформіть у вигляді звіту з лабораторної роботи.

Завдання для самостійного опрацювання

Запитання, що виносяться на самостійну роботу студентів:

1. Охарактеризуйте сфери використання *MS Excel* та проаналізуйте його можливості.

2. Наведіть аналіз можливостей пакету *Аналіз даних MS Excel*.

3. У чому полягає відмінність між категоріями інформаційний простір й інформаційний ресурс, інформація та данні, індикатор і показник?

4. Які завдання вирішує економетрична модель?

5. Які існують оптимізаційні моделі в теорії ринкової економіки?

Розкрийте особливості використання оптимізаційних задач на мікрорівні та макрорівні.

Завдання 1. Використовуючи пакет *Аналіз даних у MS Excel*, побудуйте багатофакторну регресійну модель залежності ВВП від наведених у табл. 10 соціально-економічних показників.

Вихідні дані

Період	ВВП, млн грн	Індекс споживчих цін	Експорт, млн грн	Імпорт, млн грн
1	1 937,3	102,6	2 339 402,7	2 712 852,43
2	1 820,2	101,2	2 518 534,3	3 171 353,67
3	2 054,4	101,3	3 129 139,2	3 872 083,7
4	2 029,8	100,4	2 953 856,7	3 283 634,6
5	2 108	100,5	3 104 624,5	3 632 466
6	2 221,8	100,3	3 317 653,5	3 610 201,2
7	2 433,1	100,7	3 344 490,8	3 686 831,1
8	2 030	100,2	3 511 110,9	3 839 422,4
9	2 836,1	100,1	3 676 317,9	4 147 061,3
10	2 567,6	100,3	3 437 449,5	4 021 637,6
11	2 467,7	100,6	3 344 121,8	3 922 054
12	2 358	100,8	3 691 002,6	5 134 893,1
13	2 249,9	101,7	3 208 484,9	3 700 647,9
14	2 049,5	101,7	3 409 760,2	4 297 569,8
15	2 487,4	100,8	4 108 205,4	4 953 434,5
16	2 488,3	100,6	4 067 299,9	4 820 236,3
17	2 576	100,6	4 083 445,1	4 852 211,6
18	2 692,1	101	4 235 294,1	4 682 039,7
19	2 768,2	100,9	4 258 746,5	5 315 298,1
20	2 978,5	100,1	4 167 498,9	4 872 915,4
21	3 168,2	100,8	4 114 710,7	4 851 865,6
22	3 230,8	101,6	4 345 291,8	5 872 680,3
23	3 447,3	101,2	4 450 168,3	5 822 204
24	3 027,9	101,1	4 799 157,8	6 628 819,8
25	2 963,7	102,3	3 663 214,9	4 627 526,5
26	2 854	101,2	4 682 418,3	6 465 057,5
27	3 087,4	101,2	5 444 491,8	7 712 994,2
28	3 298,8	101,4	5 571 314,2	7 936 247,4
29	3 397,5	101,4	6 284 581,1	7 710 564
30	3 455	101	6 896 161,2	7 934 961,4
31	3 679,8	100,5	7 616 856,6	8 822 925,9
32	3 866,6	100,4	6 718 135,9	8 155 969,3
33	3 997,5	100,8	6 685 131,4	8 479 144,1
34	3 858,1	100,9	5 861 332,6	7 647 194,6
35	3 649,8	100,8	3 622 525,2	5 264 462,7

Завдання 2. Використовуючи інструмент *Пошук рішення*, розв'яжіть задачу:

2.1. Для виготовлення чотирьох видів продукції використовують три види сировини. Запаси сировини, норми витрат і прибуток від реалізації кожного продукту наведено в табл. 11. Складіть математичну модель на максимум загальної вартості продукції, що випускається.

Таблиця 11

Вихідні дані

Вид сировини	Норми витрат сировини на 1 виріб, кг				Запаси сировини, кг
	A	Б	В	Г	
1	1	2	1	0	18
2	1	1	2	1	30
3	1	3	3	2	40
Прибуток від реалізації 1 виробу, грн	12	7	18	10	

2.2. Сільськогосподарське підприємство купує три різних види зерна й виготовлює з них різні види сумішей. Кожний вид зерна містить чотири інгредієнти. Відповідні дані наведено в табл. 12. Складіть математичну модель, що відповідає мінімальній вартості суміші та розв'язати її.

Таблиця 12

Вихідні дані

Інгредієнти	Види зерна, кг			Необхідний мінімум інгредієнта, кг
	I	II	III	
A	2	3	7	1 250
B	1	1	0	250
C	5	3	0	900
D	0,6	0,25	1	232,5
Ціна 1 кг зерна, грн	41	35	96	

Тема 3. Побудова баз даних MS Access

Лабораторна робота 3 Побудова баз даних MS Access

Мета: надбання практичних навичок у створенні реляційної бази даних за допомогою програми MS Access.

Завдання: ознайомитися з аналітичними можливостями MS Access; навчитися проектувати та створювати таблиці, запити, форми, звіти і макроси у програмі MS Access.

Методичні рекомендації

База даних (БД) – це систематизоване сховище даних. Сучасні бази даних створюються на електронних носіях. Це забезпечує надійне збереження даних і зручні засоби роботи з ними.

БД бувають двох видів:

за характером даних, що зберігаються:

- фактографічні (БД, які містять інформацію, що відноситься до певної предметної сфери або галузі діяльності);
- документальні (БД, які відображають конкретні документи та містять їхній бібліографічний опис);

за способом об'єднання даних:

- ієрархічні (в БД усі елементи розташовуються залежно від їхнього підпорядкування);
- мережні (в БД між елементами теж існує зв'язок, але зв'язки можуть бути довільними);
- реляційні (в БД відомості зберігаються у вигляді таблиць, між якими встановлені зв'язки).

База даних Microsoft Access – це сукупність даних і об'єктів, що відносять до певної задачі. БД Microsoft Access є реляційною системою управління базами даних, що підтримує всі засоби та можливості аналітичного оброблення даних, які властиві реляційним моделям.

СУБД (система управління базами даних) – це сукупність програмних і лінгвістичних засобів загального або спеціального призначення, що забезпечують управління створенням і використанням баз даних (MS Access).

Під час розв'язання багатьох статистичних задач сьогодні широко використовуються системи управління базами даних (СУБД) реляційного

типу. В цих СУБД реалізується реляційна модель даних – зображення їх у табличному вигляді. Рядок такої таблиці еквівалентний запису файлу бази даних (БД), а графа – полю запису. Доступ до елемента даних здійснюється з допомогою зв'язку потрібного рядка (запису) з потрібною графою (полем).

Існує два типи систем управління БД: файлами і реляційними базами даних. У системах управління файлами (file management systems), іноді їх ще називають плоскими файловими базами даних, дані вводяться і зберігаються без індексування, а під час пошуку і створенні вихідних форм обробляються послідовно. Тому ці системи недостатньо гнучкі у процесі маніпулювання даними. Системи управління реляційними базами даних (Relational database management systems) дозволяють користувачам маніпулювати даними більш вільно завдяки збереженню всіх даних тільки один раз і опису зв'язків між ними. Зв'язки є загальними елементами даних, таких як прізвище покупця або номер його чекової книжки. Дані кожного набору інформації можна знайти, відобразити і відновити на підставі даних іншого набору.

До складу багатьох СУБД, призначених для роботи на персональних комп'ютерах, входять три основні компоненти: командна мова, інтерпретуюча система або компілятор для перетворення команд до виду, придатного до виконання, і засоби взаємодії користувача із СУБД (інтерфейс користувача).

Командна мова слугує для виконання і забезпечення потрібних операцій з даними, дає змогу створювати прикладні програми, екранні форми введення та виведення інформації, структури файлів БД і маніпулювати даними. Структура та можливості мови визначають вигляд і можливості конкретної СУБД.

У СУБД можна виконувати команди по одній або групами в автоматичному режимі, заздалегідь записаними у спеціальний (програмний) файл. Команди мови СУБД записуються в текстовій формі, близькій до звичайної мови. Для того щоб персональний комп'ютер міг виконати таку команду, її треба перетворити на команду, що виконується комп'ютером.

Структура та зміст компонентів СУБД визначає її призначення та коло потенційних користувачів. Останні версії сім'ї СУБД Access 2010 для Windows, розроблені фірмою Microsoft, відповідають вимогам користувачів.

Основні функції СУБД:

- створення і модифікація структури бази даних (визначення кількості таблиць, встановлення зв'язків між таблицями, визначення типів полів тощо);

- введення даних у базу та їхнє збереження;
- редагування даних;
- відбір даних та їхнє оброблення;
- розроблення вихідних документів.

Основними сферами використання СУБД є бібліотечна справа, готельний бізнес, бухгалтерський облік, документообіг, управління підприємством, підтримка вебсайтів, електронні довідники тощо.

Працюючи з MS Access, користувач може розв'язувати такі завдання: розбивати дані на логічно пов'язані між собою частини; вводити, змінювати і знаходити необхідні дані; знаходити підмножини даних за заданими умовами; створювати форми та звіти; автоматизувати виконання стандартних завдань; графічно встановлювати зв'язки між даними; вставляти рисунки у форми та звіти; створювати особисті, готові до роботи з БД програми, які вміщують меню, діалогові вікна і командні кнопки.

База даних – це файл спеціального формату, який містить інформацію, структуровану певним чином. Комплекс програмних засобів, які призначені для створення структури нової бази, наповнення її змістом, редагування змісту та візуалізації інформації.

До об'єктів реляційної СУБД Microsoft Access відносять: таблиці, запити, форми, звіти, вебсторінки, макроси і модулі (табл. 13).

Таблиця 13

Призначення об'єктів реляційної СУБД Microsoft Access

Об'єкти СУБД Microsoft Access	Призначення
Таблиця	Зберігання структурованих даних
Запит	Вибір даних відповідно до умов
Форма	Подання даних у зручному вигляді
Звіт	Створення звітних документів для друку
Вебсторінка	Подання даних у вигляді вебсторінки
Макрос	Автоматизація стандартних дій
Модуль	Створення процедур опрацювання даних на мові програмування VBA

Створення бази даних завжди має починатися з розроблення структури її таблиць. Таблиці – це основні об'єкти бази даних, у них зберігаються дані. У табличній структурі адрес даних визначається перетином рядків і стовпців. У базах даних стовпці називаються полями, а рядки – записами. Поля утворюють структуру бази даних, а записи становлять інформацію, яка в ній міститься.

Поля – це основні елементи структури бази даних. Вони мають свої властивості. Від властивостей полів залежить, які типи даних можна вносити у поле, а які ні, а також те, що саме можна робити з даними, які містяться у полі.

Основні властивості полів таблиць баз даних на прикладі СУБД Microsoft Access:

ім'я поля – визначає як варто звертатися до даних цього поля при автоматичних операціях з базою (за замовчуванням імена полів використовуються як заголовки стовпців таблиць);

тип поля – визначає тип даних, що можуть міститися в заданому полі (текстовий, числовий, дата/час, поле MEMO, грошовий, лічильник, логічний, поле об'єкта OLE, гіперпосилання);

розмір поля – визначає граничну довжину (у символах) даних, що можуть розміщатися в заданому полі, основна властивість текстового поля;

формат поля – визначає спосіб форматування даних у клітинках, що належать полю, наприклад, для типу поля дата/час – короткий формат 19.06.18;

маска введення – визначає форму, в якій вводяться дані в поле (засіб автоматизації введення даних);

підпис – визначає заголовок стовпця таблиці для заданого поля (якщо підпис не зазначений, то як заголовок стовпця використовується властивість *Ім'я поля*);

значення за замовчуванням – те значення, що вводиться в клітинки поля автоматично (засіб автоматизації введення даних);

умова на значення – обмеження, використовуване для перевірки правильності введення даних (засіб автоматизації введення, що використовується переважно для даних, що мають числовий тип, грошовий чи тип дата/час);

повідомлення про помилку – текстове повідомлення, що видається автоматично під час спроби введення в поле помилкових даних (перевірка помилковості виконується автоматично, якщо задана властивість *Умова на значення*);

обов'язкове поле – властивість, що визначає обов'язковість заповнення даного поля при наповненні бази.

Вищевказані властивості полів відносяться загалом до полів текстового типу. Поля інших типів можуть мати або не мати ці властивості, але можуть додавати до них і свої. Наприклад, для даних, що представляють дійсні числа, важливою властивістю є кількість десяткових знаків.

Microsoft Access працює з такими типами даних, які можуть міститися у полі:

- текстовий – тип даних, який використовується для збереження звичайного неформатованого тексту обмеженого розміру (до 255 символів);
- поле MEMO – спеціальний тип даних для збереження великих обсягів тексту (до 65 535 символів);
- числовий – тип даних для збереження числових даних. Він також має свій розмір (байт, ціле, довге ціле, дійсне). Залежно від розміру поле має свої властивості, наприклад, для дійсного необхідно вказати розмір десяткової частини числа;
- дата/час – тип даних для збереження календарних дат і поточного часу. Мають свої формати (повний, довгий, середній, короткий);
- грошовий – тип даних для збереження грошових сум. У цьому випадку комп'ютер відображає числа разом з грошовими одиницями;
- лічильник – спеціальний тип даних для унікальних (не повторюваних у поле) натуральних чисел з автоматичним нарощуванням. Природне використання – для порядкової нумерації запису.
- поле об'єкта OLE – спеціальний тип даних, який призначений для зберігання об'єктів OLE, наприклад, мультимедійних об'єктів (картинки, музичні кліпи та відеозаписи).

Розглянемо покрокове створення бази даних у Microsoft Access 2010 на прикладі побудови БД "Склад".

Під час запуску MS Access з'явиться діалогове вікно (рис. 48).

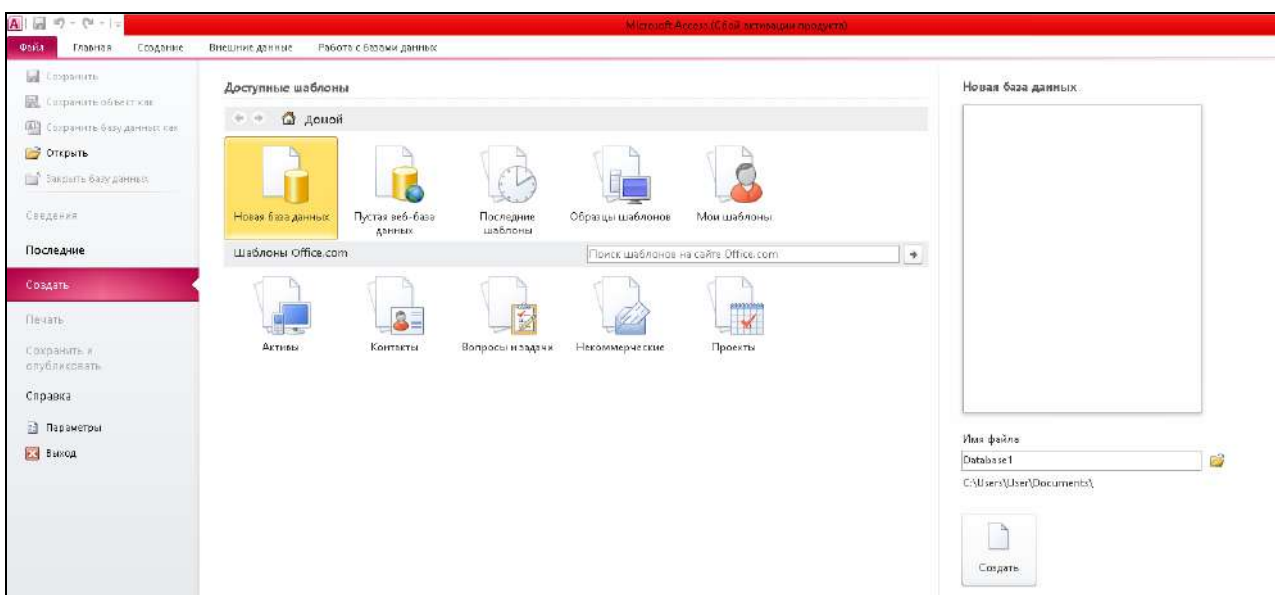


Рис. 48. Вікно запуску БД Microsoft Access 2010

Головне вікно MS Access – це робоча область, де користувач маніпулює різними об'єктами MS Access. Натисніть на кнопку *Нова база даних*. У вікні введіть назву БД і оберіть кнопку *Створити* (рис. 49).

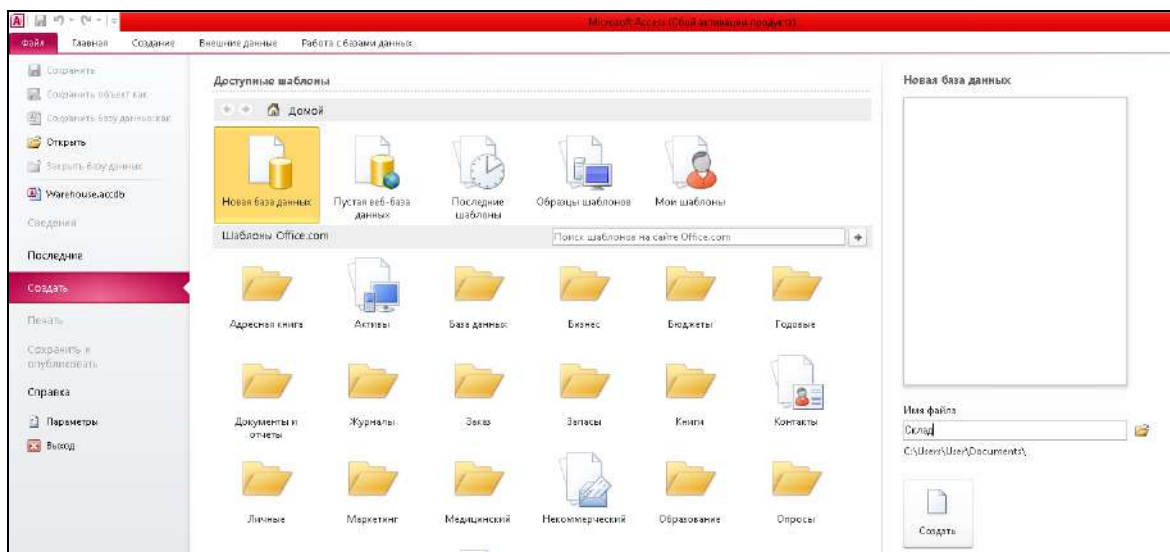


Рис. 49. Створення нової БД *Склад (Warehouse)*

З метою ефективного управління діяльністю магазину необхідно координувати діяльність складу. Тому база даних має складатися з трьох таблиць: "Товар", "Постачальник" і "Закупівля". Таблиці мають таку структуру:

"Товар":

код товару (ключове поле);

найменування;

ціна;

одиниці виміру.

"Постачальник":

назва постачальника (ключове поле);

адреса;

телефон.

"Закупівля":

назва постачальника (ключове поле);

код товару (ключове поле);

кількість;

дата.

На наступному кроці необхідно створити об'єкти бази даних, для чого слід натиснути правою кнопкою миші на поле *Таблиця1*, викликати контекстне меню та обрати команду *Конструктор* (рис. 50).

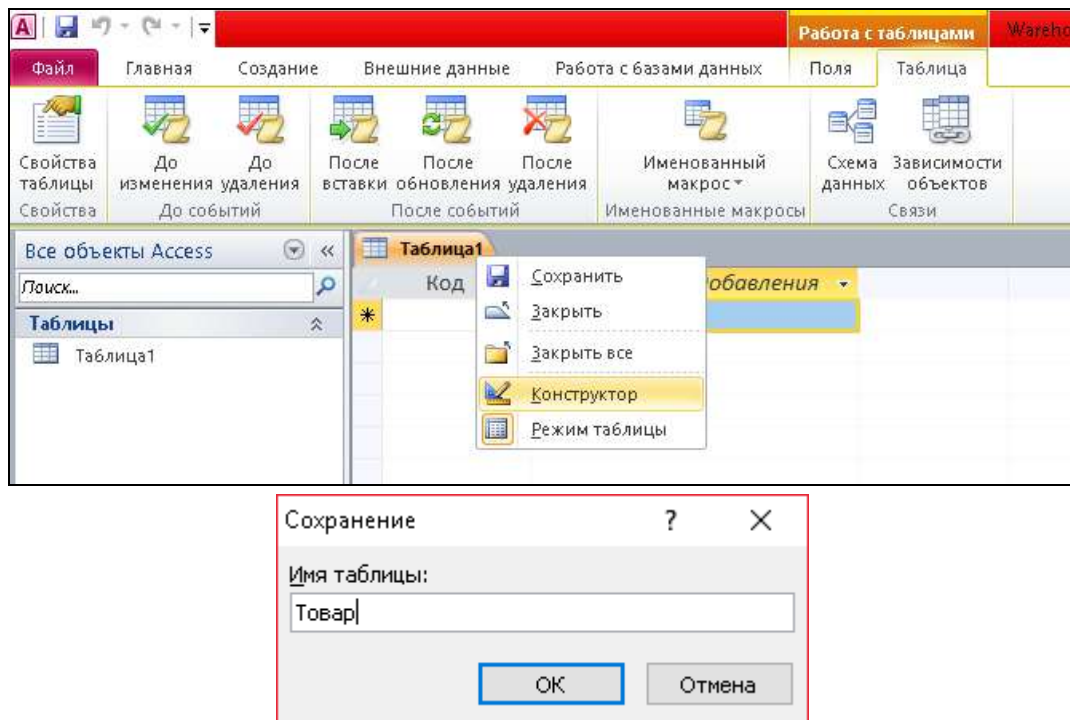


Рис. 50. Створення таблиці

Тепер починаємо заповнювати назви полів та задавати відповідний їм тип даних, який буде використовуватися. Створюючи структури таблиць одне поле (чи одну комбінацію полів), треба назначити ключовим. З ключовими полями комп'ютер працює особливо. Він перевіряє їхню унікальність і швидше виконує сортування за такими полями.

Після цього в розділі *Властивості поля* необхідно задати необхідні властивості поля. Наприклад, для текстових полів основною властивістю є його довжина.

Наступним кроком необхідно задати ключове поле. Для цього виділяємо рядок поля таблиці та помічаємо його як ключове, натиснув на панелі інструментів *Ключове поле*. У лівій частині рядка з'явиться значок ключа. Роль ключового поля можуть грати декілька полів таблиці, для цього необхідно виділити ці поля за нажаті клавiші Ctrl і натиснути кнопку *Ключове поле*.

Завершити опис структури таблиці можливо закриттям вікна опису структури. Якщо ключове поле не було задано, то Access запропонує задати ключове поле. Якщо воно не заплановане, то відповідайте "Ні". Якщо відповісти "Так", то Access автоматично вставляє ключове поле типу лічильник. У цій ситуації більш кращою є відміна закриття вікна опису структури та вказівка ключового поля вручну (рис. 51).

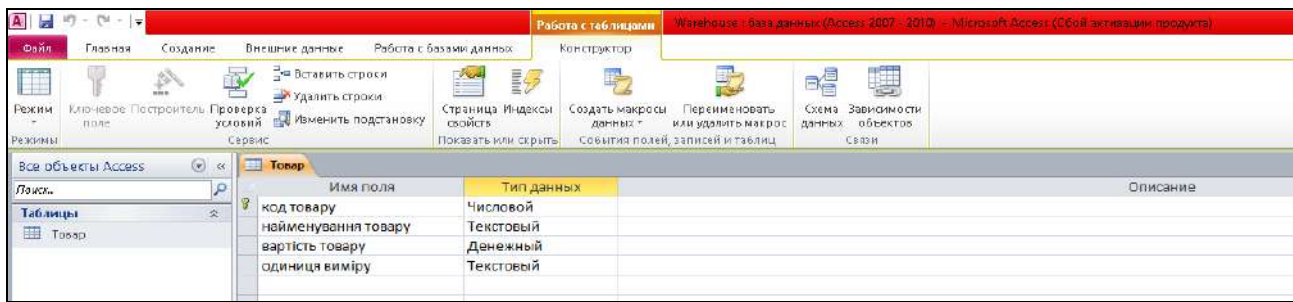


Рис. 51. Заповнення полів таблиці "Товар"

Заповніть таблиці записами. Для цього необхідно виділити необхідну таблицю у вікні БД і натисніть *Відкрити*. Введіть дані в поля таблиці (10 – 20 записів, залежно від змісту таблиці). Для ключових полів, тип яких не є лічильником, розробіть систему кодів. Код має бути коротким та певною мірою відображати зміст запису. Найпростіший код для ключа – це лічильник, що може бути вставлений автоматично (рис. 52).

Таблиця №1: ТаблицяТовар

Ім'я поля	Тип даних	Примітки
ТоварНазва	Текстовий	КЛЮЧ. Довжина: 30. Обов'язкове поле: Так. Пусті строки: Ні. Індексоване поле: Так (Співпадіння не допуск.). Підпис: «Назва товару»
Ціна	Грошовий	
ОдВиміру	Текстовий	Довжина: 10. Підпис: «Од. виміру»

Таблиця №2: ТаблицяПостачальник

Ім'я поля	Тип даних	Примітки
НазваПостачальник	Текстовий	КЛЮЧ. Довжина: 30. Обов'язкове поле: Так. Пусті строки: Ні. Індексоване поле: Так (Співпадіння не допуск.). Підпис: «Назва постачальника»
Адреса	Текстовий	Довжина: 50
Телефон	Текстовий	Довжина: 16

Таблиця №3: ТаблицяЗакупка

Ім'я поля	Тип даних	Примітки
НазваПостачальник	Текстовий	Довжина: 30. Обов'язкове поле: Так. Пусті строки: Ні. Підпис: «Назва постачальника»
НазваТовар	Текстовий	Довжина: 30. Обов'язкове поле: Так. Пусті строки: Ні. Підпис: «Назва товару»
Кількість	Числовий	
Дата	Дата/час	Формат поля: ДД.ММ.РРРР (Короткий формат)

Рис. 52. Опис типів даних

Знову натисніть правою кнопкою миші на вкладку таблиці "Товар" та оберіть *Режим таблиці* для заповнення таблиці необхідними значеннями. Після заповнення таблиця має вигляд, наведений на рис. 53.

Перед тим, як розпочинати зв'язування двох таблиць, за аналогією з попереднім пунктом потрібно створити і заповнити таблиці "Постачальник" та "Закупівля" (рис. 54).

код товару	найменува	вартість тоє	одиниця ви	Щелкните для добавления
101	Борошно	9,50 ₴	кг.	
102	Цукор	13,00 ₴	кг.	
103	Олія	32,00 ₴	літр	
104	Масло	150,00 ₴	кг.	
105	Макаронні ви	17,00 ₴	кг.	
106	Рис	18,00 ₴	кг.	
107	Бобові	10,00 ₴	кг.	
108	Гречка	16,00 ₴	кг.	
109	Сіль	9,00 ₴	кг.	
110	Чай	27,00 ₴	пачка	

Рис. 53. Заповнена таблиця "Товар"

Назва постачальника	Адреса	Телефон	Щелкните для добавления
ПАТ "Лани Херсонщини"	м. Каховка, Хе	055367-5-67-44	
ВАТ "Смерека"	м. Полтава	03327-3-45-21	
ПП "Старт"	м. Миколаїв	05120-61-45-45	
ПП Данько	м. Херсон	05501-6-87-45-	
ПАТ "Домогосподарочка"	м. Запоріжжя	06062-5-78-45	

Назва поста	Код товару	Кількість	Дата	Щелкните для добавления
ПАТ "Лани Хер	103	150	30.09.2018	
ПП Данько	105	60	02.10.2018	
ВАТ "Смерека"	102	600	02.10.2018	
ПП "Старт"	106	50	17.10.2018	
ПП "Старт"	107	40	17.10.2018	
ПП "Старт"	108	58	17.10.2018	
ПАТ "Лани Хер	101	1100	07.11.2018	
ПАТ "Домогос	110	20	07.11.2018	
ВАТ "Смерека"	109	55	08.11.2018	
ВАТ "Смерека"	104	7	08.11.2018	

Рис. 54. Заповнення таблиць "Постачальники" та "Закупівлі"

Якщо початкова інформація оброблялася до цього в іншій програмі управління БД чи програмою розрахунку електронних таблиць, то можливо перенести вже створені набори даних у MS Access – таблиці, що заощадить

час, особливо якщо мова йде про досить великі БД, що містять велику кількість записів.

Для імпорту набору даних у меню *Файл* вибираємо підменю *Зовнішні дані* команду *Імпорт*. Ця команда дозволить перенести дані з текстового файлу, програми оброблення електронних таблиць в Access-таблицю. Вона також дозволяє перенести об'єкти з іншої БД Access в активну БД.

Окрім імпорту даних у MS Access можливо експортувати таблиці в текстові файли, електронні таблиці, файли інших прикладних програм управління БД, а також в іншу БД. Для виконання експорту маркуйте експортовану таблицю та активуйте в меню *Зовнішні дані* команду *Експорт*. У вікні слід обрати формат файлу чи бази даних, у яку необхідно експортувати дані. Формати експорту ідентичні форматам імпорту. Залежно від того, який формат був обраний користувачем, майстер експорту, що запускається автоматично або в діалозі з користувачем, проводить експорт обраної таблиці.

Також існує можливість зберегти таблицю в наявній БД, але у вигляді іншого об'єкта. Для цього таблицю необхідно виділити й активувати команду *Зберегти як* головного меню *Файл*. У діалоговому вікні збереження необхідно ввести ім'я копії файлу, а також вибрати, в який об'єкт під час збереження буде перетворена таблиця. Наприклад, копію таблиці "Закупівля" збережемо як запит (рис. 55).

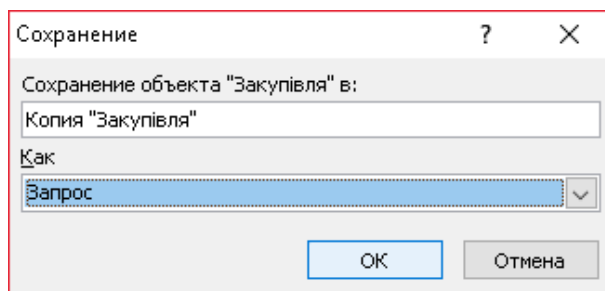


Рис. 55. Вікно перетворення таблиці в запит

Наступним кроком є встановлення зв'язків між створеними таблицями. Для цього необхідно перейти у вкладку *Робота з базами даних*, натиснути на кнопку *Схема даних*, обрати кожну з таблиць, по черзі натискаючи кнопку *Додати* та натиснути кнопку *ОК*. У вікні задайте необхідні параметри і натисніть *ОК*. Якщо схема не була створена автоматично, потрібно натиснути правою кнопкою миші на порожній області та вибрати команду *Додати таблиці* (рис. 56).

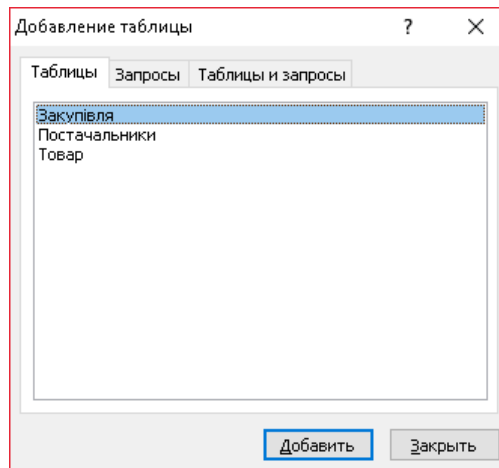
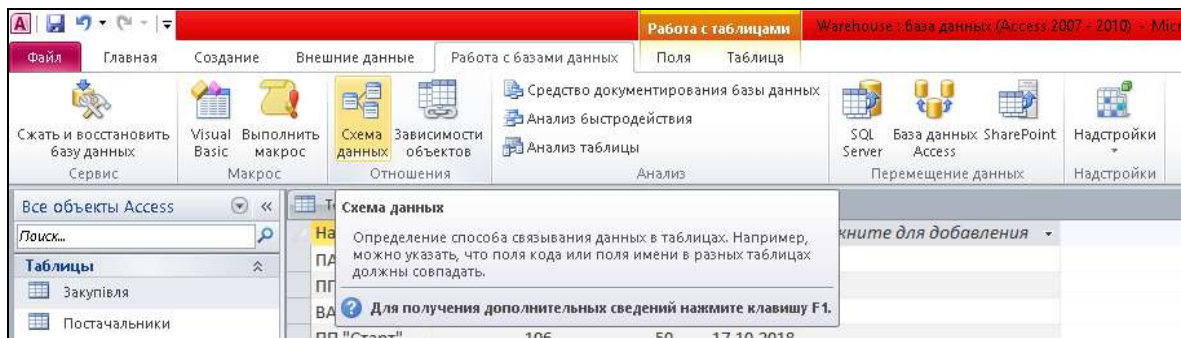


Рис. 56. Вікно додавання таблиці

У результаті з'явиться вікно з таблицями, які входять до БД (рис. 57).

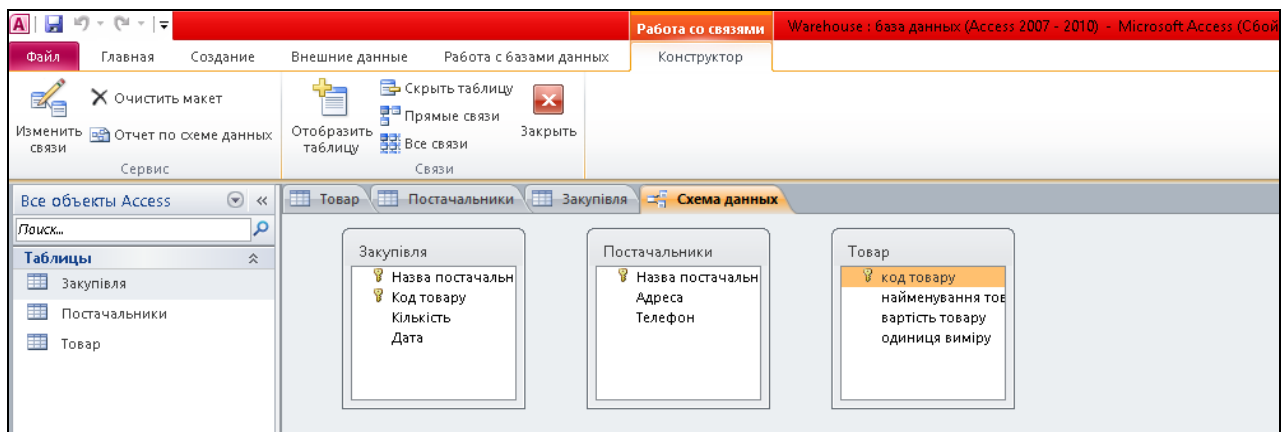


Рис. 57. Вікно схеми даних

У вікні *Схема даних* зв'язки встановлюються по черзі між парами таблиць: спочатку необхідно визначити головну таблицю в парі, потім виділити в схематичному зображенні головної таблиці ключове поле та перетягнути його на поле у підпорядкованій таблиці, з якою встановлюється зв'язок. У вікні *Зміна зв'язків* установіть прапорець *Забезпечення цілісності даних* (рис. 58).

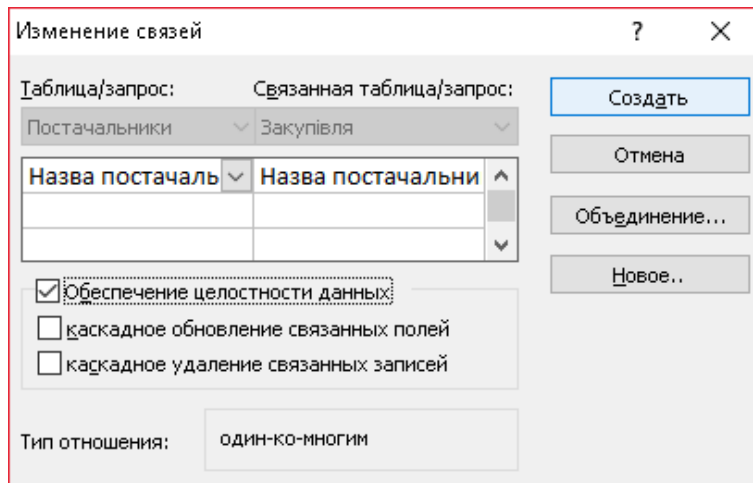


Рис. 58. Вікно встановлення зв'язків між таблицями

Отже, в основному вікні схеми даних зображення таблиць будуть пов'язані лініями, що показують тип зв'язку (рис. 59). Бази даних, які мають зв'язані таблиці, називають реляційними базами даних.

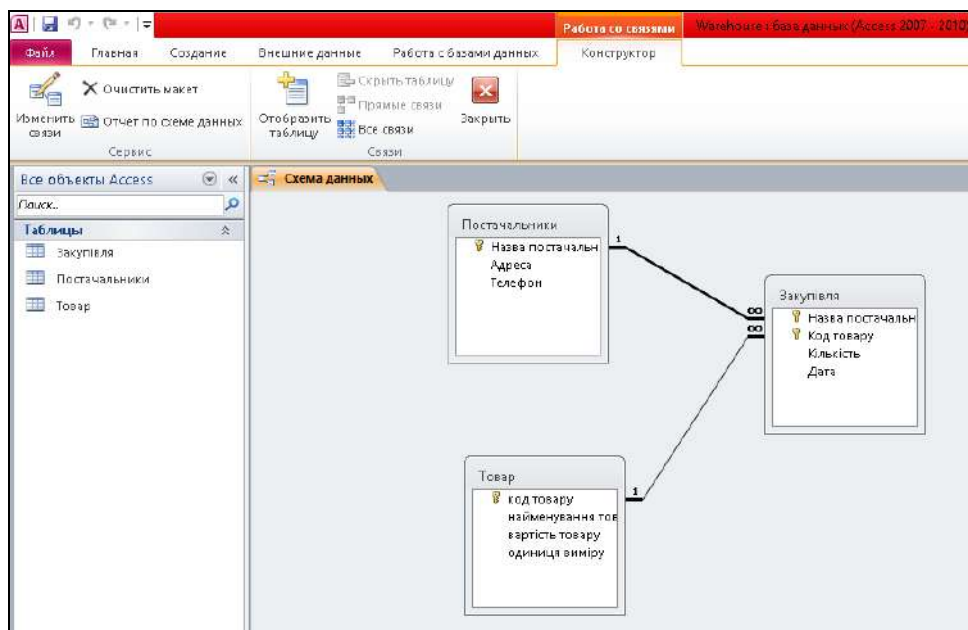


Рис. 59. Готова схема даних БД "Склад"

Зв'язок між таблицями має дві характеристики: ступінь зв'язку та клас належності таблиці до зв'язку. Значення цих характеристик можуть бути визначені з аналізу зв'язку між полями таблиць. Ступінь зв'язку показує, скільки полів однієї таблиці можуть бути зв'язані з кожним екземпляром іншої таблиці. Виділяють три ступеня зв'язків між таблицями у схемі:

"один до одного" (1:1) – одному значенню поля в одній таблиці відповідає тільки одне значення поля в іншій;

"один до багатьох" (1:∞) – одному значенню поля в одній таблиці відповідає декілька (одній чи більше) значень в іншій, а кожне поле другої таблиці може бути зв'язано тільки з одним полем першої таблиці;

"багато до багатьох" (∞:∞) – кожне поле першої таблиці може бути зв'язано з декількома екземплярами другої таблиці та навпаки.

Клас належності таблиці до зв'язку може бути обов'язковим і необов'язковим. За обов'язкового класу належності кожне поле таблиці має бути зв'язано з іншим полем, за необов'язкового класу належності ця умова не є обов'язковою. Інакше кажучи, в БД зберігається інформація про всі поля таблиць.

Тільки найпростіші бази даних містять винятково таблиці. Якби інформація зберігалась у таких простих структурах, то для роботи з нею можливо було б обійтись без спеціальних систем управління базами даних. Основні переваги СУБД реалізуються під час роботи не з окремими таблицями, а з групами взаємопов'язаних таблиць. Тому наступним кроком у створенні БД є формування запитів.

Залежно від подальшого використання або способу отримання даних, отриманих за допомогою запиту, в Access існує **4 типи запитів**:

1. *Запити на вибірку* відбирають потрібні поля даних з записів, що задовольняють заданій умові з однієї або декількох таблиць та/або інших запитів БД.

2. *Перехресні запити* формують дані з таблиць і запитів у вигляді електронної таблиці (матриці). У такій таблиці заголовки рядків утворюються з унікальних значень одного поля, заголовки стовпців – з унікальних значень іншого поля, а на їхньому перетині обчислюються які-небудь сумарні значення для відібраної множини записів.

3. *Запити на зміну* діляться на 4 види:

на створення нової таблиці;

на додавання нових записів у таблицю;

на видалення відібраних записів з таблиці;

на зміну значень яких-небудь полів у відібраних записах таблиці.

4. *Запити з параметрами* – це такі запити, в яких критерії відбору задаються користувачем під час кожного запуску запиту. Тобто запит з параметрами можна отримати з будь-якого з трьох попередніх типів запитів, ввівши в нього параметри для відбору записів. Під час запуску такого запиту на екран виводиться вікно, в яке слід ввести значення параметра вибірки, для кожного з параметрів.

Наприклад, для створення запиту на вибірку необхідно у вікні БД перейти на вкладку *Створення* та натиснути *Майстер запитів* (рис. 60).

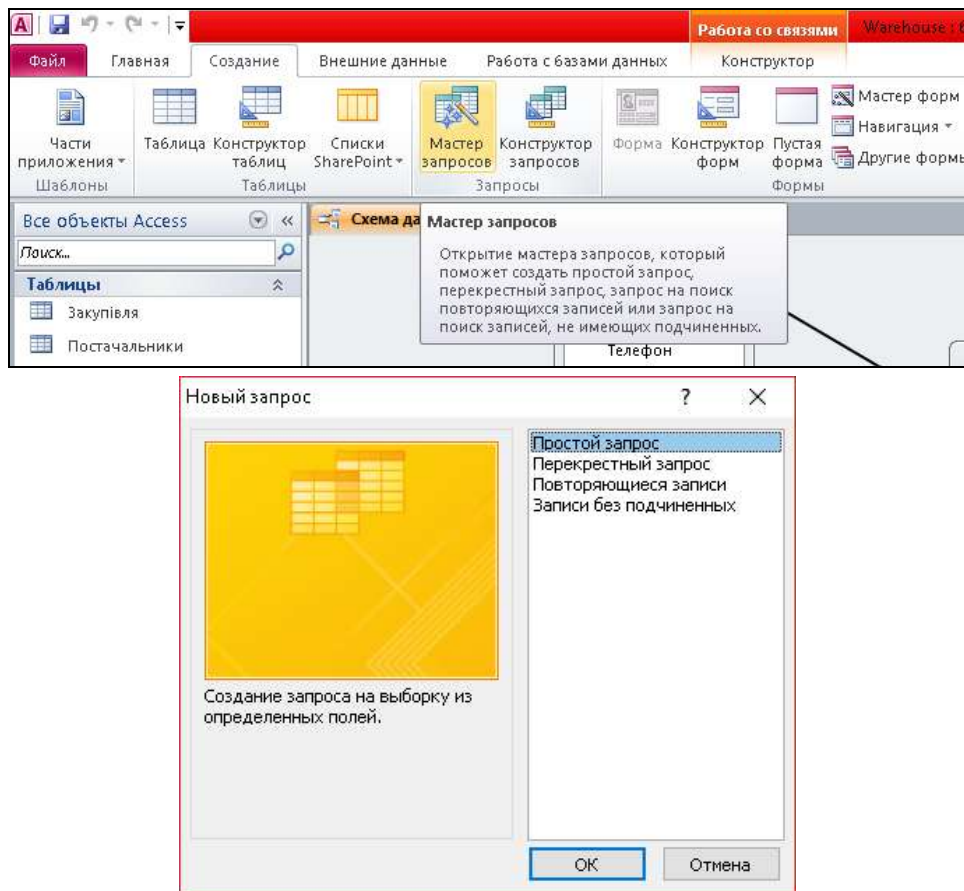


Рис. 60. Створення запиту

Потім необхідно додати всі таблиці в запит (рис. 61).

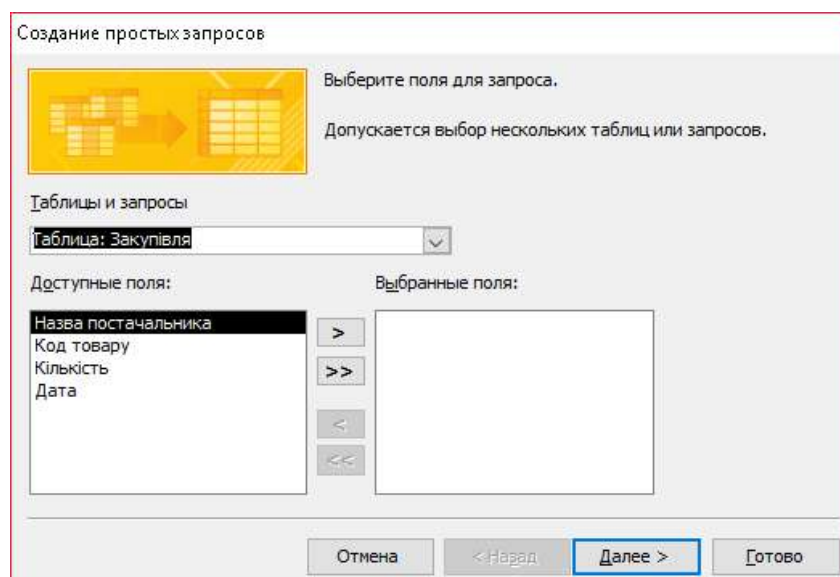
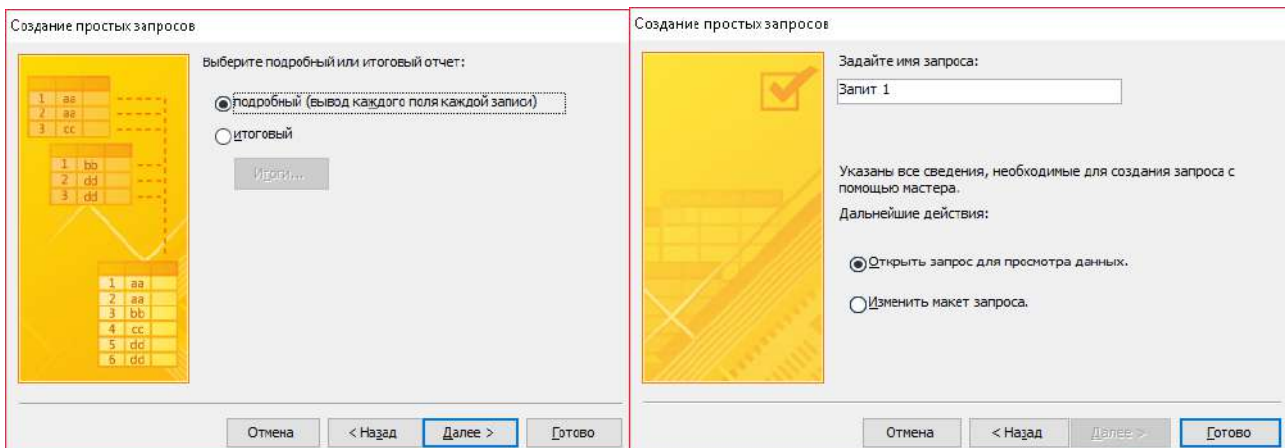


Рис. 61. Додавання таблиць (полів) до запиту

Покрокове створення запиту наведено на рис. 62.



Имя	Идентификатор	Идентификатор	Дата	Поставщик	Адрес	Телефон	Товар	Цена	Единица
ПАТ "Лани Херсонщини"	103	150	30.09.2018	ПАТ "Лани Херсонщини"	м. Каховка, Хв 055967-5-67-41	055967-5-67-41	103 Сілія	32,00 Р	літр
ПП "Данько"	105	60	02.10.2018	ПП "Данько"	м. Херсон	05501-6-87-45	105 Макароны ви	17,00 Р	кг.
ВАТ "Смерека"	102	600	02.10.2018	ВАТ "Смерека"	м. Полтава	03327-3-45-21	102 Цукор	13,00 Р	кг.
ПП "Старт"	100	50	17.10.2018	ПП "Старт"	м. Миколаїв	05120-61-45-41	106 Рис	18,00 Р	кг.
ПП "Старт"	107	40	17.10.2018	ПП "Старт"	м. Миколаїв	05120-61-45-41	107 Бобови	10,00 Р	кг.
ПП "Старт"	108	58	17.10.2018	ПП "Старт"	м. Миколаїв	05120-61-45-41	108 Гречка	16,00 Р	кг.
ПАТ "Лани Хв"	101	1100	07.11.2018	ПАТ "Лани Херсонщини"	м. Каховка, Хв 055267-5-67-41	055267-5-67-41	101 Борщ	9,50 Р	кг.
ПАТ "Домогос"	110	20	07.11.2018	ПАТ "Домгоспідприємства"	м. Запоріжжя	06062-5-78-45	110 Чай	27,00 Р	пачка
ВАТ "Смерека"	109	35	08.11.2018	ВАТ "Смерека"	м. Полтава	03327-3-43-21	109 Сіль	5,00 Р	кг.
ВАТ "Смерека"	104	7	08.11.2018	ВАТ "Смерека"	м. Полтава	03327-3-45-21	104 Масло	150,00 Р	кг.

Рис. 62. Результаты створення запиту

Також існує інший спосіб створення запитів – через *Конструктор запитів* (рис. 63).

