

МІНІСТЕРСТВО ОСВІТИ І НАУКИ УКРАЇНИ
НАЦІОНАЛЬНИЙ ТЕХНІЧНИЙ УНІВЕРСИТЕТ УКРАЇНИ
"КІЇВСЬКИЙ ПОЛІТЕХНІЧНИЙ ІНСТИТУТ
ІМЕНІ ІГОРЯ СІКОРСЬКОГО"

**СОЦІАЛЬНО-ЕКОНОМІЧНИЙ ПОТЕНЦІАЛ
УПРАВЛІННЯ СТАНОМ ДОВКІЛЛЯ**

КОМП'ЮТЕРНИЙ ПРАКТИКУМ

*Рекомендовано Методичною радою КПІ ім. Ігоря Сікорського
як навчальний посібник для студентів,
які навчаються за спеціальністю 122 «Комп'ютерні науки»,
спеціалізацією «Інформаційні технології моніторингу довкілля»*

Київ
КПІ ім. Ігоря Сікорського
2020

Соціально-економічний потенціал управління станом довкілля. Комп'ютерний практикум [Електронний ресурс] : навч. посіб. для здобувачів ступеня магістра за освітньою програмою «Комп'ютерний моніторинг та геометричне моделювання процесів і систем» спеціальності 122 «Комп'ютерні науки» / КПІ ім. Ігоря Сікорського ; уклад.: Караєва Н.В., Варава І.А. – Електронні текстові дані (1 файл: 6,39 Мбайт). – Київ : КПІ ім. Ігоря Сікорського, 2020. – 67 с.

*Гриф надано Методичною радою КПІ ім. Ігоря Сікорського (протокол № 7 від 27.02.2020 р.)
за поданням Вченої ради Теплоенергетичного факультету (протокол № 8 від 24.02.2020 р.)*

Електронне мережне навчальне видання

Соціально-економічний потенціал управління станом довкілля

Комп'ютерний практикум

Укладачі: Караєва Наталія Веніамінівна, к.е.н., доцент;
Варава Іван Андрійович, к.т.н., доцент

Відповідальний редактор: Коваль О.В., канд. техн. наук, в.о. зав. кафедри автоматизації проєктування енергетичних процесів і систем ТЕФ КПІ ім. Ігоря Сікорського

Рецензент: Фуртат Ірина Едуардівна, канд. техн. наук, доц. каф. кафедра теоретичної і промислової теплоенергетики ТЕФ КПІ ім. Ігоря Сікорського

Посібник розроблений на підставі робочої програми кредитного модуля з дисципліни "Соціально-економічний потенціал управління станом довкілля" та призначений для якісної організації виконання практичних робіт студентами обсягом 50 годин, підвищення розуміння основ обґрунтування та прийняття управлінських соціально-екологічних рішень на основі застосування математичних методів аналізу.

Призначений для студентів, які навчаються за освітньою програмою "Комп'ютерні науки" підготовки магістрів за спеціальністю 122 – "Інформаційні технології моніторингу довкілля" денної форми навчання.

Спрямований на формування у студентів умінь та навичок з основ розробки засобів прогнозування екологіко-соціальних наслідків господарської діяльності. Забезпечує студентів необхідними теоретично-методичними знаннями для опанування відповідної теми комп'ютерного практикуму та виконання завдань, запланованих впродовж семестру.

ЗМІСТ

	Стр.
Вступ	4
1. Основні концепції відтворення соціально-економічного потенціалу сталого розвитку	6
2. Інформаційне забезпечення побудови системи моніторингу соціально-економічного потенціалу управління станом довкілля.....	11
3. Математичні моделі економічного зростання із врахуванням факторів людського капіталу.....	19
4. Основи кореляційно-регресійного аналізу	23
5. Кластерний аналіз інструментами Data Mining (SPSS)	54
6. Представлення результатів комп'ютерних практикумів у протоколі	62
Список використаної літератури	65

Вступ

Головним об'єктом діяльності магістрів за спеціальністю 122 "Комп'ютерні науки" спеціалізації "Інформаційні технології моніторингу довкілля" є підтримка та створення комп'ютерних систем моніторингу.

Досвід розвинених країн свідчить, що реальні зрушення в напрямі формування передумов сталого розвитку (СР) як держави в цілому, так і окремих регіонів можливі лише за умови нарощення не лише виробничого, а й природно-ресурсного, демографічного, науково-технічного, інформаційного та соціально-культурного потенціалу. Комплексне поняття "соціально-економічний потенціал управління станом довкілля" слід розглядати як систему факторів (умов, джерел, засобів), котрі забезпечують досягнення цілей СР. Перехід економіки України на модель СР значною мірою залежить від наявного потенціалу розширеного відтворення ресурсної бази соціально-економічного піднесення на новій інституціональній і технологічно-інноваційній основі.

Наукові та практичні розробки щодо глобальної проблеми СР, з огляду на її комплексність і міждисциплінарний характер, вимагають застосування величезного масиву різноманітних даних, вивчення і застосування провідного вітчизняного та зарубіжного досвіду, широкої міжнародної співпраці, використання світових інформаційних ресурсів. Завдяки зростаючій інтенсифікації інформаційного забезпечення виробництва, знижується потреба в багатьох традиційних видах сировини, що, у свою чергу, сприяє збереженню природи та вирішенню екологічних проблем.

Для вимірювання СР, моніторингу його динаміки розробляється і використовується багатопараметричні системи відповідних статистичних показників та інтегральних індикаторів. Система моніторингу розглядається як інформаційна, діагностична, прогностична система, реалізація якої здійснюється в рамках управлінської діяльності. Ефективним інструментом аналізу тенденцій та прогнозування перспектив СР являються економетричні методи та моделі. Економетрична модель – це окрема функція чи система функцій (рівнянь), що описує кореляційно-регресійний зв'язок між соціально-економічними показниками, один чи декілька з яких є залежною змінною, а усі інші – незалежними [1, с. 14].

Кожна країна (або регіон, область) має свої особливості, але в однорідних групах країн діють близькі за змістом закономірності розвитку. З метою ранжування та виділення однорідної сукупності країн (області, підприємства тощо) за рівнем соціально-економічного розвитку національної економіки застосовується кластерний аналіз.

Тому завданням даних методичних вказівок, відповідно до робочої навчальної програми курсу "Соціально-економічний потенціал управління станом довкілля", є надання студенту-магістру теоретико-методичних знань, щодо обґрунтування інформаційно-аналітичного забезпечення розробки системи моніторингу, яка може бути використана як «радник» особи, що приймає рішення, для надання максимально об'єктивної інформації при плануванні і здійсненні організаційних рішень у напряму підвищення соціально-економічного потенціалу управління станом довкілля.

Особливістю виконання комп'ютерних практикумів є комплексність виконання завдань. Тобто, 12 практикумів поділено на два блоки: 1) десять комп'ютерних практикумів являє собою єдине завдання, що спрямовано на побудову багатофакторної регресійної моделі в Excel із усунення мультиколінеарності і розробки власної системи прогнозування ключових показників соціально-економічного потенціалу управління станом довкілля; 2) два комп'ютерних практикумів спрямовано на кластерний аналіз інструментами Data Mining (SPSS).

Структура навчального посібника складається з надання теоретичних основ формування структурних компонентів інформаційно-аналітичної системи моніторингу. Методична же основа побудови багатофакторної регресійної моделі в Excel із усунення мультиколінеарності та кластерного аналізу інструментами Data Mining (SPSS) наведено у вигляді протоколів, що містять таку інформацію: завдання дослідження, джерела інформації, обґрунтування вибору метода, детальний опис алгоритму проведення того чи іншого аналізу з наведенням відповідних скріншотів, висновки. Викладення матеріалу в такому вигляді, дозволяє студенту набути вміння надавати прозоро обґрунтовану аналітичну інформацію і формувати доступну інструкцію розробленою системи для користувача.

1. Основні концепції відтворення соціально-економічного потенціалу сталого розвитку

Домінантою концепції СР є концепція людського розвитку (ЛР), що виникла на базі ноосферної концепції В. Вернадського і теорії людського капіталу.

Основні положення ноосферної концепції:

- у книзі "Наукова думка як планетарне явище" (1936-1938) Володимир Вернадський зазначав, що «еволюція видів переходить в еволюцію біосфери». При цьому *еволюційний процес отримує "особливе геологічної значення завдяки тому, що він створив нову геологічну силу – наукову думку соціального людини.* Саме під впливом наукової думки і людської праці біосфера переходить в новий стан – в ноосферу" [2, с. 48];
- *ноосфера (гр. "ноос" – розум і "сфера" – куля) – якісно новий стан біосфери Землі (її перебудова) і навколоzemного простору (а отже, нової оболонки Землі), який формується під впливом розумової та фізичної діяльності людства (що її можна зіставити з могутніми природними геологічними процесами) в інтересах мислячої істоти як єдиного цілого. Людина в ноосфері стає найбільшою геологічною силою завдяки своїй думки, праці та техніки [3, с. 162]»*
- *вектором спрямованості еволюції життя на Землі є поява і вдосконалення нервової тканини, зокрема мозку [3, с. 186]. Без створення мозку не було би появі і розвитку наукової думки в біосфері, а без наукової думки не було б геологічного ефекту – перебудови біосфери людством [4, с. 87-88];*
- *зростання ролі людського розуму в перетворенні компонентів і процесів природного середовища призвело до появі людської, соціальної культури, яка розглядається В. Вернадським в якості "культурної біогеохімічної енергії, яка є тією формою енергії, що створює в даний час ноосферу" [3, с. 149, 163-166];*
- *основні шляхи розвитку наукової думки визначаються запитами виробничої діяльності людини і потребами людського суспільства. "Із потреб землеробства і пов'язаної з ним іригації при створенні культурних товариств були тоді ж сформовані основи геометрії, а з потреб складного*

побуту великих держав – торгівлі, військових і фіскальних потреб – розвивалися основи арифметики" [2, с. 91];

– на відміну від положення марксистської теорії про створення людини-розумної завдяки виснажливої праці, В. Вернадський розглядав людину як біо-соціальну істоту. Людина як біологічна істота, наділена природними властивостями, що притаманні будь-якій тварині, і в той же час, людина як соціальна істота є закономірним результатом розвитку Всеєвіту, загальнопланетарним явищем, яке реалізує у праці свою думку [2, с. 72];

– завдання людини полягає в надані максимальної користі суспільству. І в той же час базовою потребою людини при житті є "досягнення можливо більшого щастя". "Таке складається як в розумовому і художньому кругозорі, так і в матеріальній забезпеченості; розумовий кругозір – наука; художній – витончені мистецтва, поезія, музика, живопис, скульптура і навіть релігія – світ людської фантазії, світ ідеалів і найприємніших слів; матеріальна забезпеченість – необхідна в меншій мірі, так як її задоволення, по грубості, відходять на другий план, але необхідність їх занадто чутлива і без неї обйтися не можна і не треба. Всього цього досягає людина – тільки завдяки крові, страждання минулих поколінь і сотень тисяч людей нашого часу" [3, с. 153];

– першою передумовою переходу біосфери в ноосферу в ХХ столітті є створення численних держаних наукових центрів на всій планеті Земля [2, с. 119-120]. Також зростає роль вищої освіти [2, с. 162];

– наукові знання і техніка, що створені людиною, перетворилися в планетарне явище. Вони нарівні з силами природи визначають майбутнє планети [3, с. 188-189];

– ноосферний розвиток неможливий без інтерналізації науки, свободи думки, моральної відповідальності вчених за використання наукових відкриттів [3, с. 167-168];

– людський розум і праця мають бути спрямовані не на самознищення, а на вихід в космічний простір. Ноосфера не може бути створена до припинення війн між народами.

Ноосферний розвиток ХХ століття характеризується зростанням спеціалізованих наукових знань в результаті швидкої руйнації меж між

окремими науками. "Ми все більше спеціалізуємося не за науками, а за проблемами. Це дозволяє, з одного боку, надзвичайно заглиблюватися в явище, що досліджується, а з іншого – розширювати охоплення його з усіх точок зору... У нас часто ставляється до спеціалізації негативно, але в дійсності спеціалізація, взята по відношенню до окремої особистості, надзвичайно посилює можливості її знань, розширює наукову сферу, яка їй доступна" [2, с. 105].

Основні положення теорії людського капіталу. Людський капітал є одним з найважливіших джерел економічного зростання і без значних інвестицій у нього ніяка економіка не може розвиватися успішно.

Поняття людський капітал увійшло в науку на початку 60-х років ХХ століття в західній економічній літературі. Виникнення і розвиток теорії людського капіталу пов'язані з іменами відомих американських учених-економістів Теодора Шульца (американський економіст, лауреат Нобелівської премії з економіки 1979 року), Гері Беккера (професор економіки й соціології Чикагського університету, одержав у 1992 р. Нобелівську премію за "поширення сфери мікроекономічного аналізу на цілий ряд аспектів людського поводження й взаємодії, включаючи неринкове поводження"). Першим опублікував наукові праці з теорії людського капіталу Т. Шульц: "Формування капіталу освіти" [5] і "Інвестиції в людський капітал" [6]. На думку автора однією з форм капіталу є освіта, людським його називають тому, що ця форма стає частиною людини, а капіталом є внаслідок того, що являє собою джерело майбутніх задоволень або майбутніх заробітків, або того й іншого разом. Людський капітал це розвиток знань і здібностей, які людям дають "шкільна освіта, навчання на робочому місці, зміцнення здоров'я і зростаючий запас економічної інформації" [6, с. 125–126]. Тобто, Т. Шульц обґрунтував необхідність капіталовкладення в людський капітал для отримання майбутніх доходів та необхідність інвестування в освіту, що дасть змогу підвищити рівень продуктивності працівників та зростання добробуту всього суспільства. Інвестування коштів в людський капітал забезпечить істотне зростання випуску продукції та стане джерелом зростання майбутніх заробітків чи майбутніх задоволень.

Г. Беккер розглядає людський капітал як наявний запас знань, здібностей і мотивації, що є в кожній людині. Він формується за рахунок інвестицій у людину, серед яких можна назвати витрати на навчання, підготовку на виробництві, витрати на охорону здоров'я, міграцію і пошуки інформації про ціни та доходи [7, с. 39]. Г. Беккер здійснив обрахунок ефективності капіталовкладень в освіті, обґрунтував визначення людського часу як основного економічного ресурсу, описав структуру розподілення особистих доходів. *Освіту як самий важливий актив людського капіталу*, неможна розглядати тільки як засіб збільшення майбутніх доходів. *Освіта – благо тривалого користування, яке має до того ж здатність створювати режим самогенерації або самоінвестування людського капіталу.*

Сьогодні найважливішою складовою системи формування людського капіталу є інвестиції в людський капітал. *Інвестування в розвиток та формування людського капіталу* необхідно проводити в наступних напрямках: *первинна освіта; охорона здоров'я; перекваліфікація; підвищення кваліфікації; забезпечення психологічного фактору людського капіталу; посилення мотивацій працівників до підвищення якості своєї праці; міграція працівників; пошук економічно важливої інформації.*

Здоров'я людей є не тільки медико-біологічною, але й соціальною категорією. Всесвітня організація охорони здоров'я визначає здоров'я як стан повного фізичного, психічного і соціального благополуччя. У розгорнутому формулюванні здоров'я населення – це сукупність елементів, що відображають як абсолютні показники, так і характер і ступінь різних відхилень стану організму в результаті впливу антропогенних та інших чинників навколошнього середовища на рівні індивідуума, групи, колективу і популяції. Стан здоров'я не є щось незмінне, він може поліпшуватися або погіршуватися. Звичайне поліпшення означає розширення активності в усталеному стилі життя, тоді як погіршення стану здоров'я – скорочення потенційної активності.