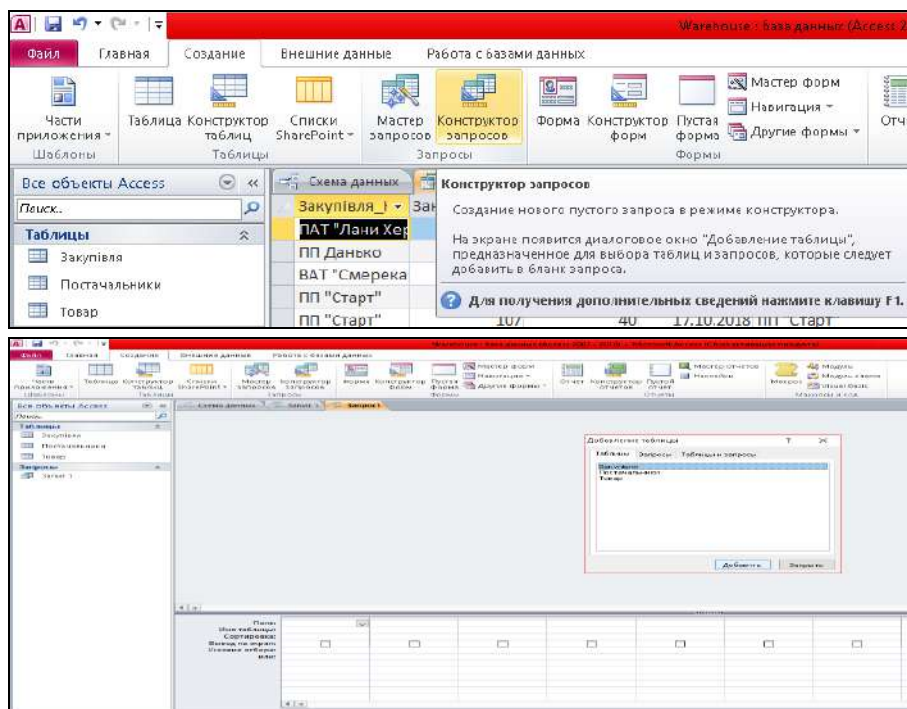


Рис. 63. Запуск Конструктора запитів

Додаємо всі таблиці до запиту. Унаслідок вони відобразяться у верхньому полі створеного запиту (рис. 64).

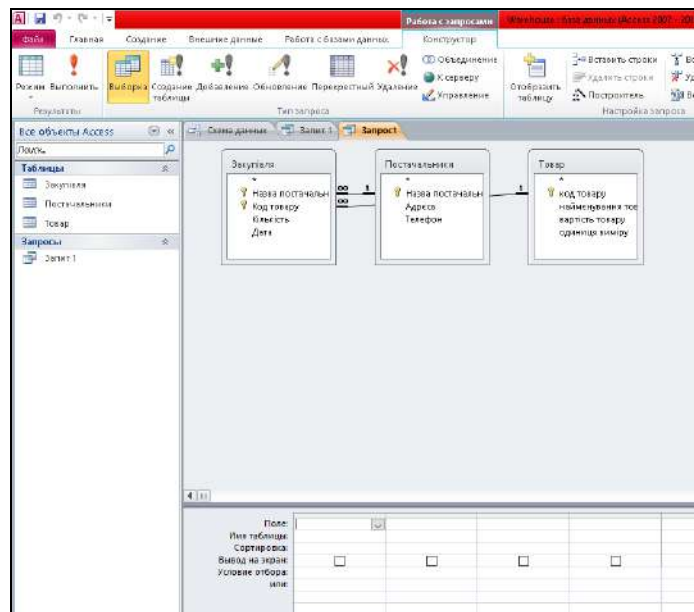


Рис. 64. Додавання таблиць до запиту

Після цього необхідно додати в рядок *Поле* в нижній частині вікна запиту імена полів таблиць БД, що відбираються цим запитом. Для цього клікніть у клітинці першого рядка та зі списку оберіть потрібне поле. В переліку зірочками відмічені назви таблиць, де з них – поля таблиць. Рекомендується обирати поля. Порядок проходження полів зліва направо має бути за зменшенням важливості інформації (рис. 65).

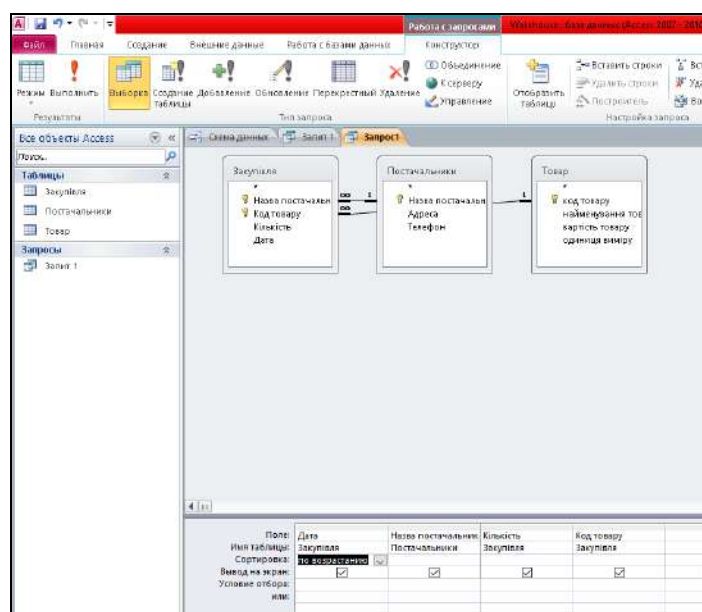


Рис. 65. Формування запиту

Після натиснення на команду *Виконати!* (кнопка червоного кольору на панелі управління) можливо переглянути готовий запит у вікні БД (рис. 66).

Дата	Назва постачальника	Кількість	Код товару
30.09.2018	ПАТ "Лани Херсонщини"	150	103
02.10.2018	ВАТ "Смерека"	600	102
02.10.2018	ПП Данько	60	105
17.10.2018	ПП "Старт"	58	108
17.10.2018	ПП "Старт"	40	107
17.10.2018	ПП "Старт"	50	106
07.11.2018	ПАТ "Домогосподарочка"	20	110
07.11.2018	ПАТ "Лани Херсонщини"	1100	101
08.11.2018	ВАТ "Смерека"	7	104
08.11.2018	ВАТ "Смерека"	55	109

Рис. 66. Сформований запит

Наступним об'єктом бази даних є форма.

Форми – це об'єкти, за допомогою яких у базу вводять нові дані, переглядають наявні, а також здійснюють редагування даних, розміщених у таблиці. Форма подібна до бланка чи карточки. Прикладами форми є будь-який бланк, картка з досьє чи бібліографічна картка на книжку в бібліотеці. Одна форма містить дані лише з одного запису.

MS Access дозволяє швидко створювати форми з допомогою *Майстра форм*. Можна також розробляти форми вручну, відкривши пусту форму та перетягнувши об'єкти (наприклад, поля, текст і графіку) в потрібне місце форми. Можна поєднувати автоматизоване та ручне розроблення. Наприклад, за допомогою *Майстра форм* можна створити основу форми, а потім відкрити форму в режимі *Конструктор форми* та внести зміни вручну.

Якщо користувачеві потрібно управляти процесом створення форм і витратити на це менше зусиль, то краще скористатися допомогою *Майстра форм*. Як й інші надбудови Access, *Майстер форм* задає низку запитань про форму, яка створюється, а потім будує форму залежно від відповідей користувача.

На формі можливо розташувати:

- поля типу OLE з картинками, фотографіями тощо;
- елементи керування: кнопки, перемикачі тощо;

надписи: заголовки форми, рубрик (а також розмалювати форму чи задати фоновий рисунок-заставку);

обчислювальні поля (це також елементи керування) для відображення результатів обчислень, виконаних на базі наявних полів;

закладки (багатосторінкові форми, де поля групують за змістом на різних закладках);

підпорядковані форми тощо.

Наприклад, для створення простої форми необхідно через пункт меню *Створення* вкладку *Майстер форм* створити макет форми, натиснувши на відповідну команду групи *Форма*. Потім необхідно обрати, які саме поля увійдуть до створеної форми та покроково за допомогою *Майстра форм* виконати дії, показані на рис. 67.

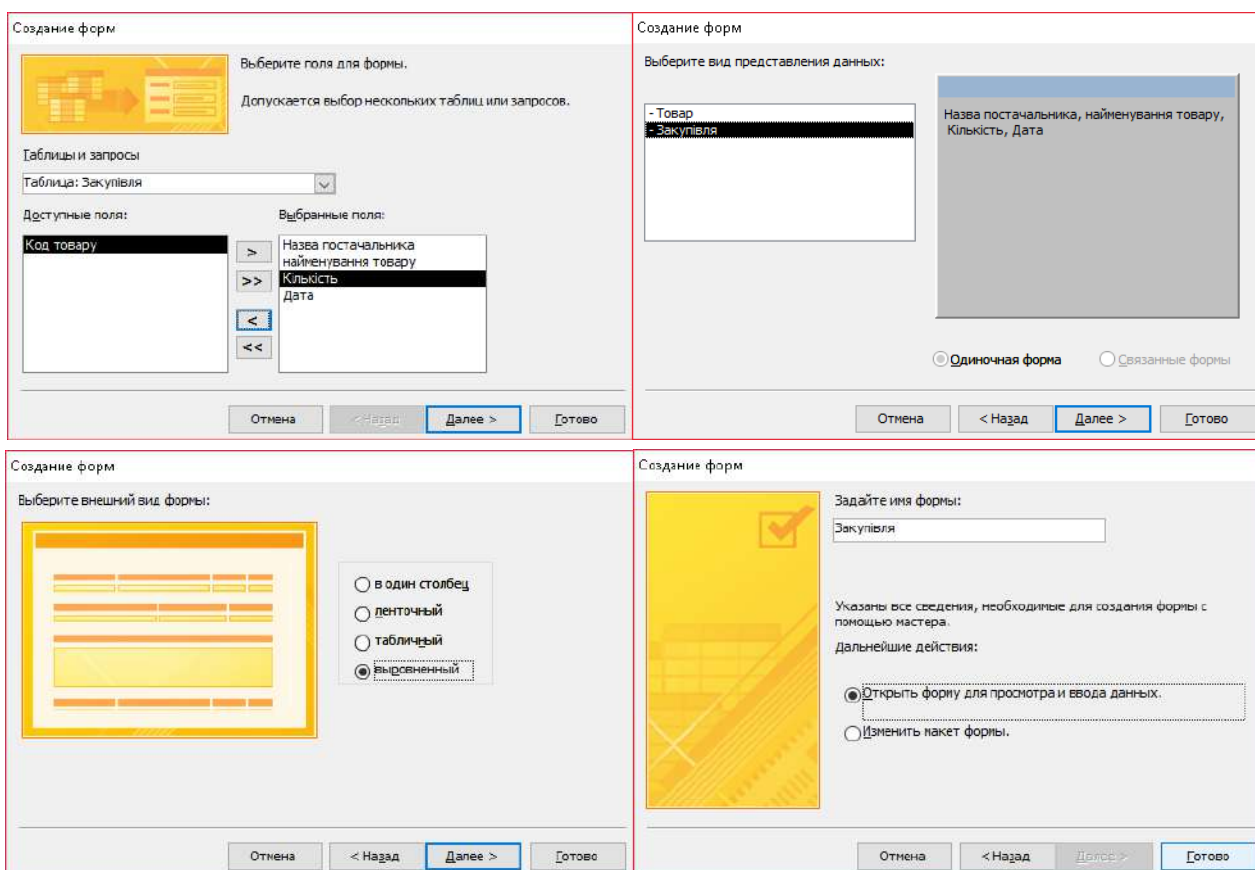


Рис. 67. Створення форми

Унаслідок отримаємо форму "Закупівля" (рис. 68).

За допомогою навігаційних кнопок користувач може переходити до наступного запису, а також вносити зміни (рис. 69).

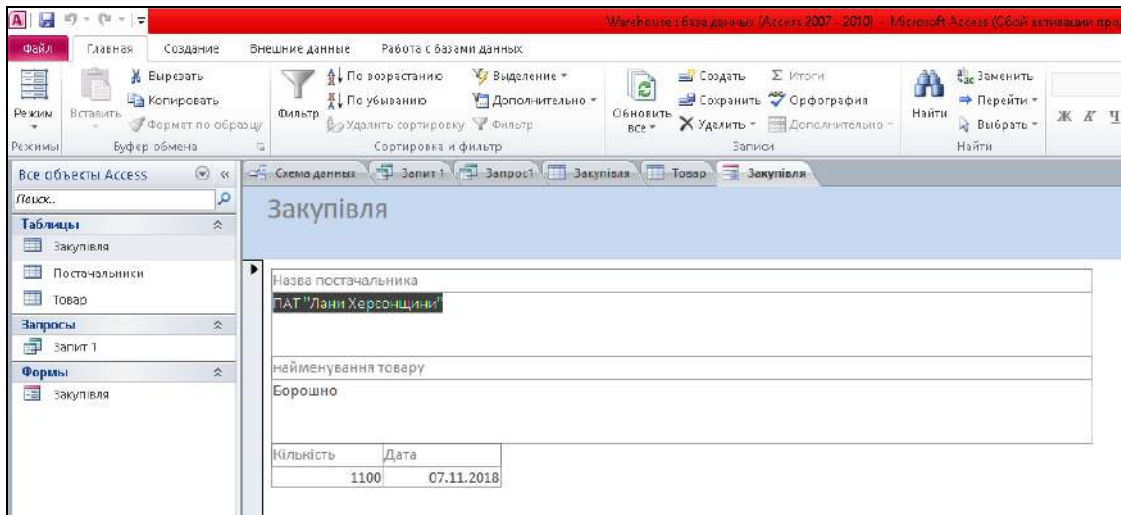


Рис. 68. Форма "Закупівля"

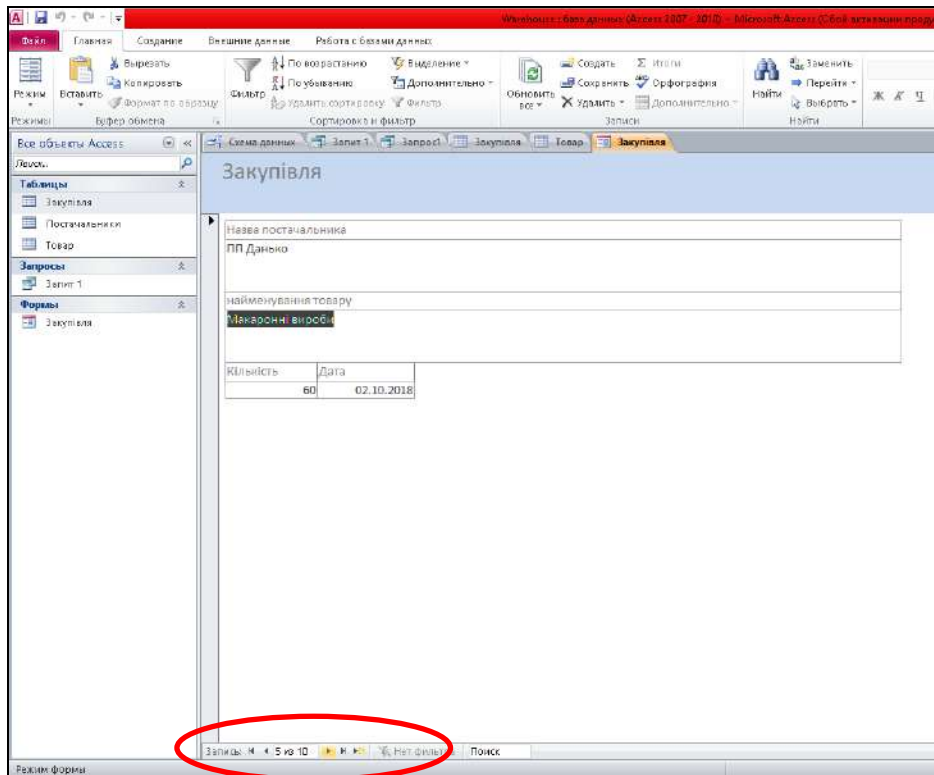


Рис. 69. Форма "Закупівля"

Фактично за допомогою форми створюється графічний інтерфейс доступу до БД, який може містити різні елементи управління (текстові поля, кнопки, перемикачі і т. д.), а також записи. Зазвичай на формі розміщуються записи, які є іменами полів бази даних, і текстові поля, що містять дані з бази даних. Користувач може змінювати дизайн форми (розмір, колір і т. д.), елементів управління та записів.

Форми створюються на основі таблиць і запитів. Під час кожного відкриття збереженої форми оновлюються дані запиту, на основі якого створюється форма. Тому зміст форми завжди відповідає інформації в таблицях і запитах.

Змінювати дизайн форми можливо за допомогою вкладки *Формат* (рис. 70).

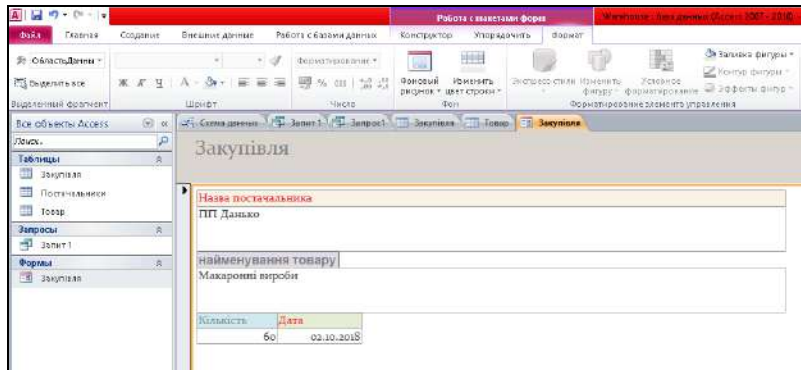


Рис. 70. Робота з вкладкою *Формат*

Порада! Рекомендуємо використовувати режим *Роздільна форма*, оскільки крім самого шаблону, в нижній частині вікна буде відображатися мініатюра таблиці, яка зробить процес редагування ще більш наочним (рис. 71).

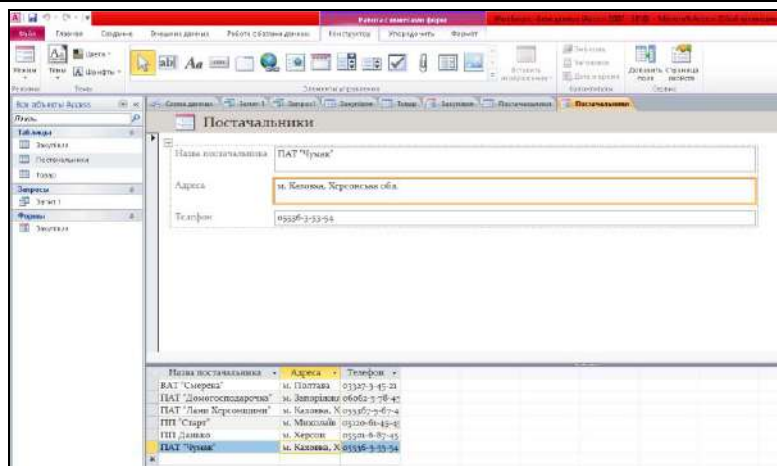
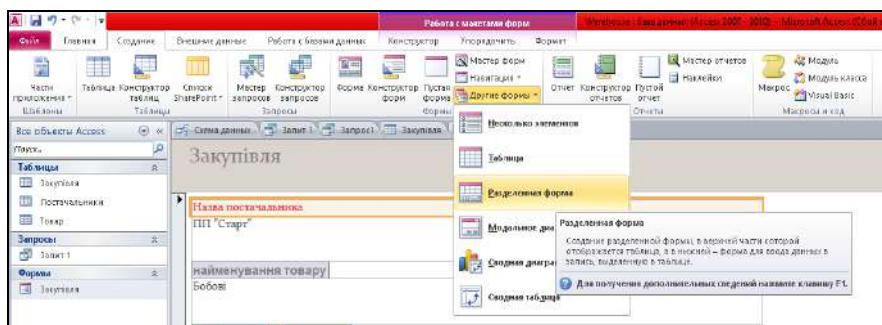


Рис. 71. Редагування дизайну форми

Звіти бази даних MS Access – об'єкти бази даних, які використовують для підведення яких-небудь підсумків на підставі наявних даних, і виведення цих підсумків у певному форматі на друк.

Звіти дають широкі можливості для групування й обчислення проміжних і загальних підсумків для великої кількості даних. Звіти можуть бути використані для одержання гарно оформлених рахунків, замовлень на закупівлю, поштових наліпок тощо.

Програма MS Access дає можливість користувачу створити звіт декількома способами, основними з яких є:

- конструктор звітів;
- майстер звітів;
- порожній звіт.

Найчастіше звіт створюють за допомогою майстра звіту, а пізніше дооформлюють власноруч за допомогою конструктора звітів.

Звіт складається з декількох елементів, які створює конструктор, а саме: заголовок звіту;

верхній і нижній колонтитули (повторюються на всіх сторінках);
область даних.

За допомогою елементів керування у звіт можна додавати рисунки, діаграми та інші об'єкти. У колонтитулі можна розмістити будь-яку обчислювальну функцію бази даних.

Примітка. Редагувати дані у звіті не можна.

Для створення звіту слід перейти на вкладку *Створення* вибрати команду *Майстер звітів*. У вікні *Створення звітів* слід обрати відповідний об'єкт, на підставі якого буде створено звіт, натиснути кнопку *Далі*, на наступному кроці за необхідністю розподілити рівні групування, після знов натиснути кнопку *Далі*, обрати порядок сортування записів, натиснути кнопку *Далі*, потім обрати вид макету звіту, натиснути кнопку *Далі* задати ім'я звіту та натиснути кнопку *Готово*. Після раніше зазначених дій буде створено звіт (рис. 72).

Якщо відображення звіту вас не влаштовує, його можна трохи підкоригувати. Для цього необхідно натиснути правою кнопкою миші на вкладці звіту і вибрати вкладку *Конструктор*. Потім вручну редагуйте відповідні стовпці (рис. 73).

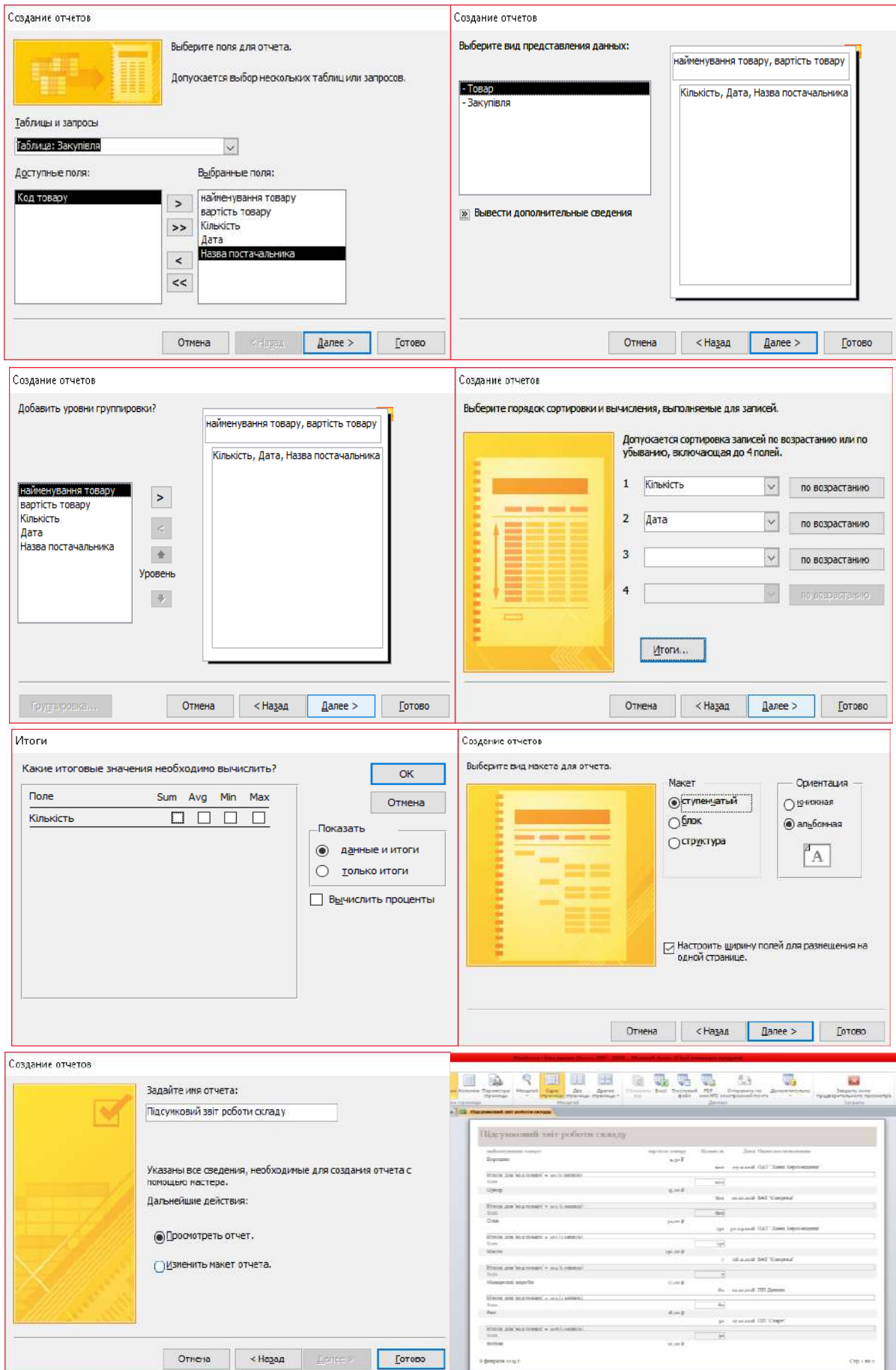


Рис. 72. Этапы створення звіту

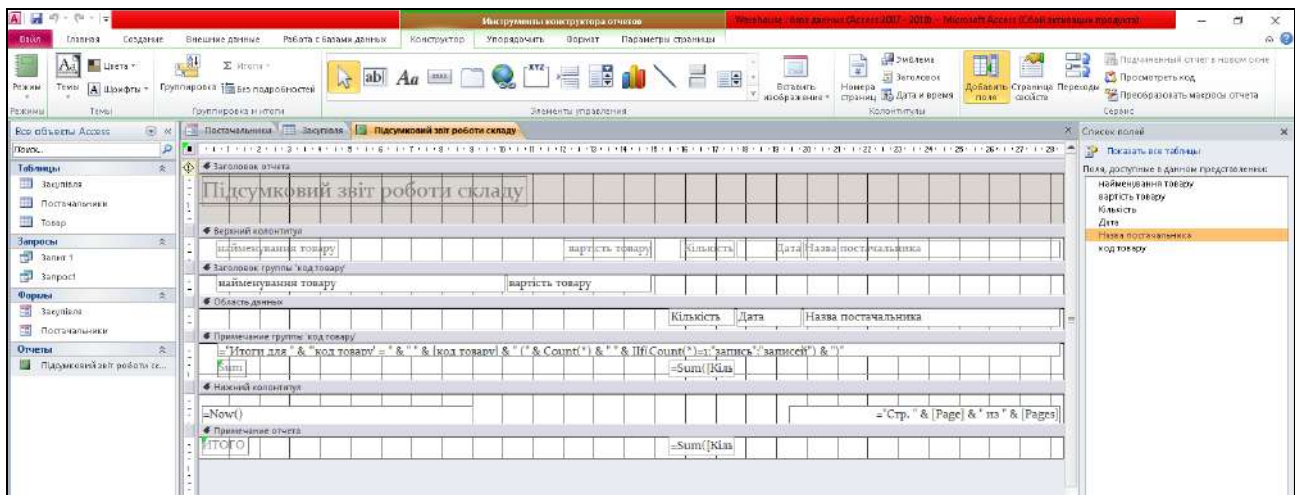


Рис. 73. Редагування звіту

Підсумовуючи, можна сказати, що в лабораторній роботі наведено основні рекомендації щодо проектування бази даних, створення таблиць, запитів, форм і звітів. Також подано приклад сформованої бази даних з інструкціями користувача, який є наглядним помічником під час створення нової БД.

Завдання:

1. Розробіть СУБД "Абітурієнт" для автоматизації роботи регіонального центру оцінювання якості освіти. БД має містити три таблиці: анкети абітурієнтів, дані з навчальної дисциплін і результати іспитів.

Анкета містить наступні дані про абітурієнта:

- реєстраційний номер (ключове поле);
- прізвище, ім'я, по батькові;
- дату народження;
- навчальний заклад, який абітурієнт закінчив (назву, номер, населений пункт);
- дату закінчення навчального закладу;
- наявність червоного диплома або золотої/срібної медалі;
- адресу проживання (місто, вулиця, номер будинку, телефон).

Дані з навчальної дисципліни містять:

- шифр навчальної дисципліни (ключове поле);
- назву навчальної дисципліни.

Результати іспитів містять:

- реєстраційний номер абітурієнта;
- шифр навчальної дисципліни;

- екзаменаційну оцінку;
- середній бал атестату.

Результати роботи оформіть у вигляді звіту з лабораторної роботи.

Завдання для самостійного опрацювання

Запитання, що виносяться на самостійну роботу студентів:

1. Охарактеризуйте сферу використання MS Access, її можливості.
2. У чому полягає прикладний характер СУБД? Наведіть приклади.
3. Які БД, окрім MS Access, вам відомі?
4. Назвіть компоненти бази даних Access. Які саме переваги вони надають у роботі бізнес-аналітика?
5. Модулі та макроси MS Access.

Завдання 1. Розробіть інформаційну систему "Споживчий кошик" для аналізу рівня життя родині. Рівень життя залежить від співвідношення доходів сім'ї та цін на споживані продукти.

БД має містити як мінімум 3 таблиці: "Продукти", "Доходи" і "Споживання".

Таблиці мають таку структуру:

"Продукти":

код продукту (ключове поле);
найменування;
одиниці виміру.

"Доходи":

рік, місяць (ключове поле);
сукупний дохід за місяць.

"Споживання":

рік, місяць;
код продукту;
кількість;
ціна.

Завдання 2. Розробіть статистико-логістичну систему "Перевезення" для транспортного підприємства. На підприємстві є пакет заявок від інших організацій на перевезення різних вантажів. БД системи складається з трьох таблиць: "Транспорт", "Заявки", "Доставка", що мають таку структуру:

"Транспорт":

марка автомобіля;
державний номер (ключове поле);
витрата палива (літрів на 100 км).

"Заявки":

код заявки (ключове поле);
дата;
пункт відправлення;
пункт призначення;
назва вантажу;
одиниця виміру;
кількість вантажу.

"Доставка":

дата і час відправлення;
дата і час повернення;
державний номер автомобіля;
код заявки;
одиниця вимірювання;
кількість фактично перевезеного вантажу;
пройдена відстань.

Тема 4. Побудова проєктів за допомогою MS Project

Лабораторна робота 4

Побудова проєктів за допомогою MS Project

Мета: придбання навичок побудови проєкту за допомогою MS Project.

Завдання: ознайомитися з теоретичною базою побудови проєкту; побудувати різні види проєктів; навчитися визначати час початку та закінчення работ, критичний час проєкту; вміти оптимізувати проєкт за критеріями часу, вартості та кількості виконавців.

Методичні рекомендації

1. Знайомство з MS Project.

Microsoft Project – це комплексне програмне забезпечення, система управління проєктами та спосіб оптимізації управління портфелями, який

дозволяє планувати й контролювати проектну діяльність організацій. За допомогою використання вбудованих шаблонів, інструментів для різного рівня аналітики та статистики, засобів управління робочим часом і т. п.

Microsoft Project створений, щоб допомогти менеджеру проекту в розробленні планів, розподілі ресурсів за завданнями, відстеженні прогресу й аналізі обсягів робіт. Microsoft Project створює розклад критичного шляху. Розклади можуть бути складені з урахуванням використовуваних ресурсів.

Роботу в Microsoft Project рекомендується починати з освоєння проектного підходу, ознайомлення з його принципами та методами проектування. Це дозволить правильно виконувати такі процеси: розділяти великі проекти на частини, коригувати тимчасові оцінки, враховувати і закладати ризики, відстежувати командну роботу та користуватися мотиваційними прийомами.

У разі реалізації проектного підходу програма Project допомагає вирішувати такі завдання:

1. Покроково розробляти проект, користуючись інтелектуальними підказками на різних рівнях.

2. Створювати завдання, розподіляти ресурси, зазначати терміни виконання та отримувати модель, яка буде максимально наближена до реальності.

3. Створювати мережеву діаграму проекту, тобто встановлювати зв'язки завдань і пріоритетів з вирівнюванням ресурсів у міру реалізації проекту.

4. Привласнювати ієрархічні коди структур для ресурсів і завдань.

5. Використовувати консолідований аналіз груп проектів і загальний пул ресурсів, що дозволяє оптимізувати проектний план на рівні організації загалом.

6. Розраховувати критичний шлях та змінювати вказаний період виконання завдань у разі потреби

7. Представляти за допомогою графічних індикаторів проблемні завдання.

8. Аналізувати як поточний стан справ, так і тенденції розвитку проекту за допомогою автоматизованих засобів підготовки звітів.

9. Створювати шаблони проектів.

10. Додавати макроси і VBA-програми.

11. Моделювати варіанти будь-яких рішень, а також аналізувати наслідки гіпотетичного прийняття тієї чи іншої моделі (стратегії, рішення).

Для роботи в програмі використовують поняття "завдання", "ресурс" і "призначення". Для досягнення мети проекту робота розбивається

на завдання. Поняття "ресурс" частіше застосовується до співробітника, але може застосовуватися і до нерухомості, устаткування, матеріалів. У Microsoft Project призначення виникають у той момент, коли на виконання завдання виділяються ресурси. Саме призначення визначають обсяг необхідного на рішення завдань часу і, як наслідок, загальний час проєкту.

Для відображення, аналізу та введення даних існують представлення завдання (діаграма Ганта, форма завдань й ін.), ресурсів (графік ресурсів, лист ресурсів) і призначень (наприклад, використання ресурсів), які бувають графічними, табличними та у вигляді форм.

Щоб на одному екрані повніше відобразити необхідну інформацію, замість одиночного (звичайного) режиму представлення шляхом встановлення "прапорця" в меню, виводять комбінований режим. У цьому випадку екран ділиться по горизонталі, що дає змогу бачити одночасно два складника проєкту.

Поділ робіт проєкту формує структуру декомпозиції робіт, у якій завдання представлені різними типами: окремими завданнями; сумарним завданням (фазою), що складається з групи пов'язаних завдань (підзавдань); віхою – опорної відміткою – точкою важливої події, за якою контролюють хід виконання проєкту та повторювальними завданнями (регулярно виникають по ходу реалізації проєкту (наприклад, "ранкові планерки").

Починаючи з версії Project 2010, завдання можна планувати як в автоматичному, так і в ручному режимах, розміщуючи, в цьому випадку, завдання в будь-яке місце розкладу.

Діаграма Ганта (англ. Gantt chart, також стрічкова діаграма, графік Ганта, календарний графік) – це популярний тип стовпчастих діаграм (гістограм), який використовується для ілюстрації плану, графіка робіт з якогонебудь проєкту. Є одним з методів планування проєктів. Використовується в додатках з управління проєктами. Перший формат діаграми був розроблений Генрі Л. Гантом у 1910 році.

По суті, діаграма Ганта складається з смуг, орієнтованих уздовж осі часу. Кожна смуга на діаграмі представляє окреме завдання в складі проєкту (вид роботи), її кінці – моменти початку і завершення роботи, її протяжність – тривалість роботи. Вертикальною віссю діаграми слугує перелік завдань. Крім того, на діаграмі можуть бути відзначені сукупні завдання, відсотки виконання, показники послідовності та залежності робіт, мітки ключових моментів (віхи), мітка поточного моменту часу та ін.

Ключовим поняттям діаграми Ганта є "віха" – це мітка значимого моменту в ході виконання робіт, спільний кордон двох або більше завдань.

Віхи дозволяють наочно відобразити необхідність синхронізації, послідовності у виконанні різних робіт. Віхи, як і інші кордони на діаграмі, не є календарними датами. Зрушення віхи призводить до зрушення всього проекту. Тому діаграма Ганта не є графіком робіт. Крім того, діаграма Ганта не відображує значущості або ресурсоємності робіт, не відображує сутності робіт (області дії). Для великих проєктів діаграма Ганта стає надмірно великовою та втрачає будь-яку наочність.

Вікно Microsoft Project 2013 складається з таких елементів:

1. Рядок меню.
2. Панелі інструментів.
3. Рядок введення.
4. Панель демонстрацій.
5. Робоча область.
6. Рядок стану.

Рядок меню, панелі інструментів і рядок стану є стандартними для всіх Windows-додатків. Прийоми роботи з ними такі самі, як і в Microsoft Office.

Рядок введення призначений для введення та редагування даних в осередку таблиць (на зразок рядка формул в Excel).

Панель демонстрацій використовується для перемикання між представленнями робочої області вікна. Всі дані про проєкт зберігаються в єдиній базі даних, що складається з великої кількості полів.

Представлення – це спосіб відображення частини пов'язаних між собою даних із загальної бази даних проєкту. В системі реалізовано досить велика кількість представлень – діаграма Ганта, сітьовий графік, календар, графік ресурсів і т. д. При бажанні можна змінювати стандартні представлення.

Робоча область призначена для відображення обраного представлення. Вона може містити таблиці, діаграми, графіки, форми і використовується як для перегляду, так і для редагування даних проєкту.

Для створення нового проєкту слід вибрати пункт меню *Файл / Створити*. Буде створено порожній проєкт з порожньою базою даних. Також Microsoft Project 2013 пропонує використовувати вже створені шаблони проєктів, а саме:

1. Простий план проєкту – це дуже простий шаблон, який допомагає новачкам створювати план. За допомогою цього шаблону можна легко створювати завдання і кінцеві результати проєкту та управляти ними.

2. Освоєний обсяг – цей шаблон допоможе вам створити проєкт для відстеження освоєного обсягу.

3. Гнучке управління проєктами – цей шаблон призначений для тих, хто хоче використовувати Microsoft Project для управління "agile-проєктами". Він містить кілька подань і фільтрів для того, щоб слідувати цій методології.

4. Створіть бюджет – цей шаблон допоможе вам використовувати ресурси бюджету і містить звіти, що працюють з ресурсами бюджету. За допомогою цього шаблону можна відстежувати витрати за проєктом.

5. Бізнес-план для стартапу – цей шаблон допоможе почати бізнес з нуля: створити тимчасову шкалу, визначити стадії проєкту та впорядкувати ресурси, необхідні для виконання всіх завдань. Також дозволить визначити перспективи росту компанії, підхід до ведення справ, оцінити ризики та переваги, створити бізнес-план, маркетинговий план, організувати офісний простір і підібрати персонал.

6. Підготовка річного звіту – цей шаблон враховує загальні фінансові та бухгалтерські види діяльності, які використовуються для підготовки фінансового річного звіту. Сюди входить отримання даних з декількох підрозділів, робота з іншими підрозділами, щоб забезпечити точність представлення, друку та поширення необхідної інформації.

7. План маркетингової компанії – у цьому шаблоні наведено різні рекомендації з планування рекламної кампанії в організації, включно з визначенням положення на ринку, збором даних і аналізом ефективності кампанії.

8. Цикл шести сигм (OIAСK) – у цьому шаблоні може визначати фази, ключові дії та результати стандартного циклу шести сигм.

2. Створення проєкту засобами MS Project 2013.

Використовуючи програму MS Project 2013, розглянемо покроковий алгоритм побудови проєкту на прикладі проєкту "Відкриття нового магазину".

На першому кроці запускаємо MS Project 2013 та обираємо команду *Новий проєкт (Blank Project)* (рис. 74).

У результаті перед користувачем відкривається робочий аркуш MS Project 2013 (рис. 75).

Перед початком роботи над створенням нового проєкту необхідно задати певні параметри, що суттєво спростять подальшу роботу, а саме:

1. Змінити планування завдань (за замовчуванням стоїть режим планування завдань *Вручну*, змінюємо на *Автоматичне планування*). Це дозволить в автоматичному режимі виставляти дати завдань відповідно до встановлених зв'язків) (рис. 76).

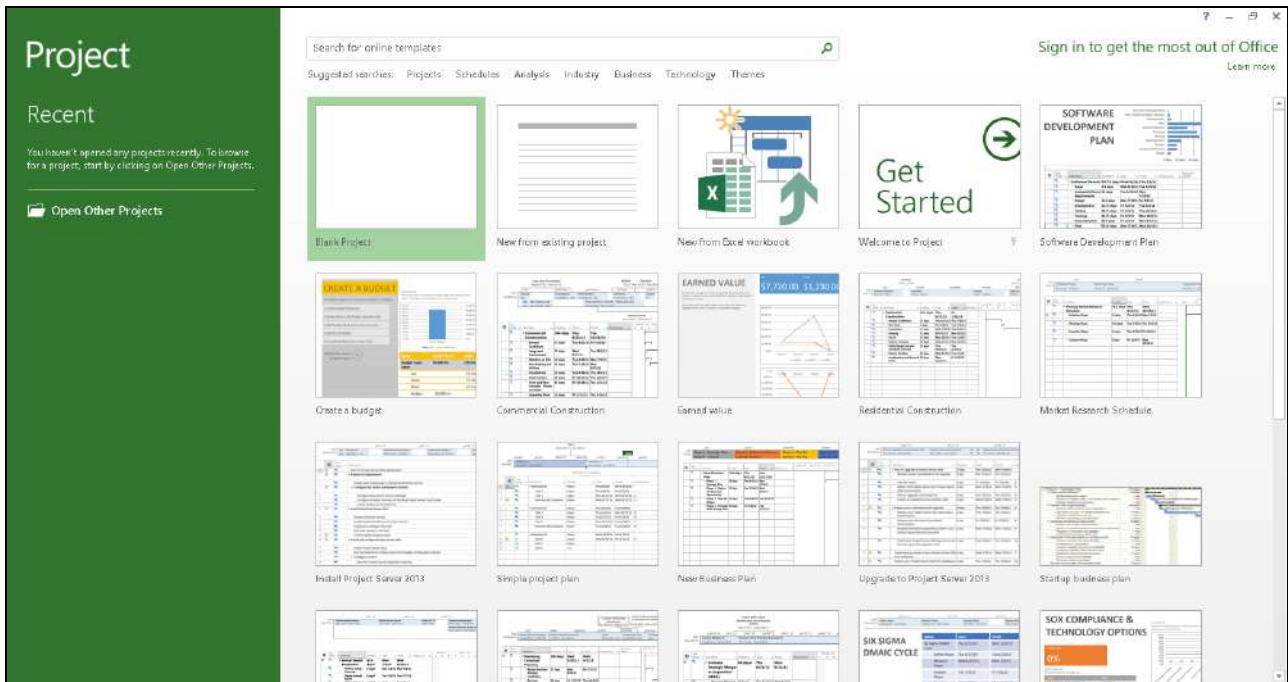


Рис. 74. Запуск програми MS Project 2013

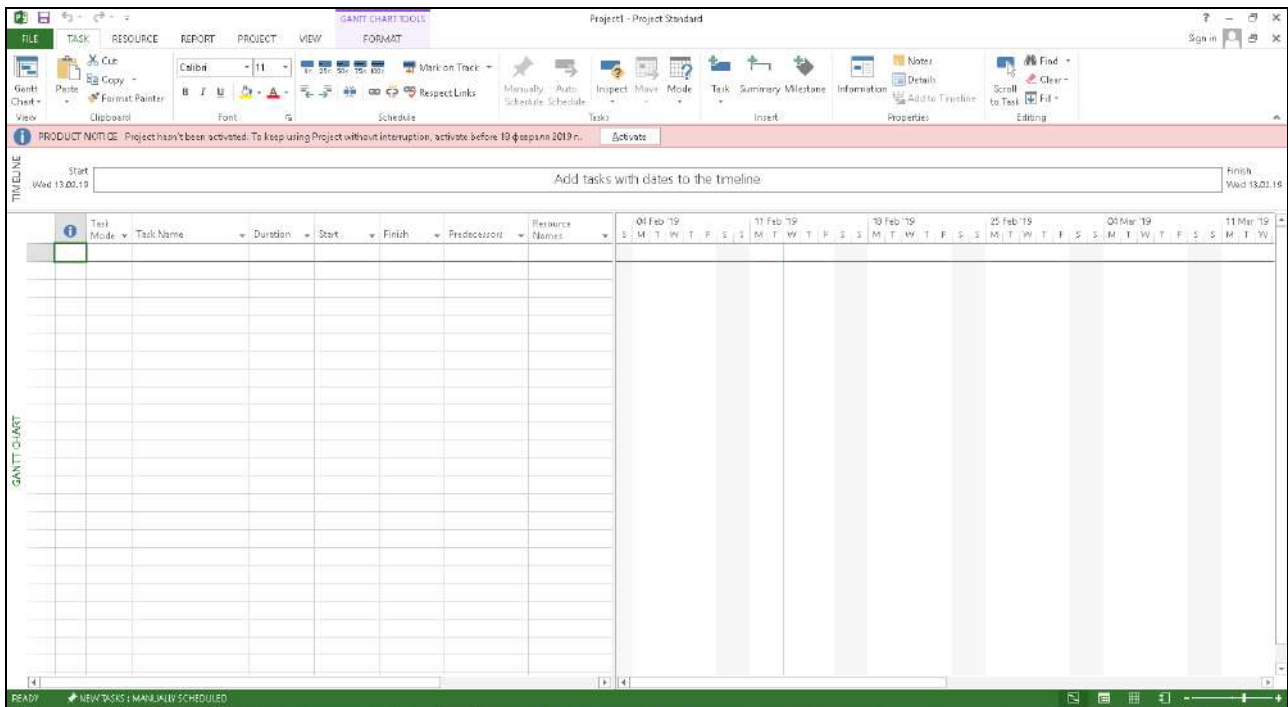


Рис. 75. Вікно нового проєкту MS Project 2013

2. У вкладці панелі інструментів *Формат* обираємо команду *Сумарне завдання проєкту (Project Summary Task)* (рис. 77). Це дозволить бачити підсумкове завдання проєкту.

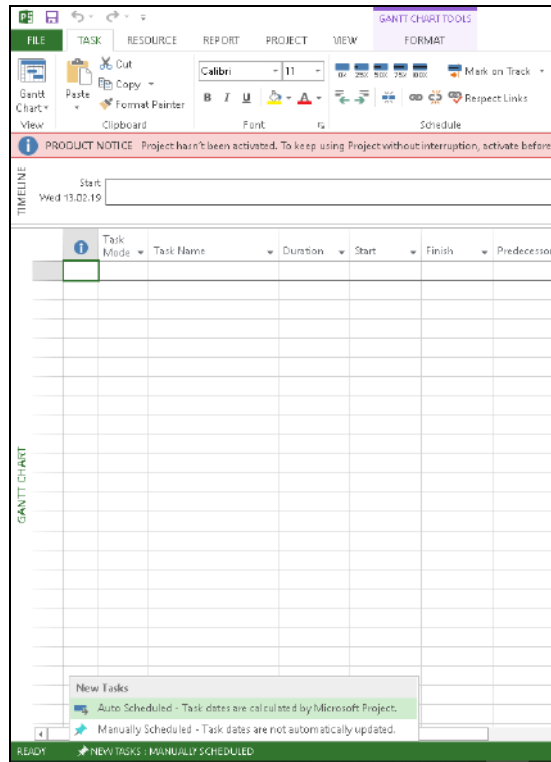


Рис. 76. Зміна режиму планування завдань

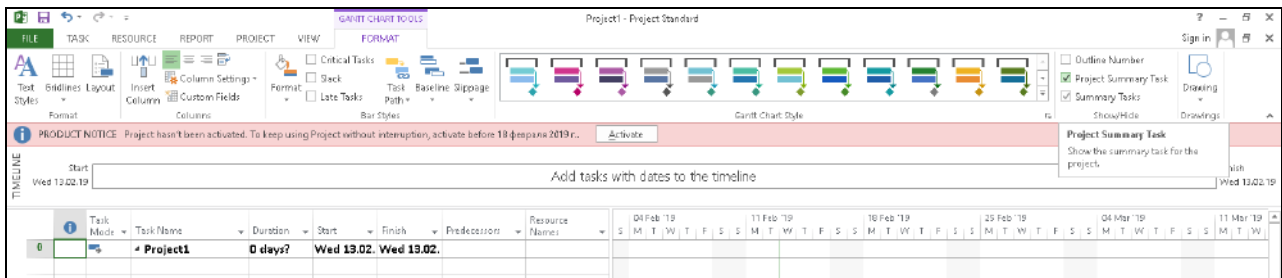


Рис. 77. Активація команди *Сумарне завдання проєкту*

3. Також, з метою контролю виконання всіх критичних завдань, встановлюємо відповідний прапорець у вкладці *Формат* (рис. 78).

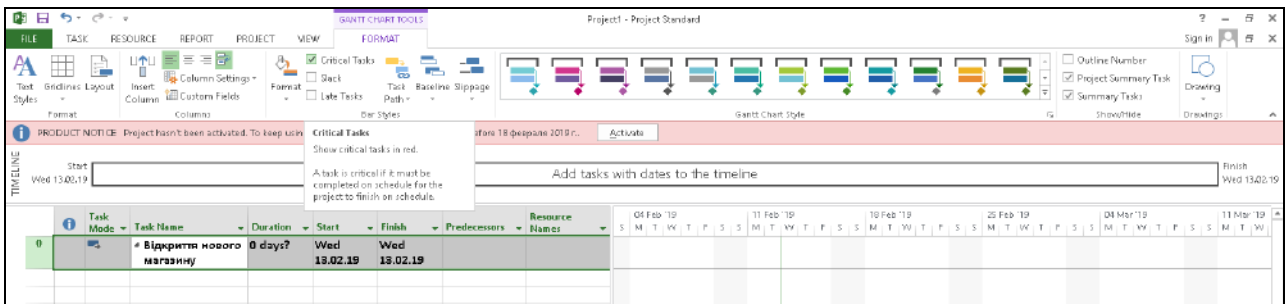


Рис. 78. Активація команди *Контроль виконання завдань*

Наступним кроком обираємо назву проекту. Для цього двічі клікаємо в полі *Назва задачі (Task Name)* та вносимо назву проекту у вікно, що відкрилося, та натискаємо *OK* (рис. 79).