Загалом, теорія людського капіталу обґрунтувала необхідність та дала поштовх аналізу поведінки людини, функціонування сфери освіти, науки, охорони здоров'я, міграції трудових ресурсів з позицій ринкової економіки. Людський капітал було визнано складовою національного

багатства. Це дало змогу активізувати дослідження якісного вдосконалення людського капіталу, розвитку інтелектуального капіталу в зв'язку з їх перетворенням на інноваційні чинники економічного зростання.

Перехід до концепції людського розвитку. Загалом, в теорії людського капіталу людина розглядалася як найважливіший засіб виробництва і об'єкт найефективніших інвестицій, а зростання валового національного продукту ВНП на душу населення було кінцевою метою розвитку [8, с. 26]. Але соціально-економічна криза, що загострилася наприкінці 1970-х – початку 1980-х років ХХ століття, стало серйозним підґрунтям формування уявлень про людський розвиток як процес реалізація прав та потреб людей, розвиток та реалізація їхніх здібностей і талантів.

Засновниками концепції людського розвитку є зарубіжні вчені-економісти: Махбуб-уль-Хак (1934–1998) (Пакистан) і Амартія Сен (Індія) (професор Гарвардського університету, лауреат Нобелівської премії 1998 р.), які разом із іншими дослідниками проблем розвитку заклали їх теоретичні основи. Значний вплив на формування сучасного погляду на розвиток як на процес, у центрі якого знаходиться людина, максимальна реалізація її прав і потреб, розвиток і реалізація її здібностей, здійснили теоретичні розробки А. Сена, зокрема його концепція розширення вибору. Згідно запропонованої автором концепції розширення можливостей вибору людини [9], максимізацію прибутковості виробничої діяльності не слід розглядати в якості кінцевої мети розвитку. Це є лише засіб розширення можливостей вибору людини у сферах здобуття освіти, професійного розвитку, зміцнення здоров'я тощо. Надання людям більшої свободи вибору такого способу життя, який вони вважають належним і є базовим вектором розвитку. А. Сен підкреслює, що в кінцевому рахунку розвиток – це питання свободи. Розвиток – це процес розширення можливостей, а не підвищення економічного добробуту. Центральна ідея концепції ЛР полягає в тому, що добробут має оцінюватись за можливостями людей вести той спосіб життя, який більшість вважає достойним, а не за рівнем ВВП на душу населення чи іншими макропоказниками [8, с. 7].

Поняття "людський потенціал" і «людський розвиток» та основні положення концепції людського розвитку було розроблене групою експертів Програми розвитку Організації Об'єднаних Націй (ПРООН) і, починаючи з 1990 р. щорічно готовиться глобальна Доповідь про людський розвиток. У Звіті про людський розвиток за 1990 р. було проголошено: «Справжнє багатство народів – це люди» [10, с.10].

Таким чином, в концепції людського капіталу, інвестиції в людину виступають засобом підвищення продуктивності праці та доходу, а в концепції ЛР розвитку високорозвинена благополучна людина визнається за мету, а забезпечення її добробуту – це кінцеве завдання розвитку. Найбільш елементарними передумовами людського розвитку є довге та здорове життя, освіта, гідний спосіб життя та володіння політичними та громадськими свободами, необхідними для участі в житті суспільства.

Людина є носієм людського потенціалу, вирішальним компонентом національного багатства, найважливішим стратегічним ресурсом кожної держави. Людський потенціал визнається не засобом досягнення добробуту людей, а обов'язковим компонентом їх добробуту. *Людський потенціал* охоплює не тільки професійно-кваліфікаційні характеристики, а й рівень освіченості, креативності, мобільності, можливість адаптуватися до соціально-економічних змін у суспільстві та інші кількісно-якісні характеристики їх носіїв [11, с. 93–94].

2. Інформаційне забезпечення побудови системи моніторингу соціально-економічного потенціалу управління станом довкілля

Ефективність еколого-соціально-економічних управлінських рішень може бути досягнута тільки в тому випадку, коли буде розроблено довершену декомпозиційну модель мети у вигляді ієрархії цілей [12]. Ієрархія цілей відповідає на запитання про відповідальність за досягнення різних цілей завдяки встановленню взаємозв'язку між їх ієрархічними рівнями (рис. 1).

Ієрархія цілей – це декомпозиція (розукрупнення) цілей більш

високого рівня в цілі нижчого рівня, тобто це "дерево цілей", де встановлюються конкретні задачі, що лежать в основі конкретних видів робіт. При цьому цілі верхнього рівня не можуть бути досягнуті, поки не будуть реалізовані цілі найближчого нижнього рівня. В основі методики побудови "дерева цілей" лежать філософські концепції подання системи, які певною мірою гарантують повноту структуризації цілей з урахуванням прийнятої концепції. Як уже було зазначено, провідною для сучасного розвитку світового співтовариства є концепція СР та її складові (наприклад, концепція ЛР). Концепція СР – мета світового господарювання, а стратегія СР – це засоби досягнення цієї мети, що передбачають такі складові (завдання):

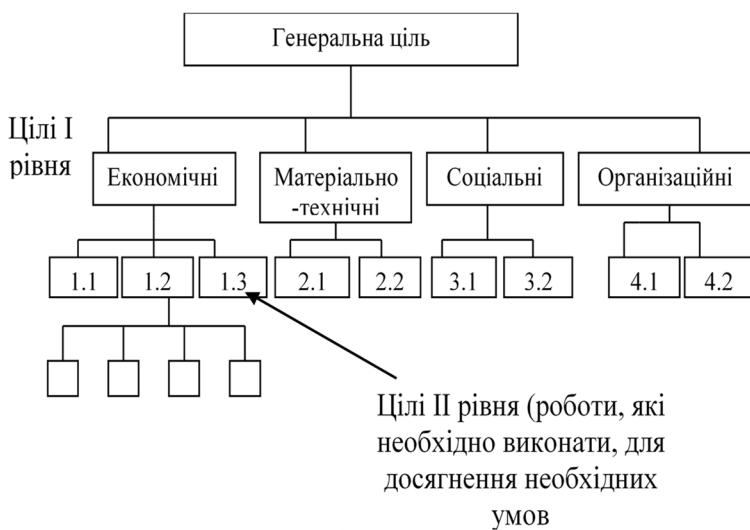


Рис. 1. Приклад побудови "дерева цілей"

- досягнення науково обґрунтованих параметрів якості життя людей; збільшення середньої тривалості життя населення; розвиток соціальної активності міських жителів; планування сім'ї; раціоналізація масштабів і структури особистого споживання; забезпечення для всіх громадян рівних можливостей в здобутті освіти та отриманні медичної допомоги; соціальний захист людей похилого віку, інвалідів та інших соціально-вразливих груп населення;

- збереження і відновлення природних екосистем; стабілізація і поліпшення якості довкілля; зменшення викидів забруднюючих речовин до атмосферу і водойм; скорочення утвореної маси відходів (надзвичайно токсичних); організація їх переробки та утилізації;

- формування нової етики шляхом виховання свідомого ставлення до біосфери як до фундаменту життя; дотримання законів її розвитку, а також зумовлених ними обмежень і заборон;
- створення соціально і екологічно ефективної економіки, що забезпечить гідний рівень життя людей, конкурентоспроможність продукції, гнучкість, швидку адаптацію до змін у ринковій кон'юнктурі та екологічній ситуації;
- формування етики ефективного господарювання, раціонального споживання і природокористування, здорового способу життя.

Наприклад, у вересні 2015 року в рамках 70-ї сесії Генеральної Асамблеї ООН у Нью-Йорку відбувся Саміт ООН зі СР та прийняття Порядку денного розвитку після 2015 року, на якому було затверджено нові орієнтири розвитку. Підсумковим документом Саміту "Перетворення нашого світу: порядок денний у сфері сталого розвитку до 2030 року" було затверджено 17 Цілей Сталого Розвитку (ЦСР) та 169 завдань. Україна, як і інші країни-члени ООН, приєдналася до глобального процесу забезпечення СР і Уряд України представив Національну доповідь "Цілі сталого розвитку: Україна" [13], яка визначає базові показники для досягнення ЦСР (2015-2030 рр.). Тобто, згідно рис. 1 цілями І рівня є 17 цілей, що наведено на рис. 2.



Рис. 2. Скріншот ПРООН, де наведено глобальні цілі СР, що відображені в Національній доповіді "Цілі сталого розвитку: Україна"

Істотна складність побудови системи цілей пов'язана з поняттями "мета" і "засіб досягнення мети". *Мета* – це реальний стан системи, якого намагаються досягти у процесі її функціонування, а засоби досягнення мети – це його шляхи (цілі) і способи (завдання). *Ціль* – це результат, якого можна досягти на протязі доступного для огляду і передбачуваного періоду. Цілі можуть виникати на основі взаємодії між різноманітними зовнішніми та внутрішніми факторами, що впливають на поведінку та розвиток системи.

Як показано на рис. 3–*a*, кожна ціль (на прикладі цілі подолання бідності) головної мети СР реалізується через низку завдань.



Рис. 3 Скріншот Національної доповіді "Цілі сталого розвитку: Україна", щодо формування завдань цілі подолання бідності і значення цільових індикаторів [13]

В міру переходу з верхнього рівня ієрархії цілей на нижній відбувається своєрідний зсув розглянутої вище шкали від цілі-напряму (цілі-ідеалу, мрії) до конкретних цілей, які на нижньому рівні ієрархії (завдань) виражатися у формі очікуваних результатів конкретної роботи з указанівкою показників (значень індикаторів) оцінки її виконання (рис. 3–*b*).

З метою формалізації цілей ЛР використовується багатопараметрична система відповідних статистичних показників та інтегральних індикаторів. На початку 1990-х років експертами Програми розвитку Організації Об'єднаних Націй (ПРООН) було запропоновано розраховувати індекс ЛР (Human Development Index – HDI). Як зазначається в Доповіді ПРООН "Індекси та індикатори людського розвитку: Оновлені статистичні дані 2018" [14], розроблений індекс ЛР втілює в собі прогрес людства, об'єднуючи в одному показнику інформацію про здоров'я, освіту і доходи людей.

На базі Доповіді ПРООН "Індекси та індикатори людського розвитку: Оновлені статистичні дані 2018" *розроблено статистичний Web-додаток "Human Development Indices and Indicators: 2018 Statistical update"* [14]. На сайті представлено моніторинг і аналіз (за період 1990-2018 рр.) складових індексів та індикаторів ЛР за 189 країнам (у тому числі й Україна) і територіями (рис. 4.).

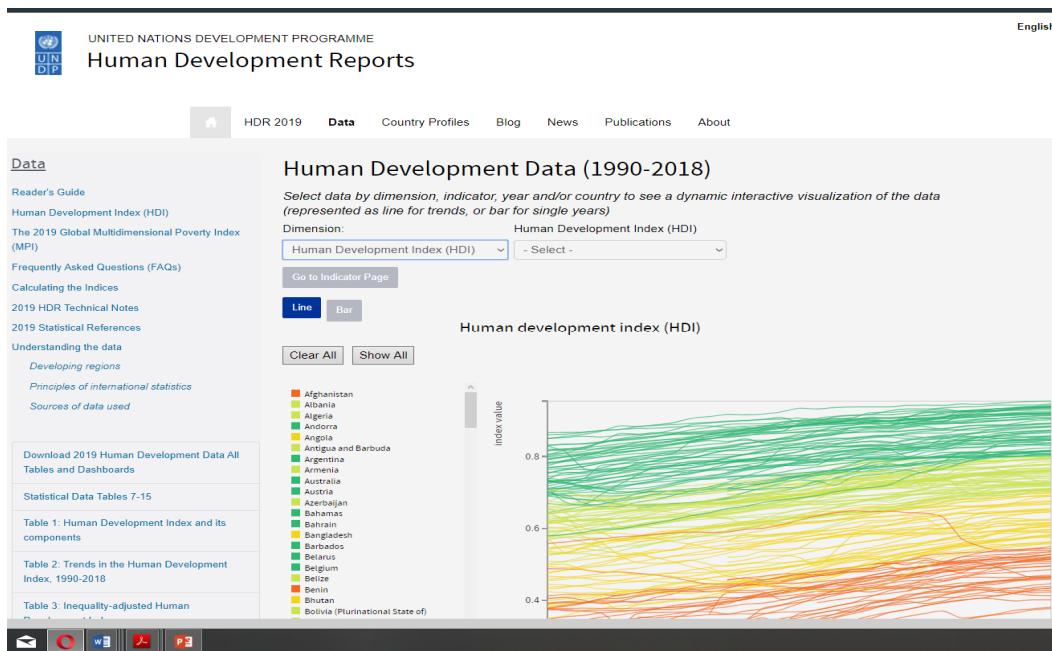


Рис. 4. Головна сторінка Web-додатку "Human Development Indices and Indicators: 2018 Statistical update"

Як показано на рис. 2.4 лінія, що відображає динаміку (1990–2018 рр.) змін значень індексу ЛР для кожної країни позначена тим чи іншим кольором, в залежності від рівня розвитку (табл. 1).

Таблиця 1. Шкала оцінки рівня розвитку країн залежно від значення Індексу ЛР

Значення індексу ЛР	Рівень розвитку країни
0,800 – 1,000	Дуже високий
0,700 – 0,799	Високий
0,555 – 0,699	Середній
0,350 – 0,559	Низький

У дослідженнях проблем сталого і людського розвитку також використовують класифікацію Світового банку, яка базується на головному економічному критерії – суспільній продуктивності праці. За цієї схемою виокремлюються чотири типи країн [16, с.283-284]:

- 1) високорозвинуті країни (ВВП на душу населення понад 20 тис. дол.);
- 2) країни вище середнього рівня розвитку (10-20 тис. дол.);
- 3) країни нижче середнього рівня розвитку (5-10 тис. дол.);
- 4) бідні країни (менше 5 тис. дол.).

Візуалізацію результатів аналітичної обробки даних для кожної країни і території представлено у вигляді таблиць, графіків та профілів, які характеризують позиції (рейтинг) певного країні в групі інших країн за складовими індексу ЛР (рис. 5).



Рис. 5. Приклад візуалізації позиції України за складовими індексу ЛР

Крім того, користувачеві надається можливість ознайомитися з текстовим файлом аналітичної записки по кожній країні, який містить характеристику країни за ключовими індексами ЛР (зокрема: Індекс людського розвитку (ІРЧ), Індекс людського розвитку з урахуванням нерівності (ІHDI), індекс гендерного розвитку (GDI), гендерний характер Індексу нерівності (GII) та розділ за п'ятьма інформаційними панелями.

Щодо України, починаючи з 1999 року, науковцями Інституту демографії та соціальних досліджень ім. М.В. Птухи НАН України та фахівцями Державної служби статистики України здійснюється підготовка й видання щорічного статистичного бюллетеня "Регіональний людський розвиток" як Аналітична доповідь. В 2012 році рішенням Президії НАН України та колегії Державної служби статистики України від 13.06.2012 № 123-м затверджено нову Методику вимірювання регіонального ЛР [15].

Текст Методики у вільному доступі розміщено на офіційному веб-сайті Державної служби статистики України (www.ukrstat.gov.ua) у розділі "Методологія" та "Класифікатори"/"Статистична методологія"/"Демографічна та соціальна статистика" / "Соціальний захист" та на сайті Інституту демографії та соціальних досліджень імені М. В. Птухи (www.idss.org.ua).

До розрахунку регіонального індексу людського розвитку включено 33 показники, об'єднані у 6 блоків відповідно до основних аспектів людського розвитку [15]:

- Відтворення населення;
- Соціальне становище;
- Комфортне життя;
- Добробут;
- Гідна праця;
- Освіта.