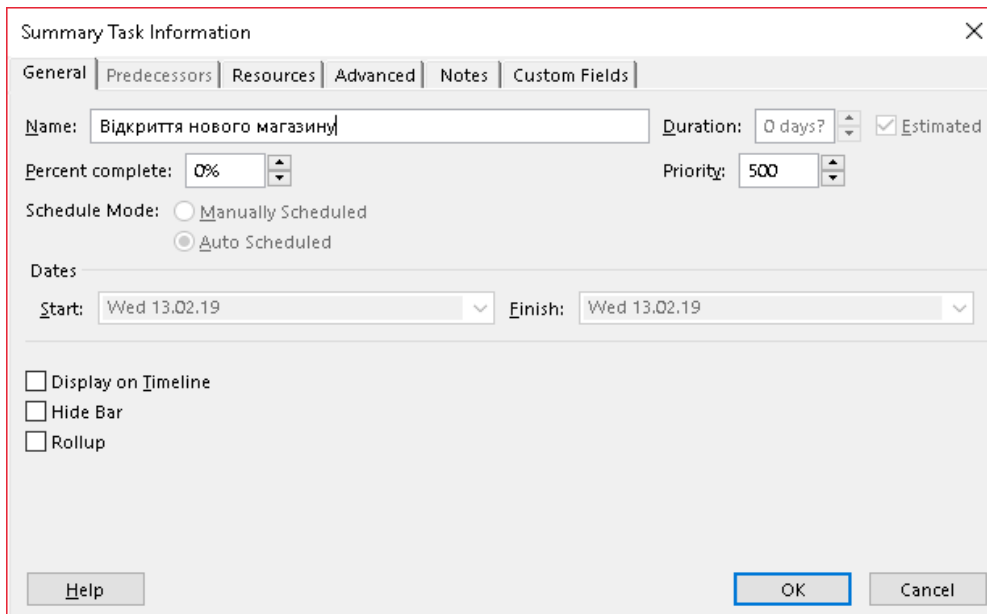


Рис. 79. Введення назви проекту

Проект "Відкриття нового магазину" раціонально розділити на етапи (сумарного завдання) проекту. Так перед початком створенням магазину необхідно провести розрахунки та побудувати бізнес-план, який, своєю чергою, складається з підзавдань (зібрати необхідну інформацію та провести моніторинг конкурентного середовища, провести оцінювання ризиків, презентувати та захистити бізнес-план перед інвесторами тощо). Всі завдання створюються простим додаванням у стовпець *Назва завдань* (рис. 80).

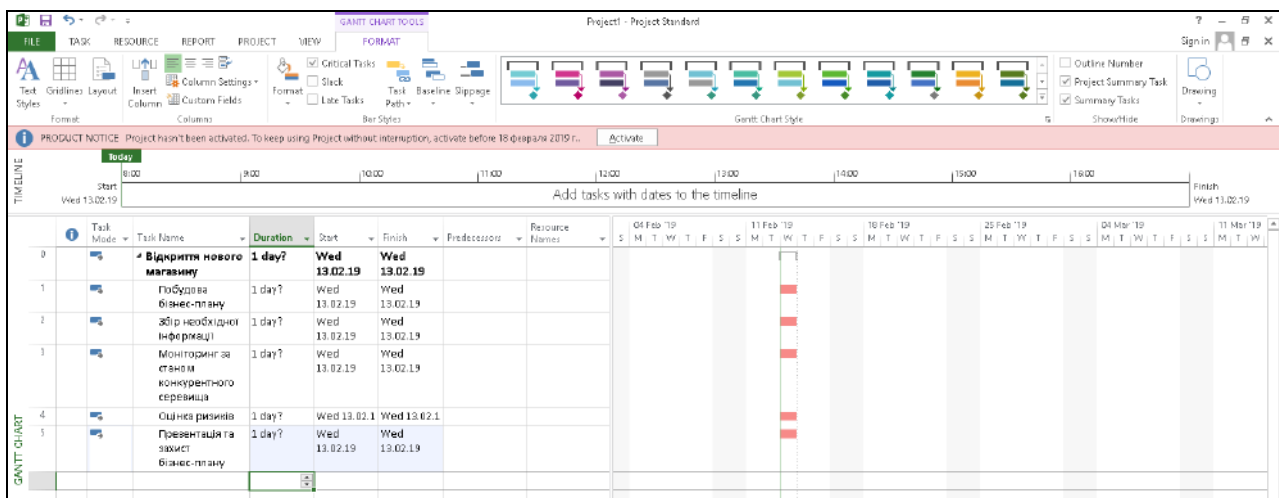


Рис. 80. Додавання завдань у проект

Доцільно зазначити, що користувач має можливість самостійно задати дату початку та завершення проекту, тим самим встановити критичний час виконання проекту.

Для того щоб структурувати завдання та показати, які саме завдання є похідними від сумарного, необхідно виділити всі похідні завдання та на панелі інструментів у вкладці *Завдання* натиснути на кнопку *Indent task* (рис. 81).

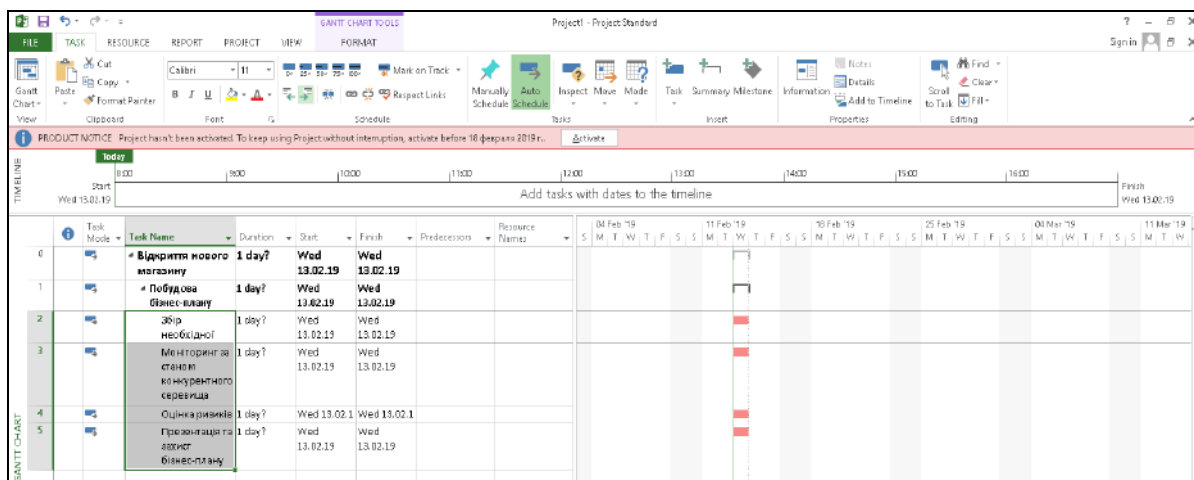


Рис. 81. Вибір похідних завдань

Потім вводимо наступне сумарне завдання – запуск роботи магазину, в автоматичному режимі це завдання знижується до рівня підзавдань побудови БД, тому необхідно його вирівняти до рівня сумарного завдання, натиснувши кнопку *Outdent task* (рис. 82).

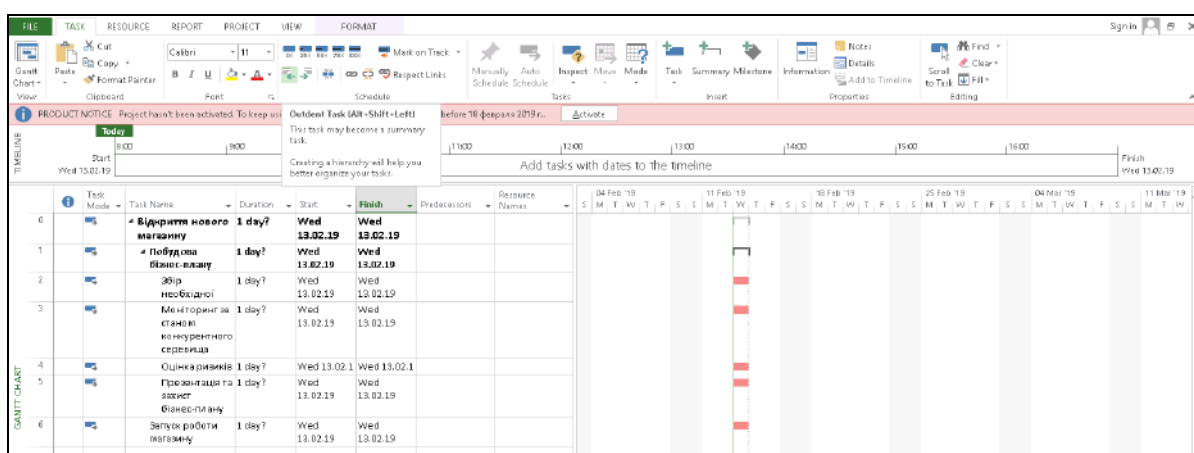


Рис. 82. Встановлення підзавдань до сумарного завдання

Додаємо відповідні підзавдання до сумарного завдання *Запуск роботи магазину* (рис. 83).

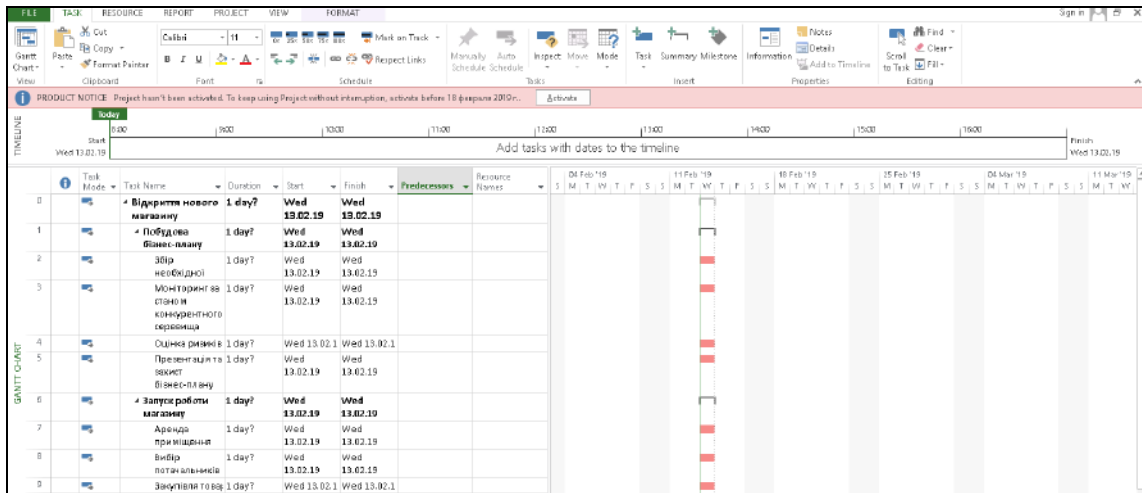


Рис. 83. Заповнення похідних завдань підзавдання проекту

З метою розуміння структурної декомпозиції робіт необхідно їх відобразити, для цього додаємо до робочого поля новий стовпець WBS (структурна декомпозиція робіт) (рис. 84). Тепер користувач бачить номер кожного завдання у загальній структурі.

На наступному етапі додаємо до проекту віхи (так позначаються всі важливі події в плані вашого проекту, такі як завершення основної фази проекту). Віхи є значними подіями, які увійшли в план або нав'язані їм. Оскільки зазвичай віха не містить виконання будь-яких робіт, вона відображається завданням з нульовою тривалістю.

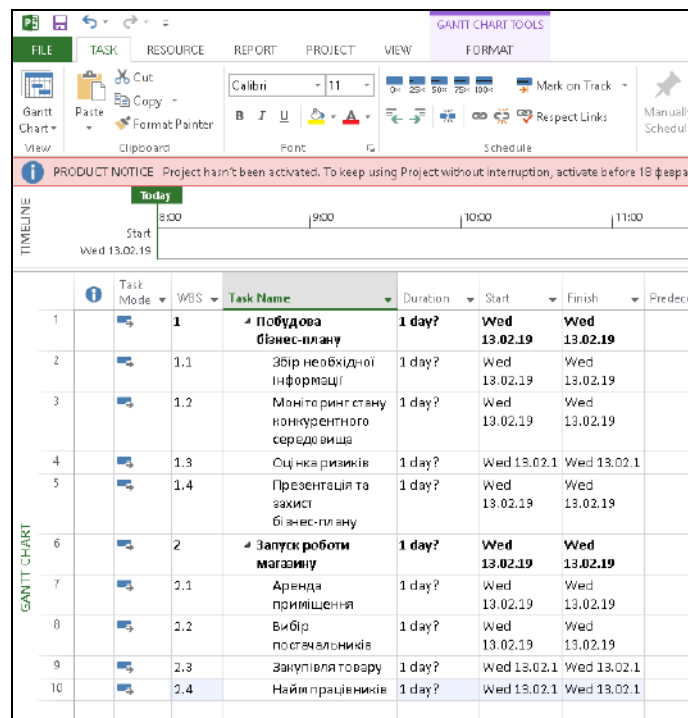


Рис. 84. Додавання функції WBS

Віхи додаються як звичайні завдання через контекстне меню (Insert task) та мають особливу візуалізацію на діаграмі праворуч – у вигляді мітки-ромбу (рис. 85).

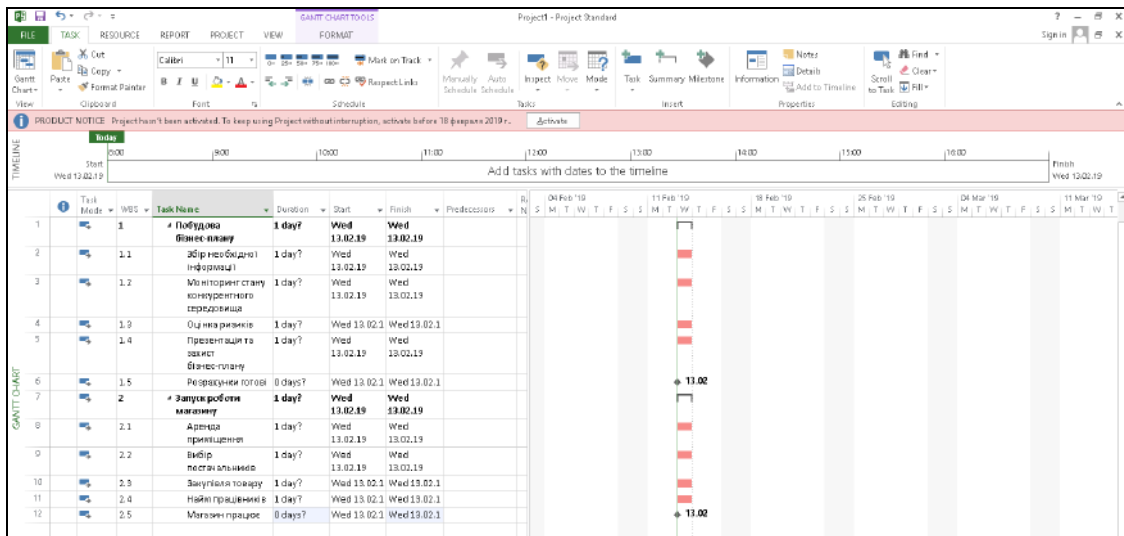


Рис. 85. Додавання віх проєкту

Наступним кроком є встановлення зв'язків між завданнями. Для цього користувач має розуміти черговість виконання робіт, які саме завдання можуть виконуватися паралельно одне до одного, а які є залежними від попередніх завдань.

Для встановлення зв'язків необхідно на діаграмі виділити відповідне завдання та протягнути лінію до наступного завдання. Всі критичні завдання будуть відмічені червоним кольором. Від зміни терміну виконання цих завдань, змінюється загальний час виконання проєкту. Наприклад, від своєчасного збору актуальної інформації про стан зовнішнього середовища залежить оцінювання ризиків проєкту (рис. 86).

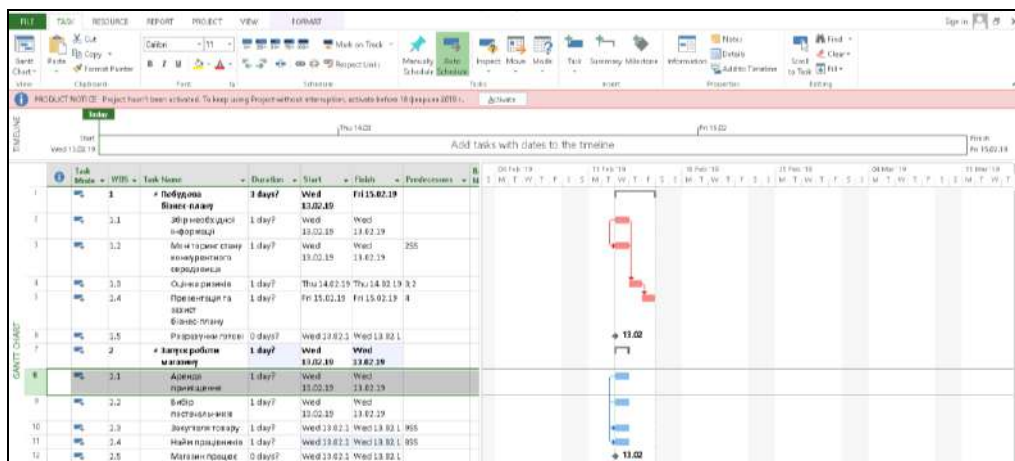
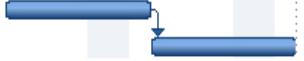
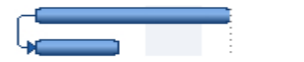




Рис. 86. Встановлення зв'язків між завданнями

Доцільно зазначити, що зв'язки бувають 4 типів (табл. 14).

Таблиця 14

Типи зв'язків в MS Project 2013

Тип зв'язку	Характеристика	Візуалізація зв'язку
Finish-Start (FS) (закінчення-початок)	Виконання завдання починається не раніше, ніж завершиться виконання попереднього завдання	
Start-Start (SS) (початок-початок)	Виконання завдання починається одночасно з попереднім завданням	
Finish-Finish (FF) (закінчення-закінчення)	Виконання завдання завершується одночасно з завершенням виконання попереднього завдання	
Start-Finish (SF) (початок-закінчення)	Виконання завдання завершується після початку виконання попереднього завдання	

Важливо знати, що:

- 1) всі завдання мають бути пов'язані між собою та знаходитися у певній залежності;
- 2) у кожного завдання має бути вхідний і вихідний зв'язок;
- 3) з'єднувати між собою можливо лише підзавдання, сумарні завдання не можна з'єднувати;
- 4) виконання критичних завдань немає часового резерву, тобто затримка виконання критичного завдання призведе до значної затримки виконання всього проєкту загалом.

Найбільш розповсюдженими типами зв'язку є початок-початок та кінець-початок. Також не потрібно змінювати тривалість сумарних завдань, оскільки в автоматичному режимі планування вона розраховується програмою самостійно з огляду на структуру та тип зв'язків між підзавданнями.

Після завершення робіт всередині сумарного завдання з'єднуємо останнє підзавдання з віхою, тим самим автоматично визначається тривалість сумарного завдання (рис. 87).

Потім з'єднуємо між собою підзавдання сумарних завдань та отримуємо остаточні терміни виконання проєкту (рис. 88).

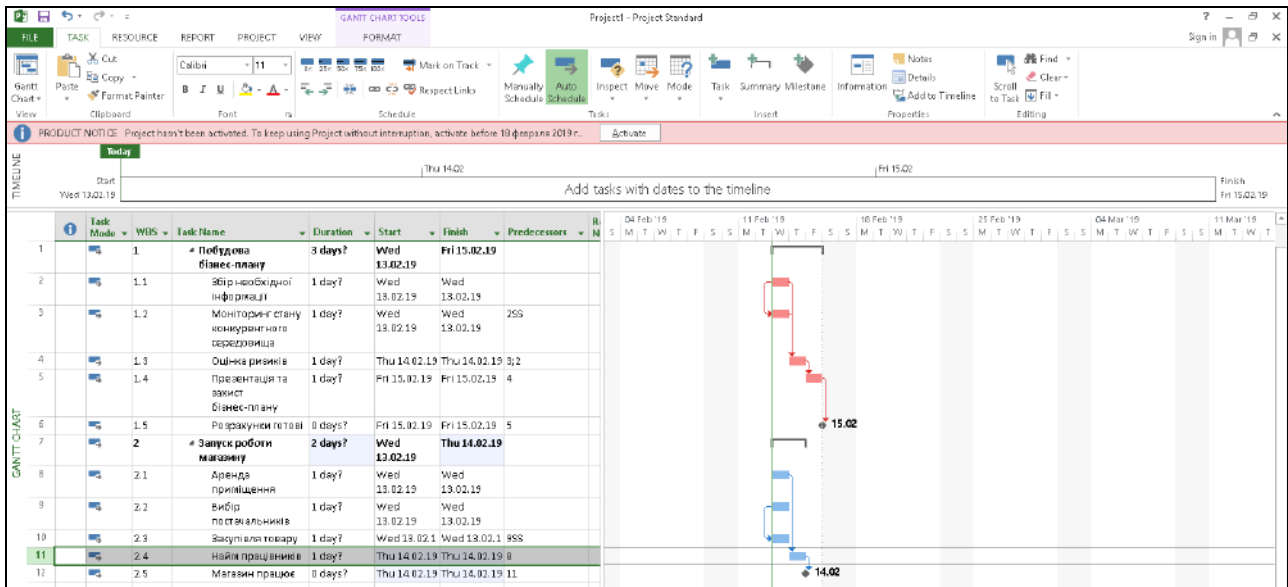


Рис. 87. Визначення тривалості сумарного завдання

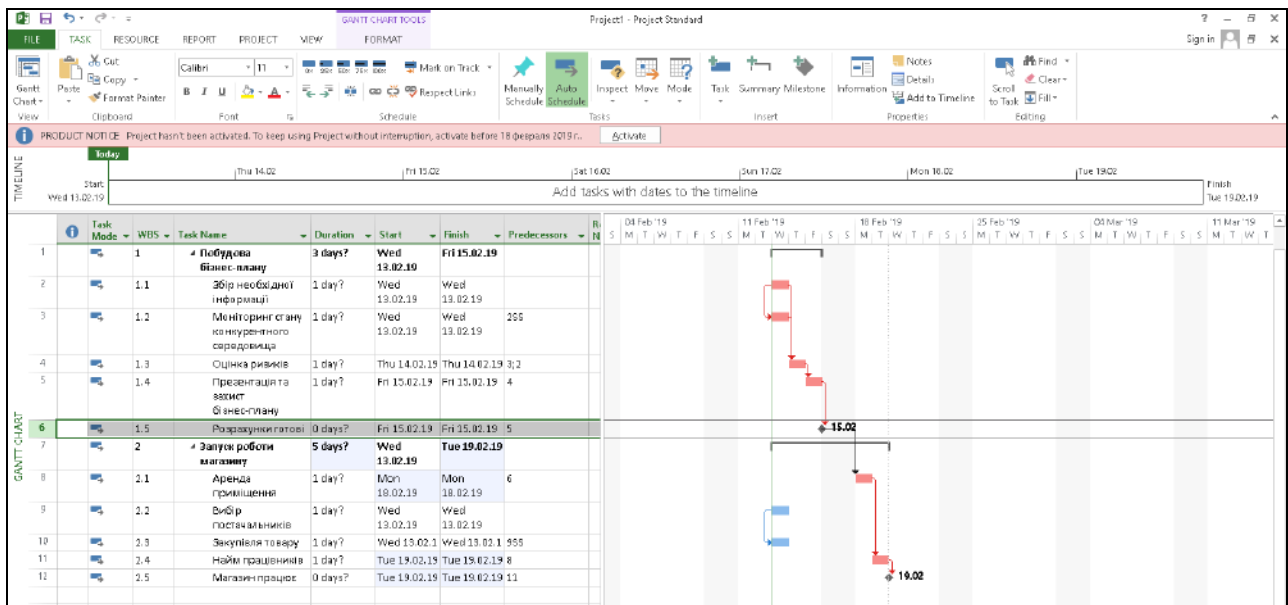


Рис. 88. Внесення корективів щодо термінів виконання завдань

Після того, як ви самостійно скорегували тривалість кожного підзавдання, MS Project автоматично розрахував тривалість сумарного завдання та загальну тривалість проєкту (рис. 89).

Тепер ми можемо призначити виконавців завдань (Resource names) та побачити, які саме роботи (завдання) є попередніми для відповідного підзавдання (Predecessors) (рис. 90). Призначаючи виконавців проєкту, постійно перевіряйте, чи не співпадають їхні роботи між собою, тобто чи не виникає перевантаженості серед виконавців.

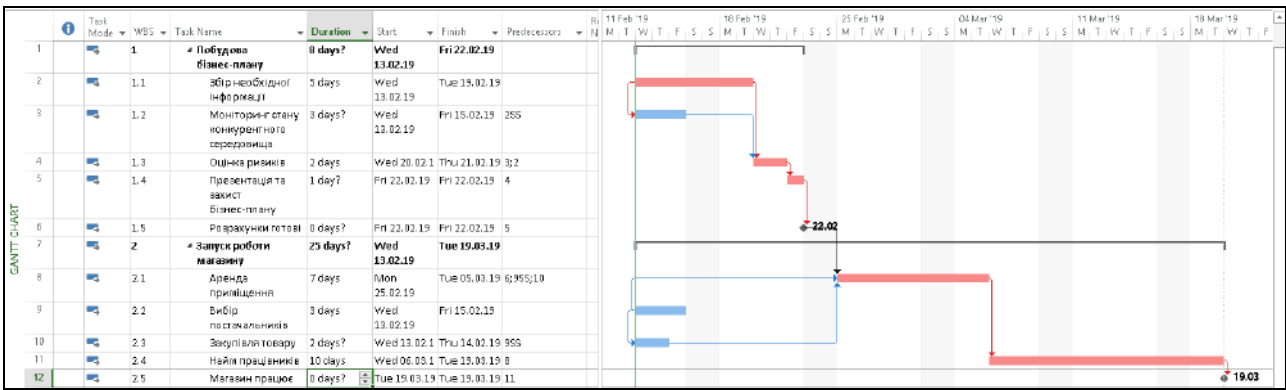


Рис. 89. Визначення остаточного терміну виконання проєкту

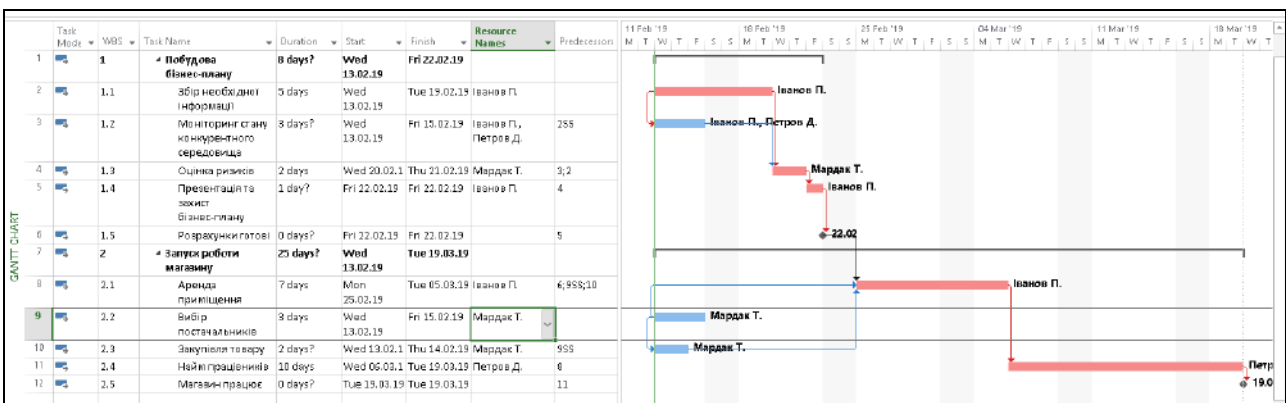


Рис. 90. Призначення виконавців завдань проєкту

Для цього у вкладці *Ресурси* оберіть команду *Вирівняти все (Level all)* та ви побачите, які саме працівники одночасно виконують два та більше підзавдання (рис. 91).

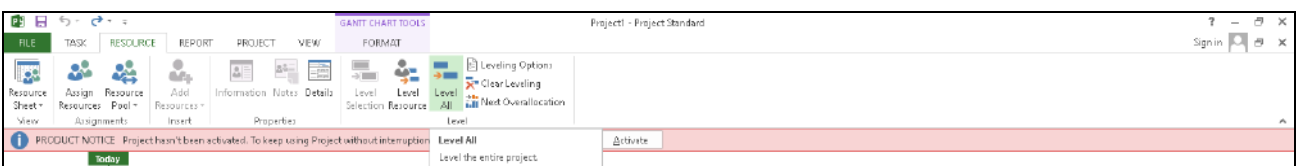


Рис. 91. Вибір команди *Вирівняти все*

У нашому прикладі – це Мардак Т., тому програма автоматично пересуває черговість завдань цього працівника, в результаті чого отримуємо нову діаграму Ганта (рис. 92).

3. Оптимізація проєкту.

З метою оптимізації проєкту за кількістю виконавців необхідно додати стовпець *Пріоритет*, де від 0 до 500 буде визначено важливість черговості виконання завдань (рис. 93).

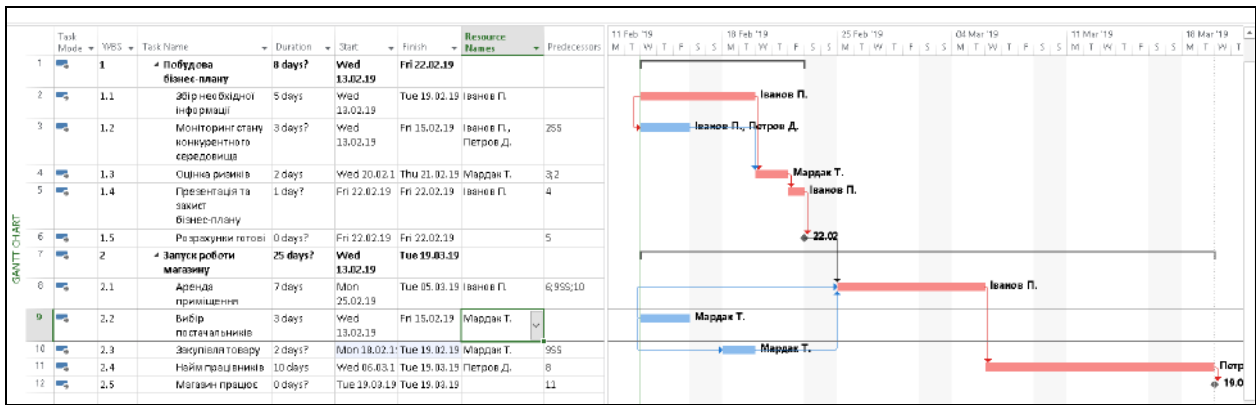


Рис. 92. Корегування черговості робіт виконавців

Task Mode	WBS	Task Name	Duration	Start	Finish	Priority	Resource Names
1	1	Побудова бізнес-плану	8 days?	Wed 13.02.19	Fri 22.02.19	Priority	
2	1.1	Збір необхідної інформації	5 days	Wed 13.02.19	Tue 19.02.19		
3	1.2	Моніторинг стану конкурентного середовища	3 days?	Wed 13.02.19	Fri 15.02.19		
4	1.3	Оцінка ризиків	2 days	Wed 20.02.19	Thu 21.02.19		
5	1.4	Презентація та захист бізнес-плану	1 day?	Fri 22.02.19	Fri 22.02.19	500	Іванов П.

Рис. 93. Встановлення пріоритетів виконання завдань

При цьому необхідно у *Параметрі вирівнювання* встановити прапорець на *Автоматичному режимі* (рис. 94).

Resource Leveling [X]

Leveling calculations

Automatic Manual

Look for overallocations on a basis

Clear leveling values before leveling

Leveling range for 'Project1'

Level entire project

Level From: To:

Resolving overallocations

Leveling order:

Level only within available slack

Leveling can adjust individual assignments on a task

Leveling can create splits in remaining work

Level manually scheduled tasks

Help Clear Leveling... Level All **OK** Cancel

Рис. 94. Вибір додаткових параметрів оптимізації проєкту

Наступним параметром оптимізації проекту є вартість робіт. Тобто користувач має можливість визначити бюджет проекту. Після того як визначені обсяг та завдання проекту, можна створити в полі *Лист ресурсів* перелік ресурсів (людей, устаткування та матеріалів), які беруть участь у виконанні завдань проекту.

Для цього необхідно додати стовпець *Витрати (Cost)* (рис. 95).

Task Mode	WBS	Task Name	Duration	Start	Finish	Prio	Resource Names	Prede	Cost	Resource Names
0	0	Відкриття нового магазину	25 days?	Wed 13.02.19	Tue 19.03.19	500			0,00 P	
1	1	Побудова бізнес-плану	8 days?	Wed 13.02.19	Fri 22.02.19	500			0,00 P	
2	1.1	Збір необхідної інформації	5 days	Wed 13.02.19	Tue 19.02.19	700	Іванов П.		0,00 P	Іванов П.
3	1.2	Моніторинг стану конкурентного середовища	3 days?	Wed 13.02.19	Fri 15.02.19	600	Іванов П., Петров Д.	2SS	0,00 P	Іванов П., Петров Д.
4	1.3	Оцінка ризиків	2 days	Wed 20.02.19	Thu 21.02.19	600	Мардак Т.	3;2	0,00 P	Мардак Т.
5	1.4	Презентація та захист бізнес-плану	1 day?	Fri 22.02.19	Fri 22.02.19	500	Іванов П.	4	0,00 P	Іванов П.
6	1.5	Розрахунки готові	0 days?	Fri 22.02.19	Fri 22.02.19	500		5	0,00 P	
7	2	Запуск роботи магазину	25 days?	Wed 13.02.19	Tue 19.03.19	500			0,00 P	
8	2.1	Аренда приміщення	7 days	Mon 25.02.19	Tue 05.03.19	800	Іванов П.	6;9SS;11	0,00 P	Іванов П.
9	2.2	Вибір постачальників	3 days	Wed 13.02.19	Fri 15.02.19	500	Мардак Т.		0,00 P	Мардак Т.
10	2.3	Закупівля товару	2 days?	Mon 18.02.19	Tue 19.02.19	800	Мардак Т.	9SS	0,00 P	Мардак Т.
11	2.4	Найм працівників	10 days	Wed 06.03.19	Tue 19.03.19	500	Петров Д.	8	0,00 P	Петров Д.
12	2.5	Магазин працює	0 days?	Tue 19.03.19	Tue 19.03.19	500		11	0,00 P	

Рис. 95. Додавання функції *Витрати*

У полі *Витрати* відображається загальна величина запланованих витрат для завдання, ресурсу або призначення, яке складається з витрат на трудовитрати, вже за виконаними ресурсами завдання, і витрат, що плануються. Якщо користувач попередньо не вніс дані про трудовитрати та вартість виконання робіт, програма автоматично проставить нулі у відповідному стовпці *Витрати*.

Для заповнення *Витрат* необхідно відкрити лист *Ресурси (Resource Sheet)* за допомогою контекстного меню та заповнити таблицю (рис. 96).

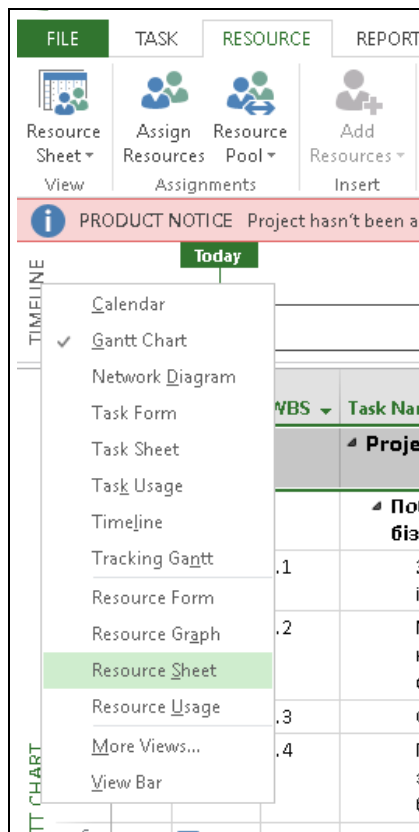


Рис. 96. Відкриття листа *Ресурси*

Так вносимо дані щодо тарифної ставки працівників відповідно до виконуваних робіт (рис. 97).

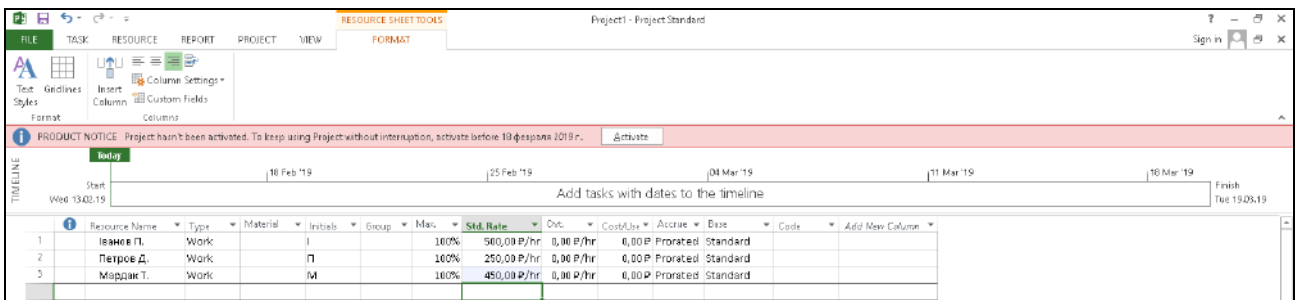


Рис. 97. Заповнення листа *Ресурси* відповідними даними

Унаслідок отримаємо вартість проєкту за 100 % завантаженості працівників (рис. 98).

Додатковою функціональною можливістю MS Project є співставлення планово-фактичних робіт. Тобто, користувач після запуску проєкту має можливість постійно вимірювати результати виконання плану. Для цього, необхідно перейти у вкладку *Проект (Project)* та обрати команду *Створити базовий план* (рис. 99).

	Task Mode	WBS	Task Name	Duration	Start	Finish	Cost	Resource Names
1		1	Побудова бізнес-плану	8 days?	Wed 13.02.19	Fri 22.02.19	37 200,00 P	
2		1.1	Збір необхідної інформації	5 days	Wed 13.02.19	Tue 19.02.19	20 000,00 P	Іванов П.
3		1.2	Моніторинг стану конкурентного середовища	3 days?	Wed 13.02.19	Fri 15.02.19	6 000,00 P	Петров Д.
4		1.3	Оцінка ризиків	2 days	Wed 20.02.19	Thu 21.02.19	7 200,00 P	Мардак Т.
5		1.4	Презентація та захист бізнес-плану	1 day?	Fri 22.02.19	Fri 22.02.19	4 000,00 P	Іванов П.
6		1.5	Розрахунки готові	0 days?	Fri 22.02.19	Fri 22.02.19	0,00 P	
7		2	Запуск роботи магазину	25 days?	Wed 13.02.19	Tue 19.03.19	66 000,00 P	
8		2.1	Аренда приміщення	7 days	Mon 25.02.19	Tue 05.03.19	28 000,00 P	Іванов П.
9		2.2	Вибір постачальників	3 days	Wed 13.02.19	Fri 15.02.19	10 800,00 P	Мардак Т.
10		2.3	Закупівля товару	2 days?	Mon 18.02.19	Tue 19.02.19	7 200,00 P	Мардак Т.
11		2.4	Найм працівників	10 days	Wed 06.03.19	Tue 19.03.19	20 000,00 P	Петров Д.
12		2.5	Магазин працює	0 days?	Tue 19.03.19	Tue 19.03.19	0,00 P	

Рис. 98. Розрахунок вартості проєкту

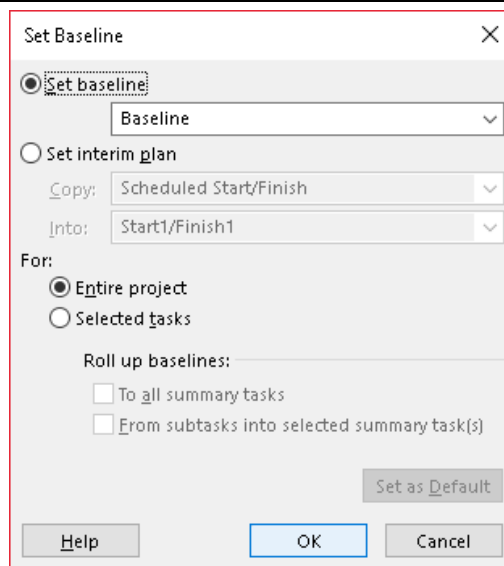
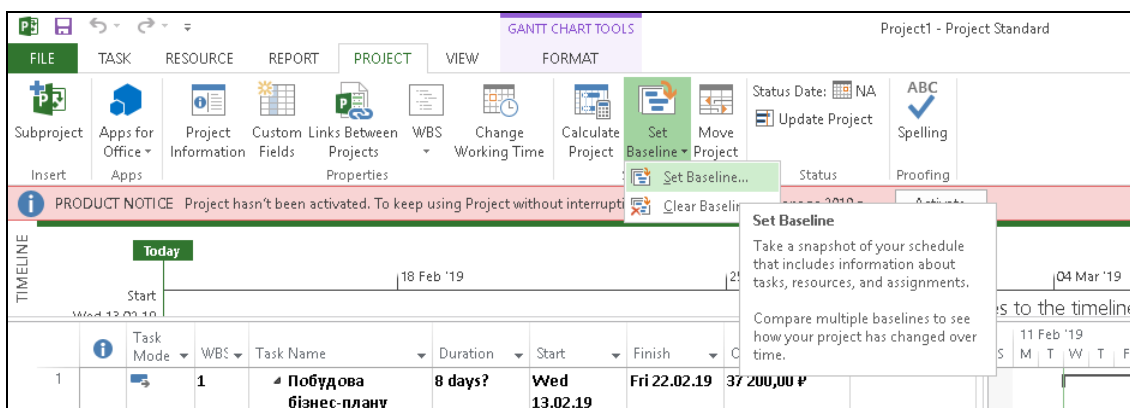


Рис. 99. Запуск команди Створити базовий план

Після натиснення на кнопку **ОК** ніякі візуальні зміни не відбудуться. Однак після додавання додаткового стовпця *Базові витрати*, користувач під час зміни тривалості виконання робіт одночасно побачить відхилення від планових обсягів (рис. 100).

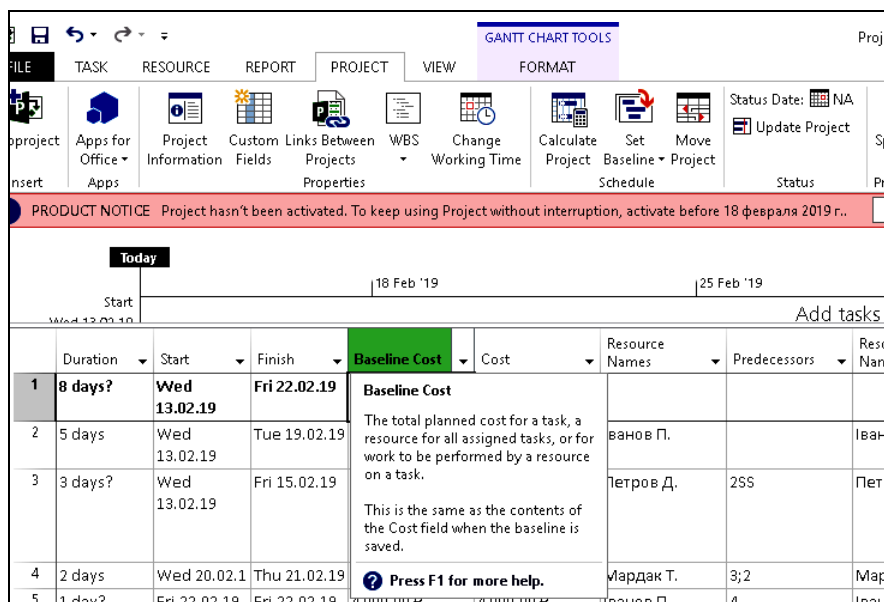


Рис. 100. Додавання функції *Базові витрати*

У нашому випадку збільшилася тривалість збору інформації та зменшилося кількість днів, відведених на пошук персоналу. Унаслідок змінилася вартість сумарних завдань і загальна вартість проекту (рис. 101).

	WBS	Task Name	Duration	Start	Finish	Baseline Cost	Cost	Resource Names	Predec
	0	Project1	25 days?	Wed 13.02.19	Tue 19.03.19	103 200,00 P	107 200,00 P		
	1	Побудова бізнес-плану	10 days?	Wed 13.02.19	Tue 26.02.19	37 200,00 P	45 200,00 P		
	1.1	Збір необхідної інформації	7 days	Wed 13.02.19	Thu 21.02.19	20 000,00 P	28 000,00 P	Іванов П.	
	1.2	Моніторинг стану конкурентного середовища	3 days?	Wed 13.02.19	Fri 15.02.19	6 000,00 P	6 000,00 P	Петров Д.	2SS
	1.3	Оцінка ризиків	2 days	Fri 22.02.19	Mon 25.02.19	7 200,00 P	7 200,00 P	Мардак Т.	3;2
	1.4	Презентація та захист бізнес-плану	1 day?	Tue 26.02.19	Tue 26.02.19	4 000,00 P	4 000,00 P	Іванов П.	4
	1.5	Розрахунки готові	0 days?	Tue 26.02.19	Tue 26.02.19	0,00 P	0,00 P		5
	2	Запуск роботи магазину	25 days?	Wed 13.02.19	Tue 19.03.19	66 000,00 P	62 000,00 P		
	2.1	Аренда приміщення	7 days	Wed 27.02.19	Thu 07.03.19	28 000,00 P	28 000,00 P	Іванов П.	6;9SS;1
	2.2	Вибір постачальників	3 days	Wed 13.02.19	Fri 15.02.19	10 800,00 P	10 800,00 P	Мардак Т.	
	2.3	Закупівля товару	2 days?	Mon 18.02.19	Tue 19.02.19	7 200,00 P	7 200,00 P	Мардак Т.	9SS
	2.4	Найм працівників	8 days	Fri 08.03.19	Tue 19.03.19	20 000,00 P	16 000,00 P	Петров Д.	8

Рис. 101. Результати співставлення планово-фактичних робіт

Для більшої наочності можливо додати стовпці *Відхилення за вартістю (Cost Variance)* та *% відхилення за виконанням робіт (% Work Complete)* (рис. 102).

WBS	Task Name	Duration	Start	Finish	Baseline Cost	Cost	Cost Variance
0	Project1	25 days?	Wed 13.02.19	Tue 19.03.19	103 200,00 P	107 200,00 P	4 000,00 P
1	Побудова бізнес-плану	10 days?	Wed 13.02.19	Tue 26.02.19	37 200,00 P	45 200,00 P	8 000,00 P
2	1.1 Збір необхідної інформації	7 days	Wed 13.02.19	Thu 21.02.19	20 000,00 P	20 000,00 P	0,00 P
3	1.2 Моніторинг стану конкурентного середовища	3 days?	Wed 13.02.19	Fri 15.02.19	6 000,00 P	6 000,00 P	0,00 P
4	1.3 Оцінка ризиків	2 days	Fri 22.02.19	Mon 25.02.19	7 200,00 P	7 200,00 P	0,00 P
5	1.4 Презентація та захист бізнес-плану	1 day?	Tue 26.02.19	Tue 26.02.19	4 000,00 P	4 000,00 P	0,00 P
6	1.5 Розрахунки готувати	0 days?	Tue 26.02.19	Tue 26.02.19	0,00 P	0,00 P	0,00 P
7	2 Запуск роботи магазину	25 days?	Wed 13.02.19	Tue 19.03.19	66 000,00 P	62 000,00 P	-4 000,00 P
8	2.1 Аренда приміщення	7 days	Wed 13.02.19	Thu 07.03.19	28 000,00 P	28 000,00 P	0,00 P
9	2.2 Вибір постачальників	3 days	Wed 13.02.19	Fri 15.02.19	10 800,00 P	10 800,00 P	0,00 P
10	2.3 Закупівля товару	2 days?	Mon 18.02.19	Tue 19.02.19	7 200,00 P	7 200,00 P	0,00 P
11	2.4 Найм працівників	8 days	Fri 08.03.19	Tue 19.03.19	20 000,00 P	16 000,00 P	-4 000,00 P

WBS	Task Name	Duration	Start	Finish	% Work Complete	Baseline Cost	Cost
0	Project1	25 days?	Wed 13.02.19	Tue 19.03.19	84%	103 200,00 P	107 200,00 P
1	Побудова бізнес-плану	10 days?	Wed 13.02.19	Tue 26.02.19	83%	37 200,00 P	45 200,00 P
2	1.1 Збір необхідної інформації	7 days	Wed 13.02.19	Thu 21.02.19	98%	20 000,00 P	20 000,00 P
3	1.2 Моніторинг стану конкурентного середовища	3 days?	Wed 13.02.19	Fri 15.02.19	85%	6 000,00 P	6 000,00 P
4	1.3 Оцінка ризиків	2 days	Fri 22.02.19	Mon 25.02.19	67%	7 200,00 P	7 200,00 P
5	1.4 Презентація та захист бізнес-плану	1 day?	Tue 26.02.19	Tue 26.02.19	0%	4 000,00 P	4 000,00 P
6	1.5 Розрахунки готувати	0 days?	Tue 26.02.19	Tue 26.02.19	0%	0,00 P	0,00 P
7	2 Запуск роботи магазину	25 days?	Wed 13.02.19	Tue 19.03.19	85%	66 000,00 P	62 000,00 P
8	2.1 Аренда приміщення	7 days	Wed 13.02.19	Thu 07.03.19	100%	28 000,00 P	28 000,00 P
9	2.2 Вибір постачальників	3 days	Wed 13.02.19	Fri 15.02.19	75%	10 800,00 P	10 800,00 P
10	2.3 Закупівля товару	2 days?	Mon 18.02.19	Tue 19.02.19	68%	7 200,00 P	7 200,00 P
11	2.4 Найм працівників	8 days	Fri 08.03.19	Tue 19.03.19	80%	20 000,00 P	16 000,00 P

Рис. 102. Додавання додаткових функцій оптимізації проєкту

В автоматичному режимі на діаграмі Ганта жирною лінією візуалізовано відсотки виконання плану проєкту.

За результатами реалізації проєкту користувач може проставляти та вносити відповідну інформацію, тим самим робити висновки про успішність виконання або відставання від планових показників.

Також MS Project дозволяє користувачам створювати й налаштовувати різні графічні звіти для будь-яких даних проєкту, не застосовуючи додаткового програмного забезпечення. У ході виконання проєкту звіти автоматично будуть оновлюватися, відображаючи останні дані. Класифікація видів звітів у MS Project наведено в табл. 15.