Методика також передбачає розрахунок інтегрального показника стану навколошнього природного середовища. Метою розрахунку є оцінка стану навколошнього середовища (далі – інтегральний показник), що визначається за показниками стану атмосферного повітря, земельних та водних ресурсів через антропогенне навантаження і природний стан екосистем регіонів України. Інтегральний показник є абсолютним і визначається об'єктивними результатами моніторингу стану довкілля в

певному регіоні України та не залежить від його місця відносно інших регіонів.

Розрахунки індексу ЛР дозволяють побудувати єдину шкалу, на якій у ранжованому порядку розміщуються всі регіони України. Конкретне значення інтегрального індексу (або індексів, що характеризують окремі аспекти людського розвитку) не мають економічної інтерпретації – важливим є лише місце, яке належить кожному конкретному регіону на єдиній для України шкалі.

Принципи відбору показників [15]:

- Придатність для щорічних розрахунків.
- Забезпеченість наявною інформацією Державної служби статистики України (сайт).
- Забезпечення надійності оцінок на региональному рівні.
- Відповідність специфіці проблем людського розвитку в Україні (диференціація окремих показників за віком, статтю та типом місцевості).
- Однозначність трактування щодо впливу на людський розвиток.
- Відсутність високої кореляції між окремими показниками.
- Достатність статичної та динамічної варіації.

Інформаційною базою розрахунків складових індексу ЛР є щорічна державна статистична звітність, яка представлена на сайті Державної служби статистики України у розділі "статистична інформація" (рис. 6).

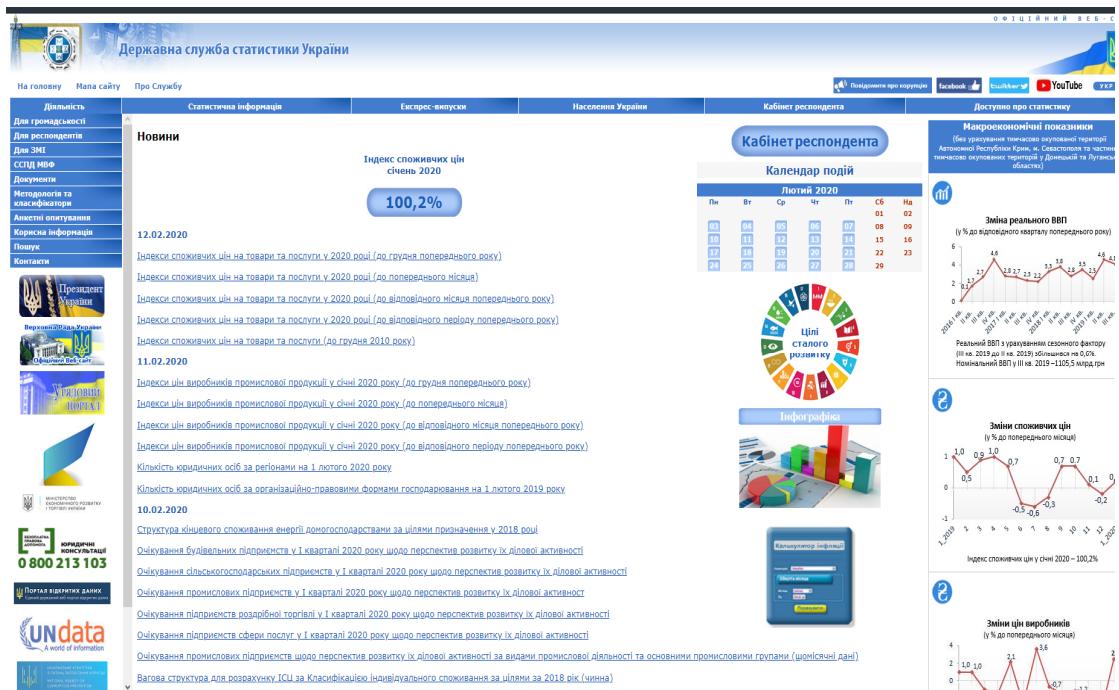


Рис. 6. Головна сторінка сайту Державної служби статистики України

3. Математичні моделі економічного зростання із врахуванням факторів людського капіталу

Важливим питанням на макро- і мікроекономічному рівнях залишаються можливості кількісного та якісного дослідження факторів людського капіталу впливу на продуктивність підприємства та на економічне зростання країн. При визначенні величини людського капіталу застосовують як вартісні (грошові), так і натуральні оцінки [16, с. 28].

Основними розробниками моделей економічного зростання із врахуванням факторів людського капіталу є: *Гері Беккер*, *Пол Ромер* (американський економіст, лауреат премії Ректенвальда 2002 року та Нобелівської премії з економіки 2018 року), *Роберт Лукас* (американський економіст, лауреат Нобелівської премії з економіки 1995 року), *Роберт Барро* (американський економіст, що спеціалізується на класичній макроекономіці, професор економіки імені Пола Варбурга в Гарвардському університеті), *Хавьєр Сала-і-Мартін* (іспанська економіст каталонського походження, професор кафедри економіки імені Джерома і Меттью Гросмана Колумбійського університету в США, автор індексу глобальної конкурентоспроможності). Розглянемо сутність моделей.

Модель оцінки людського капіталу Г. Беккера.

Г. Беккер один із перших переніс поняття людського капіталу на мікрорівень. Людський капітал підприємства він визначив як сукупність навичок, знань і умінь людини [17, с. 8]. Модель оцінки людського капіталу базується на п'яти основних припущеннях [7]:

- 1) виробництво споживчих товарів та народження і виховання дітей вимагають багато часу. Тому з ростом заробітної плати, посилюється ефект заміщення – батьки прагнуть мати менше дітей, бо народження і виховання однієї дитини вимагає все більше альтернативних витрат;
- 2) кожен дорослий індивід максимізує власну функцію корисності, виражену рівнянням

$$V_t = u(c_t) + a(n_t)n_t V_{t+1},$$

де V_t та V_{t+1} – функції корисності батьків та кожного з дітей,

$u(ct)$ –корисність споживання батьків,

nt – кількість дітей,

$a(nt)$ – корисність для батьків, що залежить від кількості дітей (батьківський альтруїзм).

Батьківський альтруїзм в розрахунку на одну дитину зменшується при збільшенні кількості дітей;

3) корисність від отримання додаткового знання пропорційна багажу знань, яким вже володіє дана особа;

4) існує два сектори – виробництво споживчих товарів та виробництво людського капіталу. Причому сектор виробництва/інвестицій в людський капітал потребує більше людського капіталу, ніж сектор виробництва споживчих товарів;

5) теоретичне суспільство складається із однакових індивідів, що живуть протягом двох періодів – дитинства і дорослого життя. Протягом дитинства кожен індивід інвестує в людський капітал, а все доросле життя працює. Індивід обирає кількість дітей, n , який він народжує і виховує протягом дорослого життя. На вихованняожної дитини витрачається v годин та f одиниць продукції. Кожна дитина володіє початковим рівнем людського капіталу H_0 .

Виробнича функція кожного індивіда має наступний вигляд

$$Y = D\{\ell(bH^0 + H)\}^\gamma K^{1-\gamma}, \text{ де } Y = c + fn + \Delta K,$$

де Y - валовий внутрішній продукт (ВВП);

D - продуктивність виробництва;

ℓ - час витрачений кожним дорослим на виробництво споживчих товарів;

b - кількість одиниць H^0 , що еквівалентні одній одиниці H ;

H^0 - запас людського капіталу, отриманий при народженні;

H - обсяг ресурсу людського капіталу;

K - валові інвестиції в фізичний капітал;

c - обсяг споживання дорослим,

f - обсяг споживання однією дитиною даного індивіда,

n - число дітей даного дорослого індивіда,

ΔK - зміна валових інвестицій в фізичний капітал;

Функція виробництва людського капіталу:

$$H_{t+1} = Ah_t(bH^0 + H_t)^\beta,$$

де A - продуктивність інвестицій у людський капітал,
 h_t - час, витрачений на інвестиції в людський капітал;
 β - ефект від масштабів щодо виробництва людського капіталу.

Результатом розв'язання моделі Г. Беккера став висновок про наявність двох рівноважних станів економіки. Перший характеризується низьким рівноважним рівнем людського капіталу, тому віддача від інвестицій для такого суспільства буде низькою – принаймні, нижчою, ніж віддача від інвестицій в фізичний капітал. Для такого рівноважного стану характерним є додатній обсяг фізичного капіталу і майже нульове значення людського капіталу. Крім того, народжуваність в даному суспільстві пояснюється малтузіанською теорією. Другий рівноважний стан економіки, на противагу першому, характеризується значним нагромадженням людського капіталу (H значно більший, ніж отриманий при народженні капітал H_0). Для даного рівноважного стану суспільства характерним є стабільний рівень народжуваності. В цьому суспільстві рівновага наступає, коли зростання рівня споживання кожного індивіда дорівнює віддачі від інвестування в фізичний капітал та віддачі від інвестування в людський капітал. Загалом рівноважний стан, до якого прийде економіка, залежить від розгортання історичних подій та від удачі. Так, наприклад, втрата значної кількості людського капіталу в результаті війни може знищити здатність суспільства до подальшого нагромадження людського капіталу і, таким чином, матиме негативний вплив на подальше економічне зростання даного суспільства в цілому.

Модель П. Ромера. У своїй моделі економічного зростання П. Ромер виділив такі категорії людського капіталу якими, за припущенням, наділений кожен суб'єкт економічної діяльності у його моделі [18]:

L_i – фізичні якості, наприклад координація зору, фізична сила тощо. Дано категорія наближена по своїй природі до поняття "здоров'я" і за припущенням вважається заданою зовні;

E_i – освіта, здобута в початковій та середній школі (базова освіта);

S_i – науковий потенціал, здобутий внаслідок отримання вищої освіти;

Z_i – нагромаджений досвід роботи економічного суб'єкта

У моделі П. Ромера людський капітал, а також фізичний капітал (сукупність засобів виробництва), використовуються для виготовлення матеріальних товарів (споживчих й інвестиційних), а також засобів виробництва.

З переліку категорій людського капіталу L_i , E_i , та Z_i використовуються для виробництва матеріальних благ, а S_i – для виробництва нематеріальних благ (ескізів). Кожен елемент людського капіталу зростає залежно від часу, затраченого на його формування і має здатність зменшуватися, тобто "амортизувати". Значення нематеріальних благ в моделі П. Ромера полягає в тому, що виробництво будь-якого засобу виробництва є неможливим без наявності відповідного йому нематеріального блага.

Теоретична модель економічного зростання Р. Лукаса враховує накопичення знання, як важливий фактор зростання. В стислому варіанті модель має наступний вигляд [19]

$$Y = A \cdot ((1 - \omega) \cdot e \cdot P)^\alpha \cdot K^{1-\alpha},$$

де Y – валовий внутрішній продукт (ВВП);

A – показник продуктивності;

ω – частка часу, що витрачається на накопичення знань;

$(1 - \omega)$ – частка часу, що витрачається на працю;

e – середня продуктивність працівника;

P – чисельність населення;

α – показник еластичності капіталу;

K – наявний капітал.

Модель Р. Барро та Х. Сала-і-Мартіна. У своїй книзі "Економічне зростання" [20] автори доповнили набір виробничих ресурсів, що традиційно складався із робочої сили та фізичного капіталу, людським капіталом. Людський капітал трактується як сукупність навиків, притаманних працівнику. Нагромадження людського капіталу стає джерелом довгострокового економічного зростання, навіть якщо відсутнє покращення технологій виробництва.

Модель попиту на здоров'я населення М. Гросмана. Вищезазначені автори розглядали здоров'я лише як одну із компонентів людського капіталу, але М. Гросман звернув увагу на те, що здоров'я потребує

особливого теоретичного підходу і першим розробив теоретичну та емпіричну модель попиту на здоров'я як важливого фактору економічного зростання країн [21]. Згідно теорії М. Гросмана, здоров'я – це товар тривалого використання, воно є одночасно і інвестиційним, і споживчим товаром, що постійно зношується (амортизується). Процес формування здоров'я на рівні індивіда виражено рівнянням [21]

$$H_t = H_{t-1} + I - \delta ,$$

де H - накопичений ресурс здоров'я;

I - обсяг валових інвестицій в здоров'я;

δ - показник обсягу амортизації здоров'я, його величина залежить від віку індивіда (для літніх індивідів – вищий).

Кожен індивід виробляє два "блага": 1) інвестиції в здоров'я; 2) сукупність усіх інших благ, що приносять йому корисність (інші блага).

Основний ресурс споживача – його час. Індивід вирішує дві задачі вибору: 1) вибір величини інвестицій в здоров'я; 2) вибір між часом, витраченим на роботу та на споживання інших благ.

Основними висновками теоретичної моделі М. Гросмана є:

– збільшення інвестицій в здоров'я знижує кількість непродуктивного часу, а отже, підвищує дохід індивіда;

– зниження вартості ресурсів для інвестицій в здоров'я (внаслідок державної політики субсидування медикаментів, спорту тощо) приводить до підвищення здоров'я населення. Ефект сильніший серед індивідів з нижчим рівнем доходу. Вони можуть навіть заміщати споживання інших благ на інвестування в здоров'я;

– підвищення рівня освіти приводить до підвищення рівня здоров'я при незмінних інвестиціях часу й інших ресурсів. Отже, зростає дохід і якість життя.

4. Основи кореляційно-регресійного аналізу

Соціально-економічне прогнозування є одним з вирішальних наукових факторів формування стратегії і тактики як сталого так і людського розвитку. За терміном дії прогнози поділяють на: короткотермінові (від місяця до року); середньотермінові (від одного до

п'яти років); довготермінові (від п'яти до п'ятнадцяти-двадцяти років); далекоглядні (триваліші за зазначені періоди). Зазначені типи прогнозів відрізняються як за змістом, так і за характером оцінок досліджуваного процесу. *Короткотермінові* прогнози ґрунтуються на припущеннях, що у прогнозованому періоді не станеться ані кількісних, ані якісних суттєвих змін досліджуваного об'єкта. *Середньотермінові* виходять із кількісних і якісних змін у досліджуваному процесі, причому в середньотермінових кількісні зміни домінують над якісними. У *середньотермінових* прогнозах здійснюють кількісно-якісне оцінювання подій, у *довготермінових* – якісно-кількісне. Далекоглядні прогнози виходять суто з якісних змін, причому йдеться переважно про загальні закономірності розвитку досліджуваного об'єкта.

Економетрична модель – це рівняння чи система рівнянь, що описує кореляційно-регресійний зв'язок між економічними показниками, один чи кілька з яких є залежними змінними, інші – незалежними.

У загальному вигляді економетрична модель запишеться так:

$$y = f(x_1, x_2, x_3, \dots, x_m, u),$$

де y – залежна змінна; x_j ($j = 1, 2, \dots, m$) – незалежні змінні; u – стохастична складова, або

$$y_s = f(x_{s1}, x_{s2}, x_{s3}, \dots, x_{sm}, u_s), \quad s = 1, 2, \dots, k,$$

де u_s – стохастична складова s -го рівняння, тобто ця економетрична модель складається з k рівнянь.

Будь-які еколого-економічні показники, зазвичай, перебувають під впливом багатьох випадкових факторів, а тому з математичної точки зору вони повинні розглядатися як випадкові величини. Внаслідок цього залежності між еколого-економічними показниками не є однозначними, не є функціональними. Це означає, що кожному фіксованому значенню однієї екологічної або економічної змінної (або фіксованому набору змінних) відповідає не одне єдине, а множина значень іншої змінної, тобто деякий ймовірнісний розподіл. Тому в еколого-економічних дослідженнях спостерігаються і розглядаються так звані статистичні (або кореляційні) залежності.

Кореляція (від лат. *correlatio* – співвідношення) – це статистична залежність між випадковими величинами, що носить імовірнісний характер. *Кореляційний зв'язок* – це не точна залежність однієї величини від іншої. Числовим значенням однієї змінної ставиться у відповідність середнє декількох значень інших.

За *формою кореляція* може бути *прямолінійною* і *нелінійною* (*криволінійною*), за *напрямком* – *прямою* і *оберненою*. При прямої кореляції залежність між величинами буде прямою: при збільшенні однієї величини, збільшується й інша. При від'ємній кореляції залежність обернена: збільшення однієї величини пов'язано зі зменшенням другої. Ступінь кореляції вимірюється різними показниками зв'язку. Такими показниками є *коефіцієнт кореляції*, *кореляційне відношення* та ін. Дослідження форми зв'язку інколи зумовлює потребу використання нелінійних (криволінійних) *рівнянь регресії*. Це пояснюється тим, що взаємодія між ознаками, що характеризують окремі явища і процеси, нерідко має більш складний характер, ніж просто пропорційні залежності. Характерною особливістю цього зв'язку є те, що рівномірна зміна однієї ознаки супроводжується нерівномірною зміною (збільшенням або зменшенням) значення іншої ознаки.

Основними задачами кореляційного аналізу є:

- вивчення сили зв'язку між двома і більше ознаками досліджуваного об'єкту;
- встановлення факторів, що найбільш суттєво впливають на результативну ознаку;
- виявлення невідомих причинно-наслідкових зв'язків між ознаками об'єкту.

Стисло розглянемо суть кореляційного аналізу, що найбільш детально представлено в роботах [22-25]. Нехай дано систему m нормально розподілених випадкових величин. Для випадкового вектору з складовими (X_1, X_2, \dots, X_m) вибірка об'єму n дає n рядків значень ознак

$$(x_{11}, x_{21}, \dots, x_{m1}), (x_{12}, x_{22}, \dots, x_{m2}), \dots, (x_{1n}, x_{2n}, \dots, x_{mn}).$$

Взаємозв'язок між окремими ознаками описується за допомогою *коваріаційної* або *кореляційної* матриці.

Кореляційна залежність проявляється тільки у масових явищах і може встановлюватися для пари показників (*однофакторна кореляція*) або для декількох показників (*багатофакторна кореляція*).

Одним з основних показників щільноті кореляційного зв'язку показника Y зі всіма факторами X_i ($i = 1, 2, \dots, m$), а також показника ступеня близькості математичної форми зв'язку до вибіркових даних є коефіцієнт багатофакторної кореляції, який має вигляд

$$R = \frac{\sum_{i=1}^n (y_i - \bar{y})(\hat{y}_i - \hat{\bar{y}})}{\sqrt{\sum_{i=1}^n (y_i - \bar{y})^2 \sum_{i=1}^n (\hat{y}_i - \hat{\bar{y}})^2}}. \quad (1)$$

Квадрат коефіцієнта багатофакторної кореляції називається коефіцієнтом детермінації і позначається через R^2 ,

$$R^2 = \frac{\sum_{i=1}^n (\hat{y}_i - \bar{y})^2}{\sum_{i=1}^n (y_i - \bar{y})^2} = 1 - \frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{\sum_{i=1}^n (y_i - \bar{y})^2}.$$

Коефіцієнт детермінації показує, скількома відсотками варіація залежної змінної визначається варіацією незалежних змінних. Недолік коефіцієнта детермінації – те, що його значення не відображає напрямку зв'язку між досліджуваними показниками.

Якщо потрібно порівняти значення коефіцієнтів детермінації в різних моделях, то необхідно коригувати коефіцієнт кореляції з урахуванням кількості факторів X , які входять в різні моделі, тобто зменшити вплив залежності значення коефіцієнта детермінації від кількості факторів. Тоді коефіцієнт детермінації з врахуванням ступенів вільності має вигляд

$$R^2 = 1 - \frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{\sum_{i=1}^n (y_i - \bar{y})^2} \cdot \frac{n-1}{n-m}.$$

Значення коефіцієнта детермінації належать множині $(0; 1)$. Чим ближче ці значення до 1, тим істотний зв'язок між змінними економетричної моделі.

У випадку однофакторної лінійної регресії коефіцієнт кореляції можна обчислити за формулою

$$R_{xy} = \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^n (x_i - \bar{x})^2 \sum_{i=1}^n (y_i - \bar{y})^2}}.$$

Коефіцієнт кореляції R_{xy} – це відносна міра зв'язку між двома факторами. Тому значення коефіцієнта кореляції завжди розташовані між -1 та +1, тобто $-1 < R_{xy} < 1$. Якщо $R_{xy} > 0$, то зв'язок називається додатнім, тобто зі збільшенням значень X значення Y також збільшуються. Якщо $R_{xy} < 0$, то зв'язок називається від'ємним, тобто зі збільшенням значень X значення Y зменшуються. Якісне оцінювання ступеня зв'язку випадкових величин може виконуватися з використанням коефіцієнта кореляції за шкалою Чеддока (табл. 2).

Таблиця 2. Шкала Чеддока

Значення коефіцієнта кореляції	Зв'язок
[0,1...0,3)	незначний
[0,3...0,5)	помірний
[0,5...0,7)	істотний
[0,7...0,9)	високий
[0,9...0,99]	дуже високий
1,0	функціональний

Статистичну значущість коефіцієнта детермінації можна перевірити за допомогою F - критерію

$$F_R = \frac{R^2}{1-R^2} \cdot \frac{n-m-1}{m}.$$

де R^2 – коефіцієнт детермінації;

n – число даних спостережень;

m – число незалежних змінних в рівнянні регресії.

Для заданого рівня значущості α і ступенів вільності $k_1 = m$, $k_2 = n - m - 1$ знаходимо в таблиці F - критерію критичне значення F_{α, k_1, k_2} .

Якщо $F > F_{\alpha, k_1, k_2}$, то можна вважати, що коефіцієнт детермінації статистично значимий і включені у регресію фактори достатньо пояснюють стохастичну залежність показнику.

Множинний коефіцієнт кореляції.

У випадку, коли досліджуваний об'єкт або явище характеризується більш ніж двома ознаками X_1, X_2, \dots, X_k , необхідно вивчати множинні

залежності. Для оцінки сили зв'язку між певною ознакою X_i та усіма іншими ознаками слугує *множинний коефіцієнт кореляції*, який позначається R_i . Для розрахунку множинного коефіцієнта кореляції необхідно:

- побудувати матрицю парних коефіцієнтів кореляції $r_{ij}, i = \overline{1, k}$ між ознаками X_i та X_j

$$A = \begin{pmatrix} 1 & r_{12} & \dots & r_{1k} \\ r_{21} & 1 & \dots & r_{2k} \\ \dots & \dots & \dots & \dots \\ r_{k1} & r_{k2} & \dots & r_{kk} \end{pmatrix};$$

- знайти визначник $|A|$ матриці A та алгебраїчне додовнення A_{ii} елемента r_{ii} цієї матриці;
- розрахувати множинний коефіцієнт кореляції за формулою

$$R_i = \sqrt{1 - \frac{|A|}{A_{ii}}}.$$

Перевірка статистичної значущості множинного коефіцієнта кореляції здійснюється за допомогою t-статистики, яка розраховується за формулою

$$t = \frac{R^2(n - k)}{(1 - R^2)(k - 1)},$$

де n – кількість взаємопов'язаних значень ознак $X_i, i = \overline{1, k}$.

Розраховане значення t-статистики порівнюється з критичним значенням $F_{крит}$. $F_{крит}$ – табличне значення розподілу Фішера, яке також можна знайти за допомогою вбудованої статистичної функції Excel FPACПОБР ($\alpha; l_1; l_2$), де α – обраний дослідником рівень значущості, $l_1; l_2$ – ступені волі, $l_1 = k - 1$; $l_2 = n - k$.

Якщо розраховане значення t-статистики більше критичного $|t| > F_{крит}$, то множинний коефіцієнт кореляції вважається значимим на обраному рівні значущості α .

Групування даних для кореляційного аналізу. Вибікові дані для вивчення кореляційного зв'язку між ознаками X та Y зазвичай мають

вигляд пар їх значень: $(x_1; y_1), (x_2; y_2), \dots, (x_n; y_n)$, x_i – значення величини X , y_i – значення Y , n – кількість пар значень, $i = \overline{1, n}$.

Якщо кількість пар значень достатньо велика (принаймні $n > 20$), то для зручності розрахунків дані групуються.

Розглянемо алгоритм групування даних.

Крок 1. Розбити множини значень X та Y на інтервали k , використовуючи формулу Стерджеса ($k = 1 + 1,4 \ln n$), кількість інтервалів для X та Y може бути різною (означення: k – кількість інтервалів для X ; m – кількість інтервалів для Y).

Крок 2. Побудувати таблицю кореляційної матриці (табл. 3).

Таблиця 3. Шаблон кореляційної матриці

X (інтервали та їх середини) Y (інтервали та їх середини)		$[a_1; a_2)$	$[a_2; a_3)$...	$[a_k; a_{k+1})$	$n_j = \sum_{i=1}^k n_{ij}$
		x_1	x_2	...	x_k	
$[b_1; b_2)$	y_1	n_{11}	n_{21}	...	n_{k1}	n_1
$[b_2; b_3)$	y_2	n_{12}	n_{22}	...	n_{k2}	n_2
...
$[b_m; b_{m+1})$	y_m	n_{1m}	n_{2m}	...	n_{km}	n_m
$n_i = \sum_{j=1}^m n_{ij}$		n_1	n_2	...	n_k	

У першому рядку, розбитому на дві частини, записуються інтервали $[a_i; a_{i+1})$ для X та їх середини x_i . У першому стовпці, розбитому на дві частини, записуються інтервали $[b_j; b_{j+1})$ для Y та їх середини y_j . В центральній частині таблиці записуються частоти n_{ij} – кількість точок, що потратили в прямокутник, обмежений по X інтервалом $[a_i; a_{i+1})$ і по Y інтервалом $[b_j; b_{j+1})$. В останньому рядку таблиці записуються частоти n_i для X – кількість точок, що потратили в прямокутники, які відповідають

інтервалу $[a_i; a_{i+1})$, тобто $n_i = \sum_{j=1}^m n_{ij}$ – сума частот n_{ij} в стовпці з номером i .

В останньому стовпці таблиці записуються частоти n_j для Y – кількість точок, що потратили в прямокутники, які відповідають інтервалу $[b_j; b_{j+1})$,

тобто $n_j = \sum_{i=1}^k n_{ij}$ – сума частот n_{ij} в рядку з номером j .

Приклад 1. Обчислення кореляційної матриці в EXCEL.

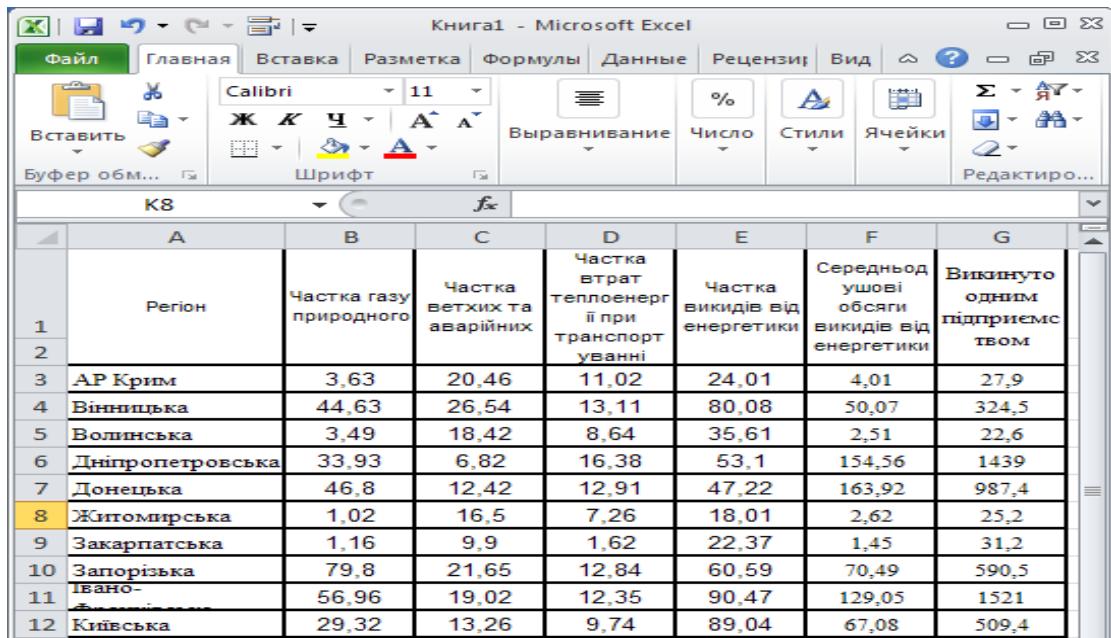
В табл. 4. наведено значення вхідних статистичних показників, що характеризують рівень еколого-енергетичних загроз соціально-економічного потенціалу регіонів України за 2013 рік.

Таблиця 4. Значення регіональних показників еколого-енергетичних загроз соціально-економічного потенціалу України за 2013 рік

Регіон	Частка газу природного, %	Частка ветхих та аварійних мереж, %	Частка втрат теплоенергії при транспортуванні, %	Частка викидів від енергетики %	Середньодушові обсяги викидів від енергетики, тонн	Викинуто одним підприємством, тонн
АР Крим	3,63	20,46	11,02	24,01	4,01	27,9
Вінницька	44,63	26,54	13,11	80,08	50,07	324,5
Волинська	3,49	18,42	8,64	35,61	2,51	22,6
Дніпропетровська	33,93	6,82	16,38	53,10	154,56	1439
Донецька	46,80	12,42	12,91	47,22	163,92	987,4
Житомирська	1,02	16,50	7,26	18,01	2,62	25,2
Закарпатська	1,16	9,90	1,62	22,37	1,45	31,2
Запорізька	79,80	21,65	12,84	60,59	70,49	590,5
Івано-Франківська	56,96	19,02	12,35	90,47	129,05	1521
Київська	29,32	13,26	9,74	89,04	67,08	509,4
Кіровоградська	6,28	14,22	9,90	20,50	3,48	н/д
Луганська	40,48	7,26	5,29	38,94	77,40	274,1
Львівська	21,17	28,70	14,11	45,88	23,77	251,9
Миколаївська	6,15	4,43	10,42	26,42	5,70	54,9
Одеська	1,24	40,22	13,91	22,61	2,67	31,9
Полтавська	0,28	13,11	9,63	29,63	14,15	77,9
Рівненська	7,68	14,27	7,69	50,68	6,53	48,4
Сумська	3,34	41,51	19,67	28,60	7,57	57,2
Тернопільська	1,23	32,17	9,12	10,29	2,00	9,6
Харківська	31,69	27,24	13,73	91,19	66,04	617,1
Херсонська	2,53	10,23	15,71	23,35	1,39	14,5
Хмельницька	11,01	20,29	9,94	53,09	6,60	43,3
Черкаська	9,47	21,21	16,85	48,71	26,72	142,6
Чернівецька	2,32	12,21	6,12	57,89	1,86	22,7
Чернігівська	17,95	27,26	14,97	51,10	21,89	146,3
м. Київ	4,77	14,62	17,25	85,38	10,02	138,3
м. Севастополь	1,60	31,42	10,82	54,83	4,60	39

Алгоритм обчислення матриці кореляції наступний.

Крок 1. Транспортувати табл.4 в книгу Excel на Лист 1 (рис. 7).