Типи звітів MS Project

Ключові параметри	Характеристика типів звітів
1	2
<i>Звіти – приладні дошки</i>	
Хід виконання	Обсяг роботи, кількість завдань, і що ще потрібно виконати для досягнення цілей проєкту
Огляд витрат	Поточний стан витрат для проєкту та його завдань верхнього рівня, зокрема заплановані витрати, залишкові витрати, фактичні витрати, накопичені витрати, базові витрати та відсоток виконання, що допомагають визначити, чи не вийде проєкт за рамки бюджету
Огляд проєкту	Яку частину проєкту виконано, майбутні проміжні етапи та прострочені завдання
Майбутні завдання	Робота, виконана за поточний тиждень, стан завдань, які ще необхідно виконати, і завдання, що починаються наступного тижня
Огляд роботи	Стан виконання роботи за проєктом і статистика роботи для всіх завдань верхнього рівня, що допомагає визначити відсоток виконання та обсяг роботи, яку ще потрібно виконати
<i>Звіти про ресурси</i>	
Перенавантажені ресурси	Стан роботи для всіх перенавантажених ресурсів, який відображає їхню фактичні трудовитрати та залишковий обсяг роботи. Потім можна спробувати усунути ці перенавантаження в поданні планувальника роботи групи
Огляд ресурсів	Стан роботи для всіх людей (робочих ресурсів), які працюють над проєктом, що дає змогу дізнатися, скільки роботи виконано, а скільки ще залишилося
<i>Звіти про витрати</i>	
Рух грошових коштів	Витрати та накопичені витрати за квартал для всіх завдань верхнього рівня. Скористайтеся списком полів, щоб відобразити інші витрати за інші періоди часу
Перевищення витрат	Відхилення від запланованих витрат для завдань верхнього рівня та робочих ресурсів, що вказує, де фактичні витрати перевищують базові витрати
Звіт про освоєний обсяг	Показники освоєного обсягу, відхилення та ефективності з часом, порівняння фактичних і запланованих витрат з основним планом та визначення, чи йде проєкт згідно з планом
Огляд вартості ресурсів	Стан витрат на робочі ресурси (людей і матеріали), що відображає відомості про витрати в таблиці та розподіл витрат у діаграмі
Огляд витрат на завдання	Стан витрат для завдань верхнього рівня, що відображає відомості про витрати в таблиці та розподіл витрат у діаграмі

1	2
Звіти про стан проєкту	
Критичні завдання	Усі заплановані у стислі терміни завдання, зазначені як критичні на критичному шляху проєкту. Це означає, що будь-яка затримка таких завдань спричинить зсув плану
Завдання з затримкою	Усі завдання, що розпочалися або завершилися пізніше дати початку або завершення за планом і які виконуються із затримкою від плану
Звіт проміжного етапу	Усі завдання у проєкті, які мають проміжні етапи, що показує, які з них мають затримку, для яких настав термін виконання та які з них виконано
Завдання, що затримуються	Усі завдання у проєкті, виконання яких триває довше, ніж очіувалося, і які мають дату завершення, пізнішу базової дати завершення

Щоб створити звіт, необхідно перейти на вкладку *Звіт (Report)* та у групі *Перегляд звітів* обрати потрібний для вас тип звіту. Наприклад, такий вигляд матиме звіт "Огляд витрат" (рис. 103):

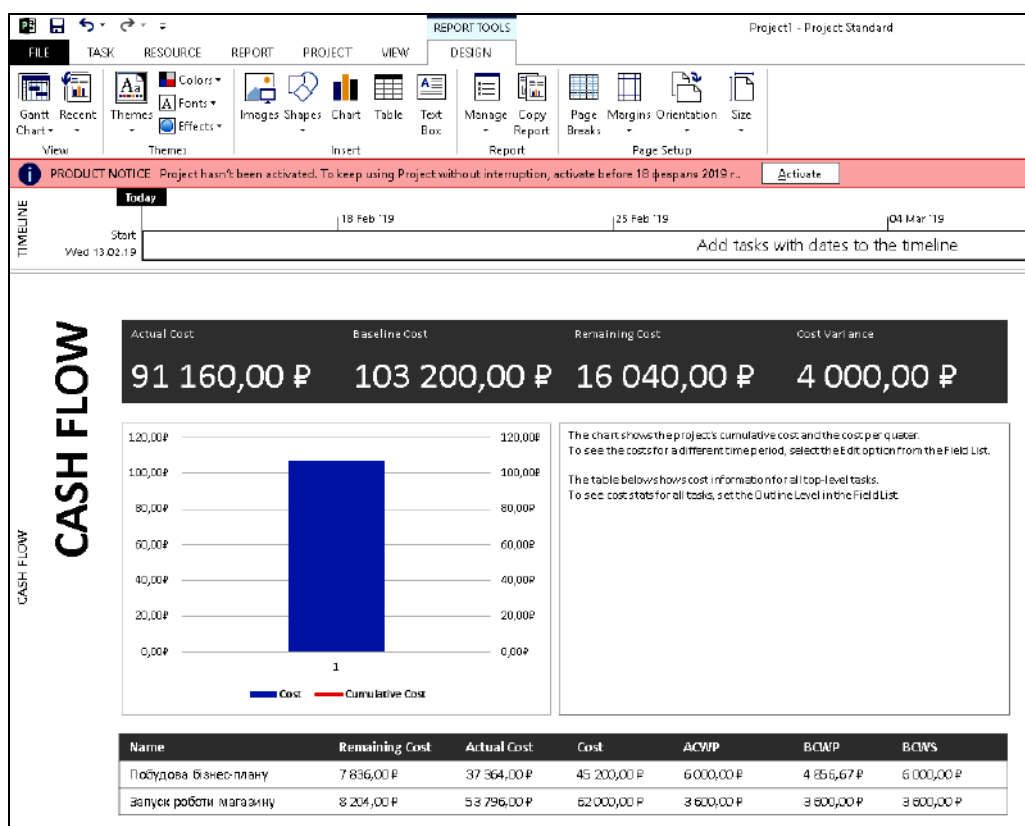


Рис. 103. Звіт "Огляд витрат"

На рис. 104 наведено звіт про виконання критичних завдань.

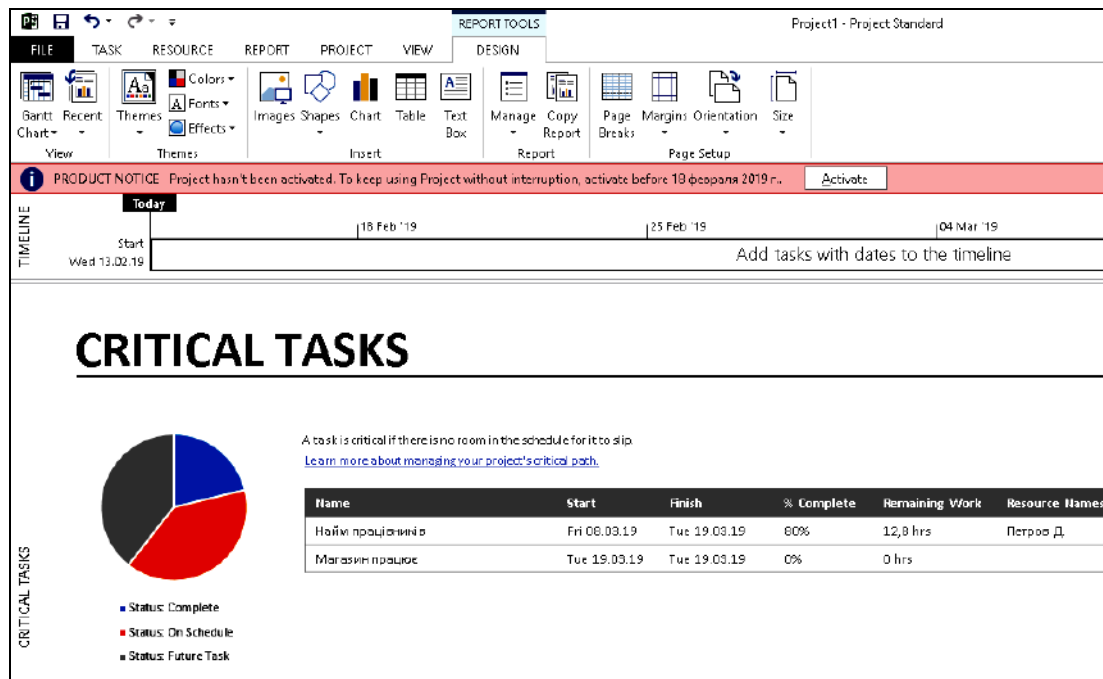


Рис. 104. Звіт про виконання критичних завдань

Також ця програма має розширені графічні можливості та дозволяє змінювати дизайн проєкту та його складників. Для цього необхідно увійти до вкладки *Дизайн* або *Формат* та обрати відповідні параметри (рис. 105).

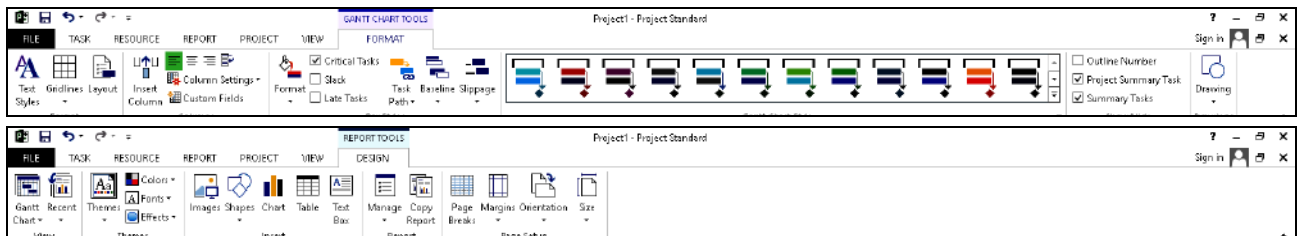


Рис. 105. Редагування дизайну проєкту

Після завершення робіт необхідно зберегти проєкт та, за необхідності, вносити відповідні зміни у ході практичної реалізації завдань проєкту.

Завдання:

1. Оберіть один із далі наведених варіантів проєктів або запропонуйте власний (за погодженням з викладачем):

- проєкт відкриття власної справи;
- проєкт розроблення бізнес-плану;
- проєкт управління діяльністю банку;
- проєкт випуску нового номеру журналу;
- проєкт розроблення маркетингової кампанії.

2. Побудуйте схематичний план на аркуші паперу, продумайте всі завдання, зробіть структурну декомпозицію проєкту, обґрунтуйте вибір ресурсів для виконання вашого проєкту та його бюджет.

3. Використовуючи програмні можливості MS Project, створіть відповідний проєкт.

Результати реалізації проєкту необхідно подати у вигляді звіту.

Завдання для самостійного опрацювання

Запитання, що виносяться на самостійну роботу студентів:

1. Охарактеризуйте прикладну програму MS Project. Назвіть її переваги та недоліки.

2. Які завдання може вирішувати користувач у ході реалізації проєктного підходу в MS Project 2013?

3. Що таке "оптимізація проєкту"? Які напрями оптимізації проєкту вам відомі?

4. Які сучасні аналоги програмних продуктів для побудови та управління проєктами вам відомі?

5. Охарактеризуйте місце та значення систем управління проєктами в роботі бізнес-аналітика.

Завдання. Підготуйте розгорнуту презентацію одного із сучасних програмних продуктів-аналогів MS Project, що використовується для розробки та управління проєктами. Зокрема, сучасними альтернативними додатками project management є:

1. Сервіс для спільної роботи й управління проєктами TeamBridge.

2. Вебсервіс для управління проєктами Ganttter.

3. Продукт OpenProj.

4. Сервіс Planner від MS Office.

5. SCIM.ru – хмарний сервіс, що дозволяє організувати роботу з проєктами.

6. ProjectLibre – безкоштовний аналог Microsoft Project.

7. GanttProject – додаток для управління проєктами.

8. Trello – безкоштовне вебдодаток для управління проєктами.

9. ABC Roster – безкоштовний програмний додаток, розроблений для організації робочого процесу в компанії.

10. Advanta – інформаційна система управління проєктами.

Змістовий модуль 2

Спеціальні програмні продукти обробки великих масивів даних. Їхні види та сфера використання

Тема 5. Аналітичні можливості RapidMiner

Лабораторна робота 5

Аналітичні та функціональні можливості RapidMiner

Мета: надбання навичок аналізу даних за допомогою RapidMiner.

Завдання: ознайомитися з теоретичною базою, підсистемами та модулями RapidMiner; навчитися обробляти й аналізувати великі масиви даних за допомогою RapidMiner.

Методичні рекомендації

1. Знайомство з RapidMiner.

RapidMiner (раніше відомий як YALE) – напевно найпотужніша на сьогоднішній день комплексна система для Data Mining (інтелектуальний аналіз даних) і статистичного аналізу, що володіє дуже серйозним набором алгоритмів для оброблення й аналізу, включно з обробленням великих масивів даних і має дуже наглядну та багатомірну візуалізацію. Пакет має досить оригінальну концепцію. Так робота з будь-яким набором даних є процесом деревоподібного типу, в який можна додавати різні оператори введення/виведення, оброблення, візуалізації, аналізу і т. п.

Початкова версія була розроблена в 2001 році відділом Штучного Інтелекту Дортмундського технологічного університету (Artificial Intelligence Unit of Dortmund University of Technology).

RapidMiner – це потужна, відкрита та розрахована на багато користувачів платформа, що слугує для створення, передачі та обслуговування наукомістких даних. Уніфікована платформа RapidMiner прискорює створення повних аналітичних робочих процесів – від підготовки даних до моделювання та до відкриття бізнесу – в єдиному середовищі, значно підвищуючи ефективність і скорочуючи час, необхідний для проєктів в області використання даних.

Додатками RapidMiner можуть бути як дослідні (модельні), так і прикладні (реальні) завдання інтелектуального аналізу даних, включно з аналізом тексту (text mining), аналізом мультимедіа (multimedia mining), аналізом потоків даних (data stream mining).

Функціональні можливості RapidMiner:

- надає більше 400 операторів для всіх найбільш відомих методів машинного навчання, включно з введенням і виведенням, попереднім обробленням даних та їхньою візуалізацією;
- інтегрує в себе оператори WEKA;
- має вбудовану мову сценаріїв, що дозволяє виконувати масивні серії експериментів;
- концепція багаторівневого представлення даних (multi-layered data view) забезпечує ефективну і прозору роботу з даними;
- графічна підсистема забезпечує багатовимірну візуалізацію даних і моделей;
- є покроковий підручник, що містить популярний вступ у машинне навчання інтелектуальному аналізу даних.

Реалізація та технічні можливості RapidMiner:

1. Програмне забезпечення написано цілком на Java, тому працює у всіх основних операційних системах.
2. Для представлення експериментів як суперпозицій операторів застосовується мова XML.
3. Вбудовування в інші додатки здійснюється за допомогою Java API.
4. Підтримуються механізми плагінів (plugin) і розширень (extension).

Отже, найчастіше RapidMiner використовують як інструмент для:

- 1) аналізу транзакцій (наприклад, банківських) щодо протидії шахрайству;
- 2) клієнтської аналітики;
- 3) персональних рекомендацій (для роздрібної торгівлі – що та кому запропонувати);
- 4) для прогнозування поставок і продажів;
- 5) для текстової аналітики (наприклад, аналіз емоційного відтінку відгуків або коментарів в автоматичному режимі).

Для подальшого ознайомлення з аналітичними та функціональними можливостями RapidMiner необхідно його завантажити з офіційного сайту програми: <https://rapidminer.com/get-started>.

Існує два шляхи безкоштовного доступу та роботи з RapidMiner:

1. Завантажити пробну trial-версію продукту, що забезпечить 30-денне використання продукту.
2. Завантажити продукт для наукових (навчальних цілей), що надає безкоштовне 12-місячне використання програми.

Система підтримує всі етапи глибокого аналізу даних, включно з візуалізацією, перевіркою та оптимізацією.

Величезна перевага RapidMiner полягає в тому, що для його використання не потрібно знати мов програмування. Тут реалізований принцип візуального програмування, тобто сам код писати не потрібно, як і не потрібно проводити складні математичні обчислення, все відбувається у такий спосіб: користувач вносить дані на робоче поле, а потім просто перетягує оператори в GUI, формуючи процес оброблення даних.

2. Введення даних та основи роботи з RapidMiner (RM).

Під час запуску програми відкривається інтерфейс (рис. 106).

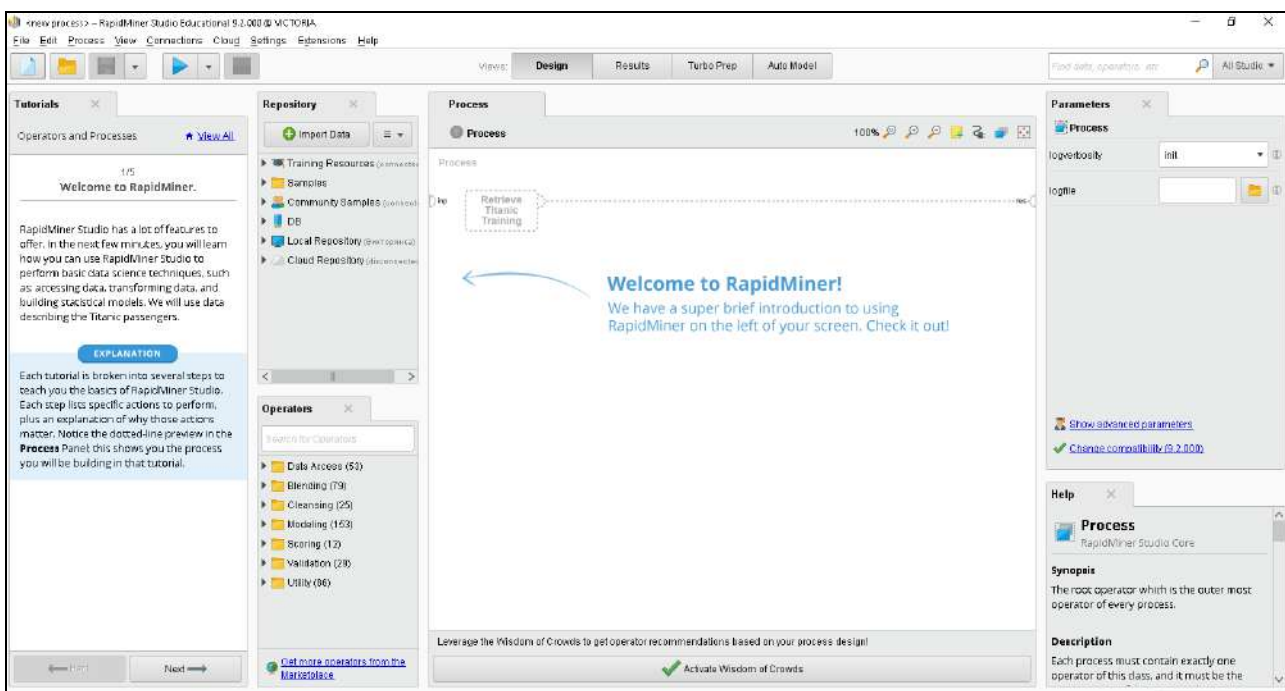


Рис. 106. Вікно запуску RapidMiner

Під час роботи з RapidMiner користувач має оперувати такими основними поняттями програми:

1. *Процес* – сукупність операторів з'єднаних між собою в заданому порядку для виконання необхідного завдання аналізу / оброблення даних.

2. *Оператор* – логічна одиниця процесу. Оператор виконує певні дії над даними, у нього є вхід-вихід (так звані "порти"), на вхід приходять дані, на вихід йдуть оброблені оператором дані. У такий спосіб ми можемо створити ланцюжок оброблення даних, наприклад, завантажити транзакції клієнтів з БД, знайти найбільші, конвертувати в долари і видати результат. Водночас можна побудувати ланцюжок паралелі, наприклад, в одній ми читаємо транзакції з різних БД, а в іншій – шукаємо дані клієнтів, потім об'єднуємо і отримуємо результат (при цьому також можливе їхнє паралельне виконання в часі). В інтерфейсі програми операторам відповідає вкладка Operators, де в ієрархії вони згруповані за функціональною ознакою. Щоб скористатися оператором, необхідно натиснути на нього і перенести в робочу область процесу. Оброблення даних у RM відбувається за допомогою операторів (Operators), які застосовуються послідовно до джерела даних.

3. *Репозиторій* – місце для зберігання процесів RM. Може бути локальним, а також віддаленим (RapidMiner Server), для якого можливо виконувати процеси на боці сервера, що розрахований на багато користувачів доступ до процесів / з'єднанням БД, запуск процесів за розкладом або передача даних як вебсервіс.

У додатку Repositories у RM тут можна побачити тільки Samples (набір процесів – прикладів), DB (поточні з'єднання до баз даних, доступних у RM (визначаються через Tools -> Manage Database Connections)) та Local Repository (місце для зберігання власних процесів на комп'ютері).

4. *Контекст процесу*. Контексту відповідає вкладка Context, де можна побачити три секції:

1. Process input – дані передаються на вхід процесу. Тут можна вказати шлях до даних теки (папки) сховища.

2. Process output – тут вказується шлях у репозиторії, куди буде збережений результат роботи процесу.

3. Macros – глобальна змінна, доступна в процесі з будь-якого місця. Може приймати дані як значення за винятком окремих виразів або чисел.

Відзначимо, що Process input і Process output позначені в процесі кружками по межі процесу з написами inr і res. Щоб скористатися даними з входу або зберегти їх, потрібно з'єднати відповідний кружок з входом/виходом операторів.

Інтерфейс RM виконаний у вигляді робочої області з вкладками, розмір і положення яких можна змінювати. Основними вкладками є:

Repository – містить дані для аналізу;

Operators – інструменти для аналізу даних;

Process – вкладка з головним процесом;

Parameters – налаштування оператора.

Для перегляду вкладок необхідно скористатися списком View та обрати команду Show Panel (рис. 107).

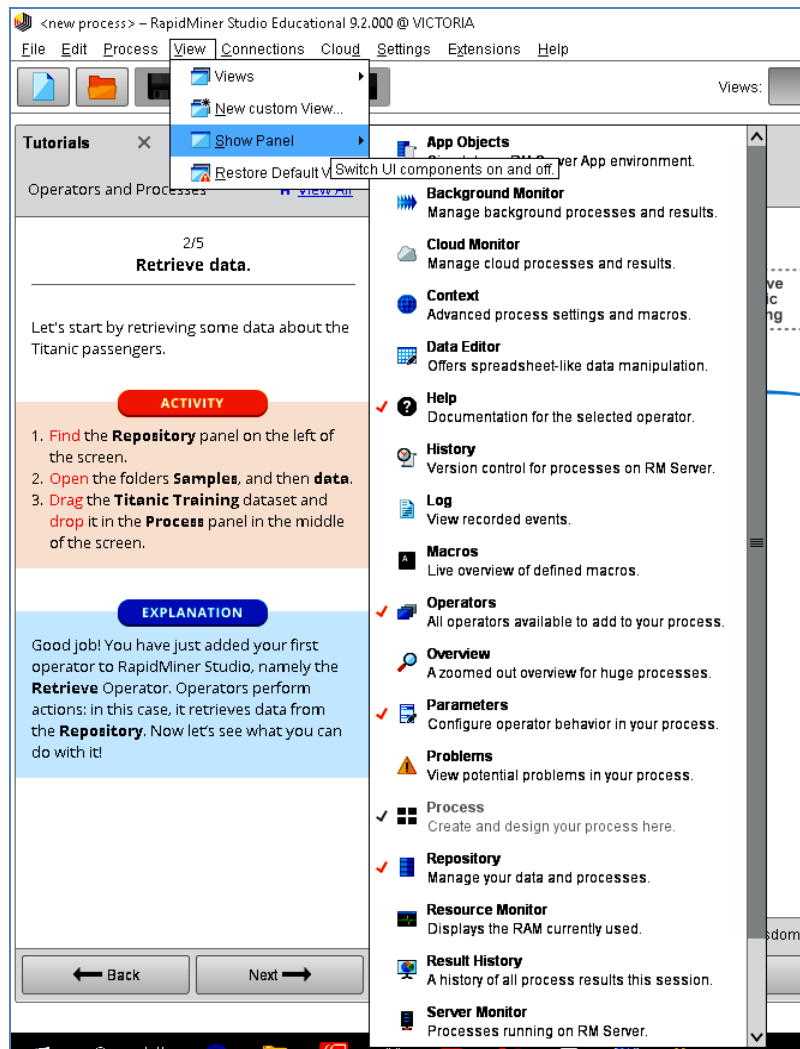


Рис. 107. Вкладки RapidMiner

У програмі пропонується візуалізувати весь процес аналізу даних у вигляді ланцюжка (графа) операторів і запустити його на виконання. Ланцюжок операторів представляється в RapidMiner у вигляді інтерактивного графа та за допомогою виразу на мові XML (eXtensible Markup Language, основної мови системи).

Розглянемо основи роботи в RapidMiner на наведеному далі прикладі.

Ви директор невеликої компанії, яка займається створенням сайтів, промисловим дизайном. Досить часто через велику кількість замовлень та нестачу співробітників ви наймаєте фрілансерів з різних країн (тому

що маєте клієнтів з усього світу) і постійно вносите інформацію про виконані роботи в таблицю Excel, вказуючи ім'я виконавця, вид робіт, дату оплати, суму та валюту оплати. В якийсь момент вам захотілося отримати суму витрат, в гривнях (по курсу Національного банку), яку ви понесли в розбивці за видами робіт на конкретну дату (рис. 108).

	A	B	C	D	E	F
1	№	Ім'я виконавця	Вид робіт	Дата оплати	Сума оплати	Валюта оплати
2	1	Петров Денис	Розробка сайту	30/10/2018	180000	RUB
3	2	Tom Peaters	Проектування об'єкту	18/01/2019	1830	USD
4	3	Yan Hu	Дизайн веб-сайту	12/05/2018	7500	JPY
5	4	Мишик Красновки	Тестування програмного забезп	14/05/2018	1000	PLN
6	5	Timotti Scott	Розробка програми	01/02/2019	1690	USD
7	6	So Lee	Дизайн веб-сайту	18/01/2018	8020	JPY
8	7	Adam Feert	Розробка сайту	11/06/2018	1200	EUR
9	8	Hellen Tomas	Розробка програми	09/09/2018	850	EUR
10	9	Соломашина Вікторія	Проектування об'єкту	14/02/2019	11700	UAH
11	10	Нефедов Віктор	Дизайн веб-сайту	16/02/2019	9500	UAH
12	11	Красноперов Олег	Розробка програми	04/07/2018	210500	RUB

Рис. 108. Вихідні дані в MS Excel

Першим кроком є додавання вихідних даних у RM. Для цього на вкладці (блоці) *Repository* натискаємо кнопку *Import data* (рис. 109).

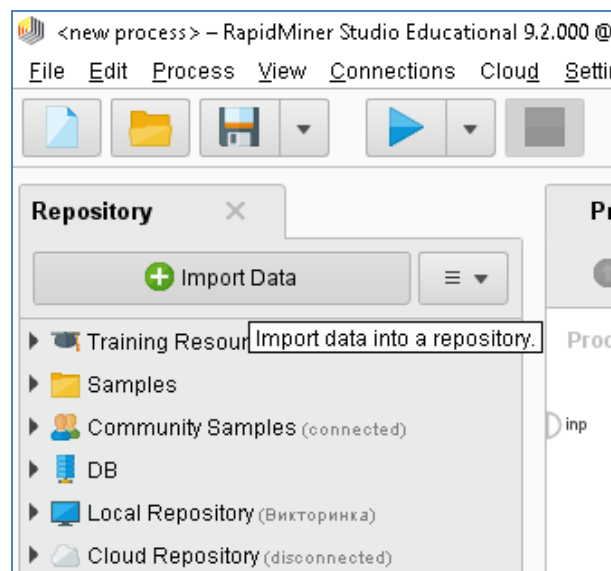


Рис. 109. Імпорт даних у RM

На наступному кроці обираємо звідки програма має імпортувати дані (у нашому випадку – це My computer) (рис. 110).

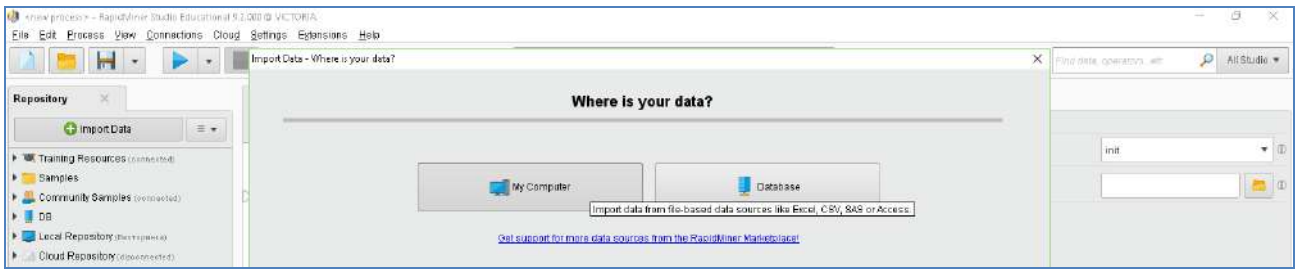


Рис. 110. Параметри імпорту даних у RM

Вказуємо шлях до файлу, вибираємо файл, натискаємо Next (рис. 111).

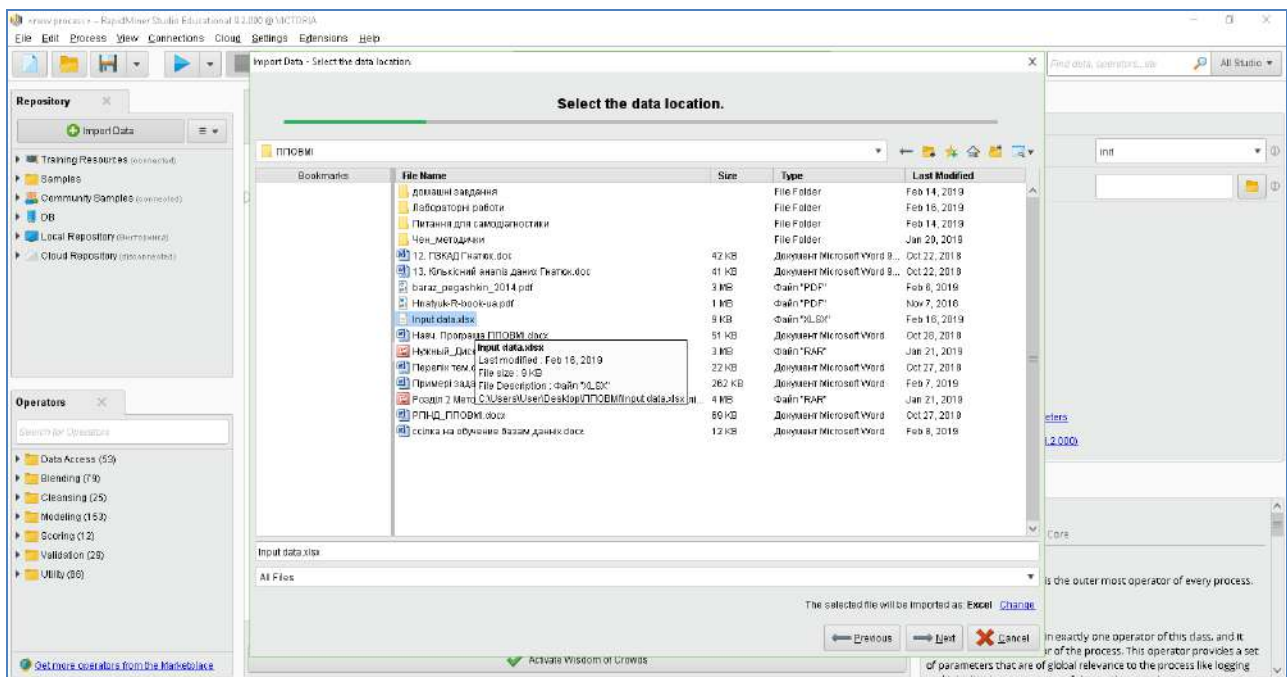


Рис. 111. Вибір файлу для імпорту даних у RM

З'явиться вікно з параметрами розпізнавання. Необхідно виставити всі параметри. Кожен атрибут має бути поміщений в окремий стовпець, а також обираємо правильний діапазон даних (рис. 112).

Потім відбувається процес форматування стовпчиків. Для зміни імені, типу або видалення стовпця необхідно натиснути на кнопку у вигляді "шестерні" в заголовку стовпця (рис. 113).

Тепер оберемо місце, куди необхідно зберегти дані. Створюємо свою папку – правим кліком миші по *Local Repository* та погоджуємося з *Директорією за замовчуванням*. Якщо всі попередні кроки були виконані правильно, то файл має з'явитися у вкладці *Repository* (рис. 114).

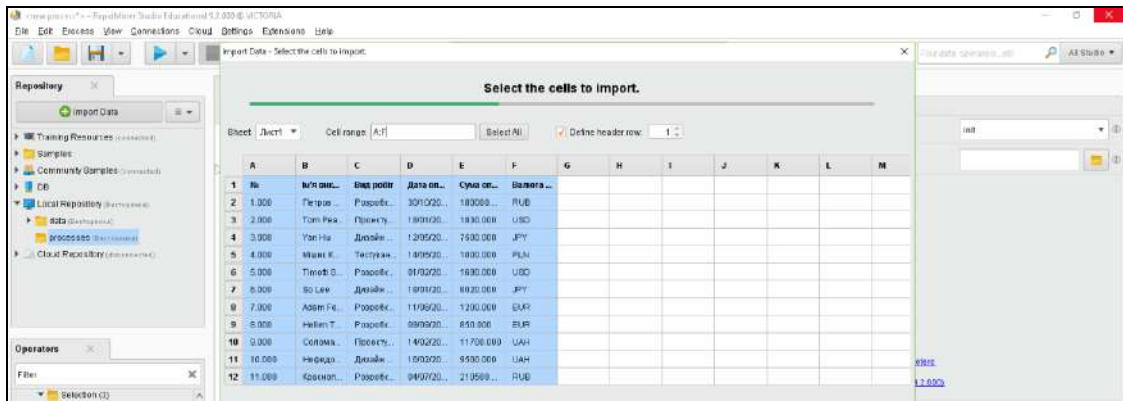


Рис. 112. Параметри розпізнавання даних у RM

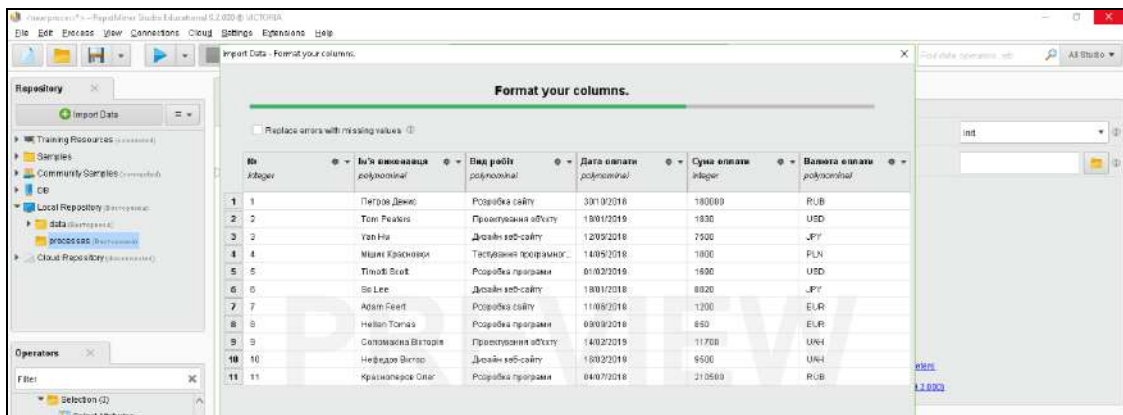


Рис. 113. Редагування формату даних у RM

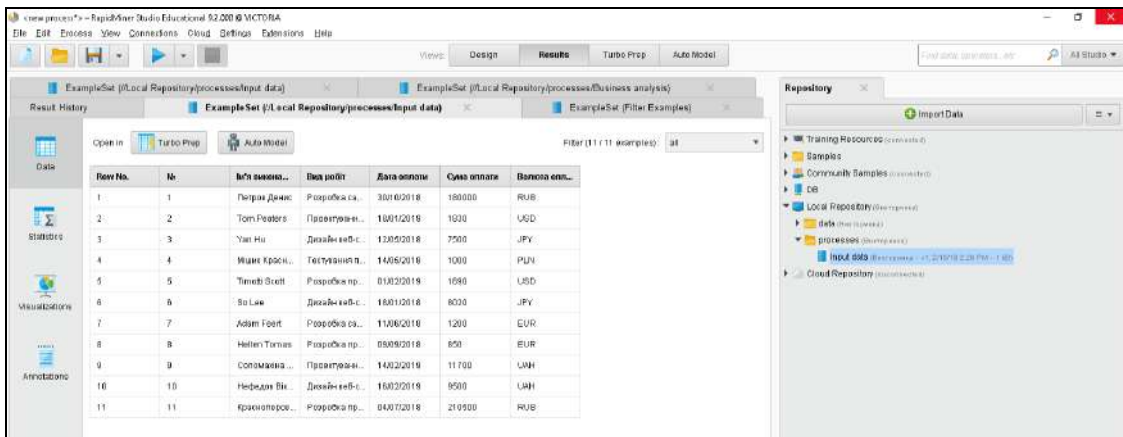
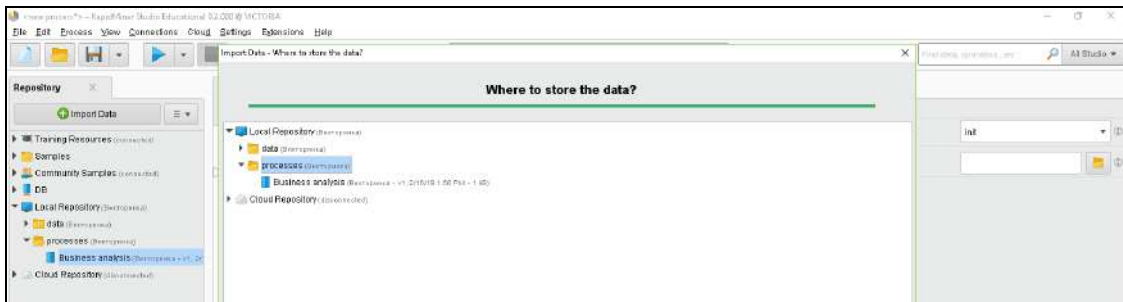


Рис. 114. Збереження даних та вигляд готового файлу у RM

Повертаємося до вкладки *Design* та перетягнемо файл на вкладку *Process*, щоб приступити до аналізу даних (рис. 115).

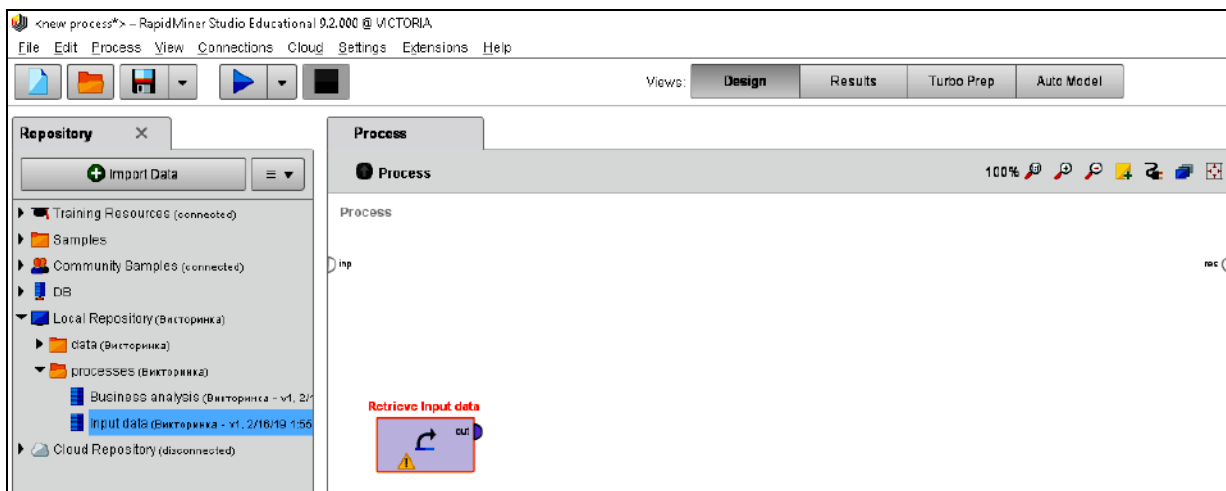


Рис. 115. Початок аналізу даних у RM

З'єднаємо порт *out* вихідних даних та *res* головного процесу, потім натиснемо на кнопку *RunProcess* (рис. 116).

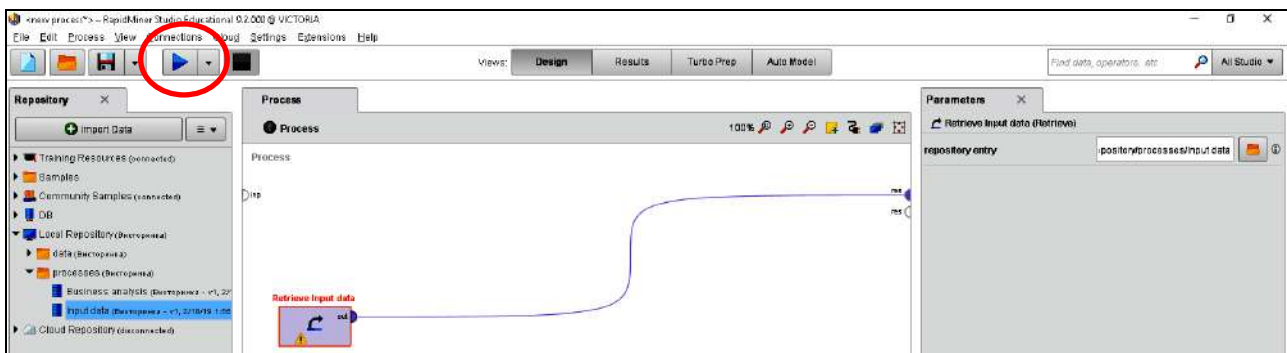


Рис. 116. Запуск процесу аналізу даних у RM

Після закінчення розрахунків програма переключиться на робочу область *Results*. Доречі, користувач можемо перемикатися між робочими областями натисканням на назви цих областей.

За замовчуванням, відкривається вкладка *Data*, яка містить отримані дані. У нашому випадку вони не змінилися, так як ми не проводили ніяких дій над вихідними даними (рис. 117).

Якщо переключитися на вкладку *Statistics*, то ми побачимо базові статистичні дані за кожним атрибутом, а також кількість відсутніх (або нерозпізнаних) атрибутів (рис. 118).

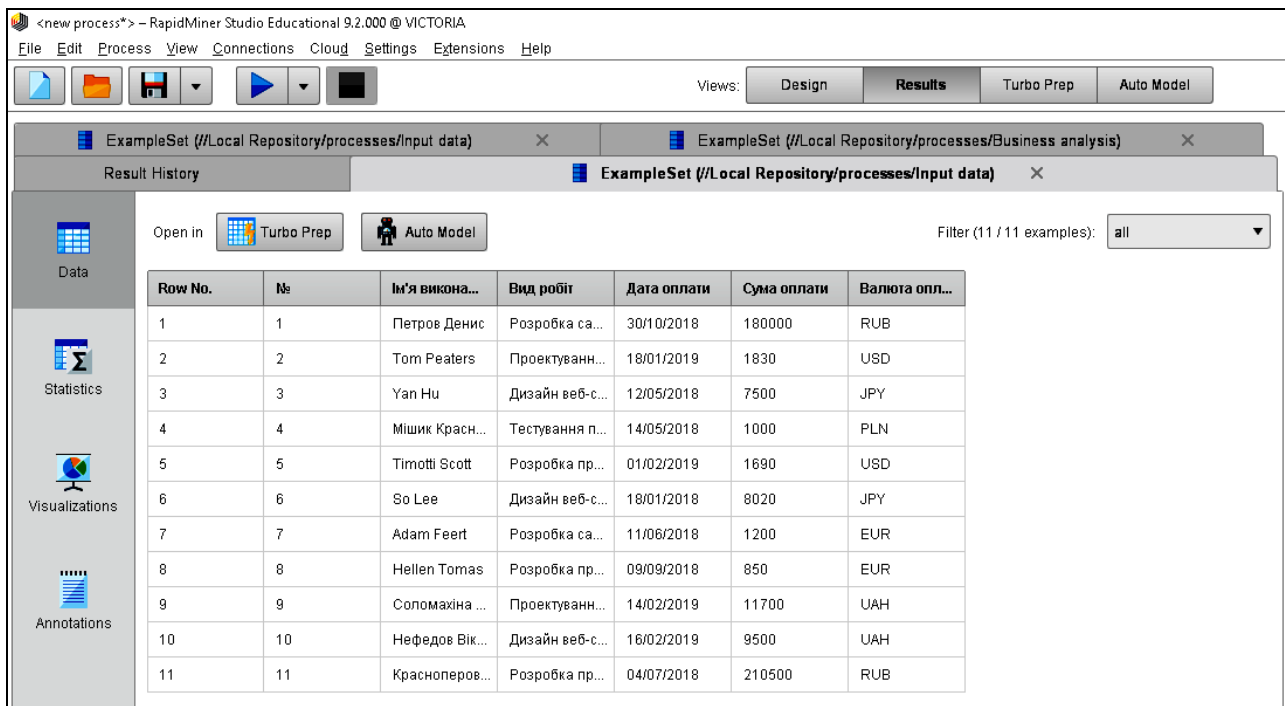


Рис. 117. Вкладка Results у RM

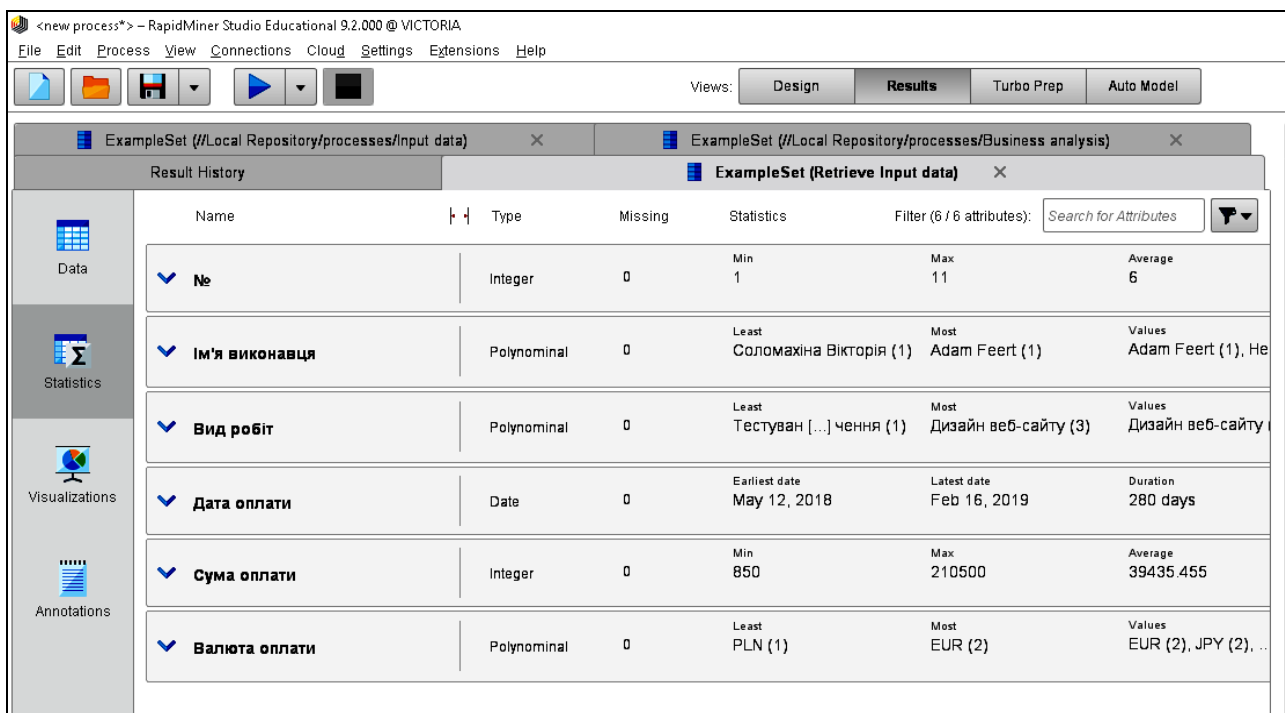


Рис. 118. Вкладка Statistics у RM

Використовуючи вихідні дані, можливо одразу їх візуалізувати. Для цього переходимо на вкладку *Visualizations* та задаємо необхідні графічні параметри (рис. 119).

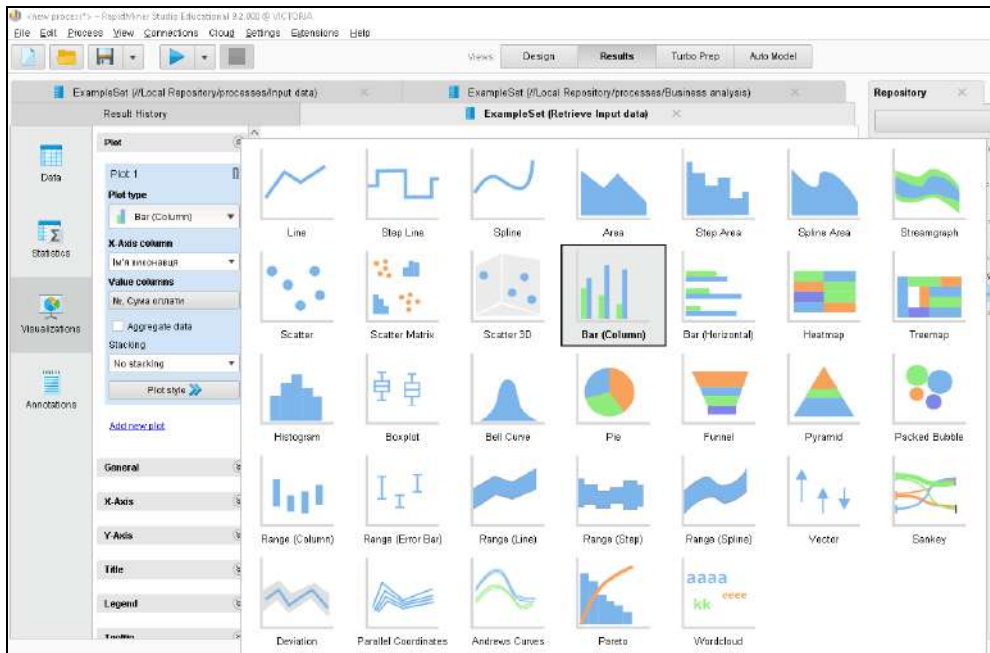


Рис. 119. Вкладка *Visualizations* у RM

Унаслідок отримаємо стовпчасту гістограму розподілу виконавців за сумою отриманих коштів (рис. 120).