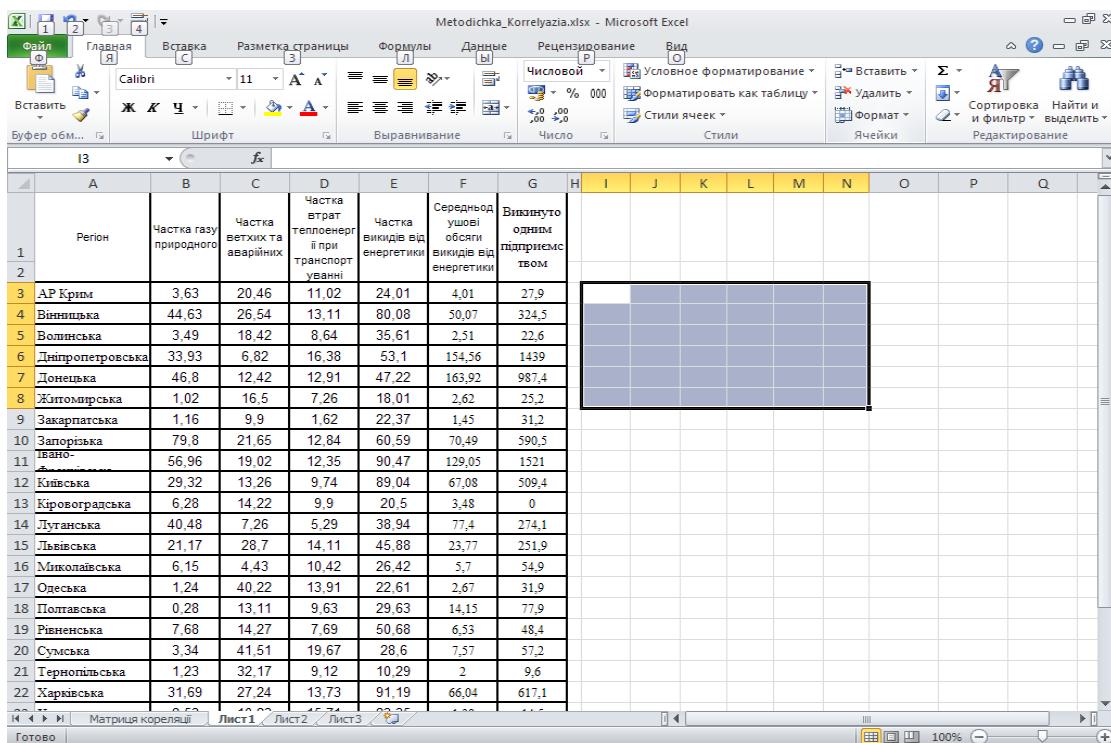


The screenshot shows a Microsoft Excel spreadsheet titled "Книга1 - Microsoft Excel". The table has columns labeled A through G. Column A is "Регіон" (Region). Columns B through G contain numerical values representing different energy metrics. Row 1 contains the column headers. Rows 2 through 12 contain data for various regions. The data includes values like 3,63, 20,46, 11,02, etc., across the columns.

	A	B	C	D	E	F	G
1	Регіон	Частка газу природного	Частка ветхих та аварійних	Частка втрат теплоенергії при транспорт увенні	Частка викидів від енергетики	Середньод ушові обсяги викидів від енергетики	Викинуто одним підприємс твом
2							
3	АР Крим	3,63	20,46	11,02	24,01	4,01	27,9
4	Вінницька	44,63	26,54	13,11	80,08	50,07	324,5
5	Волинська	3,49	18,42	8,64	35,61	2,51	22,6
6	Дніпропетровська	33,93	6,82	16,38	53,1	154,56	1439
7	Донецька	46,8	12,42	12,91	47,22	163,92	987,4
8	Житомирська	1,02	16,5	7,26	18,01	2,62	25,2
9	Закарпатська	1,16	9,9	1,62	22,37	1,45	31,2
10	Запорізька	79,8	21,65	12,84	60,59	70,49	590,5
11	Івано- франківська	56,96	19,02	12,35	90,47	129,05	1521
12	Київська	29,32	13,26	9,74	89,04	67,08	509,4

Рис. 7. Скріншот із вхідними даними

Крок 2. Визначити діапазон комірок у які будуть вводитись значення парних коефіцієнтів кореляції (для 6 показників – розмір діапазону 6×6) (рис. 8).



The screenshot shows the same Microsoft Excel spreadsheet as in Figure 7, but with a 6x6 range of cells highlighted in blue. This range starts from cell H3 and extends down to cell M18. The rest of the table remains visible below the highlighted area.

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q
1	Регіон	Частка газу природного	Частка ветхих та аварійних	Частка втрат теплоенергії при транспорт увенні	Частка викидів від енергетики	Середньод ушові обсяги викидів від енергетики	Викинуто одним підприємс твом										
2																	
3	АР Крим	3,63	20,46	11,02	24,01	4,01	27,9										
4	Вінницька	44,63	26,54	13,11	80,08	50,07	324,5										
5	Волинська	3,49	18,42	8,64	35,61	2,51	22,6										
6	Дніпропетровська	33,93	6,82	16,38	53,1	154,56	1439										
7	Донецька	46,8	12,42	12,91	47,22	163,92	987,4										
8	Житомирська	1,02	16,5	7,26	18,01	2,62	25,2										
9	Закарпатська	1,16	9,9	1,62	22,37	1,45	31,2										
10	Запорізька	79,8	21,65	12,84	60,59	70,49	590,5										
11	Івано- франківська	56,96	19,02	12,35	90,47	129,05	1521										
12	Київська	29,32	13,26	9,74	89,04	67,08	509,4										
13	Кіровоградська	6,28	14,22	9,9	20,5	3,48	0										
14	Луганська	40,48	7,26	5,29	38,94	77,4	274,1										
15	Львівська	21,17	28,7	14,11	45,88	23,77	251,9										
16	Миколаївська	6,15	4,43	10,42	26,42	5,7	54,9										
17	Одеська	1,24	40,22	13,91	22,61	2,67	31,9										
18	Полтавська	0,28	13,11	9,63	29,63	14,15	77,9										
19	Рівненська	7,68	14,27	7,69	50,68	6,53	48,4										
20	Сумська	3,34	41,51	19,67	28,6	7,57	57,2										
21	Тернопільська	1,23	32,17	9,12	10,29	2	9,6										
22	Харківська	31,69	27,24	13,73	91,19	66,04	617,1										

Рис. 8. Скріншот із виділенім діапазоном

Крок 3. Поставити курсор у верхню ліву комірку виділеного діапазону (див. рис. 1.8 – комірка I3) для розрахунку коефіцієнта кореляції (формула (1)) першого показника самого із собою.

Крок 4. В рядку формул набрати формулу «=КОРРЕЛ(« (рис. 9).

The screenshot shows a Microsoft Excel window titled 'Книга1 - Microsoft Excel'. The formula bar at the top has the text '=КОРРЕЛ(' entered. Below the formula bar is a table with columns labeled A through I. The first column (A) contains the header 'Період'. The second column (B) contains 'Частка газу природного'. The third column (C) contains 'Частка ветхих та аварійних'. The fourth column (D) contains 'Частка втрат теплоенергії при транспорт уванні'. The fifth column (E) contains 'Частка викидів від енергетики'. The sixth column (F) contains 'Середньод ушові обсяги викидів від енергетики'. The seventh column (G) contains 'Викинуто одним підприємс твом'. The eighth column (H) is empty. The ninth column (I) contains the formula '=КОРРЕЛ('.

A	B	C	D	E	F	G	H	I
1 Період	Частка газу природного	Частка ветхих та аварійних	Частка втрат теплоенергії при транспорт уванні	Частка викидів від енергетики	Середньод ушові обсяги викидів від енергетики	Викинуто одним підприємс твом		
3 АР Крим	3,63	20,46	11,02	24,01	4,01	27,9		=КОРРЕЛ(
4 Вінницька	44,63	26,54	13,11	80,08	50,07	324,5		
5 Волинська	3,49	18,42	8,64	35,61	2,51	22,6		
6 Тернопільська	33,93	6,89	16,38	53,1	15,156	1130		

Рис. 9. Скріншот вводу функції «=КОРРЕЛ(«

Крок 5. Виділити комірки з числовими даними первого показника (стовбець В3:В29) і поставити крапку з комою «;» в рядку формули (рис. 10).

The screenshot shows a Microsoft Excel window titled 'Книга1 - Microsoft Excel'. The formula bar at the top has the formula '=КОРРЕЛ(B3:B29)' entered. Below the formula bar is a table with columns labeled A through I. The first column (A) contains the header 'Період'. The second column (B) contains 'Частка газу природного'. The third column (C) contains 'Частка ветхих та аварійних'. The fourth column (D) contains 'Частка втрат теплоенергії при транспорт уванні'. The fifth column (E) contains 'Частка викидів від енергетики'. The sixth column (F) contains 'Середньод ушові обсяги викидів від енергетики'. The seventh column (G) contains 'Викинуто одним підприємс твом'. The eighth column (H) is empty. The ninth column (I) contains '(B3:B29)'. The range B3:B29 is highlighted with a blue dashed selection border.

A	B	C	D	E	F	G	H	I
1 Період	Частка газу природного	Частка ветхих та аварійних	Частка втрат теплоенергії при транспорт уванні	Частка викидів від енергетики	Середньод ушові обсяги викидів від енергетики	Викинуто одним підприємс твом		
3 АР Крим	3,63	20,46	11,02	24,01	4,01	27,9		(B3:B29)
4 Вінницька	44,63	26,54	13,11	80,08	50,07	324,5		
5 Волинська	3,49	18,42	8,64	35,61	2,51	22,6		
6 Тернопільська	33,93	6,89	16,38	53,1	15,156	1130		

Рис. 10. Скріншот виділення даних

Крок 6. Знову виділити комірки з числовими даними первого показника (стовбець В3:В29) (номер відповідає другому індексу коефіцієнту кореляції), закрити дужку і натиснути Enter (рис. 11).

The screenshot shows a Microsoft Excel spreadsheet titled "Книга1 - Microsoft Excel". The formula bar displays the formula =КОРРЕЛ(B3:B29;B3:B29). The table has columns labeled: Регіон, Частка газу природного, Частка ветхих та аварійних, Частка втрат теплоенергії при транспорт уванні, Частка викидів від енергетики, Середньодушові обсяги викидів від енергетики, and Викинуто одним підприємством. The value in cell I3 is 1,000.

	A	B	C	D	E	F	G	H	I	J
1	Регіон	Частка газу природного	Частка ветхих та аварійних	Частка втрат теплоенергії при транспорт уванні	Частка викидів від енергетики	Середньодушові обсяги викидів від енергетики	Викинуто одним підприємством			
3	АР Крим	3,63	20,46	11,02	24,01	4,01	27,9		1,000	
4	Вінницька	44,63	26,54	13,11	80,08	50,07	324,5			
5	Волинська	3,49	18,42	8,64	35,61	2,51	22,6			
6	Дніпропетровська	33,93	6,82	16,38	53,1	154,56	1439			
7	Донецька	46,8	12,42	12,91	47,22	163,92	987,4			
8	Житомирська	1,02	16,5	7,26	18,01	2,62	25,2			
9	Закарпатська	1,16	9,9	1,62	22,37	1,45	31,2			
10	Запорізька	79,8	21,65	12,84	60,59	70,49	590,5			

Рис. 11. Скріншот розрахунку першого коефіцієнта кореляції

Крок 7. Переставити курсор на наступну комірку (J3) і перейти до кроку 3, але при цьому змінювати комірку в залежності від індексів коефіцієнту кореляції, який розраховується. Повторювати кроки 3-7 до повного заповнення матриці кореляції (рис. 12).

The screenshot shows a Microsoft Excel spreadsheet titled "Metodichka_Korrelacyja.xlsx - Microsoft Excel". The formula bar displays the formula =КОРРЕЛ(B3:B29;B3:B29). The table has columns labeled: Регіон, Частка газу природного, Частка ветхих та аварійних, Частка втрат теплоенергії при транспорт уванні, Частка викидів від енергетики, Середньодушові обсяги викидів від енергетики, and Викинуто одним підприємством. The matrix of correlation coefficients is shown in the range P8:R10.

	A	B	C	D	E	F	G	H	I	J	K	L	M	N
1	Регіон	Частка газу природного	Частка ветхих та аварійних	Частка втрат теплоенергії при транспорт уванні	Частка викидів від енергетики	Середньодушові обсяги викидів від енергетики	Викинуто одним підприємством							
3	АР Крим	3,63	20,46	11,02	24,01	4,01	27,9	1,000	-0,088	0,156	0,565	0,785	0,743	
4	Вінницька	44,63	26,54	13,11	80,08	50,07	324,5	-0,088	1,000	0,422	-0,017	-0,253	-0,192	
5	Волинська	3,49	18,42	8,64	35,61	2,51	22,6	0,156	0,422	1,000	0,256	0,214	0,266	
6	Дніпропетровська	33,93	6,82	16,38	53,1	154,56	1439	0,565	-0,017	0,256	1,000	0,473	0,539	
7	Донецька	46,8	12,42	12,91	47,22	163,92	987,4	0,785	-0,253	0,214	0,473	1,000	0,941	
8	Житомирська	1,02	16,5	7,26	18,01	2,62	25,2	0,743	-0,192	0,266	0,539	0,941	1,000	
9	Закарпатська	1,16	9,9	1,62	22,37	1,45	31,2							
10	Запорізька	79,8	21,65	12,84	60,59	70,49	590,5							

Рис. 12. Скріншот побудованої матриці кореляції

Усунення мультиколінеарності. Термін "мультиколінеарність" означає, що в багатофакторній регресійній моделі дві або більше, на перший погляд, незалежні змінні (фактори) насправді пов'язані між собою лінійною залежністю, тобто мають високий ступінь кореляції ($r_{x_i x_j} \rightarrow 1, i \neq j$) [23-25]. Якщо в регресії присутнє явище

мультиколінеарності, то не виконується одна з чотирьох умов, а саме не виконується умова $|X^T X| \neq 0$, яка необхідна для оцінювання параметрів загальної лінійної моделі методом найменших квадратів. За наявності мультиколінеарності має місце $|X^T X| \rightarrow 0$, отже, при визначенні параметрів моделі має місце ділення на величини, близькі до нуля і оцінки параметрів моделі можуть дуже сильно відхилятись від реальних.

Для дослідження наявності мультиколінеарності широко застосовується метод *Фарара-Глобера*. Цей метод триступеневий, тобто має три види статистичних критеріїв, за якими перевіряється мультиколінеарність:

- 1) усього масиву незалежних змінних (χ^2 – «хі-квадрат);
- 2) кожної незалежної змінної з рештою змінних (F - критерій);
- 3)ожної пари незалежних змінних (t - критерій).

Усі ці критерії при порівнянні з їх критичними значеннями дають змогу робити конкретні висновки щодо наявності чи відсутності мультиколінеарності незалежних змінних.

Метод Фарара-Глобера зручно подати у вигляді покрокового алгоритму.

Крок 1. Стандартизація (нормалізація) змінних.

Позначимо вектори незалежних змінних економетричної моделі через $x_1, x_2, x_3, \dots, x_m$. Елементи стандартизованих векторів обчислимо за формулами

$$1) x_{ik}^* = \frac{x_{ik} - \bar{x}_k}{\sigma_{x_k}}, \quad 2) x_{ik}^* = \frac{x_{ik} - \bar{x}_k}{\sqrt{n \sigma_{x_k}^2}}, \quad (2)$$

де n - число спостережень ($i = 1, 2, \dots, n$);

m - число незалежних змінних, ($k=1, 2, \dots, m$);

\bar{x}_k - середнє арифметичне k -ї незалежної змінної;

σ_{x_k} - дисперсія k -ї незалежної змінної.

Крок 2. Побудова кореляційної матриці відбувається за формулами

$$1) r = \frac{1}{n} X^{*T} X^*, \quad 2) r = X^{*T} X^*, \quad (3)$$

де X^* – матриця стандартизованих незалежних (пояснювальних)

змінних;

X^{*T} – матриця, транспонована до матриці X^* .