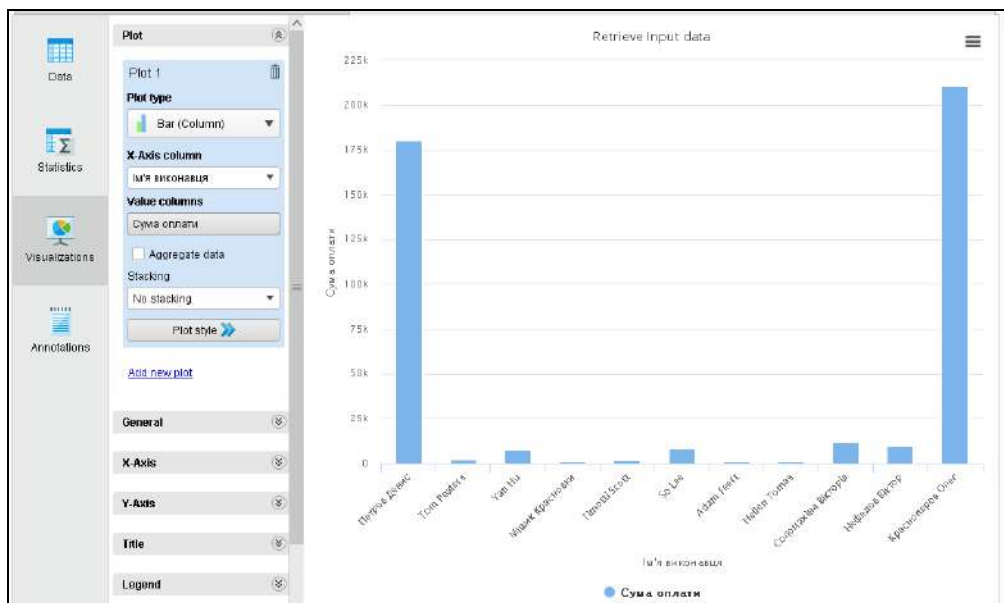


Рис. 120. Візуалізація розподілу виконавців за сумою отриманих коштів

Однак, відповідно до нашого завдання, необхідно провести аналітичні розрахунки та конвертувати зарібок кожного виконавця у національну валюту – гривню – згідно з курсом Національного Банку України. Для цього потрібно використати певні оператори.

Додамо оператор *Filter examples* (рис. 121).

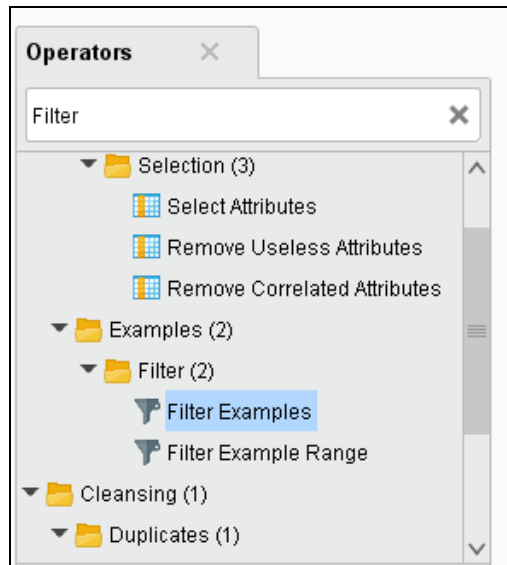


Рис. 121. Вибір оператора *Filter examples*

Його вхід з'єднаємо з виходом *Input data*, а вихід з виходом процесу, позначеним кружечком і написом *res*. У результаті отримаємо такий ланцюжок (рис. 122).

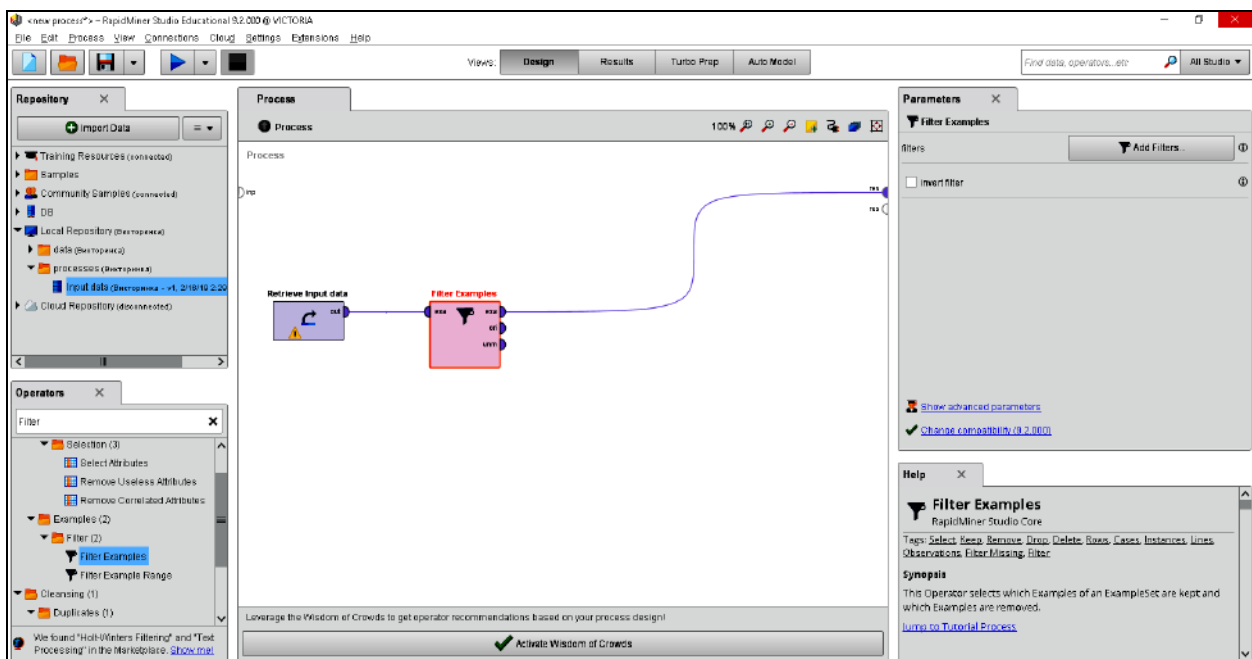


Рис. 122. Процес з використанням *Filter examples*

За допомогою доданого оператора оберемо записи тільки на зазначену дату, яку оголосимо як макрос процесу. Переходимо на вкладку *Context (Parameters)* процесу, там знаходимо секцію *Macros* і натискаємо

на відповідну кнопку. У колонці *Macro* пишемо *date*, а в *Value* бажану дату, нехай це буде 18/01/2019 (рис. 123).

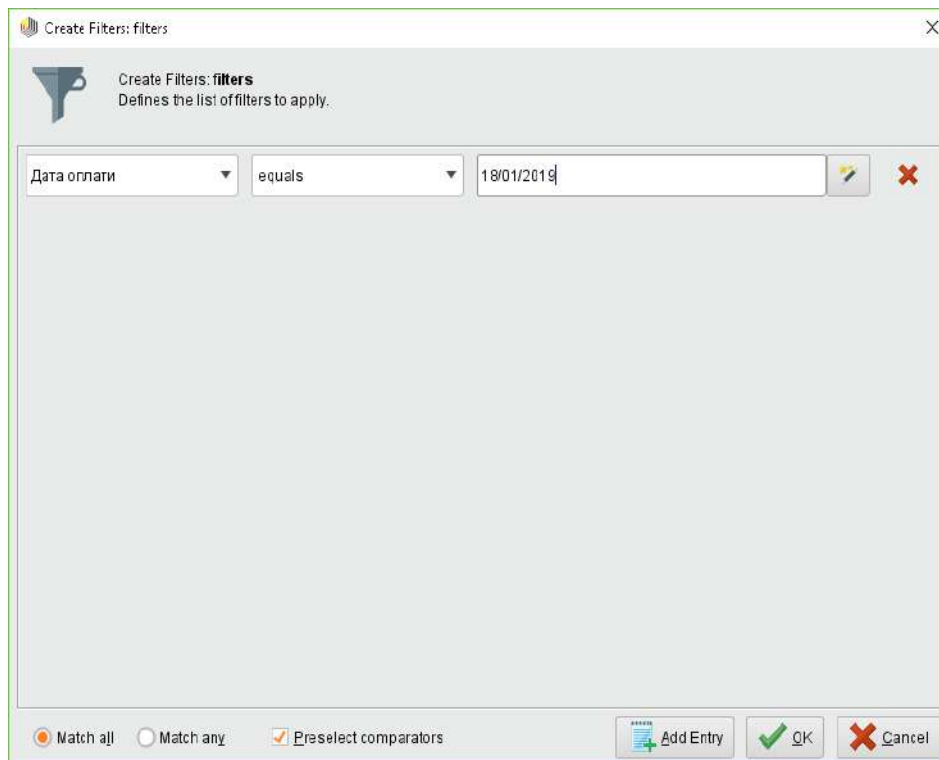



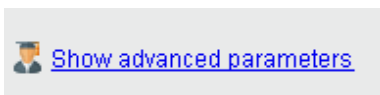
Рис. 123. Створення фільтрів

Унаслідок тиснемо на кнопку  та отримуємо відфільтроване значення з таблиці (рис. 124).

Row No.	№	Ім'я викона...	Вид робіт	Дата оплати	Сума оплати	Валюта опл...
1	2	Tom Peaters	Проектуванн...	18/01/2019	1830	USD

Рис. 124. Результат застосування фільтру

Перший результат отриманий, але нам хотілося б бачити витрати в гривнях за курсом НБ України. Для цього переходимо на вкладку *Design* і додаємо оператор *Open file (Utility / Files / Open file)*. Натискаємо на нього та виставляємо наступні параметри за допомогою команди



(рис. 125).

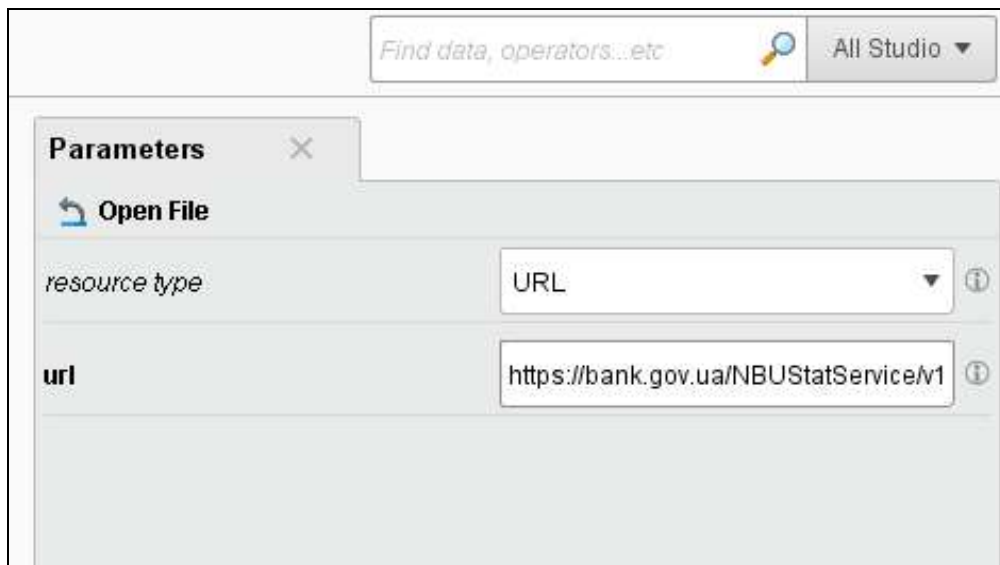
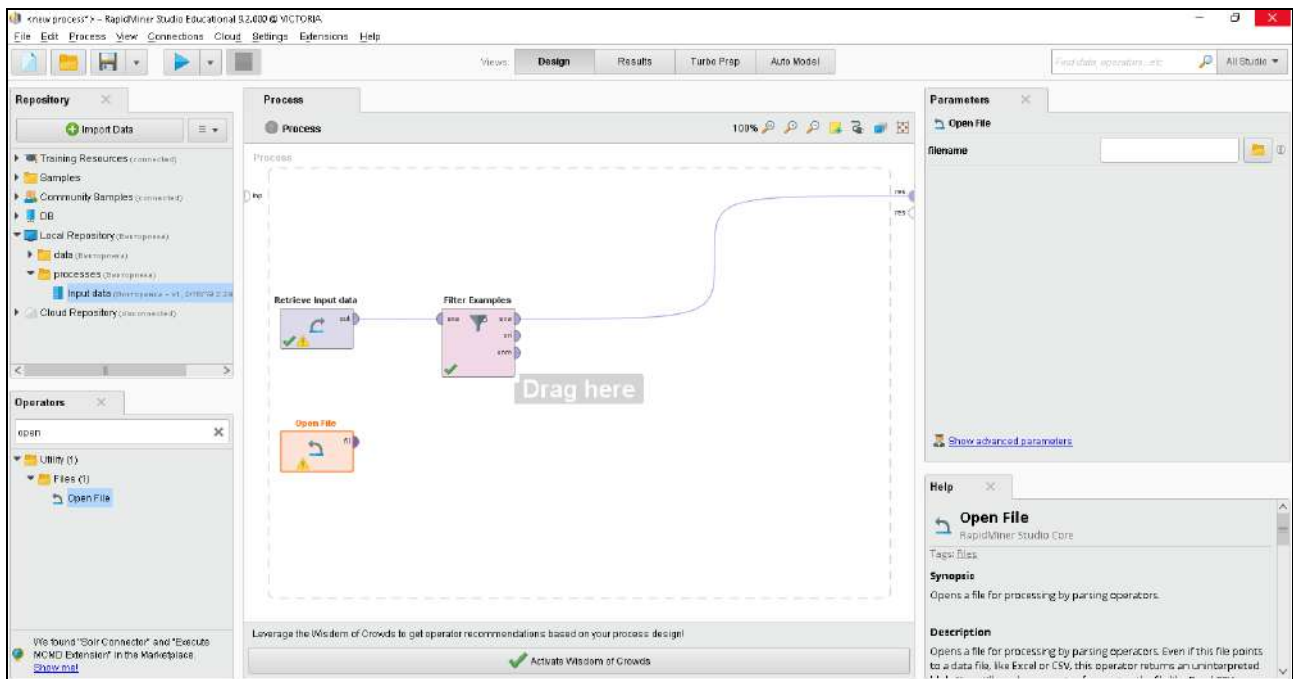


Рис. 125. Робота з оператором *Open File*

На рис. 125 у полі *url* міститься посилання <https://bank.gov.ua/NBUStatService/v1/statdirectory/exchange?date=20190118>. Тобто файл формату XML з посиланням на Курс НБ України на поточну дату, взятий з офіційного сайту: https://bank.gov.ua/control/uk/publish/article?art_id=38441973#exchange.

Безумовно, ми отримаємо дані, але щось має перетворити їх з XML в ExampleSet, тобто таблицю з даними. У нашому випадку цю роль виконає оператор *Read XML (Import / Data / Read XML)* (рис. 126).

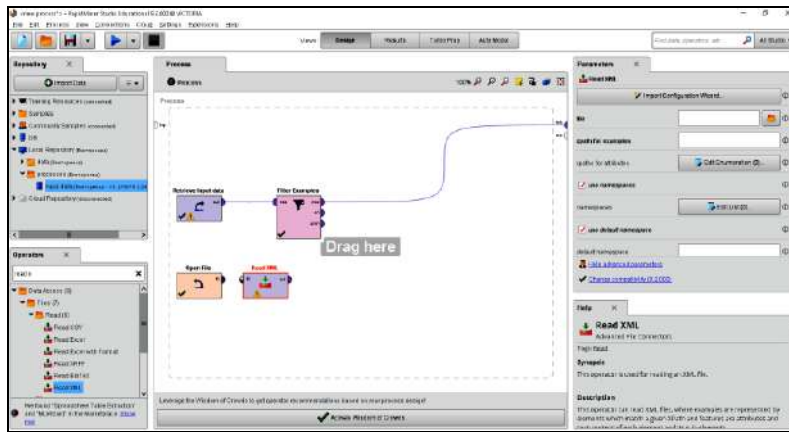


Рис. 126. Додання оператора *Read XML*

З'єднаємо вхід оператора *Read XML* з виходом оператора *Open file* та вносимо налаштування, наведені на рис. 127.

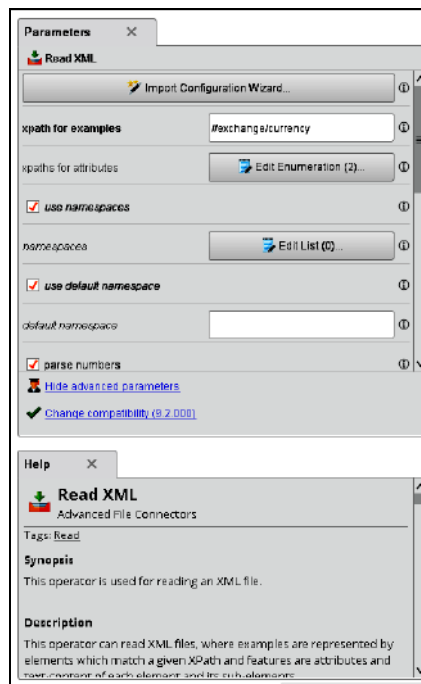
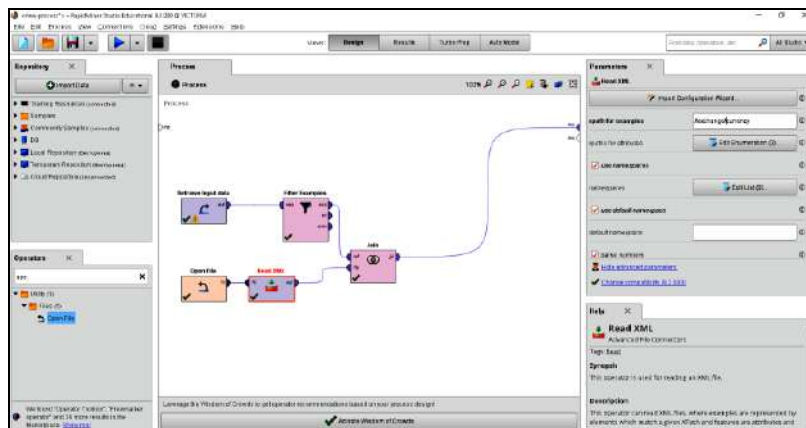


Рис. 127. Параметри оператора *Read XML*

Необхідно визначити, які атрибути RapidMiner будуть вибрані для ExampleSet. Для цього натискаємо на *Edit enumeration*, що розташовано праворуч від *xpath for attributes*, та додаємо два записи:

rate [1] / text () – вартість у гривнях одиниці валюти;

CC [1] / text () – літерний код валюти (рис. 128).

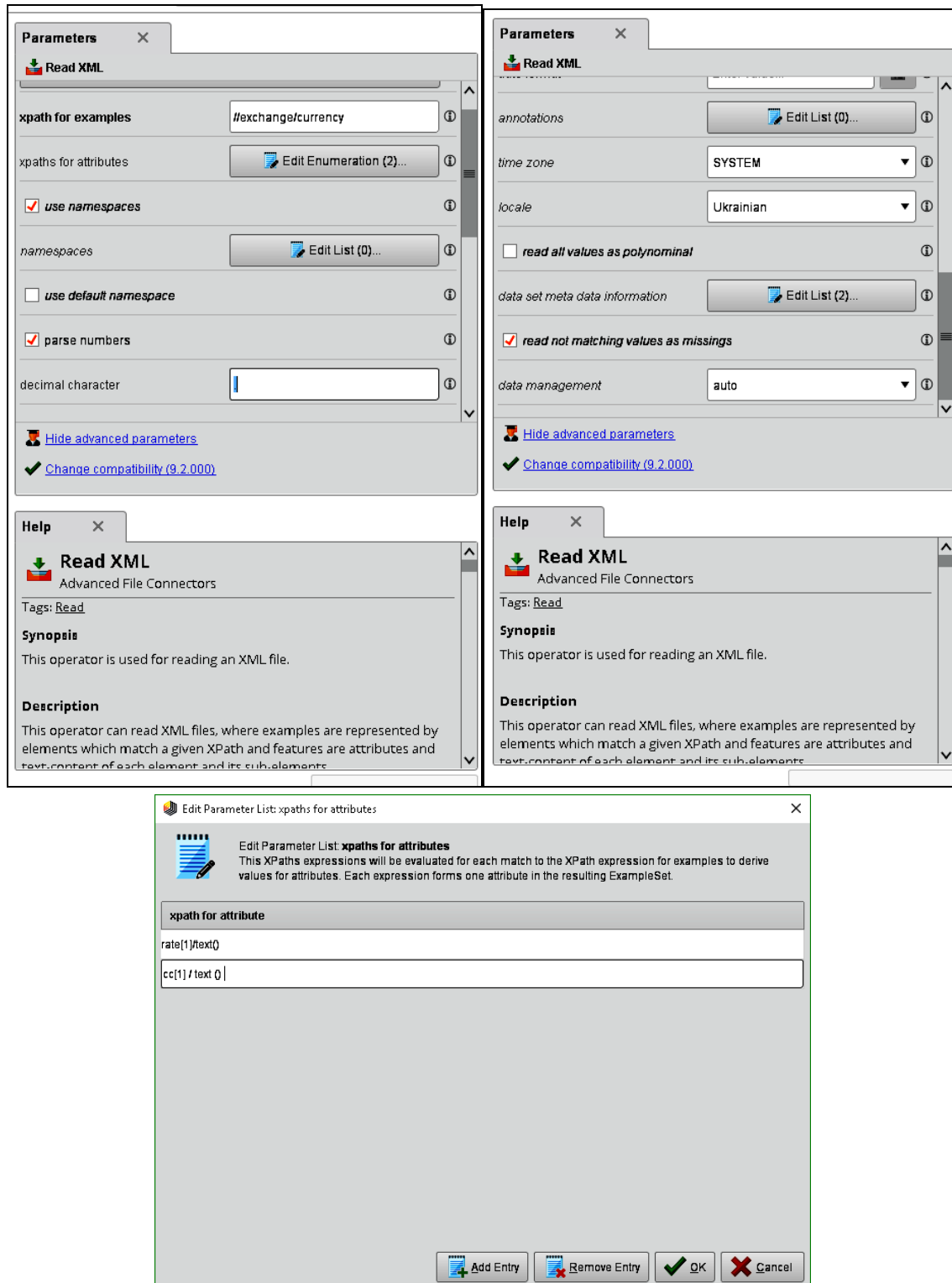


Рис. 128. Внесення параметрів *xpath for attributes*

Тепер необхідно виставити типи значень для атрибутів. Для цього натискаємо на *Edit list* праворуч від *data set meta data information* і вносимо відповідні зміни, що наведено на рис. 129.

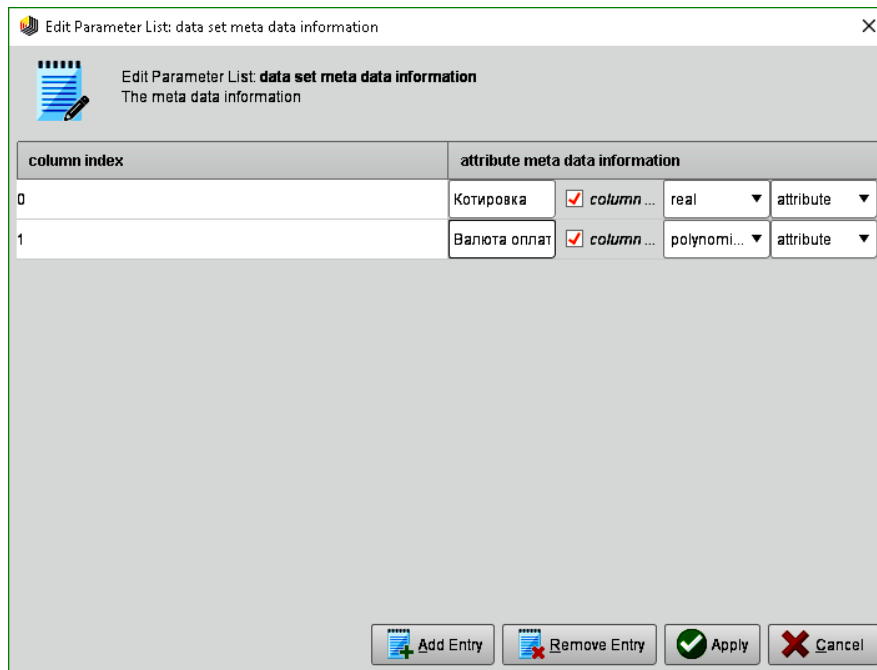



Рис. 129. Внесення параметрів *Edit list*

Після введення всіх параметрів натискаємо  та отримуємо процес, що візуалізовано на рис. 130.

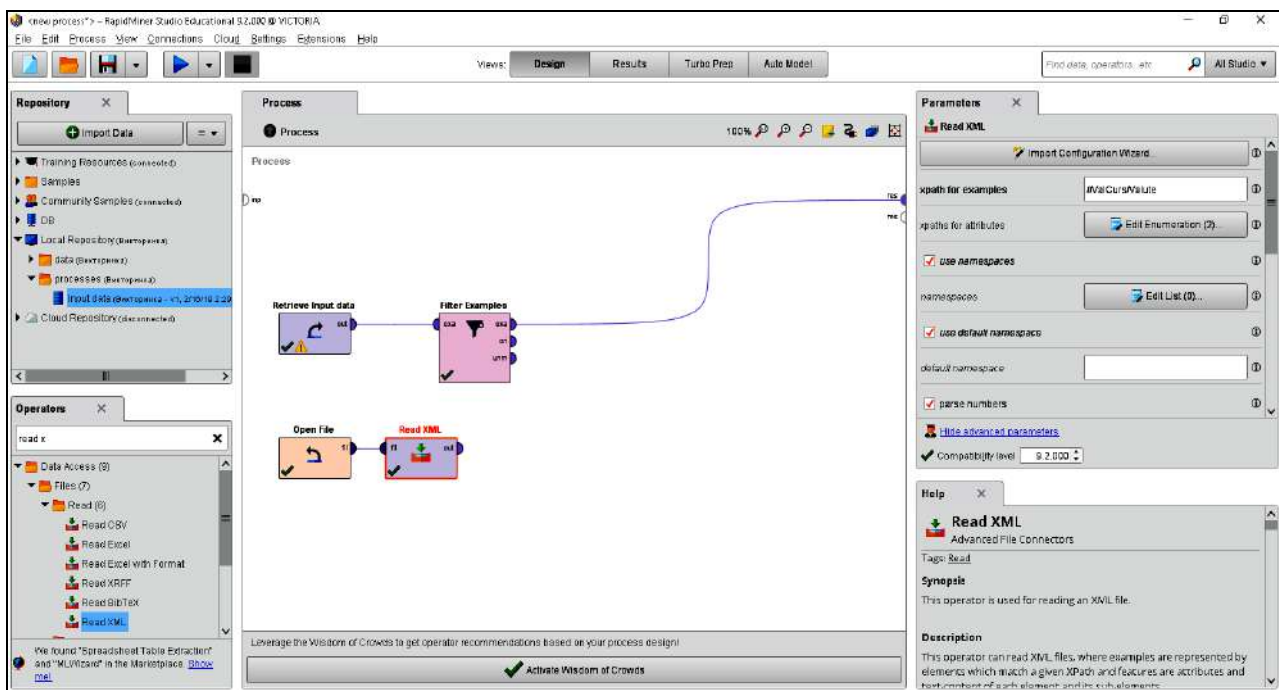


Рис. 130. Процес аналізу в RM

Наступним кроком буде конвертація валют у відфільтрованих за датою даних. Для цього спочатку нам необхідно об'єднати котирування та дані. У цьому нам допоможе оператор *Join* (*Data Transformation / Set Operations / Join*). Тепер виконуємо такі кроки: беремо вихід оператора *Filter examples*, який на заданий момент у нас з'єднаний з виходом процесу, і з'єднуємо його з оператором *Join*, робимо аналогічні дії з оператором *Read XML* (рис. 131).

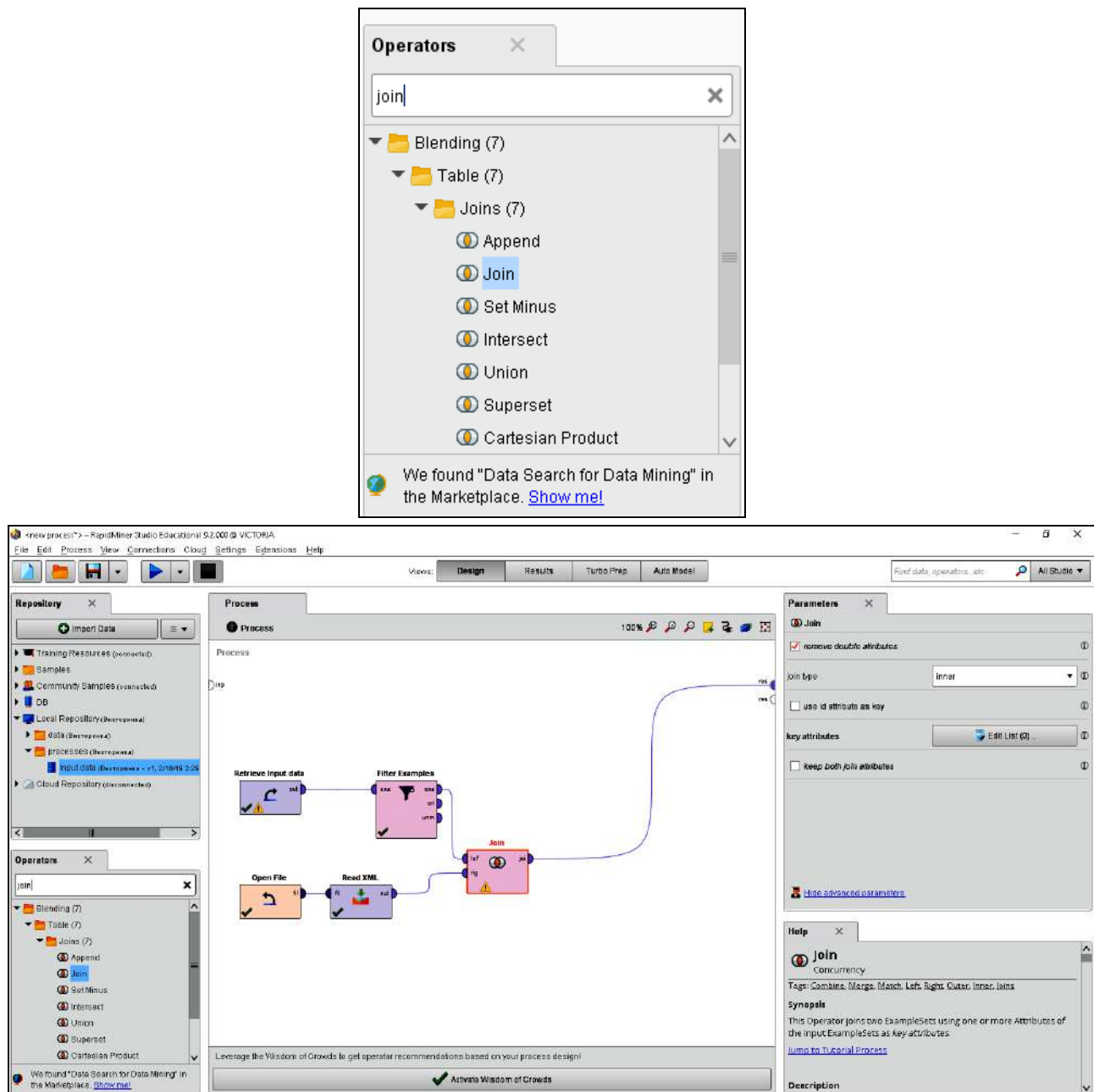


Рис. 131. Додавання оператора *Join* і запуск процесу у RM

Тепер натиснемо на оператор *Join* і визначимо, у який спосіб будуть об'єднані дані: прибираємо галочку *use id attribute as key*, так як об'єднання у нас відбувається за полем *currency*. Коли з'явиться новий параметр

key attributes, натискаємо зліва від нього на *Edit list*, у діалозі *Add entry* і в обох полях пропишемо *currency* (рис. 132).

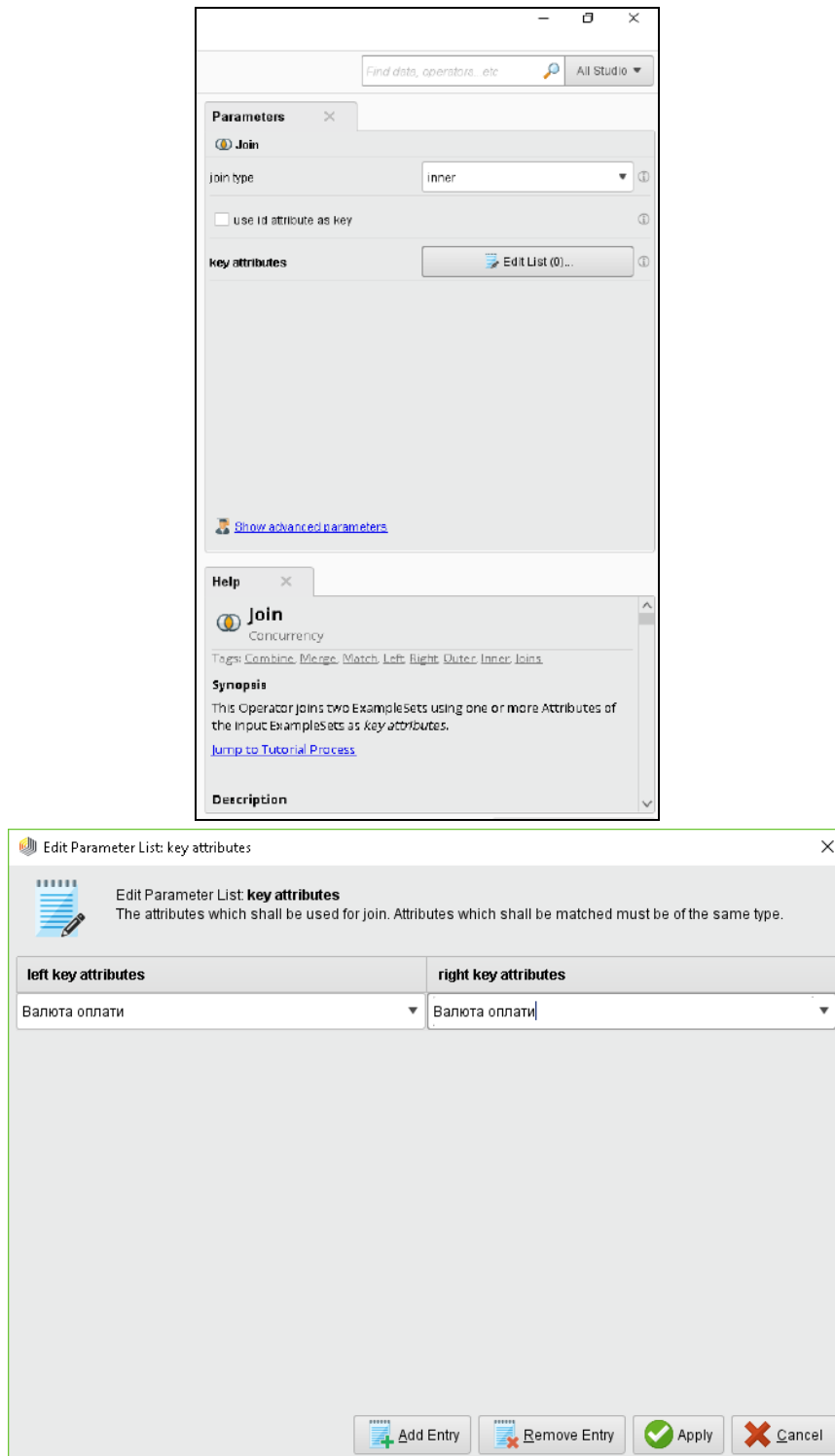



Рис. 132. Редагування параметрів оператора *Join* у RM

Зберігаємо зміни та запускаємо процес аналізу даних, натиснувши на кнопку . Отриманий результат наведено на рис. 133.

Row No.	№	Ім'я викона...	Вид робіт	Дата оплати	Сума оплати	Валюта опл...	Котировка	Валюта опл...
1	2	Tom Peaters	Проектуван...	18/01/2019	1830	USD	28.034	USD

Рис. 133. Результати аналізу даних у RM

Залишився останній крок, власне сама конвертація. Додаємо в процес оператор *Generate Attributes (Data Transformation / Attribute Set Reduction and Transformation / Generation)* і з'єднаємо його вхід з виходом оператора *Join*, а перший вихід, біля якого написано *exr* (скорочено *ExampleSet*), – з виходом усього процесу (*res*) (рис. 134).

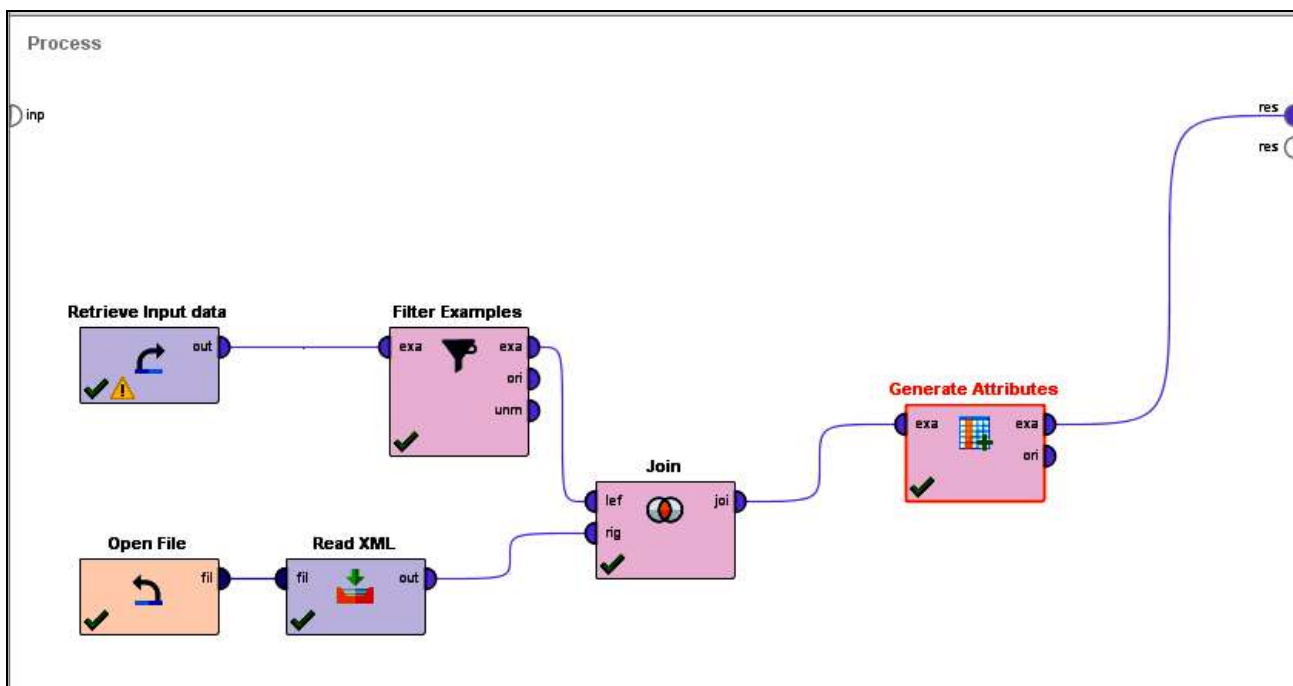


Рис. 134. Результати аналізу даних у RM

Як зрозуміло з назви оператора, його завдання додати новий атрибут, для цього натиснемо на оператор і праворуч у його налаштуваннях на *Edit list*, кнопка навпроти *function descriptions*, задаємо назву атрибуту (рис. 135).

Натискаємо кнопку  та отримуємо результат, наведений на рис. 136.

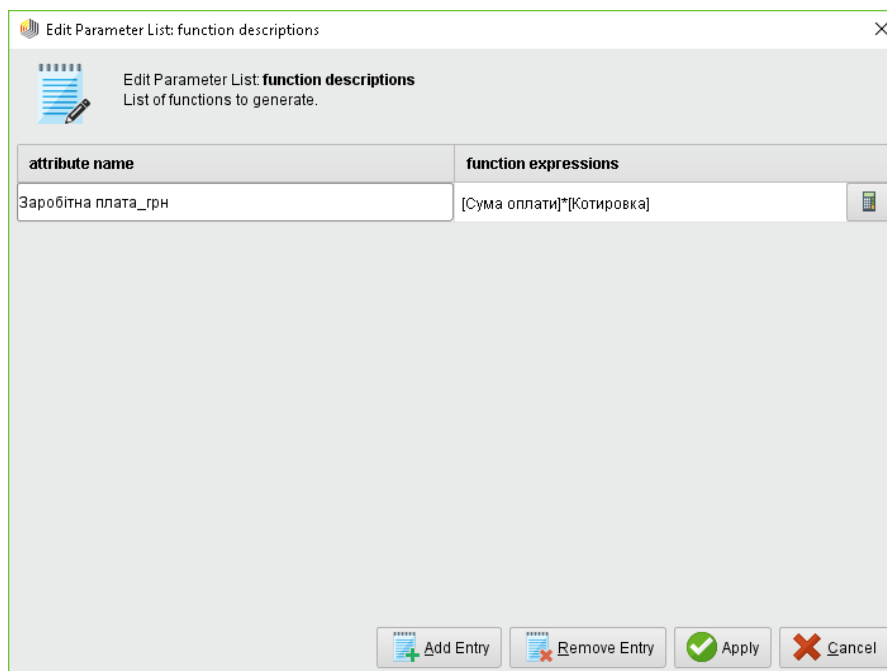


Рис. 135. Встановлення параметрів оператора *Generate Attributes*

No.	№	Ім'я викона...	Вид робіт	Дата оплати	Сума оплати	Валюта опл...	Котировка	Валюта опл...	Заробітна п...
	2	Tom Peaters	Проектуванн...	18/01/2019	1830	USD	28.034	USD	51301.504

Рис. 136. Результати виконання процесу у RM

У такий спосіб було виконано поставлене завдання та отримано витрати на оплату праці в гривнях, які понесла компанія за курсом Національного Банку України на задану дату сплати.

Завдання:

1. Використовуючи вихідні дані завдання-прикладу, зробіть аналітичний висновок щодо:

варіант 1 – витрат компанії за травень місяць 2018 р.;

варіант 2 – витрат за умови групування працівників за типом робіт;

варіант 3 – витрат за умови групування працівників за сумою оплати робіт понад 100 000 ум. од.

2. Формуючи звіт з лабораторної роботи, додайте візуалізацію витрат за варіантом завдання.

Завдання для самостійного опрацювання

Запитання, що виносяться на самостійну роботу студентів:

1. Надайте комплексну характеристику прикладній програмі RapidMiner. Назвіть її переваги та недоліки в роботі бізнес-аналітика.
2. Назвіть основні сфери використання інструментів RapidMiner в економіці.
3. Надайте розгорнутий перелік операторів RapidMiner. У чому полягають особливості їхнього використання?
4. Охарактеризуйте модулі RapidMiner.
5. Які сучасні аналоги програмних продуктів для аналізу великих масивів даних вам відомі?

Завдання 1. Використовуючи функціональні та аналітичні можливості програми RapidMiner, проведіть кореляційно-регресійний аналіз впливу показників X1 – X4 на середні ціни виробників на соняшкову олію (Y). Вихідні дані наведено на рис. 137.

	1 Уточнена посівна площа соняшника, тис.га (X1)	2 Обсяг виробництва соняшника, тис.т (X2)	3 Урожайність соняшника, ц з 1 га зібраної площі (X3)	4 Продуктивність праці в сільськогосподарському виробництві, на 1 зайнятого грн (X4)	5 Середні ціни виробників на соняшкову олію, грн./тонну (Y)
1	1601	2311	14,6	42826,6	8624,73
2	1641	2127	13	36369,7	8490,12
3	1637	2075	12,7	36707,5	8586,44
4	1784	1569	9,1	31117,5	8681,36
5	2020	2860	14,2	30254,8	8794,37
6	2107	2123	10,5	26339,8	8576,88
7	2065	2308	11,5	28174,2	8510,7
8	2531	2266	9,3	25484,8	8588,74
9	2889	2794	10	24103,8	8901,54
10	2943	3457	12,2	27066,9	8940,86
11	2502	2251	9,4	34577,2	8884,67
12	2834	3271	12	40728,8	8942,38
13	4001	4254	11,2	38007	8827,31
14	3521	3050	8,9	63376,9	8982,49
15	3743	4706	12,8	72621,9	8781,93
16	3964	5324	13,6	83854,6	8611,88
17	3604	4174	12,2	88571,3	8525,05
18	4306	6526	15,3	127372,5	8585,37
19	4232	6364	15,2	131332	8608,94
20	4572	6772	15	132680,4	8535,62

Рис. 137. Вихідні дані

Як приклад для виконання завдання можна використати алгоритм розв'язання, що наведений за посиланням: <https://studfiles.net/preview/6449015>.

Завдання 2. За допомогою системи інтелектуального аналізу даних RapidMiner побудувати процес, що дозволить аналізувати в режимі реального часу зміни курсу гривні до п'яти найбільш поширених міжнародних валют.

Тема 6. Використання пакетів SPSS та EViews для побудови моделей

Лабораторна робота 6 Аналіз великих масивів даних за допомогою спеціалізованого статистичного пакету SPSS

Мета: надбання навичок аналізу даних за допомогою SPSS.

Завдання: ознайомитися з функціональними й аналітичними можливостями спеціалізованого пакету SPSS; навчитися обробляти та аналізувати великі масиви даних за допомогою пакету SPSS.

Методичні рекомендації

1. Знайомство з пакетом SPSS.

Продукт фірми SPSS Inc. (Statistical Products and Service Solution – Статистичні продукти і сервісні рішення) є одним з найстаріших систем статистичного аналізу й управління даними. SPSS Statistics (Statistical Package for the Social Sciences (Статистичний пакет для суспільних наук) – комп'ютерна програма для статистичного оброблення даних, один з лідерів ринку в області комерційних статистичних продуктів, призначених для проведення прикладних досліджень у суспільних науках.

Функціональні можливості:

- введення та зберігання даних;
- можливість використання змінних різних типів;
- частотність ознак, побудова таблиць, графіків та діаграм, таблиці спряженості;
- первинна описова статистика;
- проведення та аналіз маркетингових досліджень.

Пакет SPSS має широкий арсенал модулів: IBM SPSS Statistics Base; IBM SPSS Advanced Statistics; IBM SPSS Bootstrapping; IBM SPSS Categories; IBM SPSS Complex Samples; IBM SPSS Conjoint; IBM SPSS Custom Tables; IBM SPSS Data Preparation; IBM SPSS Decision Trees; IBM SPSS Direct Marketing; IBM SPSS Exact Tests; IBM SPSS Forecasting; IBM SPSS Missing Values; IBM SPSS Neural Networks та IBM SPSS Regression.

Пакет SPSS побудований як традиційна база даних: накопичення масиву інформації, його формалізація та представлення результатів статистичного оброблення масиву у вигляді звіту. Зовнішніх відмінностей інтерфейсу SPSS від традиційних баз даних або електронних таблиць (MS Access, MS Excel) немає, що значно спрощує перше знайомство з пакетом і дозволяє досить швидко почати процедуру введення або імпорту даних, крім того, пакет включає довідник та глосарій статистичних термінів.

Можливості програми SPSS у плані проведення статистичного аналізу практично не відрізняються від аналізу в системі STATISTICA. Структурно SPSS є модульною системою. У базову систему закладені всі функції роботи з даними, всі основні статистичні процедури, від описових статистик і таблиць спряженості до факторного та кластерного аналізу. Масив даних для оброблення може бути експортований з текстового файлу, з файлу формату MS Excel, Lotus 1-2-3 або перенесений через буфер обміну. Пакет має значні візуальні можливості щодо побудови різних типів звичайних і спеціалізованих статистичних графіків, з можливістю редагування їхнього дизайну.

Посилання для скачування інсталяційного файлу: <http://spsstatistics.ru>.

2. Робота в пакеті SPSS.

Основою програми SPSS є SPSS Base (базовий модуль), що надає різноманітні можливості доступу до даних та управління ними. Він містить економіко-статистичні методи аналізу.

Традиційно разом з SPSS Base (базовим модулем) використовуються ще модулі: Advanced Models (розширенні моделі), Regression Models (регресійні моделі) та Tables (таблиці).

Розглянемо роботу базового модуля.

SPSS Base (Базовий модуль). SPSS Base входить у базове постачання програми та охоплює всі процедури введення, відбору та коректування даних, а також більшість пропонованих у SPSS статистичних методів. Разом з простими методиками статистичного аналізу, такими як частотний

аналіз, розрахунок статистичних характеристик, таблиць зв'язаності, кореляцій, побудови графіків, цей модуль охоплює t-тести та велику кількість інших непараметричних тестів, а також ускладнені методи, такі як багатовимірний лінійний регресійний аналіз, дискримінантний аналіз, факторний аналіз, кластерний аналіз, дисперсійний аналіз, аналіз придатності (аналіз надійності) та багатовимірне шкалювання.