Крок 3. Визначення критерію χ^2 («хі-квадрат») за формулою

$$\chi^2 = - \left[n - 1 - \frac{1}{6} (2m + 5) \right] \ln |r|, \quad (4)$$

де $|r|$ – визначник кореляційної матриці r .

Значення цього критеріїв порівнюються з табличним при $m(m-1)/2$ ступені вільності і рівні значущості α . Якщо $\chi^2_{\text{факт}} > \chi^2_{\text{табл}}$, то в масиві незалежних змінних існує мультиколінеарність.

Крок 4. Обчислення оберненої матриці

$$r^{-1} = \left(\frac{1}{n} X^{*T} X^* \right)^{-1}. \quad (5)$$

Крок 5. Обчислення F -критеріїв:

$$F_k = (z_{kk} - 1) \frac{n-m}{m-1}, \quad (6)$$

де z_{kk} – діагональні елементи матриці r^{-1} . Фактичні значення критеріїв порівнюються з табличними при $k_1=m-1$ і $k_2=n-m$ ступенях вільності та рівні значущості α . Якщо $F_k \text{ факт} > F_{\text{табл}}$, то відповідна k -та незалежна змінна мультиколінеарна з іншими.

Коефіцієнт детермінації для кожної змінної розраховується за формулою

$$R_{x_k}^2 = 1 - \frac{1}{z_{kk}}.$$

Крок 6. Знаходження частинних коефіцієнтів кореляції

$$r_{kj,1,12,\dots,m} = - \frac{z_{kj}}{\sqrt{z_{kk} \cdot z_{jj}}}, \quad (7)$$

де z_{kj} – елемент матриці r^{-1} , що міститься в k -у рядку і j -у стовпці; z_{kk} і z_{jj} – діагональні елементи матриці r^{-1} .

Крок 7. Обчислення t -критеріїв

$$t_{kj} = \frac{r_{kj,1,12,\dots,m} \sqrt{n-m-1}}{1 - r_{kj,1,12,\dots,m}}. \quad (8)$$

Фактичні значення критеріїв t_{kj} порівнюються з табличними при рівні значущості α і $n-m-1$ ступенях вільності. Якщо $t_{kj} \text{ факт} > t_{\text{табл}}$, то між незалежними змінними x_k і x_j існує мультиколінеарність.

Приклад 2. Обчислення мультиколінеарності в Excel. Алгоритм розрахунку наявності мультиколінеарності передбачає виконання 12 основних кроків.

Крок 1. Обчислення рядка середніх значень. Під кожним стовпчиком регіонального показника (індикатора) необхідно за допомогою функції СРЗНАЧ розрахувати середнє значення відповідного діапазону. Так, наприклад, для розрахунку середнього значення першого індикатора необхідно виділити комірку B31 та ввести формулу «=СРЗНАЧ(B3:B29)», де B3:B29 – діапазон значень першого індикатора (рис. 13).

The screenshot shows a Microsoft Excel spreadsheet titled "MultiKollinearist.xlsx". The active cell is B31, which contains the formula "=СРЗНАЧ(B3:B29)". The formula bar also displays this formula. The spreadsheet contains data from rows 17 to 30, with columns A through G. Column A lists regions: Одеська, Полтавська, Рівненська, Сумська, Тернопільська, Харківська, Херсонська, Хмельницька, Черкаська, Чернівецька, Чернігівська, and м. Київ. Columns B through G contain numerical values. Row 31 is highlighted in yellow and contains the formula for calculating the average of the first column's values.

Рис. 13. Скріншот розрахунку середнього значення індикатора

Виділивши комірку B31 і затиснувши ліву кнопку миші, перетягнути курсор вправо для вибору шести клітинок (відповідно до кількості індикаторів) і натиснути комбінацію клавіш «Ctrl»+«R» для копіювання формулі (рис. 14).

This screenshot shows the same Microsoft Excel spreadsheet as in Figure 13. The formula in cell B31 has been copied to cells B32 through B37. The formula bar now shows the copied formula: "=СРЗНАЧ(B3:B29)". The cells B32 through B37 now contain the calculated average values for each row, starting from 17.40481 in B32.

Рис. 14. Скріншот копіювання формулі

Крок 2. Розрахунок дисперсії. Під кожною коміркою із середнім значенням індикатора необхідно розрахувати дисперсію. Для першого стовпчика вибираємо комірку В32 і вводимо формулу «=ДИСПР(В3:В29)». Analogічно копіюємо формулу в наступні п'ять правих комірок (рис. 15).

The screenshot shows a Microsoft Excel spreadsheet titled "MultiKollinearist.xlsx". The active cell is B32, which contains the formula `=ДИСПР(В3:В29)`. The formula bar also displays this formula. The cell B31 contains the value `17,40481`. The cell A32 is highlighted with a blue border. The data range A21:G29 is visible, showing various regional names and their corresponding values across columns C through G. The ribbon at the top has tabs like "Главная", "Формулы", and "Редактирование". The status bar at the bottom shows "Лист1" and "100%".

Рис. 15. Скріншот визначення дисперсії

Крок 3. Формування нормалізованої матриці (формула (2)). Нормалізована матриця має той самий розмір як і вихідна (27 регіонів \times 6 факторів). Нормалізовану матрицю бажано розмістити на одному рівні з вихідною, наприклад, у діапазоні комірок I3:N29. Для розрахунку першого нормалізованого значення необхідно вибрести комірку I3 та задати формулу $\langle=(B3-B\$31)/\text{КОРЕНЬ}(27*B\$32)\rangle$, де $B\$31$ – зафіксоване середнє значення, $B\$32$ – зафіксована дисперсія, 27 – кількість об'єктів.

Крок 4. Розрахунок кореляційної матриці на основі нормалізованих даних (формула (3)). Для зберігання значень коефіцієнтів кореляції необхідно визначити квадратний діапазон комірок розміром 6×6 (відповідно до кількості індикаторів). Наприклад, кореляційна матриця буде зберігатися в комірках P3:U8. Далі виділимо комірку P3 і, затиснувши ліву клавішу миші, розширимо вибір до клітинки U8. В рядку формул задамо вираз $\langle=\text{МУМНОЖ}(\text{ТРАНСП}(I3:N29);I3:N29)\rangle$ (рис. 16 – а). Після цього необхідно одночасно натиснути комбінацію клавіш «Ctrl + Shift + Enter» (рис. 16 – б).

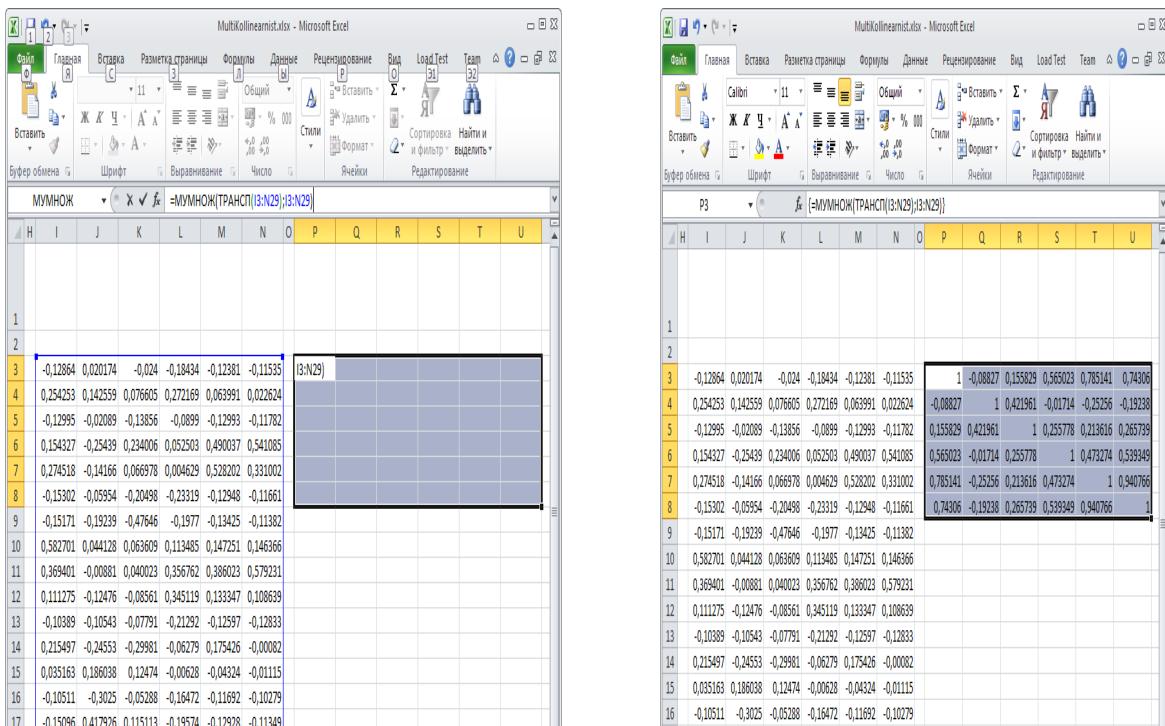


Рис. 16. Скріншот зберігання значень коефіцієнтів кореляції

Крок 5. Розрахунок детермінанта кореляційної матриці. Детермінант квадратної матриці в Excel визначається за допомогою функції МОПРЕД. Наприклад, виділимо комірку P10 і введемо формулу «=МОПРЕД(P3:U8)», де P3:U8 – діапазон комірок кореляційної матриці (рис. 17).

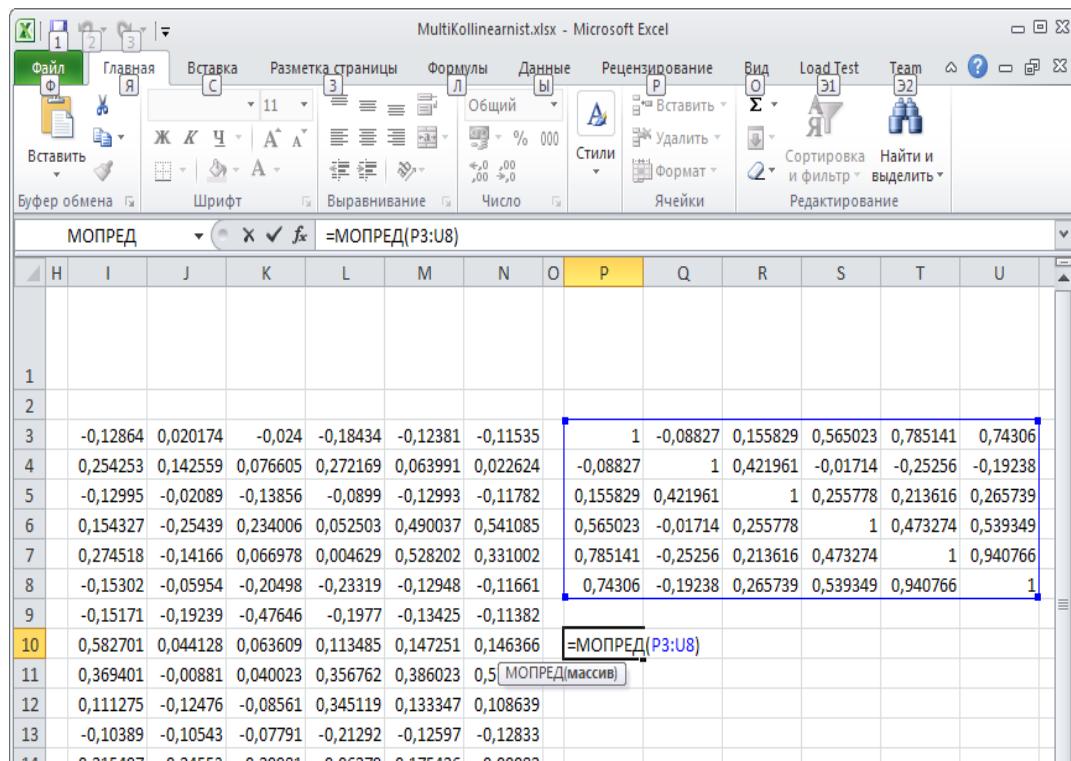


Рис. 17. Скриншот ввізначення детермінанта кореляційної матриці

Після натиснення клавіші «Enter» у комірці Р10 з'явиться значення детермінанту кореляційної матриці.

Крок 6. Обчислення критерію X^2 (формула (4)). Для розрахунку фактичного значення X^2 - критерію у вибрану комірку, наприклад Р12, вводять формулу « $=-(27-1-(2*6+5)/6)*LN(P10)$ », де 27 – кількість об'єктів, $2*6$ – подвоєна кількість факторів, Р10 – значення детермінанту кореляційної матриці (рис. 18).

The screenshot shows a Microsoft Excel spreadsheet titled "MultiKollinearist.xlsx". The formula bar at the top displays the formula $=-(27-1-(2*6+5)/6)*LN(P10)$. The main table has columns labeled I through U. Cell P12 contains the value 0,015791, which is highlighted with a blue border. The formula bar also shows the result of the formula, which is 0,015791.

I	J	K	L	M	N	O	P	Q	R	S	T	U		
1														
2														
3	-0,12864	0,020174	-0,024	-0,18434	-0,12381	-0,11535		1	-0,08827	0,155829	0,565023	0,785141	0,74306	
4	0,254253	0,142559	0,076605	0,272169	0,063991	0,022624			-0,08827	1	0,421961	-0,01714	-0,25256	-0,19238
5	-0,12995	-0,02089	-0,13856	-0,0899	-0,12993	-0,11782			0,155829	0,421961	1	0,255778	0,213616	0,265739
6	0,154327	-0,25439	0,234006	0,052503	0,490037	0,541085			0,565023	-0,01714	0,255778	1	0,473274	0,539349
7	0,274518	-0,14166	0,066978	0,004629	0,528202	0,331002			0,785141	-0,25256	0,213616	0,473274	1	0,940766
8	-0,15302	-0,05954	-0,20498	-0,23319	-0,12948	-0,11661				0,74306	-0,19238	0,265739	0,539349	0,940766
9	-0,15171	-0,19239	-0,47646	-0,1977	-0,13425	-0,11382					0,015791			
10	0,582701	0,044128	0,063609	0,113485	0,147251	0,146366								
11	0,369401	-0,00881	0,040023	0,356762	0,386023	0,579231								
12	0,111275	-0,12476	-0,08561	0,345119	0,133347	0,108639								
13	-0,10389	-0,10543	-0,07791	-0,21292	-0,12597	-0,12833								
14	0,215497	-0,24553	-0,29981	-0,06279	0,175426	-0,00082								

Рис. 18. Скріншот розрахунку фактичного X^2 - критерію

Для визначення мультиколінеарності по всім індикаторам розраховане значення необхідно порівняти з табличним при визначеному ступені вільності та заданому рівні значимості. Табличне значення розрахуємо до комірки R12 за допомогою формули «=ХИ2.ОБР(0,05;6*(6-1)/2)».

Порівнюючи фактичне значення (комірка Р12) з табличним значенням (комірка R12) X^2 - критерію, приходимо до висновку, що між індикаторами існує мультиколінеарність.

Крок 7. Визначення оберненої матриці (формула (5)). Для визначення оберненої матриці до відповідної матриці кореляції нормалізованого набору даних необхідно виділити квадратний діапазон комірок, наприклад, Р14:U19. В рядку формул ввести вираз «=МОБР(Р3:U8)» і одночасно натиснути комбінацію клавіш «Ctrl»+«Shift»+«Enter» (рис. 19).