Під час запуску програми відкривається файл даних SPSS, що нагадує робочу книгу Microsoft Excel (електронну таблицю). Загальну схему побудови файлу SPSS наведено на рис. 138.

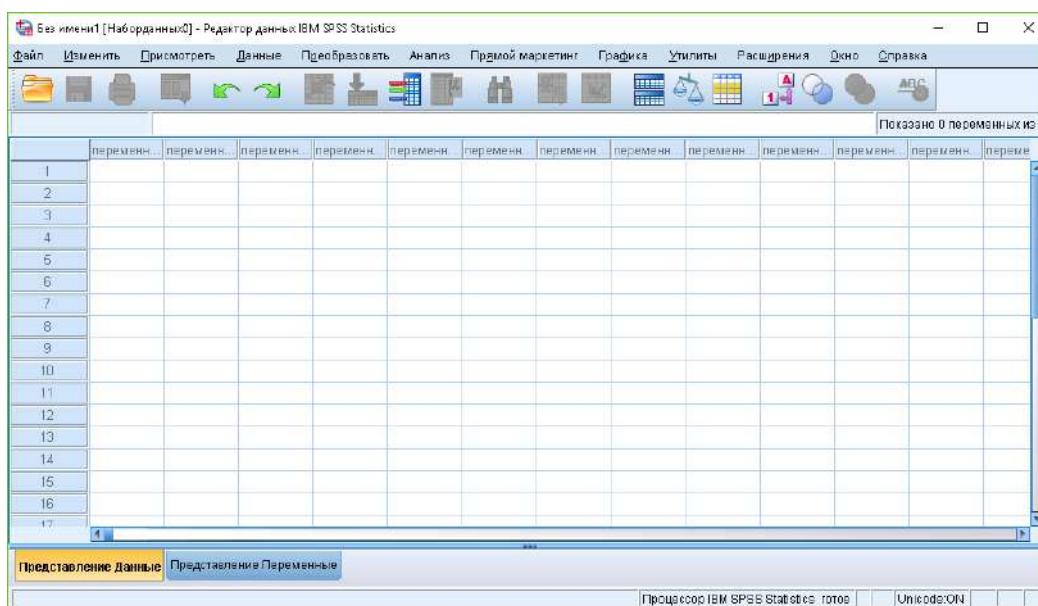


Рис. 138. Редактор даних SPSS

Існує три основні способи формування бази даних у форматі SPSS:

1. Імпорт бази даних з інших програмних джерел (Microsoft Access, Microsoft Excel, текстових файлів та інших).
2. Введення даних у SPSS за допомогою спеціалізованого програмного забезпечення (SPSS Data Entry).
3. Ручне введення даних у SPSS (ефективне за малих розмірів вибірки, а також для досягнення деяких специфічних цілей).

Імпорт даних з інших джерел. Цей спосіб створення бази даних у форматі SPSS є найбільш поширеним. Найчастіше він використовується в SPSS як допоміжний засіб для статистичного аналізу даних. Щоб здійснити імпорт даних у SPSS, необхідно сформувавши у відповідній програмі (з якої здійснюватиметься імпорт) таблицю даних, відформатовану певним способом. Таблиця даних в іншій програмі, з якої буде здійснюватися імпорт,

має відповідати саме такій схемі: заголовок змінної > значення змінної. Після того, як буде створена відповідна для імпорту таблиця даних, слід відкрити SPSS і викликати діалогове вікно імпорту даних за допомогою меню *Файл (File) > Імпорт даних (Open Database) > Новий запит (New Query)*. Відкриється майстер *Створення баз даних (Database Wizard)* (рис. 139), у якому необхідно вказати джерело даних, з якого проводитиметься імпорт даних.

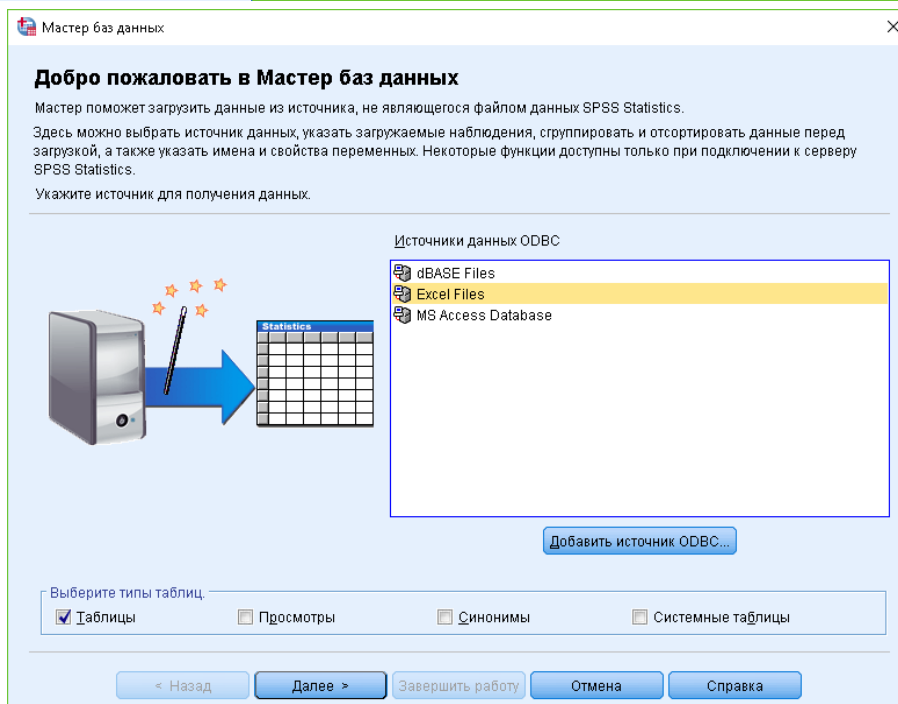
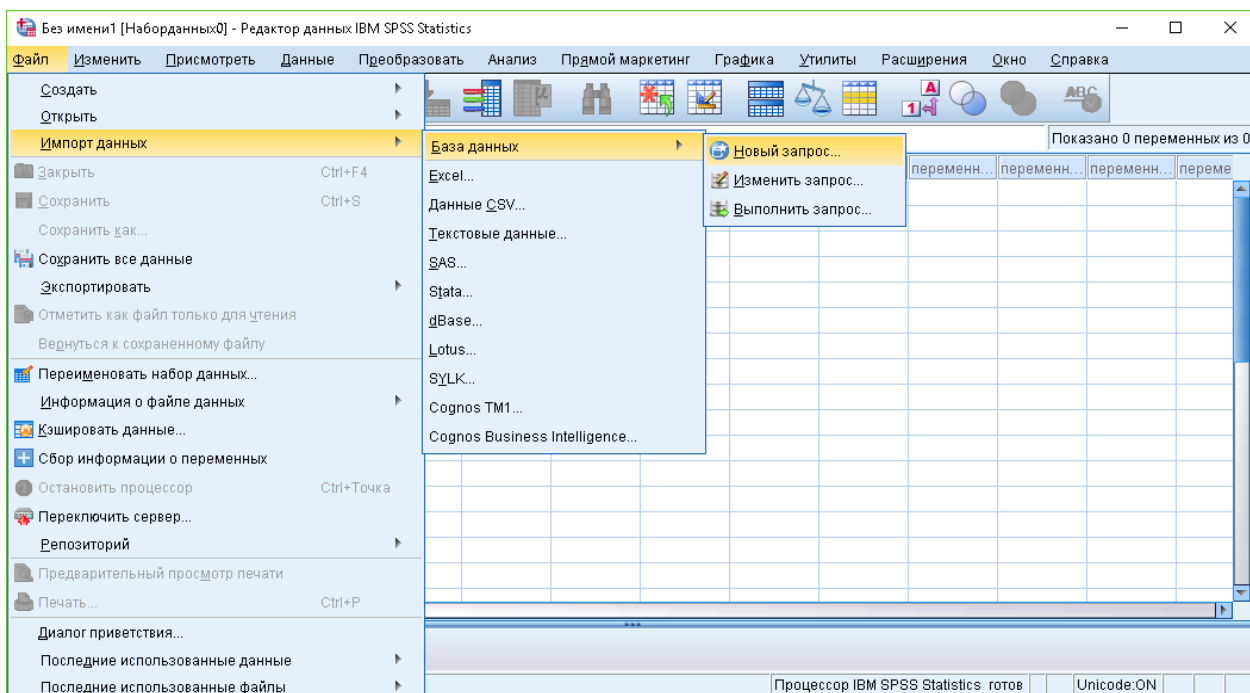


Рис. 139. Діалогове вікно *Створення баз даних*

Виберемо відповідне джерело, після чого на екрані відкриється діалогове вікно *З'єднання пристрою ODBC (ODBC Driver Login)*. У цьому діалоговому вікні слід вказати повний шлях до бази даних, з якої буде проводиться імпорт таблиці (рис. 140).

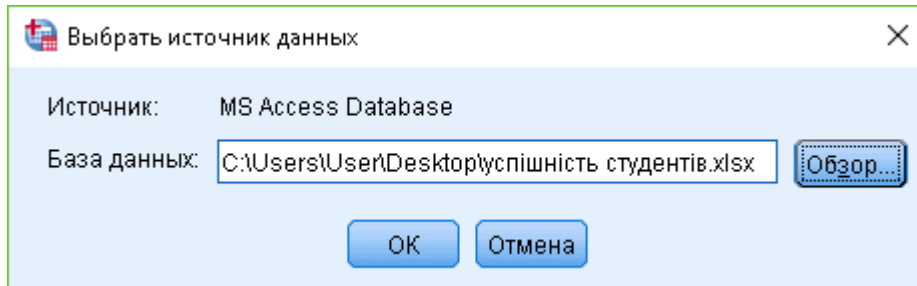


Рис. 140. Діалогове вікно *З'єднання пристрою ODBC*

Відкриється нове діалогове вікно, в ньому з лівого списку всіх таблиць, доступних у вказаному джерелі даних, виберемо ту, яка містить дані, що імпортуються, і переміщуємо її в правий список. Після чого у вікні *SPSS Data Editor* з'явиться імпортована таблиця (рис. 141).

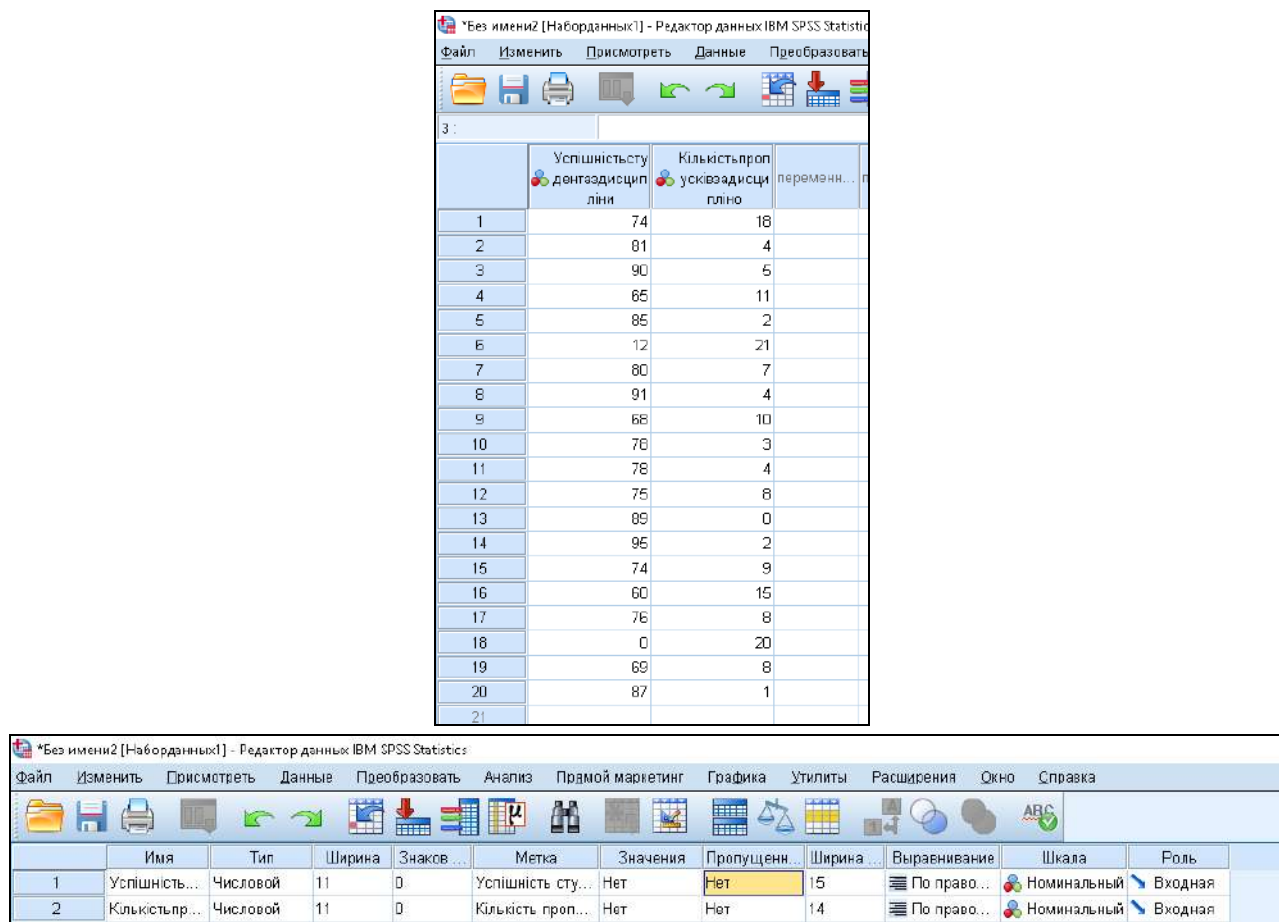


Рис. 141. Результат телепортованої таблиці

Якщо потрібно змінити тип змінної, клацніть у клітинці на кнопку з трьома крапками. Відкриється діалогове вікно *Визначення типу змінної (Define Variable Type)* (рис. 142).

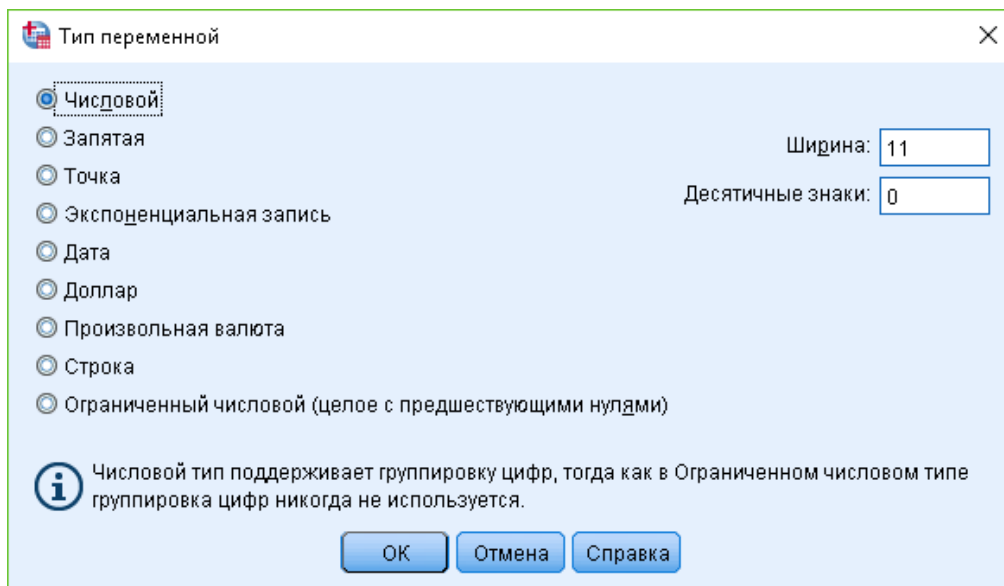


Рис. 142. Діалогове вікно *Визначення типу змінної*

Дані можна вводити за окремими спостереженнями (рядкам) або за окремими змінними (стовпцями).

Однією з переваг пакету SPSS для Windows є побудова таблиць і наявність великої кількості різноманітних графіків, які можуть бути побудовані як за допомогою процедур меню графіків, так і з різноманітних процедур меню статистик.

Меню Аналіз (Analyze) / Інші таблиці (Custom Tables) надає доступ до діалогових вікон (рис. 143), призначених для побудови одно- та багатовимірних таблиць. За допомогою цих вікон можна створювати більш презентабельні таблиці.

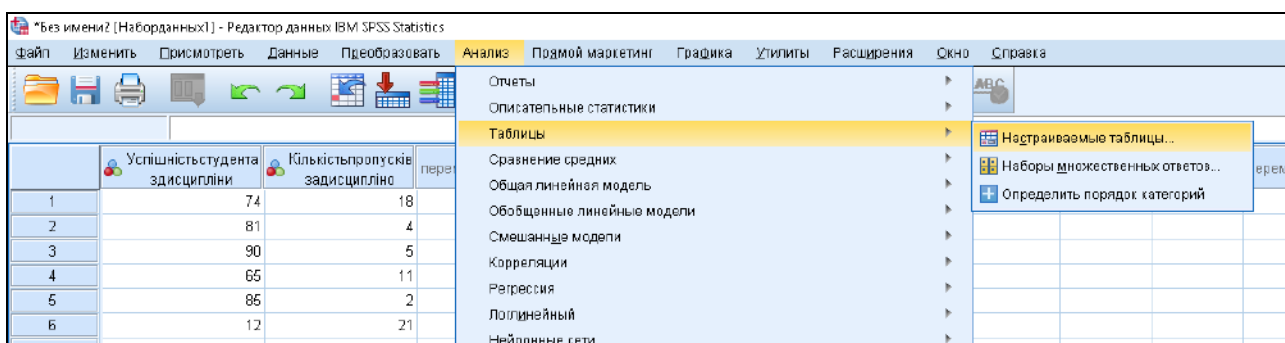


Рис. 143. Меню *Аналіз*

Побудуємо таблицю на основі введених даних. Для цього в меню *Аналіз (Analyze)* вибираємо вид таблиці, після чого відкривається діалогове вікно, в якому задаємо змінні таблиці (рис. 144).

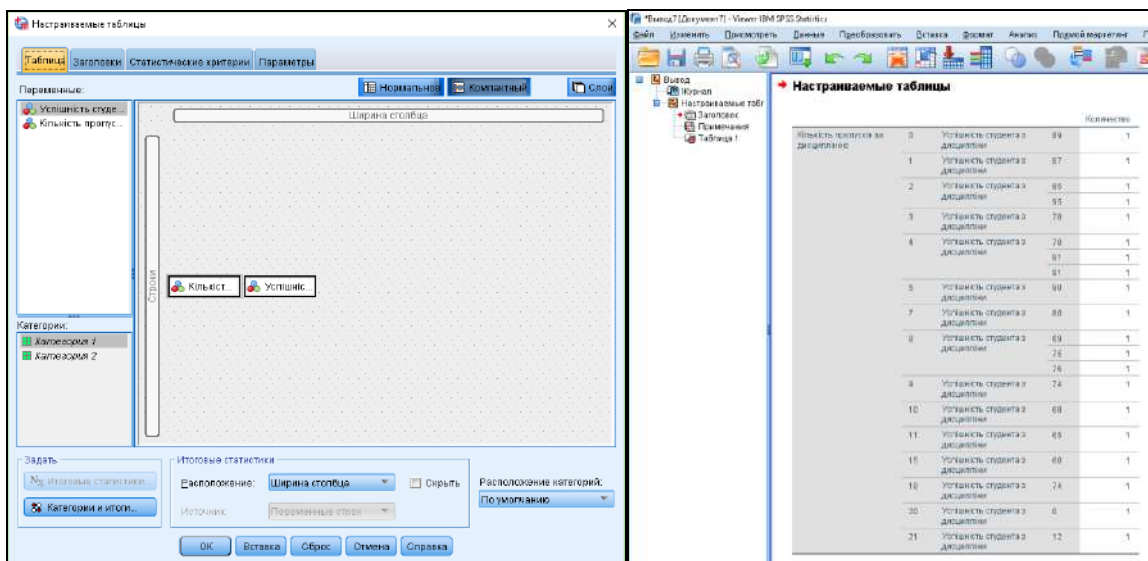


Рис. 144. Результаты построения таблицы

ППП SPSS має графічний інструментарій, що дозволяє візуалізувати результати досліджень. Для цього необхідно скористатися меню *Графіка* та обрати тип діаграми.

Для побудови графіка достатньо після вибору типу графіка вказати необхідні змінні, на підставі яких він буде побудований. Вибираємо меню *Графіка*, потім виберемо *Майстер діаграм* та тип графіка (його необхідно перемістити у верхнє поле) (рис. 145).

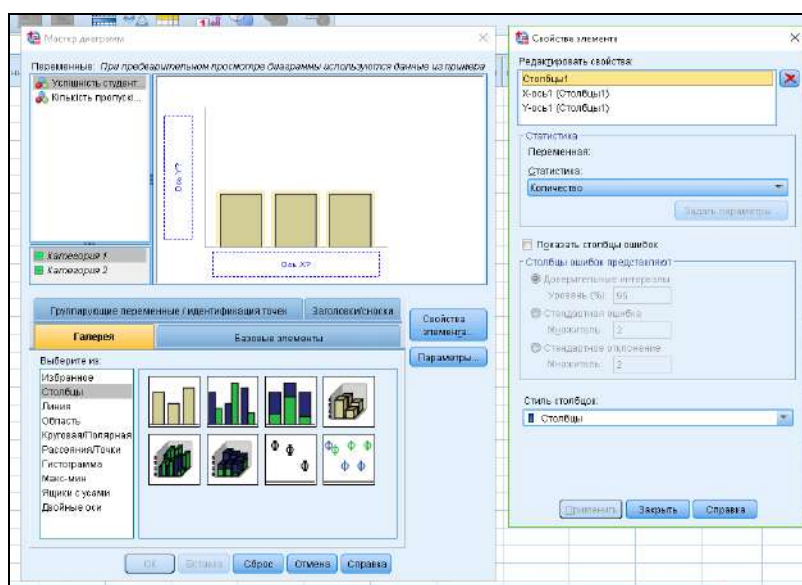


Рис. 145. Построение графика в SPSS

Після призначення змінних відкривається вікно результатів (рис. 146).

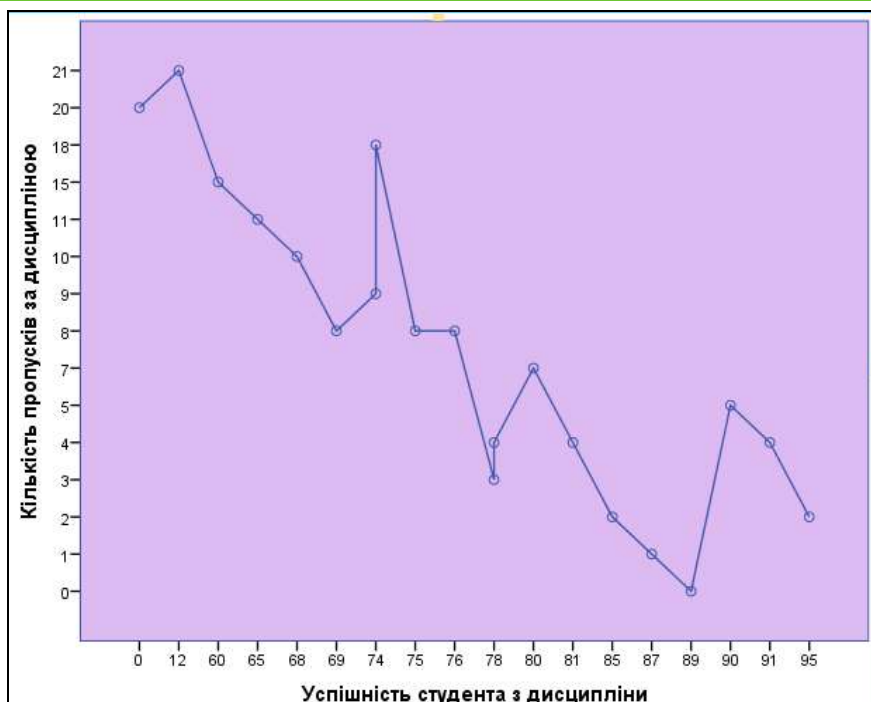
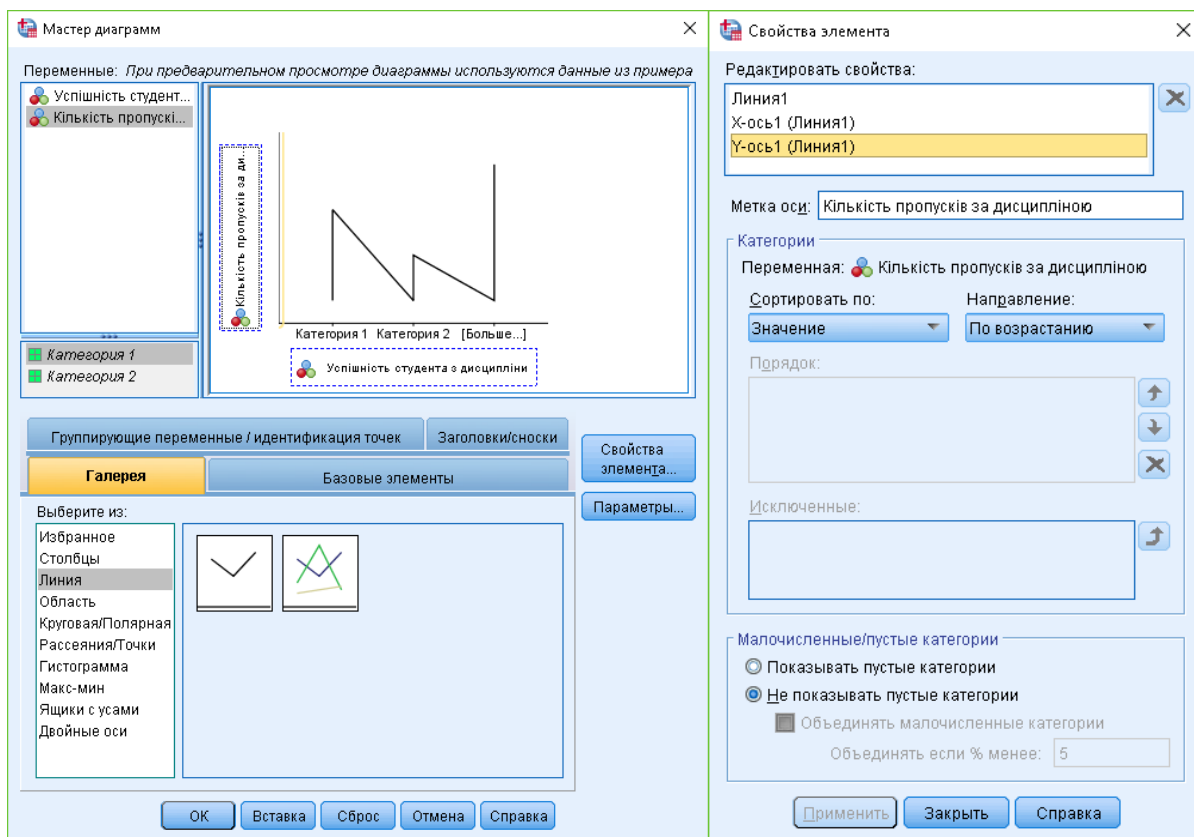
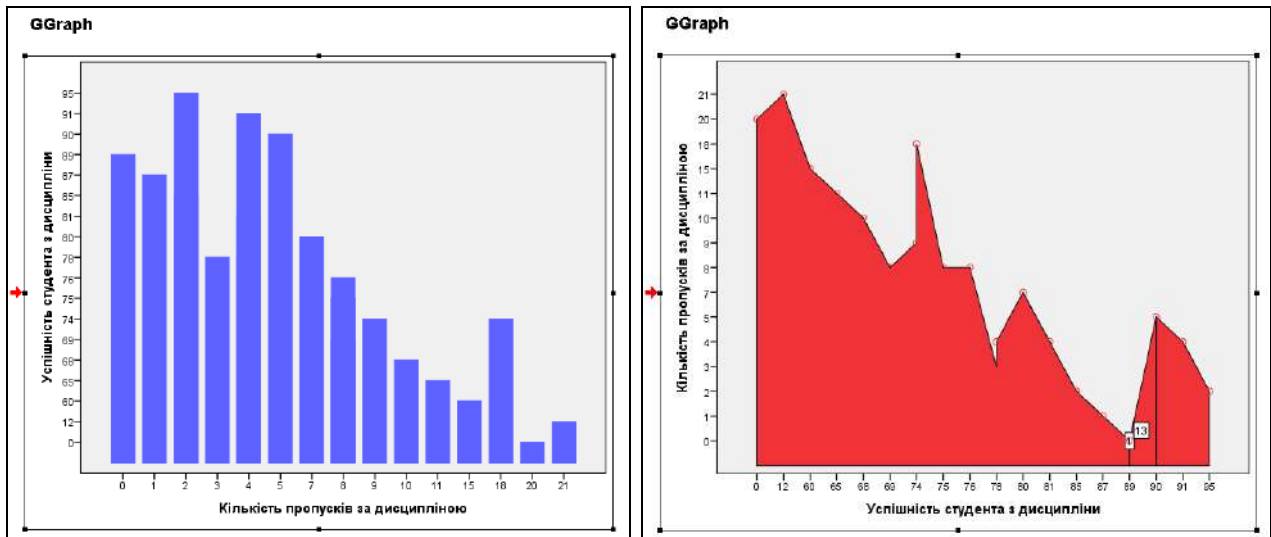


Рис. 146. Результати побудови графіка в SPSS

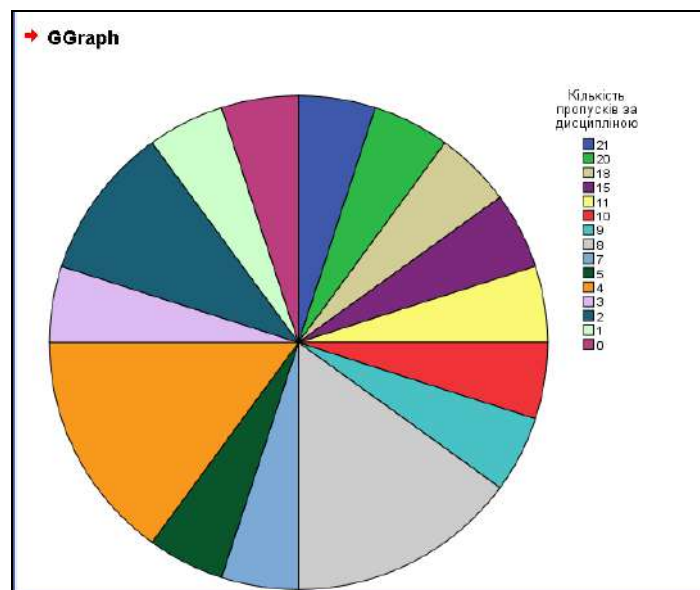
Для редагування створеного графіка необхідно двічі клацнути на полі графіка, після цього з'явиться безліч можливостей для додаткового редагування. Разом із зручними глобальними можливостями змінювати окремі

стильові елементи графіків і перетворювати змінні, що використовувалися для побудови графіка, за допомогою інтерактивних графіків стає також можливим одночасна побудова декількох графіків для окремих категорій додаткових змінних (рис. 147).



Стовпчаста гістограма

Діаграма Область



Кругова діаграма

Рис. 147. Варіанти візуалізацій у SPSS

3. Розрахунок характеристик варіаційних рядів у SPSS.

Для аналізу характеристик розподілу в SPSS використовуються такі чотири команди меню:

- 1) Analyze (Аналіз) Descriptive Statistics (Дескриптивні статистики) Descriptives..., (Описова статистика);

2) Analyze (Аналіз) Descriptive Statistics (Дескриптивні статистики) Frequencies... (Частоти);

3) Analyze (Аналіз) Descriptive Statistics (Дескриптивні статистики) Explore... (Дослідження);

4) Analyze (Аналіз) Reports (Звіти) Case summaries... (Підсумки за спостереженнями).

Для характеристики центру розподілу використовують такі статистичні критерії (показники): мода, медіана та середнє. Своєю чергою, для характеристики варіації використовують максимум, мінімум, розмах, квартильний розмах, дисперсія і середньоквадратичне відхилення, а також асиметрію та ексцес.

Розглянемо розрахунок показників ряду розподілу на прикладі оцінок, отриманих студентами за результатами вивчення однієї з навчальних дисциплін.

Відкриваємо файл з даними.

Вибираємо команду меню *Analyze (Аналіз) / Descriptive Statistics (Дескриптивні статистики) / Descriptives... (Описова статистика)* (рис. 148). Саме модуль *Descriptives...* дає найбільш повну характеристику рядів розподілу, але слід зауважити, що всі важливі характеристики можна отримати також у вкладці *Frequencies... (Частоти)*.

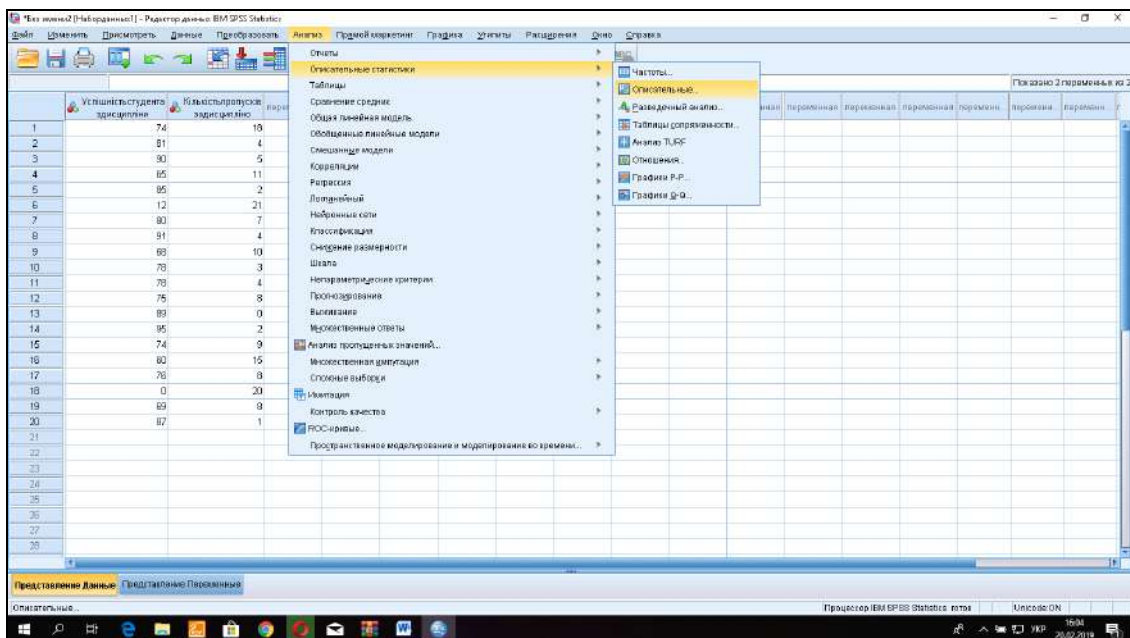


Рис. 148. Запуск модуля *Descriptives* у SPSS

Відкриється діалогове вікно *Descriptives* (рис. 149).

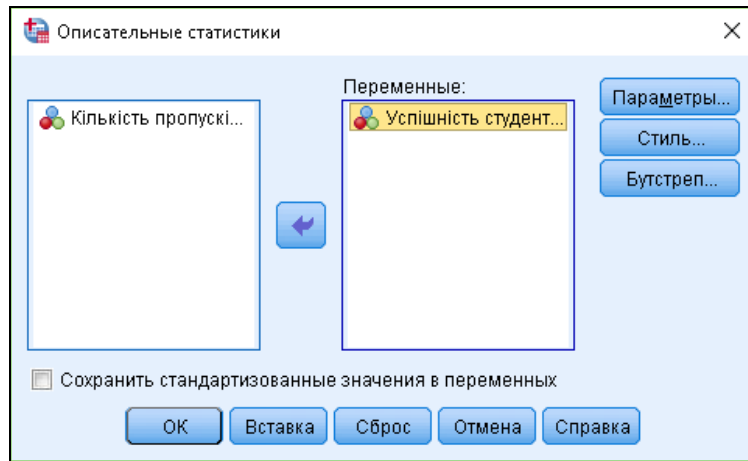


Рис. 149. **Діалогове вікно *Описові статистики (Descriptives)***

Переносимо змінну в список змінних і натискаємо на кнопку *Параметри*. Тут можна задати обчислення таких статистичних характеристик: середнього значення; суми; стандартного відхилення; стандартної помилки; дисперсії; мінімуму; максимуму; розмаху; ексцесу; асиметрії. Встановіть прапорці для виведення необхідних характеристик.

Якщо аналізується декілька змінних, можна також задати послідовність виводу характеристик ряду розподілу: в порядку зростання середніх значень; в порядку убывання середніх значень; за абеткою (за іменами змінних) та згідно зі списком вибраних цільових змінних.

За замовчуванням обирається останній варіант (рис. 150). Якщо є тільки одна змінна, як в заданому прикладі, порядок не має значення.

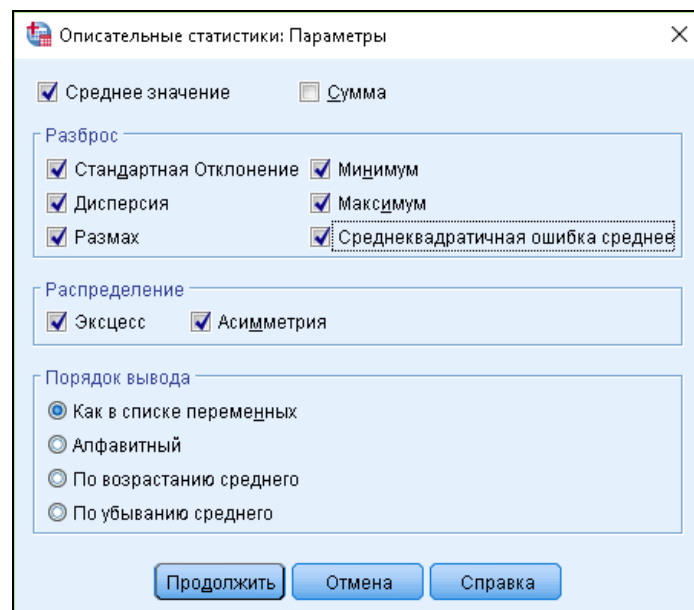


Рис. 150. **Встановлення параметрів *Описові статистики***

Натискаємо на кнопку *Продовжити*. У головному діалоговому вікні вкажіть, щоб стандартизовані значення були збережені в новій змінній відкритого файлу даних, для чого встановить прапорець *Save standardized values as variables* (рис. 151).

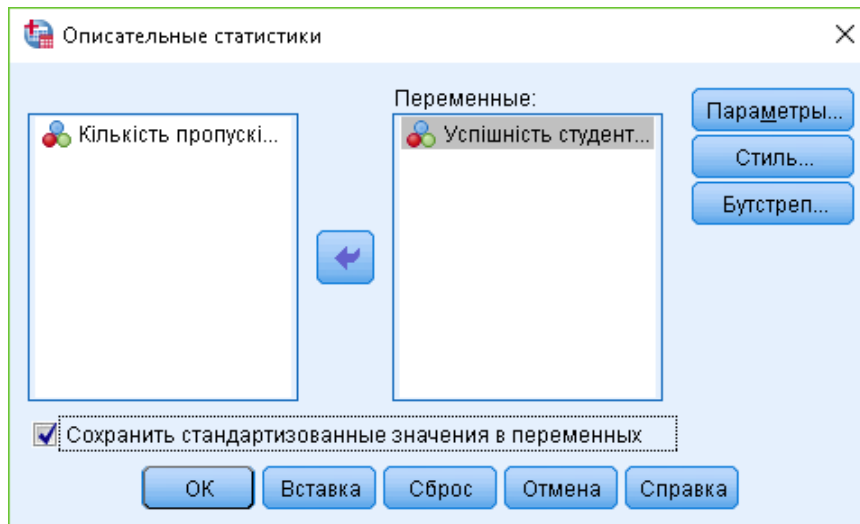


Рис. 151. Діалогове вікно *Описові статистики (Descriptives)*

Запускаємо обчислення, натиснувши на кнопку *OK*. Результат буде наведено у вікні перегляду (рис. 152).

Описательные статистики												
	N	Диапазон	Минимум	Максимум	Среднее		Средняя оплошание	Дисперсия	Асимметрия		Эксцесс	
					Статистика	Стандартная ошибка			Статистика	Стандартная ошибка	Статистика	Стандартная ошибка
Успішність студента з дисципліни	20	95	0	95	71,35	5,415	24,215	580,345	-2,194	,512	4,684	,992
N валидных (по списку)	20											

Рис. 152. Результат обчислення описових статистик

З рис. 152 видно, що мінімальна оцінка, отримана студентом – 0, а максимальна – 95. Середня оцінка по групі становить 71,35 бали. Загалом всі оцінки відхиляються від середнього в той або інший бік на 24,215 бали або на 586 %. Це свідчить про значну варіацію серед отриманих оцінок. Коефіцієнт асиметрії дорівнює $-2,194$, що говорить про значну правосторонню асиметрію. Коефіцієнт ексцесу свідчить про гостровершинність розподілу.

Щоб наочно побачити форму розподілу, необхідно побудувати гістограму з накладенням на неї кривої нормального розподілу (рис. 153).

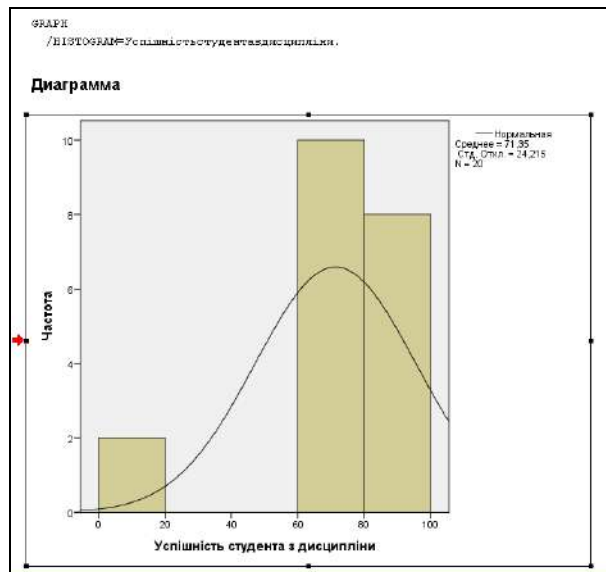
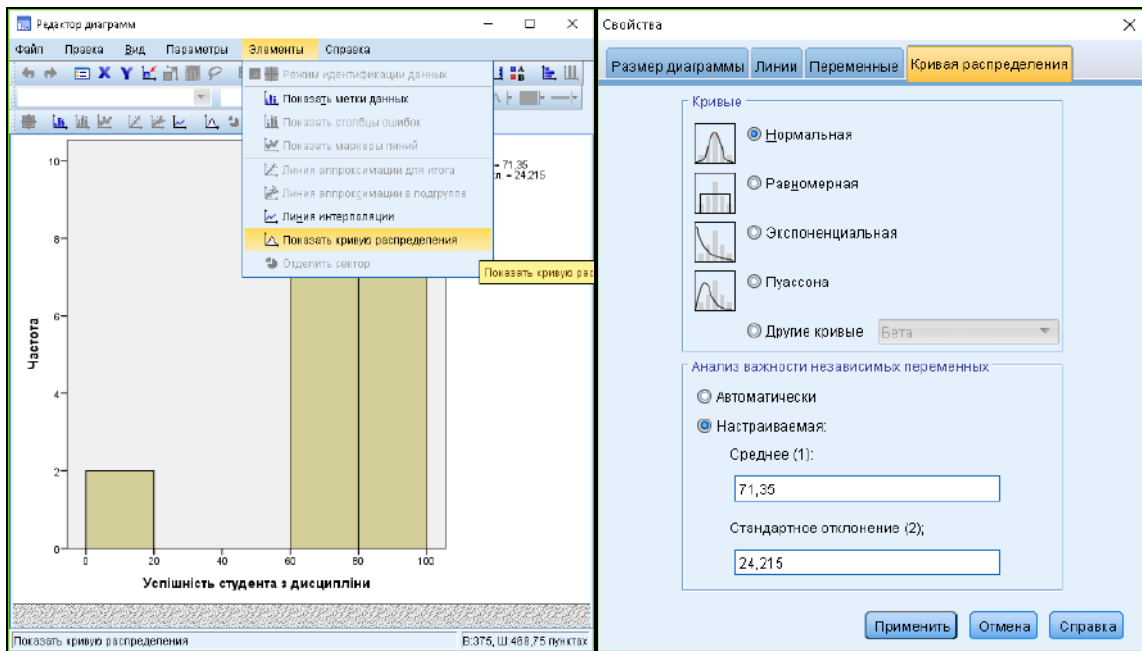
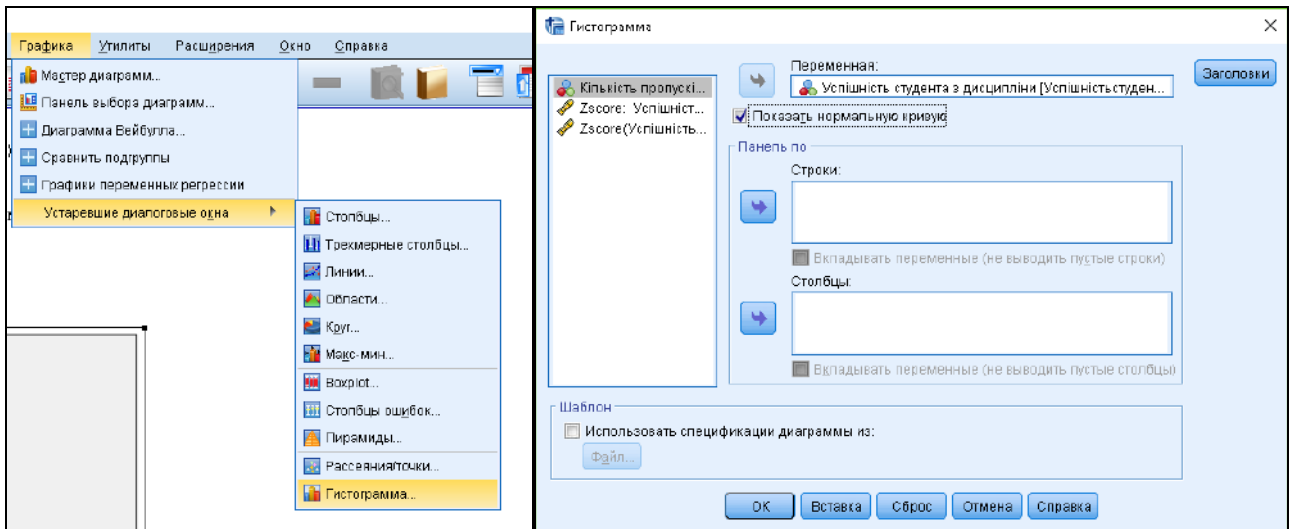


Рис. 153. Этапы построения гистограммы с кривой нормального распределения

4. Проведення дисперсійного аналізу в ППП SPSS.

Завдання: необхідно перевірити, чим зумовлена варіація даних – факторними чи випадковими ознаками за допомогою модулю *Analyze / General Linear Model / Variance Components* у ППП SPSS.

Розрахунок компонентів дисперсії в загальній лінійній моделі проводиться за наявності факторів з випадковими ефектами.

Загальна дисперсія складається з міжгрупової та середньої з групових дисперсій. Міжгрупова дисперсія характеризує варіацію, пов'язану з факторною ознакою, а середня з групових характеризує залишкову варіацію, а саме варіацію випадкових факторів.

Отже, за допомогою методу дисперсійного аналізу потрібно буде встановити чи залежить оцінка від відвідуваності або наскільки велика та частина дисперсії, причиною якої є неоднорідність оцінок. Відповіді на ці питання можна отримати за допомогою розрахунку компонентів дисперсії.

Проаналізуємо оцінки студентів, отримані за підсумками екзамену з дисципліни статистики відповідно до їх відвідувань занять протягом семестру. Для цього спочатку відкриваємо файл з даними. Потім вибираємо в меню *Analyze (Аналіз) General Linear Model (Загальна лінійна модель)* та команду *Variance Components (Компоненти дисперсії)*.

Відкриється діалогове вікно *Variance Components (Компоненти дисперсії)*. Переносимо змінну "Успішність студентів" у поле для залежної змінної, а "Кількість пропусків" у поле для незалежних (випадкових) факторів (рис. 154).

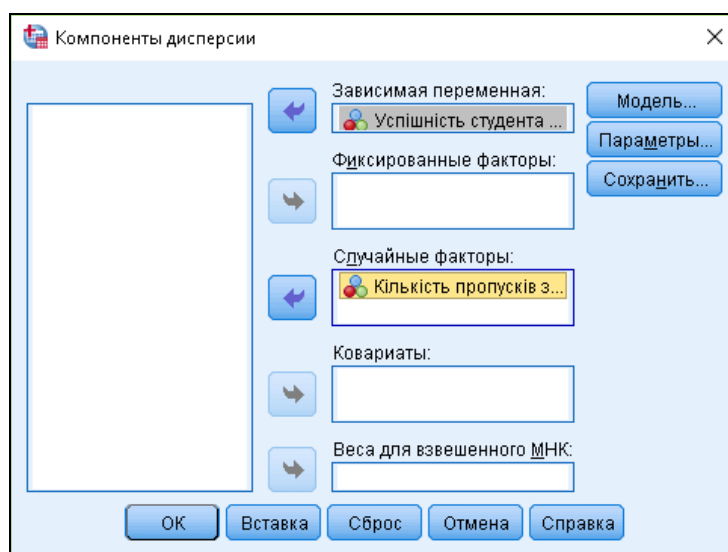


Рис. 154. Робота з модулем *Загальна лінійна модель (Компоненти дисперсії)*

Натиснувши на кнопку *Model...* (*Модель*), можна вибрати, чи буде розраховуватися повнофакторна модель (задано за замовчуванням) або будуть включені в розрахунок тільки деякі фактори. За наявності тільки одного фактора, як в наведеному прикладі, можна говорити тільки про повнофакторну модель (рис. 155).