I	J	K	L	M	N	O	P	Q	R	S	T	U		
2														
3	-0,12864	0,020174	-0,024	-0,18434	-0,12381	-0,11535	1	-0,08827	0,155829	0,565023	0,785141	0,74306		
4	0,254253	0,142559	0,076605	0,272169	0,063991	0,022624		-0,08827	1	0,421961	-0,01714	-0,25256	-0,19238	
5	-0,12995	-0,02089	-0,13856	-0,0899	-0,12993	-0,11782		0,155829	0,421961	1	0,255778	0,213616	0,265739	
6	0,154327	-0,25439	0,234006	0,052503	0,490037	0,541085		0,565023	-0,01714	0,255778	1	0,473274	0,539349	
7	0,274518	-0,14166	0,066978	0,004629	0,528202	0,331002		0,785141	-0,25256	0,213616	0,473274	1	0,940766	
8	-0,15302	-0,05954	-0,20498	-0,23319	-0,12948	-0,11661		0,74306	-0,19238	0,265739	0,539349	0,940766	1	
9	-0,15171	-0,19239	-0,47646	-0,1977	-0,13425	-0,11382								
10	0,582701	0,044128	0,063609	0,113485	0,147251	0,146366		0,015791						
11	0,369401	-0,00881	0,040023	0,356762	0,386023	0,579231								
12	0,111275	-0,12476	-0,08561	0,345119	0,133347	0,108639	96,10195		7,260944					
13	-0,10389	-0,10543	-0,07791	-0,21292	-0,12597	-0,12833								
14	0,215497	-0,24553	-0,29981	-0,06279	0,175426	-0,00082	(P3:U8)	-0,5102	0,398022	-0,88	-2,85177	0,573478		
15	0,035163	0,186038	0,12474	-0,00628	-0,04324	-0,01115		-0,5102	1,515907	-0,76385	0,107857	1,116138	-0,23448	
16	-0,10511	-0,3025	-0,05288	-0,16472	-0,11692	-0,10279		0,398022	-0,76385	1,498767	-0,25461	-0,3749	-0,35096	
17	-0,15096	0,417926	0,115113	-0,19574	-0,12928	-0,11349			-0,88	0,107857	-0,25461	1,696587	1,149283	-1,25396
18	-0,15993	-0,12778	-0,0909	-0,13859	-0,08247	-0,09209			-2,85177	1,116138	-0,3749	1,149283	11,7519	-9,24227
19	-0,09082	-0,10443	-0,18428	0,032799	-0,11354	-0,10582		0,573478	-0,23448	-0,35096	-1,25396	-9,24227	9,993164	
20	-0,13135	0,443892	0,392369	-0,14697	-0,1093	-0,10172								

Рис. 19. Скріншот визначення оберненої матриці

Крок 8. Обчислення F-критерію (Фішера). Фактичні значення F-критерію розраховується за формулою (6) і зберігається в масиві комірок P21:U21 (рис. 20).

I	J	K	L	M	N	O	P	Q	R	S	T	U	V	
11	0,369401	-0,00881	0,040023	0,356762	0,386023	0,579231								
12	0,111275	-0,12476	-0,08561	0,345119	0,133347	0,108639	96,10195		7,260944					
13	-0,10389	-0,10543	-0,07791	-0,21292	-0,12597	-0,12833								
14	0,215497	-0,24553	-0,29981	-0,06279	0,175426	-0,00082		3,203071	-0,5102	0,398022	-0,88	-2,85177	0,573478	
15	0,035163	0,186038	0,12474	-0,00628	-0,04324	-0,01115			-0,5102	1,515907	-0,76385	0,107857	1,116138	-0,23448
16	-0,10511	-0,3025	-0,05288	-0,16472	-0,11692	-0,10279		0,398022	-0,76385	1,498767	-0,25461	-0,3749	-0,35096	
17	-0,15096	0,417926	0,115113	-0,19574	-0,12928	-0,11349			-0,88	0,107857	-0,25461	1,696587	1,149283	-1,25396
18	-0,15993	-0,12778	-0,0909	-0,13859	-0,08247	-0,09209			-2,85177	1,116138	-0,3749	1,149283	11,7519	-9,24227
19	-0,09082	-0,10443	-0,18428	0,032799	-0,11354	-0,10582		0,573478	-0,23448	-0,35096	-1,25396	-9,24227	9,993164	
20	-0,13135	0,443892	0,392369	-0,14697	-0,1093	-0,10172								
21	-0,15105	0,255886	-0,11545	-0,29605	-0,13201	-0,12387			9,2529	2,166808	2,09482	2,925667	45,15799	=U21
22	0,133408	0,156649	0,106449	0,362624	0,129107	0,15874								
23	-0,13891	-0,18575	0,201755	-0,18972	-0,1345	-0,12159								
24	-0,05972	0,016752	-0,07598	0,052421	-0,11325	-0,10819								
25	-0,0741	0,035271	-0,256629	0,01676	-0,03122	-0,062								
26	-0,14088	-0,14589	-0,25986	0,091502	-0,13258	-0,11777								
27	0,005091	0,157052	0,166136	0,036219	-0,05091	-0,06027								
28	-0,118	-0,09738	0,275883	0,31532	-0,09931	-0,064								
29	-0,1476	0,240789	-0,03362	0,066588	-0,12141	-0,11019								

Рис. 20. Скріншот розрахунку значення F-критерію

При цьому всі значення F-критерію розраховуються окремо, оскільки в розрахунках необхідно використовувати діагональні елементи оберненої матриці. Перше значення F-критерію для комірки P21 обчислюється за допомогою формули «=(P14-1)*(27-6)/(6-1)» або «=(P14-1)*21/5». На рис. 20 наведено формулу для обчислення значення останнього F-критерію.

Критичне (табличне) значення F-критерію розраховується за допомогою статистичної функції «=FPACSPOB(0,05;(6-1);(25-6))».

Крок 9. Розрахунок частинних коефіцієнтів кореляції (формула (7)). Ці коефіцієнти розраховуються на основі значень елементів оберненої матриці (див. крок 7). Оскільки матриця є симетричною відносно діагоналі, то достатньо розрахувати або верхню праву, або нижню ліву трикутну частину матриці. Для розрахунку частинного коефіцієнту кореляції першого показника самого з собою необхідно виділити комірку P25 і ввести в рядку формул вираз «=-P14/КОРЕНЬ(\$P\$14*P14)». У цьому виразі один підкореневий множник є фіксованим. Скопіювавши формулу вправо на наступні п'ять комірок отримаємо в них хибні значення. Для їх виправлення необхідно виділити відповідну комірку та змінити нефіксований підкореневий множник на відповідний діагональний елемент оберненої матриці. Для кожного наступного рядка послідовність операцій виконується в тому ж порядку: виділяється комірка для діагонального елементу; вводиться формула з одним фіксованим підкореневим множником; копіюються формули до правого краю та виправляються значення. Таким чином заповнюється верхня права трикутна матриця частинних коефіцієнтів кореляції (рис. 21).

	I	J	K	L	M	N	O	P	Q	R	S	T	U
11	0,369401	-0,00881	0,040023	0,356762	0,386023	0,579231							
12	0,111275	-0,124776	-0,08561	0,345119	0,133347	0,108639	96,10195		7,260944				
13	-0,10389	-0,10543	-0,07791	-0,21292	-0,12597	-0,12833							
14	0,215497	-0,24553	-0,29981	-0,06279	0,175426	-0,00082	3,203071	-0,5102	0,398022	-0,88	-2,85177	0,573478	
15	0,035163	0,186038	0,12474	-0,00628	-0,04324	-0,01115	-0,5102	1,515907	-0,76385	0,107857	1,116138	-0,23448	
16	-0,10511	-0,3025	-0,05288	0,16472	0,11692	0,10279	0,398022	-0,76385	1,498767	-0,25461	-0,3749	-0,35096	
17	-0,15096	0,417926	0,115113	-0,19574	-0,12928	-0,11349	-0,88	0,107857	-0,25461	1,696587	1,149283	-1,25396	
18	-0,15993	-0,12778	-0,0909	0,13859	-0,08247	-0,09209	-2,85177	1,116138	-0,3749	1,149283	11,7519	-9,24227	
19	-0,09082	-0,10443	0,18428	0,032799	-0,11354	0,10582	0,573478	-0,23448	-0,35096	-1,25396	-9,24227	9,993164	
20	-0,13135	0,443892	0,393269	-0,14697	-0,1093	0,10172							
21	-0,15105	0,255886	-0,11545	-0,29605	-0,13201	-0,12387	9,2529	2,166808	2,09482	2,925667	45,15799	37,77129	
22	0,133408	0,156649	0,106449	0,362624	0,129107	0,15874							
23	-0,13891	-0,18575	0,201755	-0,18972	-0,1345	-0,12159							0,219813
24	-0,05972	0,016752	-0,07598	0,052421	-0,11325	-0,10819							
25	-0,0741	0,035271	0,256629	0,01676	-0,03122	-0,062	-1	0,231535	-0,18166	0,377493	0,464812	-0,10136	
26	-0,14088	-0,14589	-0,25986	0,091502	-0,13258	-0,11777		-1	0,506765	-0,06725	=T15/КОРЕНЬ(\$Q\$15*T18)		
27	0,005091	0,157052	0,166136	0,036219	-0,05091	-0,06027			-1	0,193998	0,08933	0,090687	
28	-0,118	-0,09738	0,275883	0,31532	-0,09931	-0,064			-1	-0,25739	0,30454		
29	-0,1476	0,240789	-0,03362	0,066588	-0,12141	-0,11019				-1	0,852851		
30													-1
31													

Рис. 21. Верхня права трикутна матриця частинних коефіцієнтів кореляції

Крок 10. Обчислення t-критеріїв Стьюдента (формула (8)). Значення t-критеріїв Стьюдента розраховується на основі частинних коефіцієнтів кореляції. На цьому етапі тестування мультиколінеарності має сенс лише для пари різних індикаторів. Тому пропускають розрахунки діагональних елементів (формула не може бути обчислена). Для розрахунку першого рядка виділимо комірку Q32 і введемо вираз «=Q25*КОРЕНЬ(27-6)/КОРЕНЬ(1-Q25^2)». Скопіюємо формулу на наступні п'ять комірок. Для другого і наступних рядків необхідно виділяти зміщену вправо й вліво комірку та копіювати в неї розміщену вище формулу (комбінація клавіш «Ctrl»+«D»), і знову, в горизонтальному напрямку, скопіювати формулу до правого краю матриці t-критеріїв (рис. 22).

The screenshot shows an Excel spreadsheet titled "MultiKollinearist.xlsx". The formula bar displays the formula $=U29*\text{КОРЕНЬ}(27-6)/\text{КОРЕНЬ}(1-U29^2)$. The spreadsheet contains a matrix of values, with rows and columns labeled L, M, N, O, P, Q, R, S, T, U, V, W, X. Cells U29 through U33 contain the formula, while other cells in the matrix are filled with numerical values. The status bar at the bottom right shows "КОРЕНЬ(число)".

Рис. 22. Скріншот розрахунку матриці t-критеріїв Стьюдента

Крок 11. Розрахунок табличного значення t-критерію Стьюдента. Табличне значення цього критерію в Excel обчислюється за допомогою статистичної формулі СТЬЮДЕНТ.ОБР.2Х (двосторонній обернений розподіл Стьюдента). В комірку Q38 введемо вираз

«=СТЬЮДЕНТ.ОБР.2Х(0,05;27-6)» (рис. 23).

L	M	N	O	P	Q	R	S	T	U
26	0,091502	-0,13258	-0,11777		-1	0,506765	-0,06725	-0,26444	0,060244
27	0,036219	-0,05091	-0,06027			-1	0,193998	0,08933	0,090687
28	0,31532	-0,09931	-0,064				-1	-0,25739	0,30454
29	0,066588	-0,12141	-0,11019					-1	0,852851
30									-1
31									
32					1,090666	-0,84655	1,868107	2,405706	-0,46691
33						2,693805	-0,3089	-1,25655	0,276573
34							0,906228	0,411006	0,417299
35								-1,22061	1,465174
36									7,484858
37									
38								2,079614	
39									

Рис. 23. Скріншот розрахунку табличного значення t -критерію Стьюдента

Крок 12. Вибір показників для регресійної моделі. На основі порівняння фактичних значень t -критеріїв із критичним значенням критерію Стьюдента робимо висновок, що між параметрами регіональних індикаторів (стовпчиків): В і F; С і D; F і G (див. рис. 12) існує мультиколінеарність, оскільки відповідні фактичні значення індикаторів більші за критичну величину критерію Стьюдента (рис. 24).

L	M	N	O	P	Q	R	S	T	U
14	-0,06279	0,175426	-0,00082	3,203071	-0,5102	0,398022	-0,88	-2,85177	0,573478
15	-0,00628	-0,04324	-0,01115		-0,5102	1,515907	-0,76385	0,107857	1,116138
16	-0,16472	-0,11692	-0,10279	0,398022	-0,76385	1,498767	-0,25461	-0,3749	-0,35096
17	-0,19574	-0,12928	-0,11349		-0,88	0,107857	-0,25461	1,696587	1,149283
18	-0,13859	-0,08247	-0,09209	-2,85177	1,116138	-0,3749	1,149283	11,7519	-9,24227
19	0,032799	-0,11354	-0,10582	0,573478	-0,23448	-0,35096	-1,25396	-9,24227	9,993164
20	-0,14697	-0,1093	-0,10172						
21	-0,29605	-0,13201	-0,12387	9,2529	2,166808	2,09482	2,925667	45,15799	37,77129
22	0,362624	0,129107	0,15874						
23	-0,18972	-0,1345	-0,12159						0,219813
24	0,052421	-0,11325	-0,10819						
25	0,01676	-0,03122	-0,062		-1	0,231535	-0,18166	0,377493	0,464812
26	0,091502	-0,13258	-0,11777			-1	0,506765	-0,06725	-0,26444
27	0,036219	-0,05091	-0,06027				-1	0,193998	0,08933
28	0,31532	-0,09931	-0,064					-1	0,30454
29	0,066588	-0,12141	-0,11019						-1
30									-1
31									
32					1,090666	-0,84655	1,868107	2,405706	-0,46691
33						2,693805	-0,3089	-1,25655	0,276573
34							0,906228	0,411006	0,417299
35								-1,22061	1,465174
36									7,484858
37									
38								2,079614	
39									
40									

Рис. 24. Скриншот порівняння фактичних і табличного значень критеріїв Стьюдента

Для виключення із складу регресійної моделі обираємо показник F за наступних причин:

- значення фактичного критерію Стьюдента для пари регіональних індикаторів до якої входить показник F приймає найбільше значення, а саме – 7, 48 (див. рис. 29 – комірка U36);
- показник F зустрічається найбільшу кількість разів у матриці критеріїв Стьюдента (2 рази).

Таким чином, до складу регресійної моделі будуть входити наступні показники: B, C, D, E та G (див. рис. 7).

Регресійний аналіз.

Основними методами регресійного аналізу є [23-24]:

- нелінійна регресія;
- множинна лінійна регресія;
- лінійна регресія;
- метод найменших квадратів.

Основним завданням регресійного аналізу є визначення впливу факторів на результативний показник.