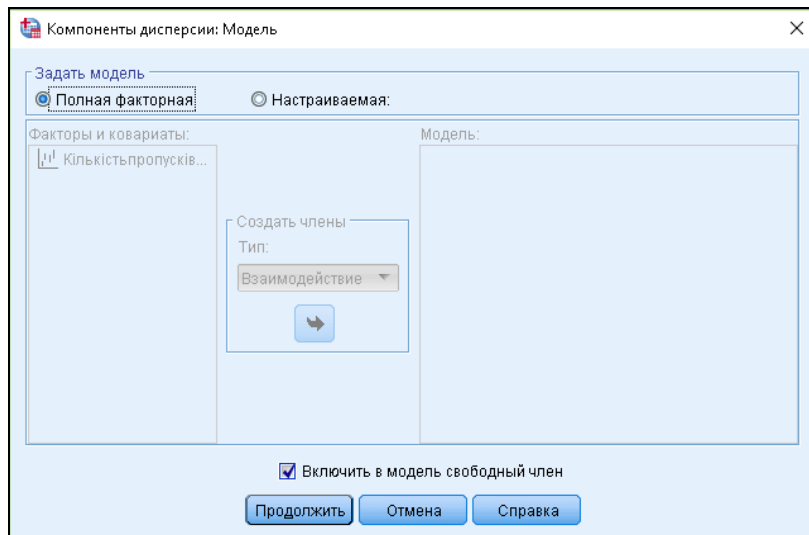


Рис. 155. Встановлення параметрів моделювання

Вибір опції *Параметры* надає можливість вибору між чотирма методами оцінювання компонентів дисперсії: MINQUE, ANOVA, максимальна подібність та обмежена максимальна подібність (рис. 156).

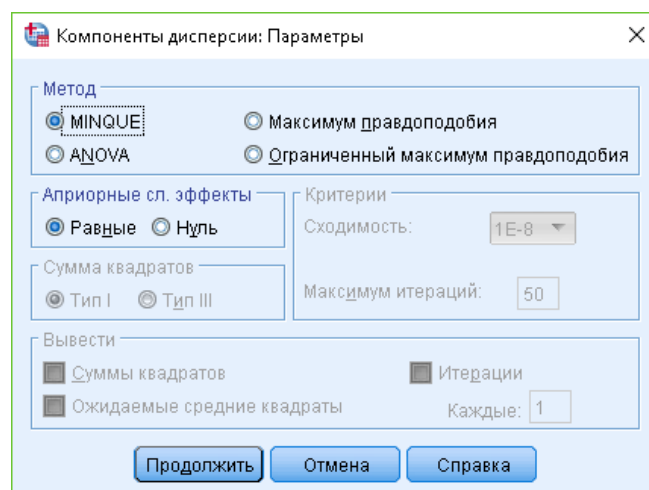


Рис. 156. Вибір методу оцінювання компонентів дисперсії

Перший підхід зводиться до розкладання відповідно до методу найменших квадратів (МНК); у однофакторном випадку сукупна дисперсія всіх спостережуваних значень розкладається на дисперсію всередині

окремих груп і дисперсію між групами. В основі узагальненої лінійної моделі, навпроти, лежить кореляційний або регресійний аналіз.

ANOVA використовує *Full rank linear model* (лінійна модель повного рангу). ANOVA також включає ряд додаткових видів аналізу (регресійний та канонічний аналіз, аналіз дискримінанта та головних компонентів і т. д.).

Виберемо метод *MINQUE* (*Minimum norm quadratic unbiased estimator*) (мінімум норми квадратичної незміщеної оцінки), тому що він найбільше підходить для аналізу рядів розподілу.

Наступним кроком є розрахунок дисперсій. У вікні перегляду з'являться оцінки компонентів дисперсії (рис. 157).

➔ **Оценка компонентов дисперсии**

Информация об уровнях факторов

	N	
Кількість пропусків за дисципліною	0	1
	1	1
	2	2
	3	1
	4	3
	5	1
	7	1
	8	3
	9	1
	10	1
	11	1
	15	1
	18	1
	20	1
	21	1

Зависимая переменная:
Успешность студента за дисциплины

Оценки дисперсии

Компонент	Оценка
Var (Кількість пропусків за дисципліною)	555,670
Дисп(Ошибка)	102,536

Зависимая переменная:
Успешность студента за дисциплины
Метод: Квадратичная несмещенная оценка с минимальной нормой
(Вес = 1 для случайных эффектов и для остатка)

Рис. 157. Результати оцінювання дисперсії

Відповідно до вікна результатів (див. рис. 157), видно, що міжгрупова дисперсія (Var (кількість пропусків)) дорівнює 555,67, а середня з групових – 102,536. На основі цих результатів можна знайти відсоткову частку дисперсії, зумовлену наявністю різних оцінок:

$$555,67 / (555,67 + 102,536) = 0,8442 \text{ або } 84,42 \%$$

Отже, можна зробити висновок, що відмінності в оцінках на 84,42 % обумовлено відвідуваністю і на 25,58 % іншими факторами.

5. Побудова багатofакторної регресійної моделі в SPSS.

Завдання: використовуючи дані, наведені на рис. 158, перевірити наявність лінійного зв'язку між відповідними показниками в модулі *Analyze / Regression / Linear*.

	Успішність_студента_з_дисципліни	Кількість_пропусків_за_дисципліно	Стать_0ч_1ж	Кількість_відвідувань_бібліотеки
1	74	18	0	2
2	81	4	0	5
3	90	5	1	10
4	65	11	0	2
5	85	2	1	1
6	12	21	0	0
7	80	7	1	3
8	91	4	1	7
9	68	10	0	0
10	78	3	1	1
11	78	4	0	0
12	75	8	0	3
13	89	0	1	4
14	95	2	1	9
15	74	9	0	0
16	60	15	0	0
17	76	8	0	2
18	0	20	0	0
19	69	8	1	0
20	87	1	0	1

Рис. 158. Вихідні дані для проведення регресійного аналізу

Регресійний аналіз – це метод моделювання даних, що використовується для прогнозу, аналізу часових рядів, тестування гіпотез та виявлення прихованих взаємозв'язків між значеннями однієї (залежної) змінної та значеннями інших (незалежних) змінних.

У разі множинного регресійного аналізу необхідно оцінити коефіцієнти рівняння:

$$y = b_1x_1 + b_2x_2 + \dots + b_nx_n + a,$$

де n – кількість незалежних змінних, позначених як x_1 та x_n ;

y – залежна змінна;

a – деяка константа.

Незалежні змінні можуть корелювати між собою, тобто між ними може існувати мультиколінійність. Цей факт необхідно обов'язково враховувати під час визначення коефіцієнтів рівняння регресії для того, щоб уникнути помилкових кореляцій. Щоб викликати регресійний аналіз у SPSS, виберіть в меню *Analyze... (Аналіз) / Regression... (Регресія)*.

Для проведення лінійного регресійного аналізу залежна змінна повинна мати інтервальну (або порядкову) шкалу. Відповідно визначаємо залежну та незалежні змінні (рис. 159).

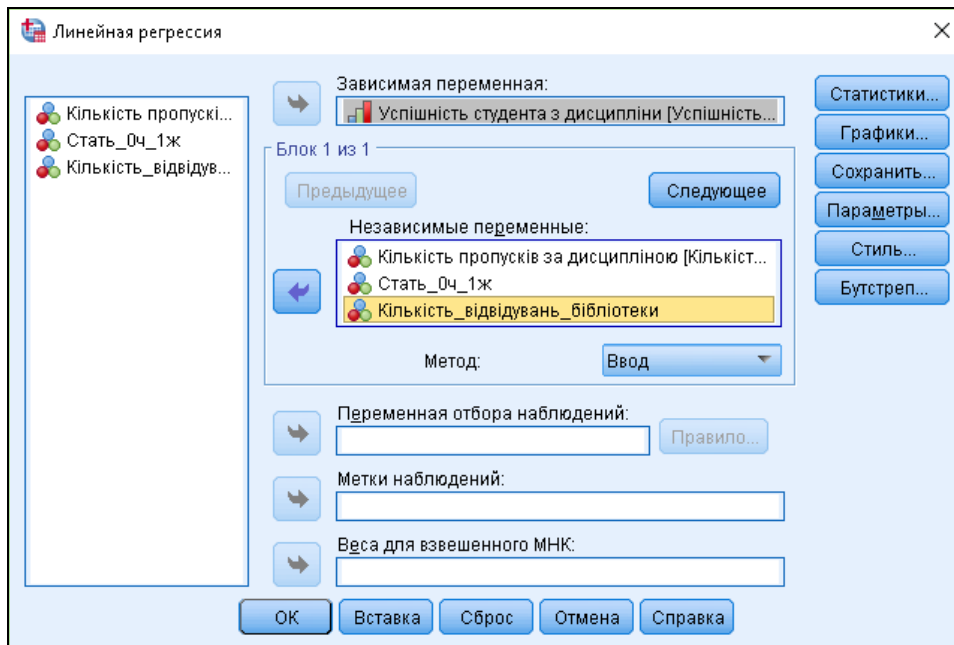


Рис. 159. Робота з модулем *Лінійна регресія*

Для множинного аналізу з декількома незалежними змінними не рекомендується залишати метод включення всіх змінних, встановлений за замовчуванням. Цей метод відповідає одночасному обробленню всіх незалежних змінних, вибраних для аналізу, і тому він може рекомендуватися для використання тільки у разі простого аналізу з однією незалежною змінною.

Для множинного аналізу слід обрати один з покрокових методів. За прямого методу незалежні змінні, які мають найбільші коефіцієнти часткової кореляції із залежною змінною покроково включаються в регресійне рівняння. За зворотного методу починають з результату, що містить всі незалежні змінні, а потім виключають незалежні змінні з найменшими частковими кореляційними коефіцієнтами, поки відповідний регресійний коефіцієнт не виявляється незначущим.

Найбільш поширеним є покроковий метод, який влаштований так само, як і прямий метод, проте після кожного кроку змінні, що використовувалися в заданий момент, досліджуються за зворотним методом. За покрокового методу можуть задаватися блоки незалежних змінних. І цьому випадку задані блоки на одному кроці обробляються спільно.

Виберіть покроковий метод, але утримайтеся від блокової форми введення даних, задавайте розрахунок усіх додаткових функцій і почніть обчислення натисненням кнопки ОК (рис. 160).

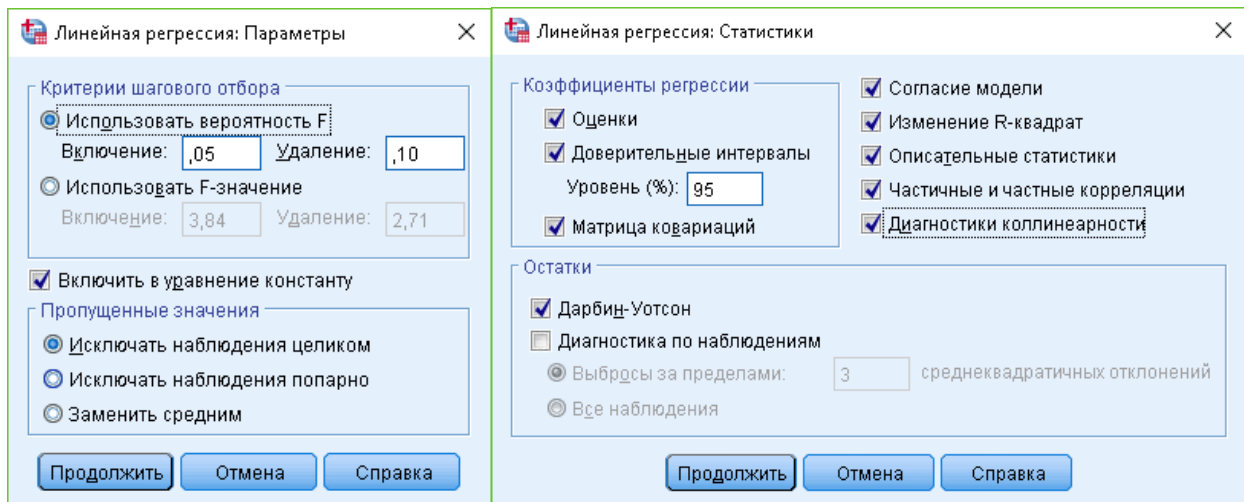


Рис. 160. Встановлення параметрів *Лінійна регресія*

У вкладці *Графіки* задайте параметри **ZRESID* і **ZPRED* як змінні, що відображаються по осях у і х відповідно. У разі лінійної регресії залишки розподіляються випадково по обидві сторони від горизонтальної нульової лінії (рис. 161).

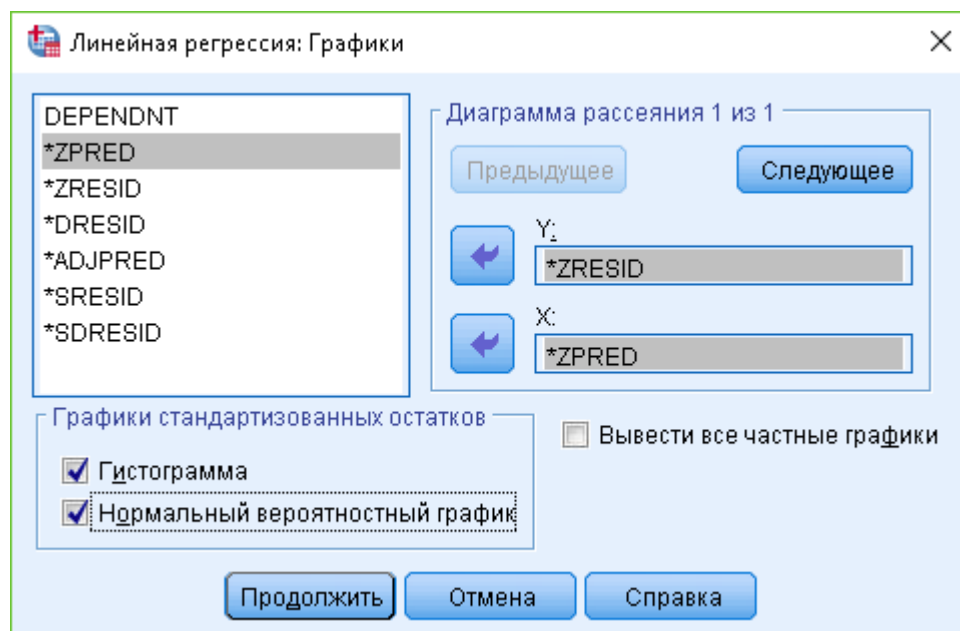


Рис. 161. Встановлення параметрів для функції *Графіки*

У такий спосіб отримуємо звіт за результатами проведеного регресійного аналізу (рис. 162).

Описательные статистики

	Среднее	Среднекв. отклонение	N
Успішність студента з дисципліни	71,35	24,215	20
Кількість пропусків за дисципліною	8,00	6,258	20
Стать_0ч_1ж	,40	,503	20
Кількість_відвідувань_бібліотеки	2,50	3,069	20

Корреляции

		Успішність студента з дисципліни	Кількість пропусків за дисципліною	Стать_0ч_1ж	Кількість_відвідувань_бібліотеки
Корреляция Пирсона	Успішність студента з дисципліни	1,000	-,841	,459	,517
	Кількість пропусків за дисципліною	-,841	1,000	-,552	-,436
	Стать_0ч_1ж	,459	-,552	1,000	,512
	Кількість_відвідувань_бібліотеки	,517	-,436	,512	1,000
Знач. (односторонняя)	Успішність студента з дисципліни	.	,000	,021	,010
	Кількість пропусків за дисципліною	,000	.	,006	,027
	Стать_0ч_1ж	,021	,006	.	,011
	Кількість_відвідувань_бібліотеки	,010	,027	,011	.
N	Успішність студента з дисципліни	20	20	20	20
	Кількість пропусків за дисципліною	20	20	20	20
	Стать_0ч_1ж	20	20	20	20
	Кількість_відвідувань_бібліотеки	20	20	20	20

Введенные/удаленные переменные^a

Модель	Введенные переменные	Удаленные переменные	Метод
1	Кількість_відвідувань_бібліотеки, Кількість пропусків за дисципліною, Стать_0ч_1ж ^b	.	Enter

a. Зависимая переменная: Успішність студента з дисципліни

b. Все требуемые переменные введены.

Сводка для модели^b

Модель	R	R-квадрат	Скорректированный R-квадрат	Стандартная ошибка оценки	Изменение R квадрат	Статистика изменений			Знач. Изменение F	Дарбин-Уотсон
						Изменение F	ст.св.1	ст.св.2		
1	,860 ^a	,740	,691	13,459	,740	15,166	3	16	,000	1,996

a. Предикторы: (константа), Кількість_відвідувань_бібліотеки, Кількість пропусків за дисципліною, Стать_0ч_1ж

b. Зависимая переменная: Успішність студента з дисципліни

ANOVA^a

Модель		Сумма квадратов	ст.св.	Средний квадрат	F	Значимость
1	Регрессия	8242,043	3	2747,348	15,166	,000 ^b
	Остаток	2898,507	16	181,157		
	Всего	11140,550	19			

a. Зависимая переменная: Успішність студента з дисципліни

b. Предикторы: (константа), Кількість_відвідувань_бібліотеки, Кількість пропусків за дисципліною, Стать_0ч_1ж

Рис. 162. Результаты регрессийного анализа

Коефіцієнти ^а													
Модель	В	Нестандартизовані коефіцієнти		Стандартно виміряні коефіцієнти	t	Значимість	95,0% Довірительний інтервал для В		Питового порядку	Кореляції		Статистика колінеарності	
		Стандартна помилка	Beta				Нижня границя	Верхня границя		Частично	Компонент	Допуск	VIF
1	(Константа)	93,497	7,998		11,693	,000	76,547	110,448					
	Кількість пропусків за дисципліною	-3,063	,808	-,797	-5,088	,000	-4,367	-1,798	-,841	-,766	-,643	,663	1,509
	Стать_0ч_1ж	-4,423	7,903	-,092	-,560	,583	-21,176	12,330	,469	-,139	-,071	,604	1,655
	Кількість відвідувань бібліотеки	1,713	1,199	,217	1,429	,172	-,828	4,254	,517	,337	,182	,704	1,420

а. Залежна змінна: Успішність студента з дисципліни

Кореляції коефіцієнтів ^а					
Модель		Кількість відвідувань бібліотеки	Кількість пропусків за дисципліною	Стать_0ч_1ж	
1	Кореляції	Кількість відвідувань бібліотеки	1,000	-,361	
		Кількість пропусків за дисципліною	,214	1,000	
		Стать_0ч_1ж	-,361	,426	1,000
	Кореляції	Кількість відвідувань бібліотеки	1,437	-3,423	
		Кількість пропусків за дисципліною	,155	,367	2,039
		Стать_0ч_1ж	-3,423	2,039	62,462

а. Залежна змінна: Успішність студента з дисципліни

Діагностика колінеарності ^а							
Модель	Измерение	Собственное значение	Показатель обусловленности	(Константа)	Доли дисперсии		
					Кількість пропусків за дисципліною	Стать_0ч_1ж	Кількість відвідувань бібліотеки
1	1	2,691	1,000	,02	,02	,03	,04
	2	,931	1,700	,01	,13	,10	,08
	3	,290	3,045	,00	,00	,55	,79
	4	,087	5,548	,97	,85	,32	,09

а. Залежна змінна: Успішність студента з дисципліни

Статистика остатков ^а					
	Минимум	Максимум	Среднее	Среднекв. отклонение	N
Предсказанное значение	28,76	98,33	71,35	20,828	20
Остаток	-31,843	32,565	,000	12,351	20
Стандартное предсказанное значение	-2,045	1,295	,000	1,000	20
Стандартный остаток	-2,366	2,419	,000	,918	20

а. Залежна змінна: Успішність студента з дисципліни

Закінчення рис. 162.

З таблиці *Кореляція* (див. рис. 162) видно, що між залежною змінною та незалежними існує зв'язок. Так, між успішністю навчання та кількістю пропусків він є сильним зворотнім (-0,841), а між успішністю студентів та статтю й кількістю відвідувань бібліотеки є помірний прямий зв'язок з (0,46 та 0,52 відповідно).

Розглянемо спочатку частину результатів розрахунку блоку *Коефіцієнти*. Тут виводяться коефіцієнт регресії b і зсув по осі ординат а під ім'ям "константа". Тобто, рівняння регресії виглядає так:

$$\text{Успішність студента} = 93,497 - 3,083 \times \text{Кількість пропусків} - 4,423 \times \text{Стать} + 1,713 \times \text{Кількість відвідувань бібліотеки}.$$

Частина розрахунків ANOVA відображає два джерела дисперсії: дисперсію, яка описується рівнянням регресії (сума квадратів, обумовлена регресією) і дисперсію, яка не враховується при записі рівняння (залишкова сума квадратів). Частка від ділення суми квадратів, обумовлених регресією і залишкової суми квадратів називається "коефіцієнтом детермінації". У таблиці результатів ця частка виводиться під ім'ям "R-квадрат". У нашому прикладі міра визначеності дорівнює:

$$8\,242,043 / 11\,140,550 = 0,739.$$

Ця величина характеризує якість регресійної прямої, тобто ступінь відповідності між регресійною моделлю і початковими даними. Міра визначеності завжди лежить у діапазоні від 0 до 1. У нашому випадку 73,9 % зміни успішності студентів пояснюється впливом незалежних факторів, що увійшли у модель.

Існування ненульових коефіцієнтів регресії перевіряється за допомогою обчислення контрольної величини F, до якої відноситься відповідний рівень значущості.

Перевірити адекватність моделі значить встановити, наскільки добре модель описує реальні процеси, наскільки якісно властивості моделі (функції, параметри, характеристики) відповідають властивостям модельованого об'єкта.

Адекватність моделі та параметрів аналізують за допомогою коефіцієнтів кореляції, детермінації, критерію Фішера та критерію Стьюдента.

Відповідно до отриманих результатів, коефіцієнт кореляції дорівнює 0,86, що свідчить про наявність сильного прямого зв'язку між досліджуваними змінними, а отже, підтверджує адекватність моделі.

Коефіцієнт детермінації дорівнює 0,74, тобто варіація У зумовлюється варіацією x_1 , x_2 , x_3 на 74 %, що також підтверджує адекватність моделі.

Якщо фактичне значення критерію Фішера (F-критерію) вище критичного (табличного), то модель адекватна, а її фактори залишаються у рівнянні. У нашому випадку табличне значення F-критерію для рівня значущості $p = 0,05$, f_1 – кількість ступенів свободи більшої дисперсії (20), f_2 – кількість ступенів свободи меншої дисперсії (3) дорівнює 3,10. Отже, $F_{\text{розр.}}(15,166) > F_{\text{табл.}}(3,10)$, тобто модель адекватна реальній дійсності та може використовуватися для подальших розрахунків (побудови прогнозу тощо).

Коефіцієнти рівняння β – це регресійні коефіцієнти, стандартизовані відповідно до області значень. Вони вказують на важливість незалежних змінних, включених у регресійне рівняння. Для множинної регресії значущість параметрів b_i перевіряється за допомогою $t_{\text{стат. Стьюдента}}$ ($t_{\text{розр.}} > t_{\text{крит.}}$). Згідно з табличними значеннями, за $df = 20 - 3 - 1 = 16$ і $p = 0,95$ знаходимо, що $t_{\text{крит.}} = 2,120$.

У нашому випадку є всі підстави вважати, що тільки b_1 (кількість пропусків) є значущими. Тоді економетрична модель буде мати остаточний вигляд:

$$\text{Успішність студента} = 93,497 - 3,083 \times \text{Кількість пропусків.}$$

Важливим моментом є аналіз залишків, тобто відхилень спостережуваних значень від теоретично очікуваних. Залишки мають з'являтися випадково (тобто не систематично) і підкорятися нормальному розподілу. Це можна перевірити за допомогою кнопки *Charts... (Діаграми)*, побудувати гістограму залишків. У наведеному прикладі спостерігається досить хороше узгодження гістограми залишків з нормальним розподілом.

Оскільки одна з основних гіпотез щодо випадкової величини говорить, що помилки мають бути розподілені за нормальним законом, представимо гістограму розподілу помилок моделі (рис. 163).

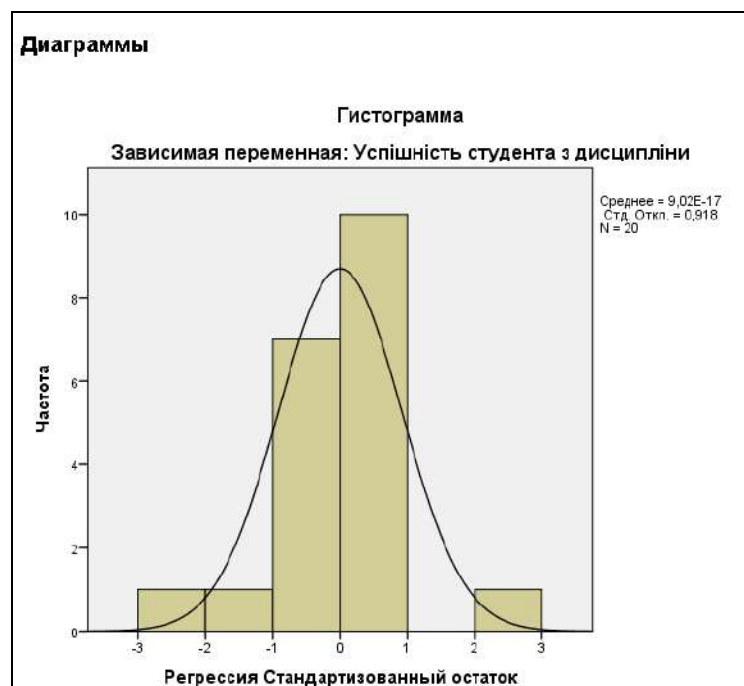


Рис. 163. Результати регресійного аналізу

Остання частина розрахунків характеризує аналіз помилок. У таблиці відображено теоретичні значення залежної змінної (*Predicted Value*) і помилки моделі (*Residual*) та їхні стандартизовані значення (див. рис. 162).

Графік розподілу залишків на нормальному імовірнісному папері наведений на рис. 164.

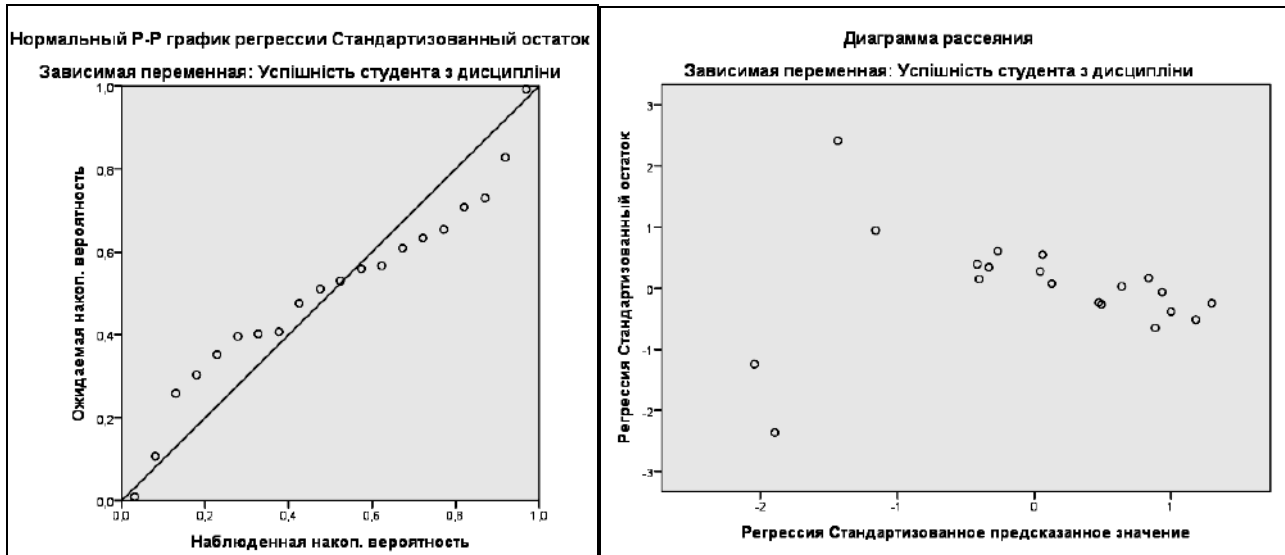


Рис. 164. Графік розподілу залишків на нормальному імовірнісному папері

Перевірка на наявність систематичних зв'язків між залишками сусідніх випадків може бути проведена за допомогою тесту Дарбіна – Уотсона (Durbin – Watson) на автокореляцію. Цей тест обчислює коефіцієнт, що лежать у діапазоні від 0 до 4. Якщо значення цього коефіцієнта знаходиться поблизу 3,35, то це означає, що автокореляція відсутня. Тест Дарбіна – Уотсона можна активувати через кнопку *Statistics (Статистичні характеристики)*. Порівнюють отримані d-статистики з табличними d-статистиками за рівнем значущості $\alpha = 0,05$, кількості факторів k , що присутні в моделі, і кількості спостережень n . Якщо розраховане значення d-статистики знаходиться в проміжку від 0 до dL ($0 < d < dL$), то це свідчить про наявність позитивної автокореляції. Якщо значення d потрапляє в зону невизначеності, тобто набуває значення $dL < d < dU$, або $4 - dU < d < 4 - dL$, то можемо зробити висновки ні про наявність, ні про відсутність автокореляції. Якщо $4 - dL < d < 4$, то маємо негативну автокореляцію. Нарешті, якщо $dU < d < 4 - dU$, то автокореляції немає.

Отже, за умови, що n (обсяг вибірки) = 20, m (кількість незалежних змінних) = 3, тоді табличні значення $dL = 1$ та $dU = 1,68$.

У заданому прикладі тест дає задовільне значення коефіцієнта, рівне 1,996, що свідчить про потрапляння у зону невизначеності, тобто зробити остаточний висновок про наявність автокореляції неможливо.

Загалом, було проведено регресійний аналіз, побудовано багатофакторну економетричну модель, що демонструє залежність успішності студентів від кількості пропусків занять.

Завдання. За власною тематикою дослідження сформууйте інформаційний простір дослідження, проведіть розрахунок характеристик варіаційних рядів і побудуйте багатофакторну регресійну модель у ППП SPSS та перевірте її на адекватність.

Результати подайте у вигляді звіту з лабораторної роботи.

Завдання для самостійного опрацювання

Запитання, що виносяться на самостійну роботу студентів:

1. Проведіть порівняльний аналіз спеціалізованих статистичних програмних продуктів для оброблення великих масивів даних: SPSS, EViews і Statistica.

2. Охарактеризуйте спеціалізований пакет EViews. З яких модулів він складається? Їхнє функціональне призначення.

3. Надайте характеристику ППП Statistica. Які функціональні, аналітичні та графічні можливості Statistica від відомі?

4. Наведіть приклади вирішення економічних завдань за допомогою аналітичних можливостей EViews та Statistica.

5. Розкрийте особливості практичного застосування ППП SPSS у соціології, маркетингових дослідження та економіці.

Завдання 1. Використовуючи власний інформаційний простір та аналітичні можливості ППП SPSS:

1) розрахуйте описові статистики для вашого ряду даних;

2) перевірте ряд даних на нормальний закон розподілу, зокрема наведіть графічне підтвердження.

Завдання 2. Використовуючи власний інформаційний простір і графічні можливості ППП SPSS, побудуйте 3 різні типи діаграм (графіків). Вибір типу інфографіки обґрунтуйте.

Тема 7. Пакет R для аналізу великих масивів даних

Лабораторна робота 7

Аналіз великих масивів даних за допомогою пакету R

Мета: надбання навичок статистичного оброблення даних і роботи з графікою за допомогою мови програмування R.

Завдання: ознайомитися з інтерфейсом і функціональними можливостями R, навчитися обробляти, аналізувати та візуалізувати різні типи даних за допомогою пакету R.

Методичні рекомендації

1. Знайомство з програмою R та RStudio.

R – мова програмування для статистичного оброблення, аналізу даних і роботи з графікою, а також вільне програмне середовище обчислень з відкритим вихідним кодом у рамках проєкту GNU. R підтримує широкий спектр статистичних і чисельних методів і потужні додаткові функціональні й аналітичні можливості (вбудована бібліотека пакетів). Пакети – це бібліотеки для роботи специфічних функцій або спеціальних областей застосування. У базову комплектацію R включений основний набір пакетів. Усього станом на 2019 рік є більш 12 000 пакетів.

R широко використовується в соціальних науках, статистиці, економіці, страховій справі, соціології, фінансах, фізиці тощо.

R доступний для всіх операційних систем, включно з Linux, Mac OS, Windows.

R – матрична, об'єктно-орієнтована мова програмування. Це означає, що теоретично все що завгодно може бути збережено як об'єкт R. Кожен об'єкт має свій клас, який описує, що містить цей об'єкт і що кожна функція може з цими даними робити. Наприклад, `plot(x)` виводить один результат, якщо `x` є регресією, і інший, якщо вектором.

Пакет R можна завантажити та встановити абсолютно вільно. Для цього необхідно відвідати сайт CRAN (<https://cran.r-project.org>) та завантажити інсталяційний пакет. Після його запуску необхідно обрати відповідні параметри встановлення (мову, компоненти пакету та інші налаштування).

Після запуску програми відкривається діалогове вікно – вікно консоль, де відображаються всі команди та результати їхнього виконання (рис. 165).

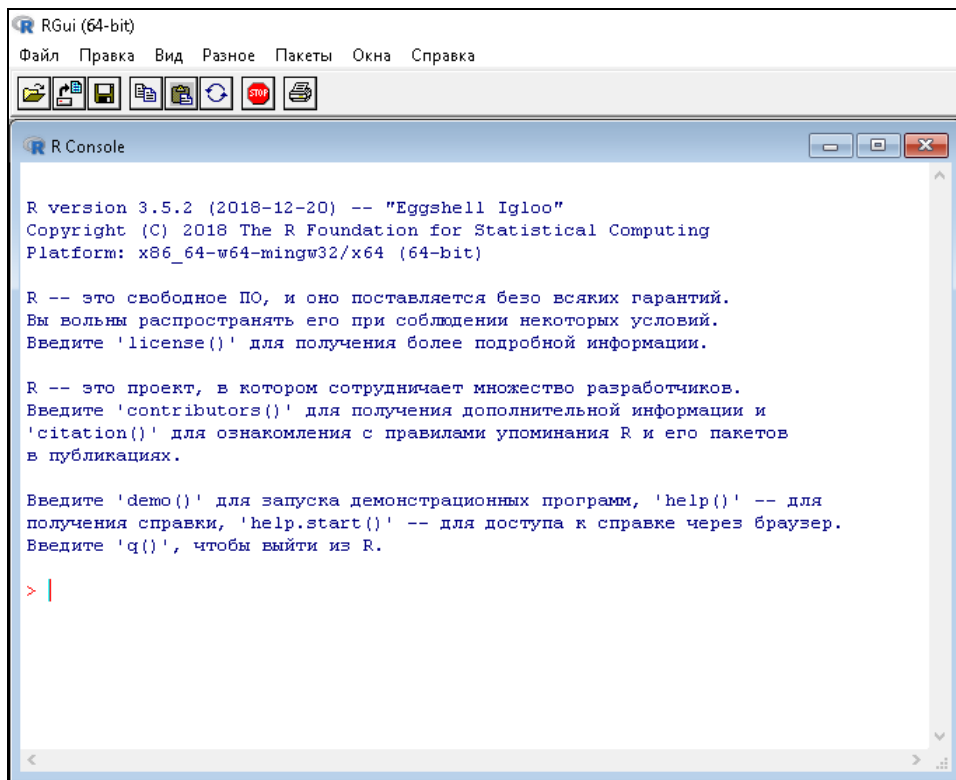


Рис. 165. Діалогове вікно (консоль) запуску пакету R

Для того щоб користувачу спростити роботу з пакетом R, було створено програму-інтерфейс під назвою RStudio. Зазначимо, що робота в пакеті R та RStudio є майже тотожною. Так у RStudio наведено безліч зручних інструментів, до того ж вона має приємний і більш зрозумілий інтерфейс. Для її завантаження необхідно спочатку встановити пакет R, а потім завантажити відповідний файл: <https://www.rstudio.com/products/rstudio/download/#download>.

Після запуску програми, відкриється вікно, що складається з 4 блоків (рис. 166).

Середовище RStudio складається з вікон:

- 1) скрипта (для написання коду та підготовки його до запуску);
- 2) консолі (відображаються всі команди та результати їхнього виконання);
- 3) робочої області (Workspace (Environment)) – для відображення всіх об'єктів та їхнього опису. На вкладці *History* можна побачити історію тих команд, що вводив користувач;
- 4) файлів (Files – відображається робоча директорія – папка), *Plots* (для візуалізації всіх графіків), *Packages* (наведено всі пакети-бібліотеки, що встановлені), *Help* (для отримання довідкової інформації).

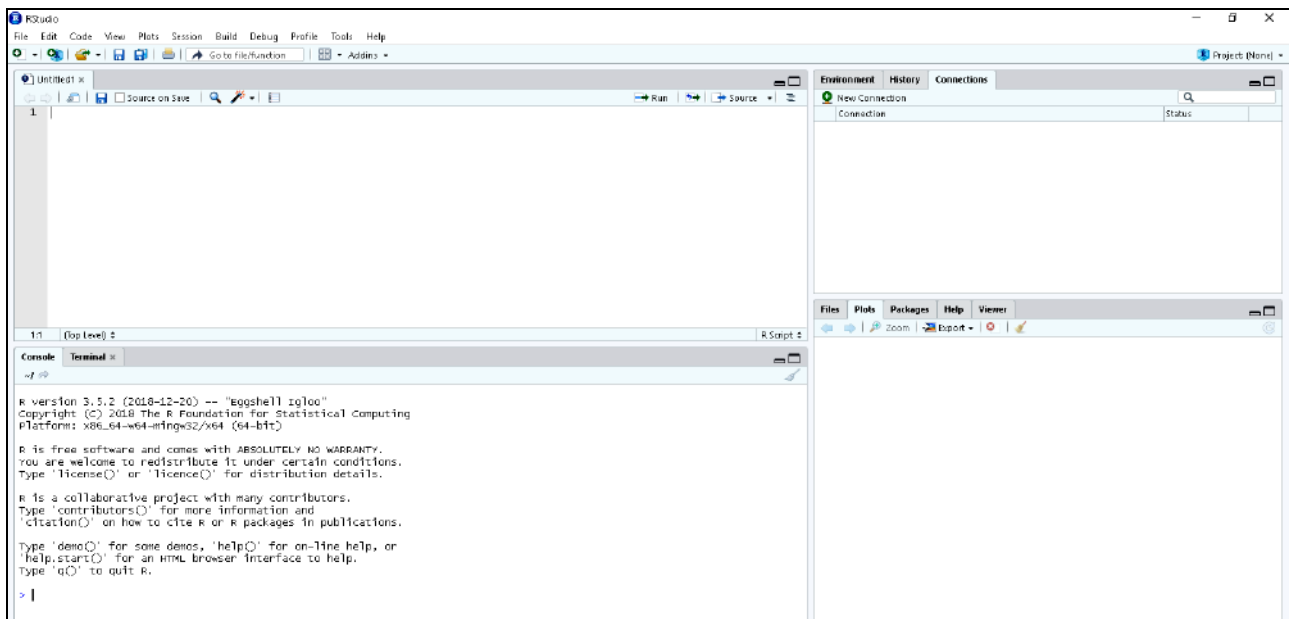


Рис. 166. Діалогове вікно запуску RStudio

Наприклад, використовуючи пакет R як звичайний калькулятор, пропишемо певні математичні дії в вікні скрипта та натиснемо кнопку *Run* або скористаємося комбінацію кнопок *Cntr + Enter* (рис. 167).

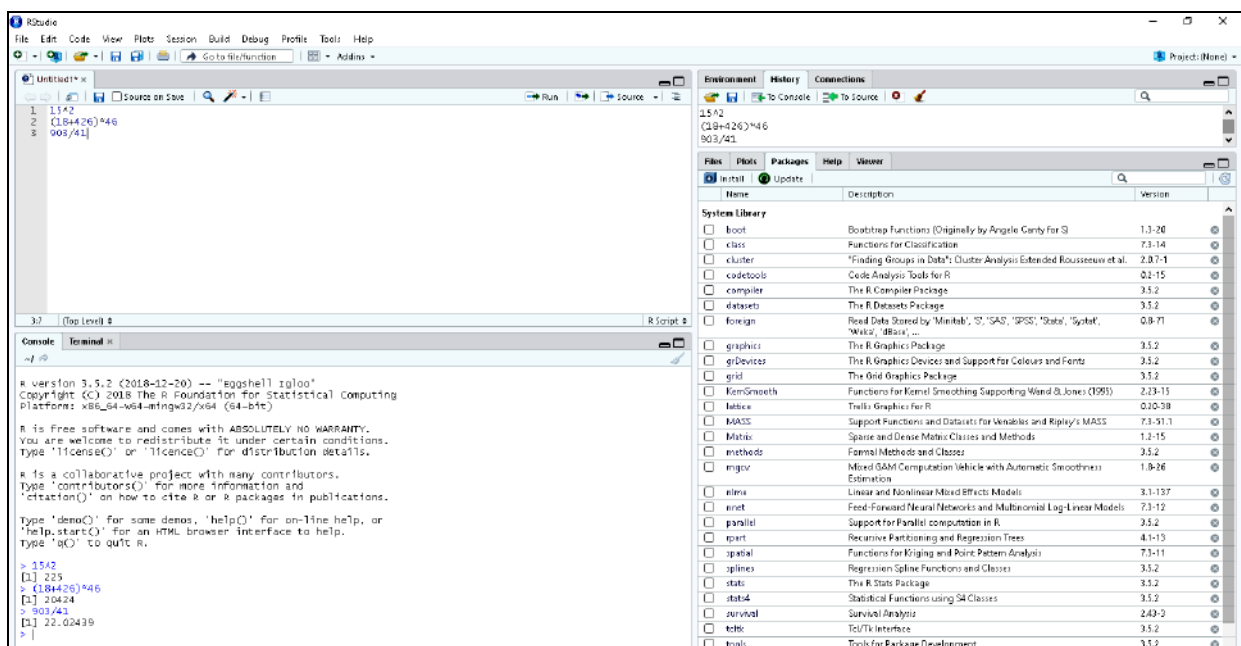


Рис. 167. Виконання простих обчислень в RStudio

Після виконання коду результат розрахунків з'явився у вікні консолі, а у вкладці History одночасно відобразилися всі ті команди, що вводив користувач.

У відповідний спосіб вводяться всі функції.

Якщо ви не знаєте, що саме за функція перед вам, вам доцільно викликати довідку. Для цього у вікні скрипта введіть перед назвою функції знак питання та натисніть кнопку *Run*. Вся необхідна довідкова інформація відобразиться у вікні (вкладці) *Help* (рис. 168).

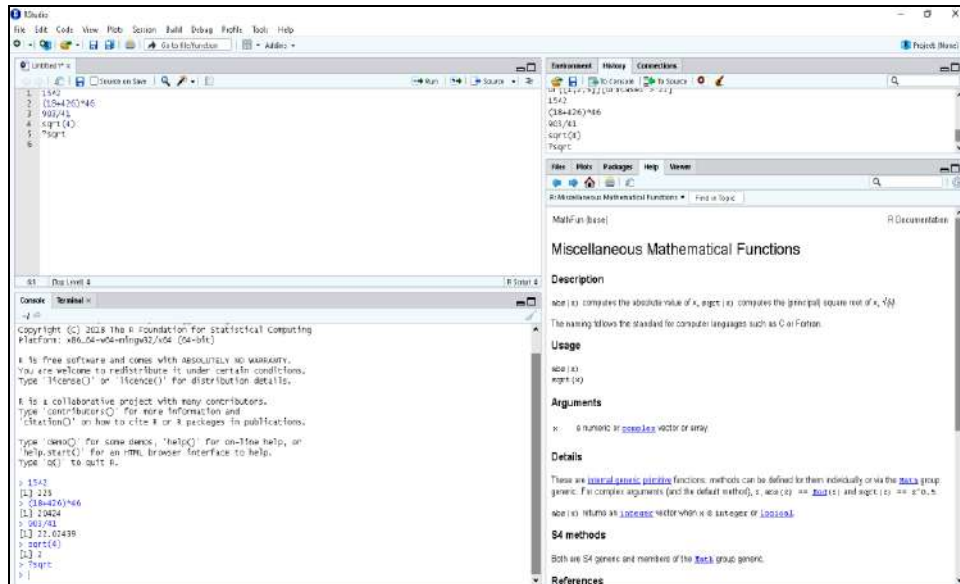


Рис. 168. Запуск довідки в RStudio

Користувач може створювати змінні. Для цього вводимо у вікні консолі позначення змінної, потім знак присвоєння (<-) та значення, якому вона відповідає. Наприклад, $y = 100$ (рис. 169).

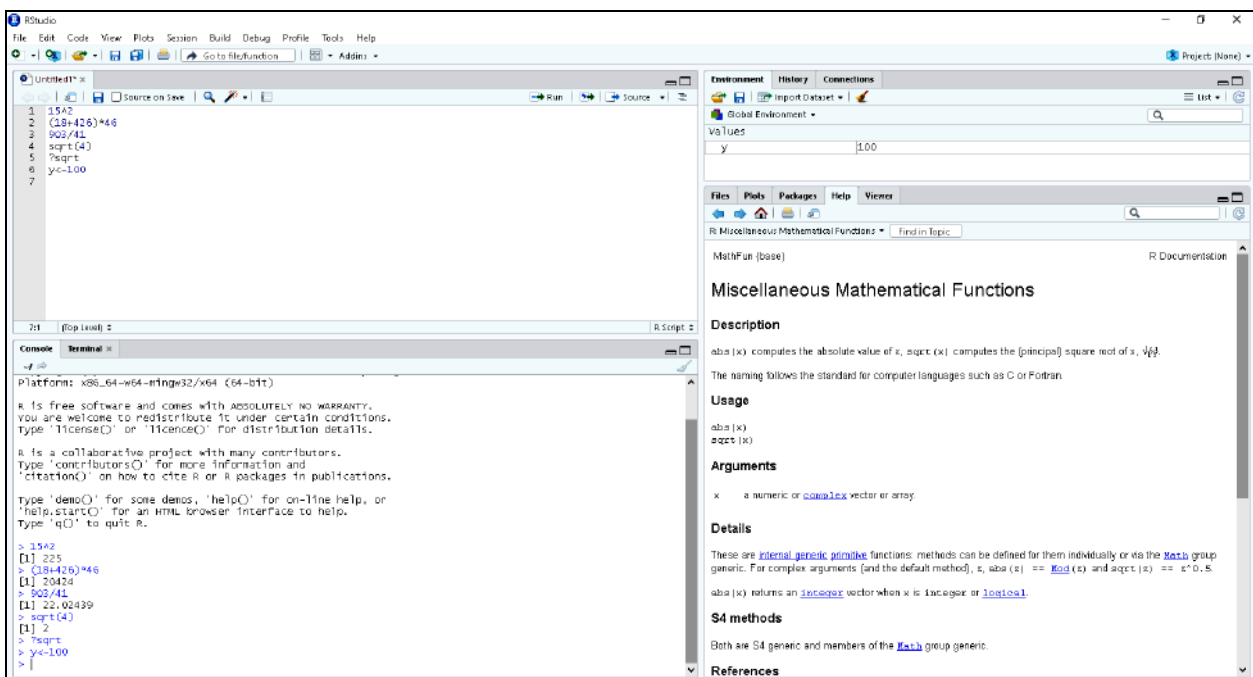


Рис. 169. Створення змінної в RStudio

У вікні *Environment* відобразився результат створення змінної. Також можна проводити розрахунки (y^2) та створювати вирази ($x = y^2$) (рис. 170).

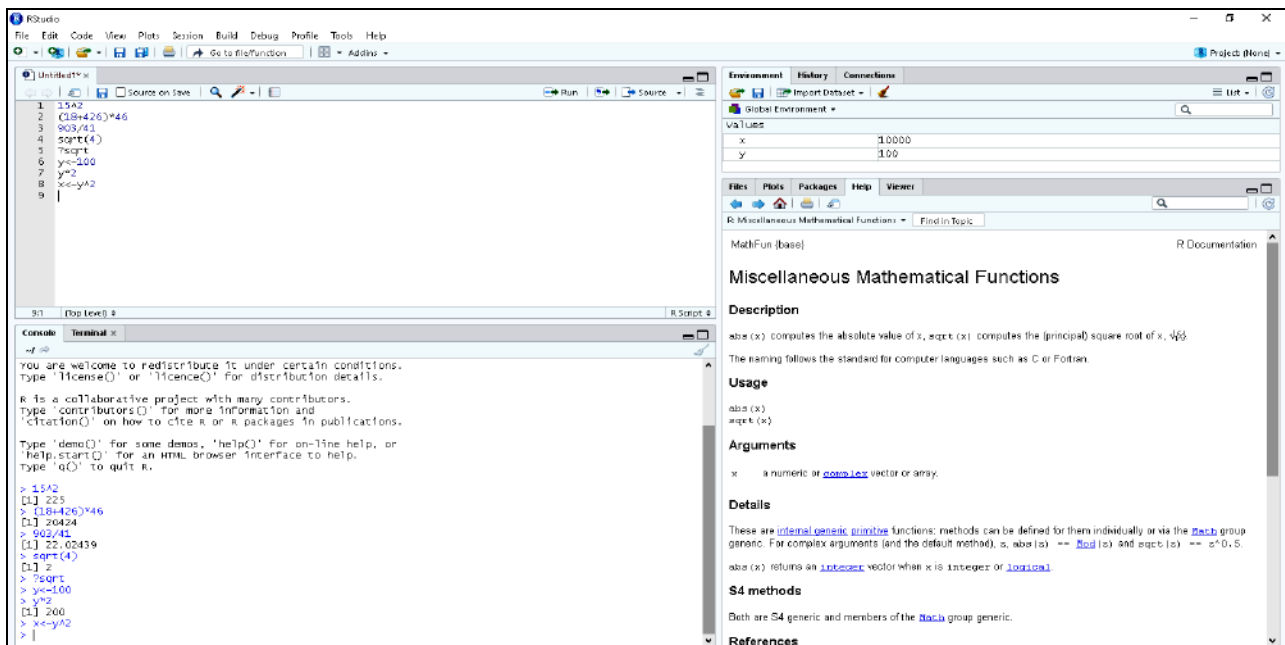


Рис. 170. Прості розрахунки в RStudio

У R будь-яка команда – це функція, яку можна передати аргумент. Функції можна просто комбінувати.

Символом присвоєння є " $<-$ ". Також можливо використовувати класичний " $=$ ". Тобто два наступних вираження є еквівалентними:

```
> a <- 2;
```

```
> a = 2.
```

Аргументи передаються в круглих дужках.

Зазвичай, краще використовувати лапки для імен, але це не завжди необхідно.

"#" використовується для коментарів.

Команди відокремлюються крапкою з комою ";" або символом переведення каретки. Якщо ви хочете розмістити в одному рядку більше одного виразу, то необхідно використовувати роздільник ";".

R чутливий до регістру: "a" і "A" є двома різними об'єктами, тому введення всіх функцій та аргументів потрібно роботи лише з малої літери.

Традиційно символ підкреслення "_" не використовується в іменах. Найчастіше краще використовувати символ точки ".". Слід уникати використання символу підкреслення як першого символу в імені об'єкта.

У R існують спецсимволи:

NA: Not Available (тобто пропущені значення);

NaN: Not a Number ("не число", наприклад, невизначеність 0/0);

Inf: (Нескінченність);

-Inf: (Мінус нескінченність).

З R можна вийти, використовуючи `q ()`. Аргумент `no` позначає, що сесію зберігати не потрібно.

Пакет R працює з такими *типами даних*: логічні (`logical`), цілочисельні (`integer`); дійсні числа (`numeric`), комплексні числа (`complex`), текстові (`character`) та списки (`list`).

Аналітичні можливості пакету R обумовлені можливістю роботи з різними об'єктами, зокрема:

1. `Vectors` (вектори) є найбільш простими з об'єктів R, що можуть містити дані тільки одного типу. Ви можете створити вектор, використовуючи функцію `c ()`, яка об'єднує кілька однотипних елементів. Можливо, також, створити послідовність використовуючи символ: або функцію `seq ()`. Наприклад, `1:5` створює вектор-послідовність чисел від 1 до 5. Функція `seq ()` дозволяє вказувати інтервал між числами. Можна повторювати зразок, використовуючи функцію `rep ()`. Також можна створити чисельний вектор з пропущеними значеннями, використовуючи функцію `numeric ()` або символний – `character ()` або логічний (тобто TRUE або FALSE) – `logical ()`.

2. `Factors` схожі на вектори, але з визначеним набором рівнів. `Factor` представляє номінальну або рангову шкалу. Використовується для подання Y в класифікаційних моделях, також функція `factor ()` трансформує вектор у `factor`. Також, `factor` може бути відсортований за допомогою опції `ordered = T` або функції `ordered ()`.

3. `Matrix` (матриці) схожі на вектори, але з специфічними інструкціями для виведення. Матриця – двовимірний набір елементів одного типу (таблиця). Якщо ви хочете створити матрицю, то однією з можливостей є використання функції `matrix ()`. Ви вводите вектор, набір рядків або стовпців, і можете вказати R, як конкретно потрібно інтерпретувати дані (за замовчуванням як стовпці).

Функції `cbind ()` і `rbind ()` комбінують вектори в матриці по стовпцях або по рядках. Розмірність матриці може бути отримана за допомогою функції `dim ()`. Інакше функції `nrow ()` і `ncol ()` повертають, відповідно, кількість рядків і стовпців.

4. `Arrays` (масиви) схожі на матриці, але можуть мати розмірність більше двох. Масив – багатовимірний набір елементів одного типу. Масив

має бути симетричним по всій розмірності. Об'єкти векторів, з яких складається масив, мають бути однакового типу, але не обов'язково числового.

5. `List` (список) є вектором для об'єктів `R`. Списки – це колекція об'єктів `R`. Функція `list ()` створює список; `unlist ()` трансформує список у вектор. Переважно у вигляді списків зручно зберігати або якісь однотипні дані, що відповідають різним ітераціям, наприклад, безліч моделей, або зберігати різнорідні дані, які мають смисловий зв'язок, наприклад, різні статистичні характеристики окремої моделі.

6. `Dataframe` (структура) схожа на матрицю, але не вимагає щоб всі колонки були ідентичні за типом. Структура є списком змінних/векторів однакової довжини. `Data.frame` – двовимірний набір даних (таблиця). На відміну від матриць, колонки в `data.frame` можуть містити дані різного типу. Однак тип даних усередині кожної колонки може бути тільки один. Це пояснюється тим, що `data.frame` – це список векторів (колонок). Тому до `data.frame` можуть бути застосовані різні функції.

7. Формули – спеціальна форма вираження відносин між змінними в рівнянні. Формули використовуються під час побудови моделей для визначення функціональної залежності між параметрами. Символ точки `(.)` підставляє всі наявні змінні.

8. `Classes` (класи) – це тип даних для змінної, а сама змінна цього типу є об'єктом – екземпляром класу). Класи приєднуються до об'єктів як `attribute` (атрибути). Всі об'єкти в `R` мають свій клас, тип і розмірність.

Усі об'єкти підтримують присвоєння імен елементів, що містяться в них. Для цього використовують символ `"` або `'`. Аналогічно векторам матриці та `data.frames` мають такі властивості, як `rownames` і `colnames`, які дозволяють змінювати імена колонок і рядків.

Для видалення імен здійснюється присвоєнням спеціального типу `NULL`.

Перетворення типів даних здійснюється через групу функцій, що починаються на `as`.

Індексація в пакеті `R` – ефективний і потужний інструмент для роботи з даними. Індокси можуть бути числовими, логічними та текстовими.

Для індексування використовується три типи виразів:

`[` – вибирає елементи вектора/списку/масиву і т. д.;

`$` – вибирає один елемент з `data.frame`/списку по його імені;

`[[-` вибирає елементи з вектора/списку/масиву і т. д, але відкидає імена, якщо вони є.

Особливості індексування дозволяють змінювати положення елементів і дублювати їх. Для видалення елементів за значенням індексу перед ними додають знак "мінус". Щоб додати новий елемент (колонку/рядок в data.frame) до списку використовується нове ім'я або числовий індекс.

Якщо йде звернення до неіснуючого індексу, то повернеться спеціальне значення NA.

2. Робота в RStudio: графічні можливості.

Перед початком роботи в RStudio необхідно імпортувати дані. Для введення даних, окрім ручного режиму, можна використати вікно Environment кнопку Import Dataset і вказати, який тип файлу будемо імпортувати та місце звідки. Нагадуємо, що ім'я файлу має бути англійською мовою та складатися з одного слова (рис. 171).

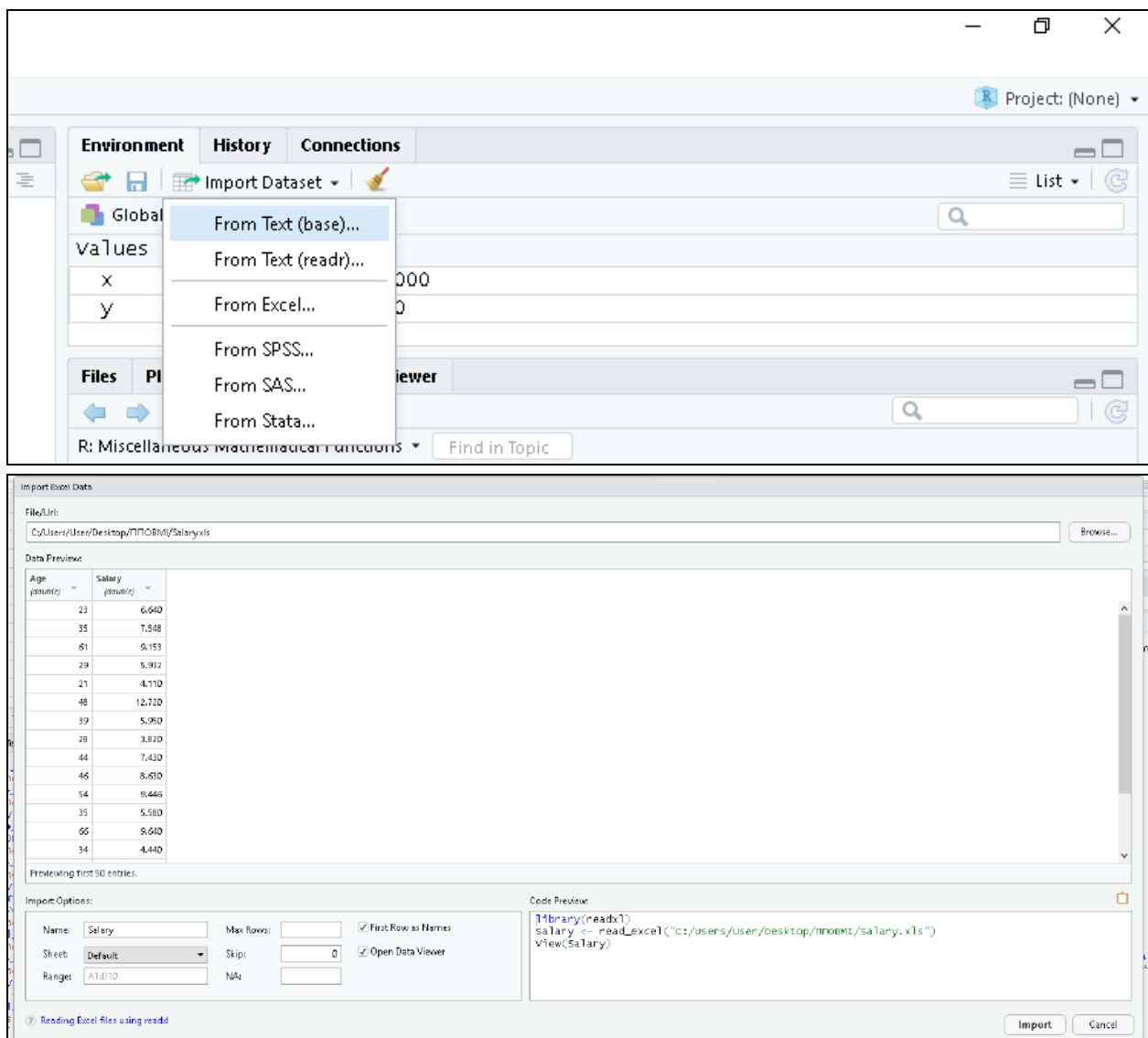


Рис. 171. Імпорт даних у RStudio

У результаті отримуємо таблицю даних, що містить інформацію про вік і розмір заробітної плати клієнтів банку (рис. 172).

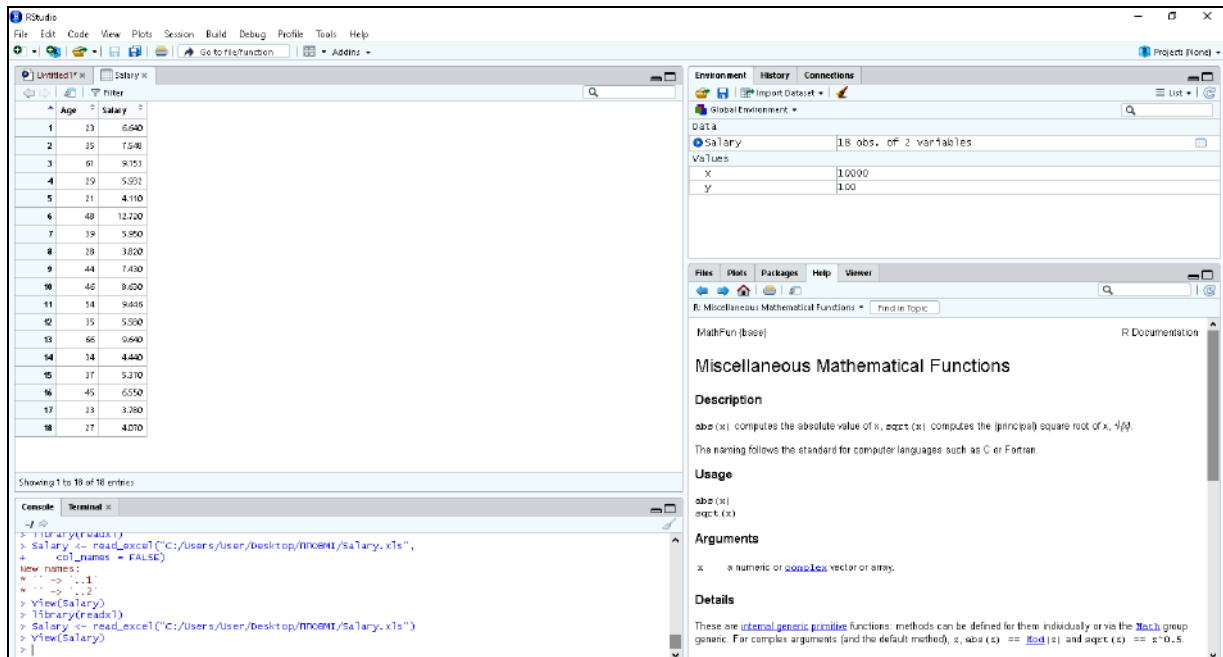


Рис. 172. Вихідні дані для аналізу у RStudio

Якщо ви хочете дати назву стовпцям або змінити її, необхідно перейти до вікна скриптів Untitled*1 та виконати команду:

`colnames(Salary)<-c('Age_years','Wage_th_UAH')`.

У результаті отримуємо нові назви стовпців таблиці (рис. 173).

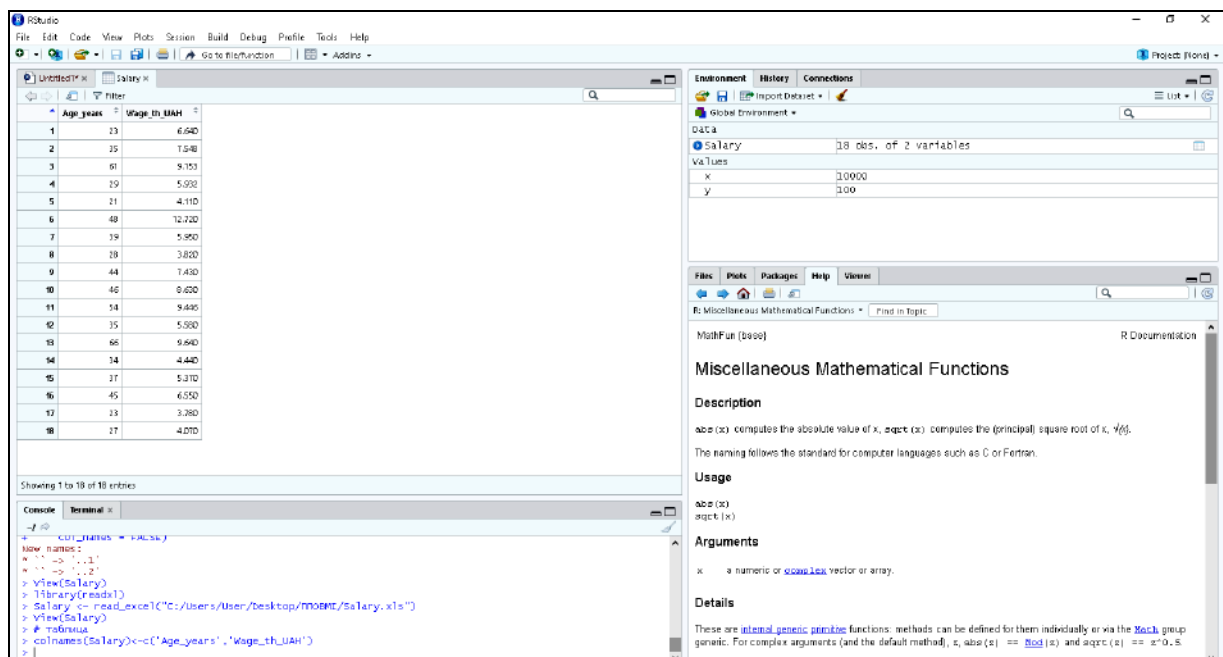


Рис. 173. Зміна назв стовпців таблиці у RStudio

Пакет R (RStudio) має потужний блок візуалізації. Для роботи з графікою в R необхідно встановити відповідний пакет – `ggplot2`, тобто виконати команду `install.packages("ggplot2")` або у бібліотеці обираємо пакет *Graphics* і натискаємо кнопку *Install*.

У `ggplot2` будь-який тип інфографіки є результатом взаємодії ряду елементів:

- 1) масиву даних (`data`);
- 2) схеми відповідності змінних масиву візуальних засобів (`aesthetic`);
- 3) геометричного об'єкта (`geom`);
- 4) статистичного перетворення (`stat`);
- 5) координатної системи (`coord`);
- 6) орієнтирів (`guide`);
- 7) панелі (`facet`);
- 8) художнього оформлення (`theme`).

Візуалізуємо вихідний інформаційний простір. Для цього використаємо команду: `plot(Salary$Age_years,Salary$Wage_th_UAH)` (рис. 174).

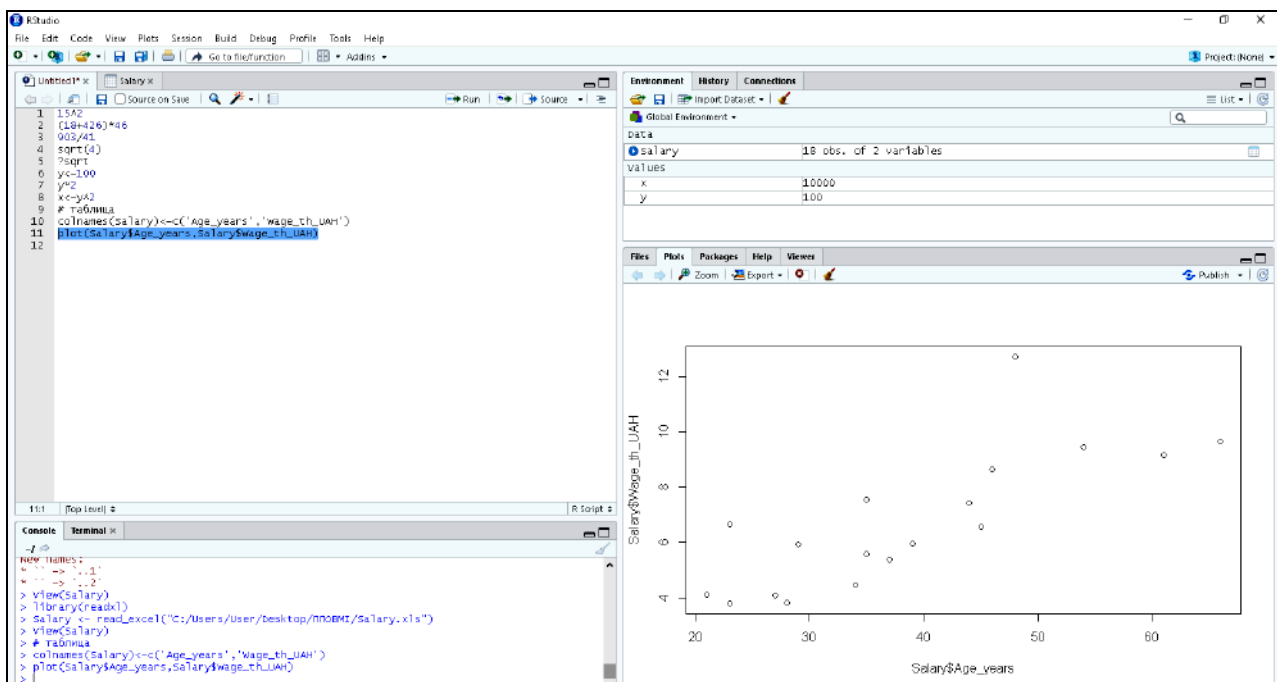


Рис. 174. Побудова діаграми у RStudio

Також користувач має можливість змінити параметри точкової діаграми: обрати тип, колір і товщину. Для цього необхідно виконати команду: `plot(Salary$Age_years,Salary$Wage_th_UAH,col='red',lwd=2)` (рис. 175).

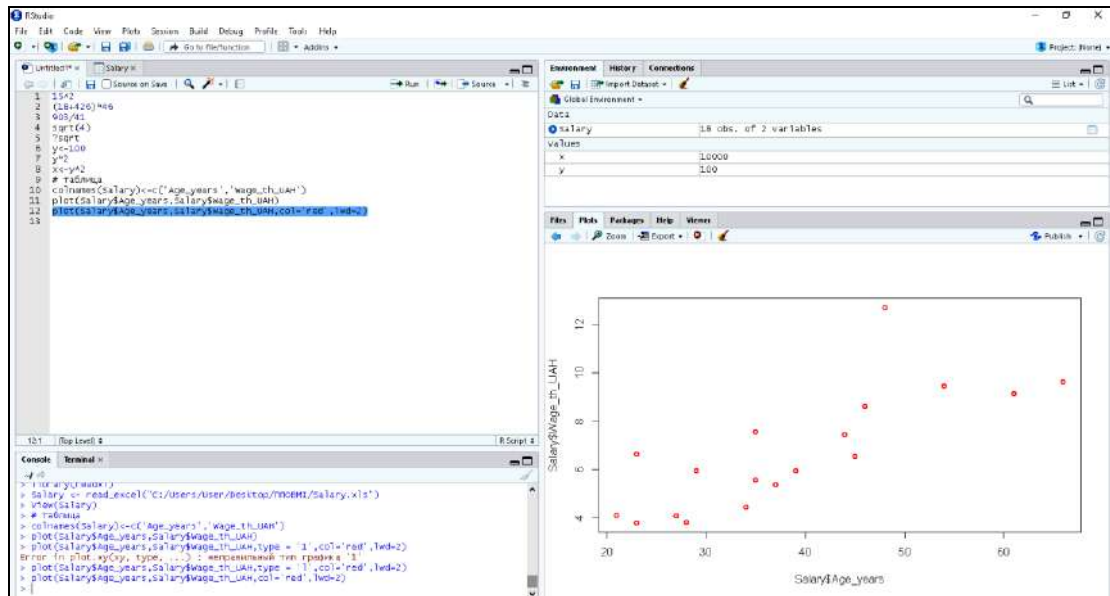


Рис. 175. Редагування графіка у RStudio

Для зміни підпису осей та присвоєння назви точковій діаграмі виконуємо команду: `plot(Salary$Age_years, Salary$Wage_th_UAH, col= 'red', lwd =2, xlab = 'Вік_клієнтів_банку', ylab = 'Заробітна_плата_тис_грн', main = 'Залежність_доходу_від_віку')` (рис. 176).

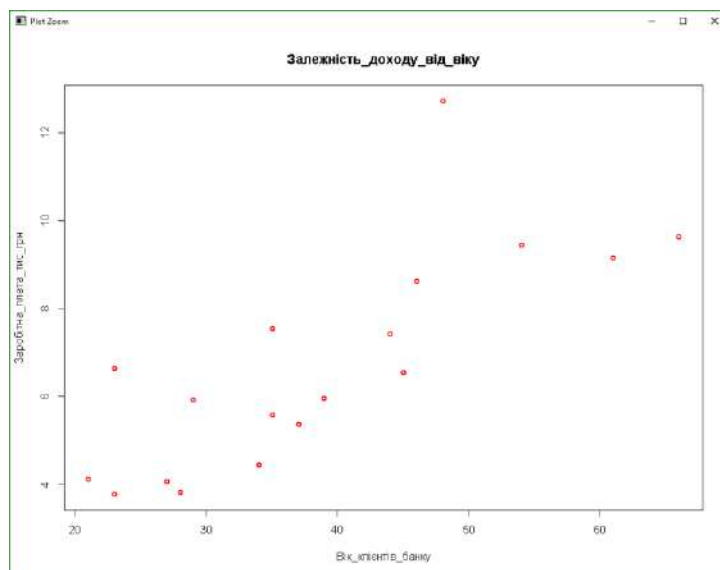


Рис. 176. Готова візуалізація у RStudio

Також користувач має можливість змінити тип діаграми. Наприклад, перетворимо точкову діаграму на стовпчасту гістограму. Для цього необхідно змінити геометричний об'єкт на `aes(xlab = 'Вік_клієнтів_банку' = cyl) + geom_bar() + coord_polar()`.

Отже, вихідні дані завантажено та візуалізовано.

3. Статистичний аналіз даних у пакеті R.

Спочатку перевіримо вихідний інформаційний простір на нормальний закон розподілу. R має чотири вбудовані функції для генерації нормального розподілу:

```
dnorm(x, mean, sd);  
pnorm(x, mean, sd);  
qnorm(p, mean, sd);  
rnorm(n, mean, sd).
```

Наведемо опис параметрів, що використовуються у раніше згаданих функціях:

x – вектор чисел;

p – вектор ймовірностей;

n – кількість спостережень (розмір вибірки);

mean – середнє значення вибірових даних. Це значення за замовчуванням дорівнює нулю;

sd – стандартне відхилення. Це значення за замовчуванням = 1.

Отже, для перевірки вихідного інформаційного простору на нормальний закон розподілу необхідно визначити загальні статистичні характеристики (мода, медана, мінімальне, максимальне значення, середнє значення, значення першого та третього квартилю) та дисперсію для кожного ряду даних. Для цього використовуємо наступні команди (рис. 177).

1) `summary(Salary)`

2) `sd(Salary$Age_years)`
`sd(Salary$Wage_th_UAH)`

```
> summary(Salary)
  Age_years      wage_th_UAH
Min.   :21.00   Min.       : 3.780
1st Qu.:28.25   1st Qu.    : 4.673
Median :36.00   Median     : 6.250
Mean    :38.61   Mean       : 6.712
3rd Qu.:45.75   3rd Qu.    : 8.360
Max.    :66.00   Max.       :12.720
> |
```

```
> sd(Salary$Age_years)
[1] 13.03452
> sd(Salary$Wage_th_UAH)
[1] 2.463514
> |
```

Рис. 177. Розрахунок характеристик центру розподілу

Тепер можемо перевірити нормальність розподілу даних. Спочатку для ряду розподілу "Вік" виконуємо певні команди та візуалізуємо гістограму розподілу (рис. 178):

```
dnorm(Salary$Age_years, mean = 38.61, sd = 13.03452)
y<-dnorm(Salary$Age_years, mean = 38.61, sd = 13.03452)
hist(y, col = 'blue', main = "Normal Distribution")
```

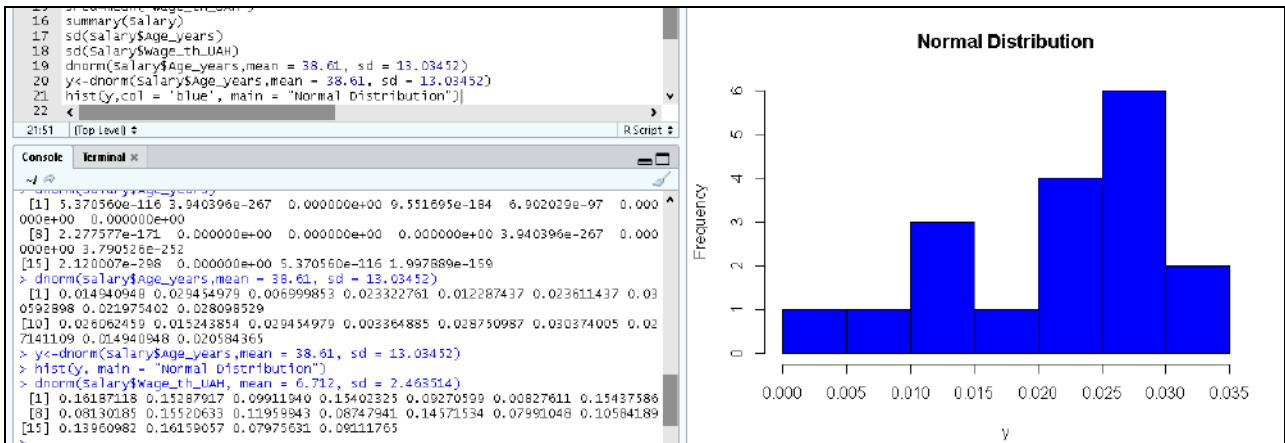


Рис. 178. Перевірка нормальності розподілу для ряду "Вік"

Відповідні дії застосовуємо й для стовпця "Заробітна плата" (рис. 179):

```

dnorm(Salary$Wage_th_UAH, mean = 6.712, sd = 2.463514)
b<-dnorm(Salary$Wage_th_UAH, mean = 6.712, sd = 2.463514)
hist(b, col = 'red', main = "Normal Distribution")

```

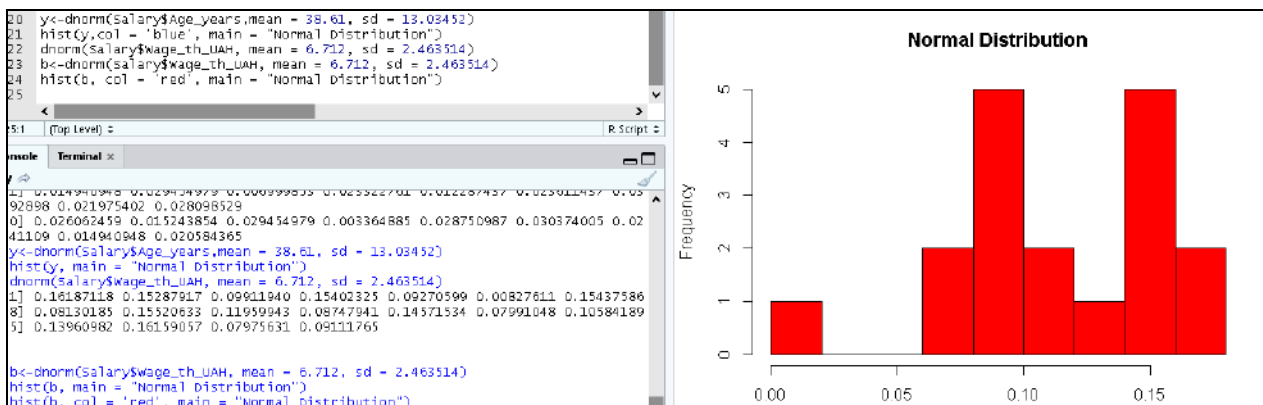


Рис. 179. Перевірка нормальності розподілу для ряду "Заробітна плата"

Результати перевірки (див. рис. 178 і 179) дозволяють стверджувати про відповідність рядів розподілу нормальному закону розподілу. Отже, можемо перейти до регресійного аналізу в пакеті R.

Регресійний аналіз є широко вживаним статистичним інструментом для встановлення моделі взаємозв'язку між двома змінними. Одна з цих змінних називається незалежною предиктор-змінною, значення якої зібрано за допомогою експериментів. Інша змінна називається змінною відповіді (залежною), значення якої впливає з змінної предиктора.

У лінійній регресії ці дві змінні пов'язані через рівняння, де показник (потужність) обох цих змінних дорівнює 1. Математично лінійна залежність

є прямою лінією на графіку. Нелінійна залежність, коли показник будь-якої змінної не дорівнює 1, створює криву.

Загальне математичне рівняння для лінійної регресії:

$$y = ax + b.$$

де y – змінна відповіді (залежна);

x – змінною предиктор (незалежна);

a і b – константи, які називаються коефіцієнтами.

З метою побудови моделі взаємозв'язку (впливу) та встановлення залежності між віком клієнтів банку та рівнем їхньої заробітної плати, використаємо функцію лінійної регресії $lm()$ у пакеті R. Синтаксис функції: $lm(formula,data)$.

Як незалежну змінну-предиктор виступає вік, тоді залежною змінною є заробітна плата клієнтів банку. Тому присвоюємо змінній вік значення "x", тоді з/п – "y". Для цього виконаємо відповідні дії:

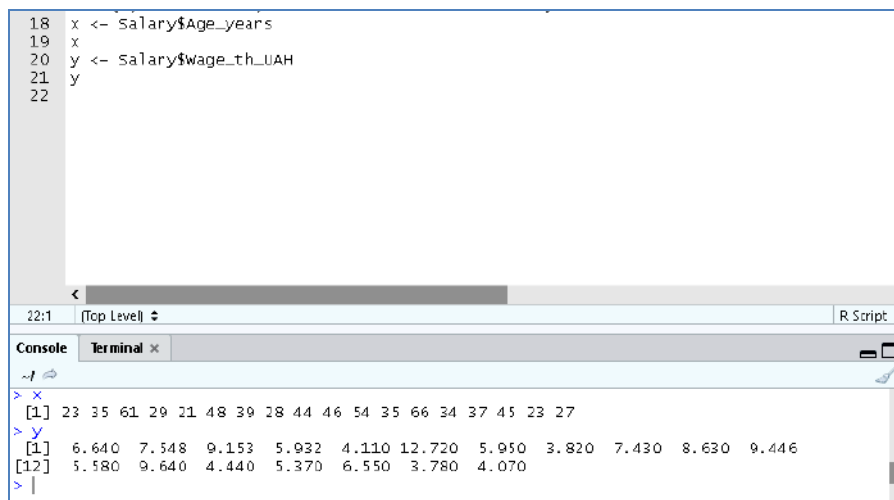
```
x <- Salary$Age_years
```

```
x
```

```
y <- Salary$Wage_th_UAH
```

```
y
```

У результаті отримаємо відповідний запис у вікні консоль (рис. 180).



```
18 x <- Salary$Age_years
19 x
20 y <- Salary$Wage_th_UAH
21 y
22
```

```
< [Progress bar] >
```

```
22:1 (Top Level) R Script
```

```
Console Terminal x
```

```
> x
[1] 23 35 61 29 21 48 39 28 44 46 54 35 66 24 37 45 23 27
```

```
> y
[1] 6.640 7.548 9.153 5.932 4.110 12.720 5.950 3.820 7.430 8.630 9.446
[12] 5.580 9.640 4.440 5.370 6.550 3.780 4.070
```

```
> |
```

Рис. 180. Присвоєння значень x та y змінним

Визначимо коефіцієнт кореляції, що надасть характеристику зв'язку між змінними. Для цього виконуємо команду $cor.test(x=Salary$Age_years, y=Salary$Wage_th_UAH)$ (рис. 181).

```
25 cor.test(x=Salary$Age_years, y=Salary$wage_th_UAH)|
26
25:51 (Top Level)
Console Terminal x
~|
Pearson's product-moment correlation
data: Salary$Age_years and Salary$wage_th_UAH
t = 4.972, df = 16, p-value = 0.0001385
alternative hypothesis: true correlation is not equal to 0
95 percent confidence interval:
 0.4908263 0.9136649
sample estimates:
      cor
0.779152
> result[["df"]]
```

Рис. 181. Розрахунок коефіцієнта кореляції

З розрахунку (див. рис. 181) видно, що між змінними існує доволі сильний прямий зв'язок ($R = 0,779$), тобто зі збільшенням віку клієнтів банку зростає їхній дохід.

Відповідно до вихідного інформаційного простору виконуємо однофакторний регресійний аналіз та отримуємо результат, наведений на рис. 182:

```
relation <- lm(y~x)
print(relation)
```

```
18 x <- Salary$Age_years
19 x
20 y <- Salary$wage_th_UAH
21 y
22 relation <- lm(y~x)
23 print(relation)
24
24:1 (Top Level)
Console Terminal x
~|
call:
lm(formula = y ~ x)
Coefficients:
(Intercept)          x
 1.0258          0.1473
```

Рис. 182. Результати виконання регресійного аналізу

Отримані значення коефіцієнтів за незалежної змінної та вільних членів трансформуємо в модель, що має вигляд:

$$Y = 1,0258 + 0,1473x.$$

Для отримання більш розгорнутого (повного) висновку використовуємо функцію `print(summary(relation))` (рис. 183).

```
18 x <- Salary$Age_years
19 x
20 y <- Salary$wage_th_UAH
21 y
22 relation <- lm(y~x)
23 print (relation)
24 Salary$Newage <- NULL
25 cor.test(x=Salary$Age_years, y=Salary$wage_th_UAH)
26 print(summary(relation))
27
```

27:1 (Top Level) ↓

Console Terminal x

```
> print(summary(relation))

Call:
lm(formula = y ~ x)

Residuals:
    Min       1Q   Median       3Q      Max
-1.5926 -1.0598 -0.6163  0.5938  4.6258

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  1.02577    1.20355   0.852 0.406630
x             0.14726    0.02962   4.972 0.000138 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 1.592 on 16 degrees of freedom
Multiple R-squared:  0.6071,    Adjusted R-squared:  0.5825
F-statistic: 24.72 on 1 and 16 DF,  p-value: 0.0001385
```

Рис. 183. Розгорнуті результати регресійного аналізу

Раніше наведене вікно (див. рис. 183) містить інформацію щодо описової статистики залишків моделі, також наведено результат розрахунку значень коефіцієнтів та змінної X, значення стандартної помилки, t-критерій та рівень значущості. Коефіцієнт детермінації дорівнює 0,607, тобто 61 % зміни Y пояснюється впливом саме X. Загалом наведені результати свідчать про те, що модель є адекватною.

Використовуючи отриману регресійну модель, можливо побудувати прогноз, а саме задаючи значення незалежної змінної, можемо визначити, як саме зміниться залежна.

Для цього використовуємо функцію `predict()`. Синтаксис функції: `predict(object, newdata)`.

Отже, знайдемо прогнозне значення заробітної плати для клієнтів банку, якщо їхній вік буде становити 19 і 65 років. Для цього створюємо новий Data.frame (рис. 184):

```
new_Age_years <- data.frame(x = c(19,65))
new_Age_years
```

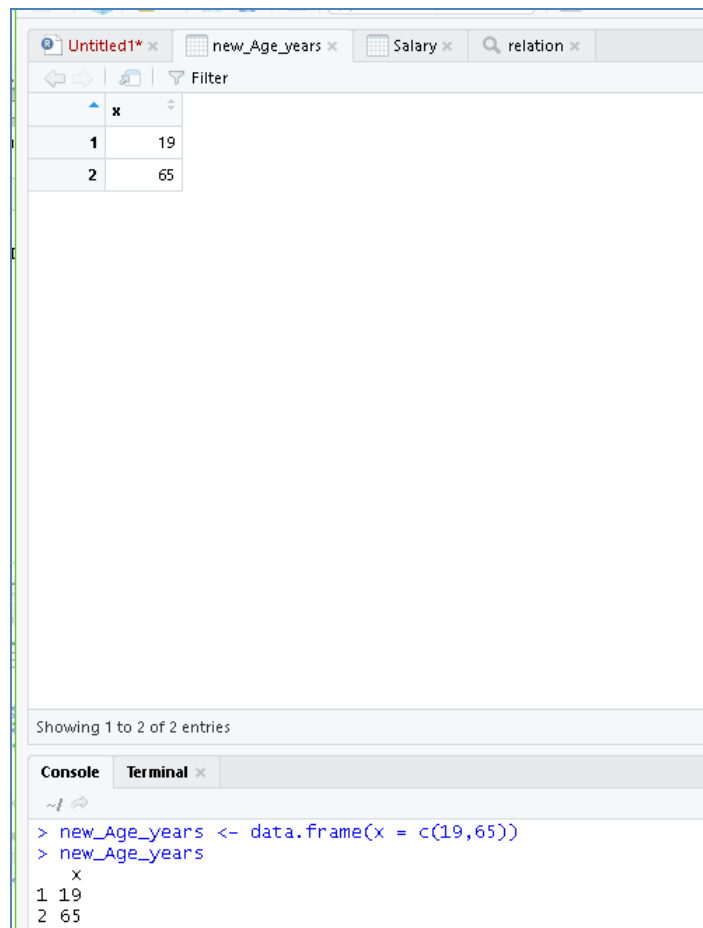


Рис. 184. Створення нового Data.frame

Виконаємо прогноз за допомогою команди `predict(relation, new_Age_years)` (рис. 185).

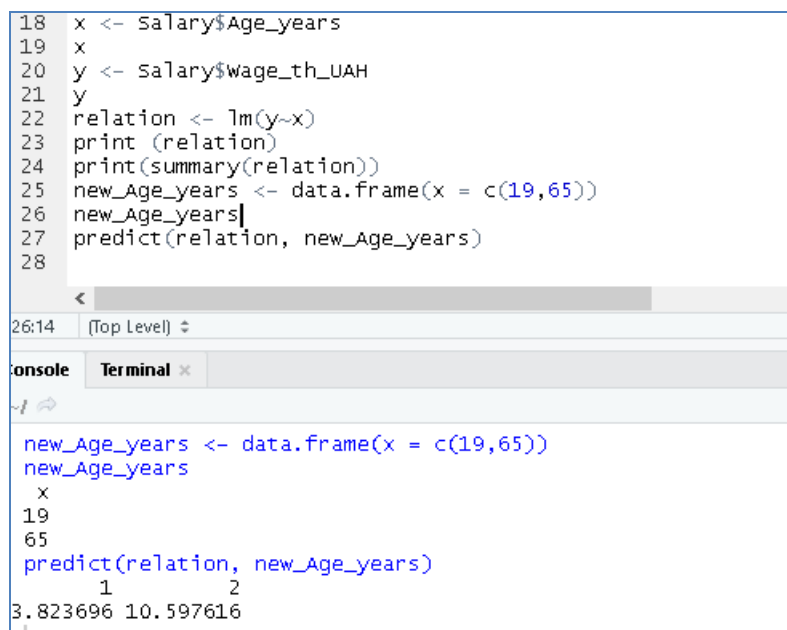


Рис. 185. Результати побудови прогнозу

У такий спосіб було отримано прогнозні значення для заданих умов. Тобто, якщо вік клієнтів банку буде дорівнювати 19 та 65 років, їхній приблизний дохід буде становити 3 823,7 і 10 597,62 грн відповідно.

Завдання:

1. Використовуючи пакет R (RStudio):

1.1. Створити символний (текстовий) вектор, що міститиме назви навчальної дисциплін, що вивчаються у поточному семестрі.

1.2. Створити список (list), що буде містити інформацію про дні тижня (у списку має бути три елемента: кількість днів (1 – 7), назва днів тижня та окремо визначені, ті дні, коли студенти мають пари).

1.3. Створити вектор, що буде відображати співставлення шкали оцінювання ECTS, яка є загальнозживаною для оцінювання якості студентських досягнень, та бальної шкали, що відповідає їй.

1.4. Створити data.frame (структура), що відобразить станом на 2019 р. найбільші міста України за кількістю населення понад 500 000 осіб.

1.5. Знайти рішення: $y = 45,02 - \sqrt{x_1} + x_2^4$, якщо $x_1 = 144$, $x_2 = 3,7$.

2. Використовуючи статистичні дані та графічні можливості пакету R, побудувати 3 різні типи діаграм і вміти їх редагувати.

Результати оформити у вигляді звіту з лабораторної роботи.

Завдання для самостійного опрацювання

Запитання, що виносяться на самостійну роботу студентів:

1. Охарактеризуйте аналітичні (статистичні) можливості пакету R.
2. Охарактеризуйте графічні редактори скриптів та IDE пакету R.
3. Надайте порівняльну характеристику функціональних можливостей пакету R та інших спеціалізованих програмних продуктів ОВМД.
4. Які надбудови та бібліотеки пакету R вам відомі?
5. Наведіть сфери застосування пакету R. Які саме переваги в роботі бізнес-аналітика надає пакет R?

Завдання 1. Є data.frame (структура):

```
df <- data.frame (var1 = c (11,21,31), var2 = c (12,22,32), var3 = c (13,23,33),  
var4 = c (14,24,34) , row.names = c ("case1", "case2", "case3"))  
df
```

```
var1 var2 var3 var4
case1 11 12 13 14
case2 21 22 23 24
case3 31 32 33 34
```

Виберіть значення var1, var2, var3 для case1.

Виберіть значення всіх змінних для case2, які більше 22.

Виберіть імена змінних для колонок 1 і 3.

Додайте колонку з ім'ям Y і значеннями -1, 0, 1.

Видаліть рядок case2.

Значення другої колонки зведіть у третю ступінь.

Завдання 2. Використовуючи власний інформаційний простір та аналітичні можливості пакету R, перевірте вихідні дані на нормальний закон розподілу, побудуйте однофакторну (багатофакторну) регресійну модель і прогноз.

Рекомендована література

Основна

1. Вуколов Э. А. Основы статистического анализа. Практикум по статистическим методам и исследованию операций с использованием пакетов STATISTICA и EXCEL : учеб. пособ. / Э. А. Вуколова. – 2-е изд., испр. и доп. – Москва : Форум, 2008. – 464 с.
2. Грох М. Microsoft Office Access 2007. Библия пользователя / М. Грох, Дж. Стокман, Г. Пауэлл. – Москва : Диалектика, 2008. – 1200 с.
3. Кабаков Р. R в действии. Анализ и визуализация данных на языке R / Р. Кабаков. – Москва : ДМК Пресс , 2014. – 588 с.
4. Когнитивная бизнес-аналитика : учебник / под науч. ред. д.т.н., проф. Н. М. Абдикеева. – Москва : ИНФРА-М, 2011. – 511 с.
5. Матюшок В. М. Основы эконометрического моделирования с использованием Eviews : учеб. пособ. / В. М. Матюшок, С. А. Балашова, И. В. Лазанюк. – Изд. 2-е, доп. и перераб. – Москва : РУДН, 2011. – 206 с.
6. Основы бизнес-анализа : учеб. пособ. / В. И. Бариленко, В. В. Бердников, Р. П. Булыга [и др.] ; под ред. В. И. Бариленко. – Москва : КНОРУС, 2016. – 272 с.
7. Просницкий А. Самоучитель "Microsoft Project 2013. Управление проектами" / А. Просницкий. – Киев : LEO Consulting, 2017. – 285 с.
8. Уикем Х. Язык R в задачах науки о данных: импорт, подготовка, обработка, визуализация и моделирование данных / Х. Уикем, Г. Гроулмунд. – Москва : Вильямс, 2017. – 592 с.

Додаткова

9. Айвазян С. А. Программное обеспечение по статистическому анализу данных: методология сравнительного анализа и выборочный обзор рынка / С. А. Айвазян, В. С. Степанов. – Москва : Наука, 1991. – 208 с.
10. Наглядная статистика. Используем R / А. Б. Шипунов, Е. М. Балдин, П. А. Волкова и др. – Москва : ДМК Пресс, 2012. – 298 с.
11. Фуллер Л. У. Access 2010 для чайников / Л. У. Фуллер, К. Кук. – Москва : Диалектика, 2010. – 384 с.

12. Blazquez D. Big Data sources and methods for social and economic analyses / D. Blazquez, J. Domenech / Technological Forecasting and Social Change. 2018. – Volume 130. – P. 99–113.

13. Hofmann M. RapidMiner: Data Mining Use Cases and Business Analytics Applications / M. Hofmann, R. Klinkenberg. – New York : Taylor and Francis Group. LLC, 2014. – 463 p.

Інформаційні ресурси

14. Герасевич В. А. Современное программное обеспечение для статистической обработки биомедицинских исследований [Электронный ресурс] / В. А. Герасевич, А. Р. Аветисов – Режим доступа : <http://miklebig.narod.ru/docum/statprog.htm>.

15. Офіційний сайт державної служби статистики України. – Режим доступу : <http://www.ukrstat.gov.ua>.

16. Программа SPSS Statistics [Электронный ресурс]. – Режим доступа : <https://spssstatistics.ru>.

17. Функції Excel (за категоріями) [Електронний ресурс]. – Режим доступу : <https://support.office.com/uk-ua/article>.

18. EViews Econometric Modeling Software [Electronic resource]. – Access mode : <https://www.ihs.com/products/eviews-econometric-modeling-analysis-software.html>.

19. Microsoft Project [Electronic resource]. – Access mode : <https://products.office.com/uk-ua/project/project-and-portfolio-management-software>.

20. The Comprehensive R Archive Network [Electronic resource]. – Access mode : <https://cran.r-project.org>.

21. RapidMiner [Electronic resource]. – Access mode : <https://rapidminer.com>.

Зміст

Вступ.....	3
Змістовий модуль 1. Класифікація програмних продуктів обробки великих масивів даних та їхні види. Вбудовані програмні продукти	5
Тема 1. Поняття програмних продуктів обробки великих масивів даних (ПП ОБМД) та їхні види	5
Лабораторна робота 1. Оброблення таблиць, редагування графіків та візуалізація в MS Excel	5
Тема 2. Обробка таблиць, графіків і побудова моделей за допомогою MS Excel	27
Лабораторна робота 2. Аналіз вибірки та регресійний аналіз у MS Excel	27
Тема 3. Побудова баз даних MS Access	53
Лабораторна робота 3. Побудова баз даних MS Access	53
Тема 4. Побудова проєктів за допомогою MS Project	77
Лабораторна робота 4. Побудова проєктів за допомогою MS Project.....	77
Змістовий модуль 2. Спеціальні програмні продукти обробки великих масивів даних. Їхні види та сфера використання.....	101
Тема 5. Аналітичні можливості RapidMiner	101
Лабораторна робота 5. Аналітичні та функціональні можливості RapidMiner	101
Тема 6. Використання пакетів SPSS та EViews для побудови моделей.....	123
Лабораторна робота 6. Аналіз великих масивів даних за допомогою спеціалізованого статистичного пакету SPSS.....	123
Тема 7. Пакет R для аналізу великих масивів даних.....	148
Лабораторна робота 7. Аналіз великих масивів даних за допомогою пакету R.....	148
Рекомендована література.....	167
Основна.....	167
Додаткова.....	167
Інформаційні ресурси	168

НАВЧАЛЬНЕ ВИДАННЯ

ПРОГРАМНІ ПРОДУКТИ ОБРОБКИ ВЕЛИКИХ МАСИВІВ ІНФОРМАЦІЇ

**Методичні рекомендації
до лабораторних та самостійних робіт
для студентів спеціальності
051 "Економіка"
першого (бакалаврського) рівня**

Самостійне електронне текстове мережеве видання

Укладач **Дериховська** Вікторія Ігорівна

Відповідальний за видання *О. В. Раєвська*

Редактор *А. С. Ширініна*

Коректор *В. Ю. Труш*

План 2021 р. Поз. № 131 ЕВ. Обсяг 170 с.

Видавець і виготовлювач – ХНЕУ ім. С. Кузнеця, 61166, м. Харків, просп. Науки, 9-А

*Свідоцтво про внесення суб'єкта видавничої справи до Державного реєстру
ДК № 4853 від 20.02.2015 р.*