Розглянемо основні види *однофакторної нелінійної моделі*, які шляхом перетворень зводяться до лінійної моделі:

- регресія $y = a / x + b$ заміною змінної величини $1/x = z$ зводиться до лінійної регресії $y = az + b$;
- регресія $y = a \ln x + b$ заміною змінної величини $\ln x = z$ зводиться до лінійної регресії $y = az + b$;
- регресія $y = ae^x + b$ заміною змінної величини $e^x = z$ зводиться до лінійної регресії;
- регресія $y = a\sqrt{x} + b$ заміною змінної величини $z = \sqrt{x}$ зводиться до лінійної регресії;

Серед *багатофакторних регресійних моделей* зустрічаються такі взаємозалежності між екологіко-економічними показниками:

– *степенева*,

$$y = a_0 x_1^{a_1} x_2^{a_2} \cdots x_m^{a_m},$$

яка зводиться до лінійної моделі:

$$\ln y = \ln a_0 + a_1 \ln x_1 + a_2 \ln x_2 + \dots + a_m \ln x_m;$$

– *гіперболічна*,

$$y = a_0 + \frac{a_1}{x_1} + \frac{a_2}{x_2} + \dots + \frac{a_m}{x_m}$$

яка зводиться до лінійної моделі:

$$y = a_0 + a_1 z_1 + a_1 z_2 + \dots + a_1 z_m,$$

$$z_j = \frac{1}{x_j}, j = 1, 2, \dots, m$$

де ;

– квадратична

$$y = a_0 + a_1 x_1^2 + a_1 x_2^2 + \dots + a_1 x_m^2,$$

яка зводиться до лінійної моделі заміною $z_j = x_j^2, j = 1, 2, \dots, m$.

Використання сучасних програмних засобів обробки даних дає змогу по новому підійти до вивчення еколого-економічних процесів, що залежать від багатьох факторів.

Можна розглядати два типи багатофакторних моделей регресійного аналізу:

- 1) лінійні відносно оцінюваних параметрів;
- 2) нелінійні.

Багатофакторні моделі *першого типу* мають вигляд

$$y = a_1 F_1(x_1) + a_2 F_2(x_2) + \dots + a_n F_n(x_n) + l,$$

де $F_i(x_i)$ можуть бути різними функціями (наприклад, $F_1(x_1) = \sqrt{x_1}, F_2(x_2) = e^{x_2}, F_3(x_3) = \ln x_3$).

Оцінки параметрів прогнозу і надійних інтервалів знаходять для лінійної моделі, а потім переходять до нелінійної моделі. окремі багатофакторні нелінійні відносно параметрів моделі можна зводити до багатофакторних лінійних регресійних моделей. Прикладом таких багатофакторних моделей може бути модель

$$y = a_0 x_1^{a_1} \cdot x_2^{a_2} \cdot \dots \cdot x_n^{a_n} l.$$

Такого вигляду регресія можна описувати процеси, що залежать від досягнутого рівня прогресу, без істотних обмежень на дані процеси. Логарифмуванням і наступною заміною змінних таку модель можна звести до лінійної. Аналогічно до лінійних можна звести нелінійні моделі відносно параметрів регресії, зокрема:

$$y = a_1 e^{a_2 x_1} e^{a_3 x_2} \cdot \dots \cdot e^{a_{n+1} x_n} l,$$

$$y = a_0 e^{a_1 x_1} \cdot \dots \cdot e^{a_n x_n} \cdot \dots \cdot x_{k+1}^{a_k+1} \cdot \dots \cdot x_{k+n}^{a_k+n} l,$$

$$y = a_0 a_1^{x_1} a_2^{x_2} \cdot x_n^{a_n} e^{a_{n+1}} l.$$

При вивченні процесів однією з важливих складових дослідження є

вибір відповідної регресійної залежності. Необхідно розглянути найбільше число різних моделей, вибрати критерії оцінки адекватності розглядуваної моделі і експериментальних даних.

Приклад 3. Побудова множинної лінійної регресійної моделі в Excel.

Побудувати множинну лінійну регресійну модель в Excel можна двома способами:

1. Засобами матричні функції МУМНОЖ та МОБР.
2. Засобом "Регресія" із надбудови "Пакет аналіза".

Перший спосіб. Використовуючи таблицю з вихідною інформацією (рис. 25) підготувати матрицю значень, у якій в першому стовпчику містяться одиниці (тобто назви регіонів), в наступних – значення стовпчиків факторів впливу, останній стовпчик містить значення залежного (прогнозованого) фактору (див. рис. 25).

A	B	C	D
1	ВРП млн грн	ДН млн грн	I млн дол.
2	1 23589,00	33602,00	344,70
3	1 14429,00	19137,00	571,50
4	1 116136,00	88922,00	5387,70
5	1 128986,00	118223,00	3067,60
6	1 18743,00	26124,00	428,30
7	1 15299,00	20841,00	1348,60
8	1 42736,00	45779,00	1363,00
9	1 20446,00	26504,00	652,00
10	1 44953,00	42732,00	3018,10
11	1 15749,00	20213,00	187,50
12	1 45541,00	51338,00	1119,00
13	1 41655,00	54838,00	2028,30
14	1 24055,00	26034,00	866,10
15	1 53878,00	52924,00	3520,30
16	1 44291,00	34462,00	1066,20
17	1 15882,00	22362,00	438,10
18	1 18333,00	24918,00	469,90
19	1 12726,00	19587,00	209,80
20	1 65293,00	67102,00	1834,10
21	1 15649,00	20978,00	161,20
22	1 18096,00	26987,00	459,40
23	1 22354,00	26194,00	378,40
24	1 9892,00	16114,00	113,10
25	1 17008,00	23179,00	380,70
26	1 196639,00	143903,00	25941,90

Рис. 25. Скріншот із вихідними даними

Матриця коефіцієнтів рівняння регресії в Excel визначається за допомогою множення попередньо обчисленої матриці та оберненої матриці, використовуючи функції МУМНОЖ. Для цього необхідно обчислити транспоновану матрицю до створеної. Спершу виділяємо квадратний діапазон комірок розміром 4×4 (відповідно до кількості стовпчиків нової матриця). Для першої комірки J2 вводимо формулу $=\text{МУМНОЖ}(\text{ТРАНСП}(A2:D26);A2:D26)$, де A2:D26 – діапазон значень матриці. Приклад розрахованої транспонованої матриці наведено на рис. 26.

J	K	L	M
Транспонована матриця			
25	1042358	1052997	55355,5
1042358	90380000742	77512576118	6,936E+09
1052997	77512576118	69204638033	5,448E+09
55355,5	6935873193	5448414276	749298413

Рис. 26. Скріншот розрахунку транспонованої матриці

Обернена матриця (рис. 27) має той самий розмір як і вихідна (4×4). Обернену матрицю бажано розмістити під вихідною, наприклад, в діапазоні комірок J8:M11.

J	K	L	M
Транспонована матриця			
25	1042358	1052997	55355,5
1042358	90380000742	77512576118	6,936E+09
1052997	77512576118	69204638033	5,448E+09
55355,5	6935873193	5448414276	749298413
Обернена матриця			
0,28230878	1,32416E-05	-1,83259E-05	-1,017E-05
1,32416E-05	1,3384E-09	-1,51608E-09	-2,343E-09
-1,83259E-05	-1,51608E-09	1,82427E-09	2,122E-09
-1,01729E-05	-2,34321E-09	2,12249E-09	8,343E-09

Рис. 27. Скріншот розрахунку оберненої матриці

Для розрахунку першого нормалізованого значення необхідно вибрати клітинку J8 та задати формулу =МОБР(J2:M5)», де J2:M5 – діапазон значень транспонованої матриці. Наступним кроком є побудова матриці, яка є результатом множення транспонованої матриці на матрицю залежної змінної. Результат виводиться в комірках J13:J18 (рис. 28). Виділяємо першу комірку і вводимо значення «=МУМНОЖ(ТРАНСП(I3:N27);I3:N27)».

J	K	L	M
1042358	9,04E+10	7,75E+10	6,94E+09
1052997	7,75E+10	6,92E+10	5,45E+09
55355,5	6,94E+09	5,45E+09	7,49E+08

Обернена матриця			
0,2823088	1,32E-05	-1,83E-05	-1,02E-05
1,32E-05	1,34E-09	-1,52E-09	-2,34E-09
-1,83E-05	-1,52E-09	1,82E-09	2,12E-09
-1,02E-05	-2,34E-09	2,12E-09	8,34E-09

6515,7
516156099
468805291
22377010

Рис. 28. Результат множення транспонованої матриці на матрицю залежної змінної

Матриця коефіцієнтів рівняння регресії в Excel визначається за допомогою множення попередньо обчисленої матриці та оберненої матриці, використовуючи функції МУМНОЖ. Виділимо діапазон комірок J20:M23 і введемо формулу «=МУМНОЖ(J8:M11;J14:J17)», де J8:M11 – діапазон комірок оберненої матриці, J14:J17 – діапазон комірок попередньо обчисленої матриці (рис. 27).

Матриця коефіцієнтів рівняння регресії			
-144,7461	-144,746	-144,746	-144,746
0,0139213	0,013921	0,013921	0,013921
0,0007824	0,000782	0,000782	0,000782
-0,094027	-0,09403	-0,09403	-0,09403

Рис. 27. Матриця коефіцієнтів рівняння регресії

Згідно даних рис. 27, рівняння регресії матиме вигляд:

$$y_r = -144.7 + 0.0139 \cdot ВРП + 0,0008 \cdot ДН - 0,084 \cdot I.$$

Другий спосіб. Для побудови лінійної регресійної моделі необхідно:

- 1) Викликати засіб "Регрессия" із надбудови "Пакет аналіза", натиснувши на кнопку "Аналіз даних" на закладці "Данные" панелі інструментів Excel (рис. 28).

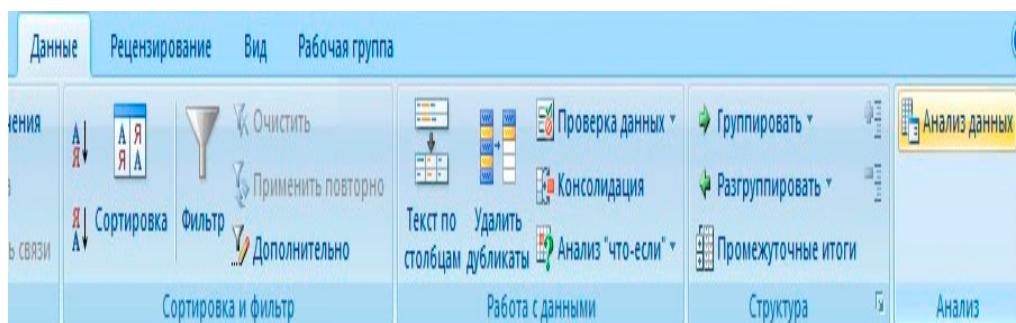


Рис. 28. Кнопка "Аналіз даних" на панелі інструментів

- 2) Викликати Сервис – Аналіз даних – Регрессия – OK. З'явиться вікно для надання вхідних даних (рис. 29). У вікні "Регрессия" вибрать вхідний інтервал для прогнозованого фактора у вікні вибору "Входной интервал Y" (наприклад, комірки G2:G11). Задати інтервал для пояснювальних факторів у вікні вибору "X" (комірки B2:F11). В поле "Новый рабочий лист" ввести коротку назву листа, наприклад Reg. Натиснути кнопку "OK".

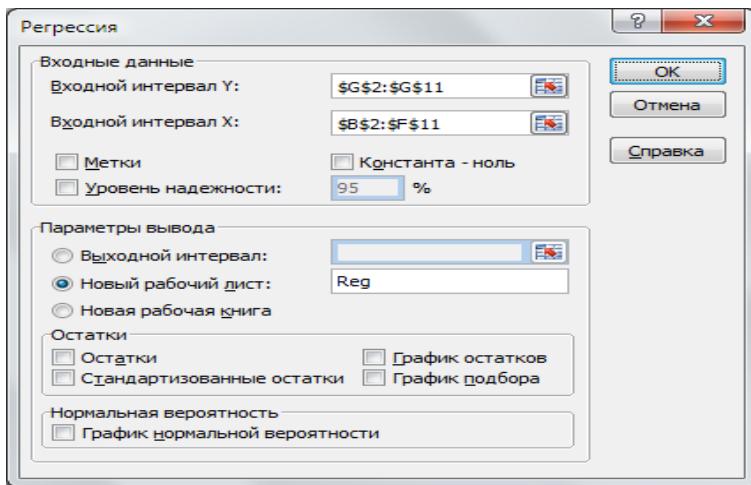


Рис. 29. Діалогове вікно функції "Регресія"

Приклад і результати роботи функції "Регресія" наведені на рис. 30.

Регресійний аналіз											
Вхідні дані			Вихідні дані								
№	Значення		Вивод ітогов								
	i	X	Y	Регресіонна статистика							
1				Mножественный R	0,972633354						
2				R-квадрат	0,946015642						
3				Нормированный R-квадрат	0,937018249						
4				Стандартная ошибка	5,786032717						
5				Наблюдения	8						
6				Дисперсионный анализ							
7				df	SS	MS	F	Значимость F			
8				Регрессия	1	3520,005952	3520,005952	105,1433059	5,01935E-05		
9				Остаток	6	200,8690476	33,4781746				
10				Итого	7	3720,875					
11				Коэффициенты Стандартная ошибка t-статистика P-значение							
12				Y-пересечение	310,6785714	4,508440371	68,91043151	6,28282E-10			
13				Переменная X1	9,154761905	0,892804231	10,25394099	5,01935E-05			

Рис. 30. Скріншот результатів регресійного аналізу

На рис. 30 у графі «Коєфіцієнти» вказані значення параметрів моделі a та b : b – в графі *Y-пересечение*, a – в графі *Переменная X1*. Отже, побудована лінійна регресійна модель має вигляд:

$$y = 69,15x + 310,68 .$$

Для перевірки статистичної значущості моделі надається значення F- статистики у графі F : $F = 105,14$. Коефіцієнт детермінації моделі R^2 надається у графі *R-квадрат*, $R^2 = 0,97$. Крім того, може бути надано: графік підбору – порівняльна діаграма, що містить емпіричну і теоретичну лінії регресії; таблиця залишків – різниць емпіричних і теоретичних значень Y (рис. 31).



Рис. 31. Додаткові результати регресійного аналізу

Приклад 4. Розрахунок прогнозів за допомогою надбудов ковзкого середнього. Використання методу ковзкого середнього прогнозу будь-якого періоду є не що інше, як отримання середнього показника декількох результатів наглядів тимчасового раду. Наприклад, якщо ви вибрали ковзке середнє за перших чотири місяці, прогнозом на червень буде середнє значення показників за лютий, березень, квітень і травень. Обчислення за допомогою цього методу достатньо точно відбивають зміни основних показників попереднього періоду. Чим менше число результатів наглядів, на підставі яких обчислене ковзке середнє, тим точніше воно відбиває зміни в рівні базової лінії. Але, якщо базою для прогнозованого ковзкого середнього є всього лише одне або два нагляди, то такий прогноз може стати дуже спрощеним.

Алгоритм складання прогнозів за допомогою надбудов ковзкого середнього.

Крок 1. Виберіть команду "Сервис – Надстройка". З'явиться діалогове вікно "Надстройка". Встановіть прапорець у списку меню "Пакет аналіза" і класніть на кнопці ОК. При необхідності активізуйте робочий лист, що містить дані про вашу базову лінію.

Крок 2. У меню "Сервис" ви знайдете нову команду "Аналіз даних". Виберіть команду "Сервис – Аналіз даних". З'явиться діалогове вікно "Аналіз даних", в якому містяться всі доступні функції аналізу даних.

Крок 3. Із списку виберіть інструмент аналізу "Скользящее среднее" і класніть на кнопці ОК. З'явиться діалогове вікно "Скользящее среднее".

Крок 4. У полі "Входний інтервал" введіть дані про вашу базову лінію або вкажіть діапазон у робочому листі, посилання на нього з'явиться в цьому полі.

Крок 5. У полі "Інтервал" введіть кількість місяців, днів які хочете включити в підрахунок ковзкого середнього.

Крок 6. У полі введення "Выходной интервал" введіть адресу комірки, в яку поміщають прогнозні дані.

На рис. 32 показаний прогноз викидів забруднюючих речовин в атмосферне повітря внаслідок діяльності модельного підприємства за допомогою надбудов ковзкого середнього та виведений його графік.

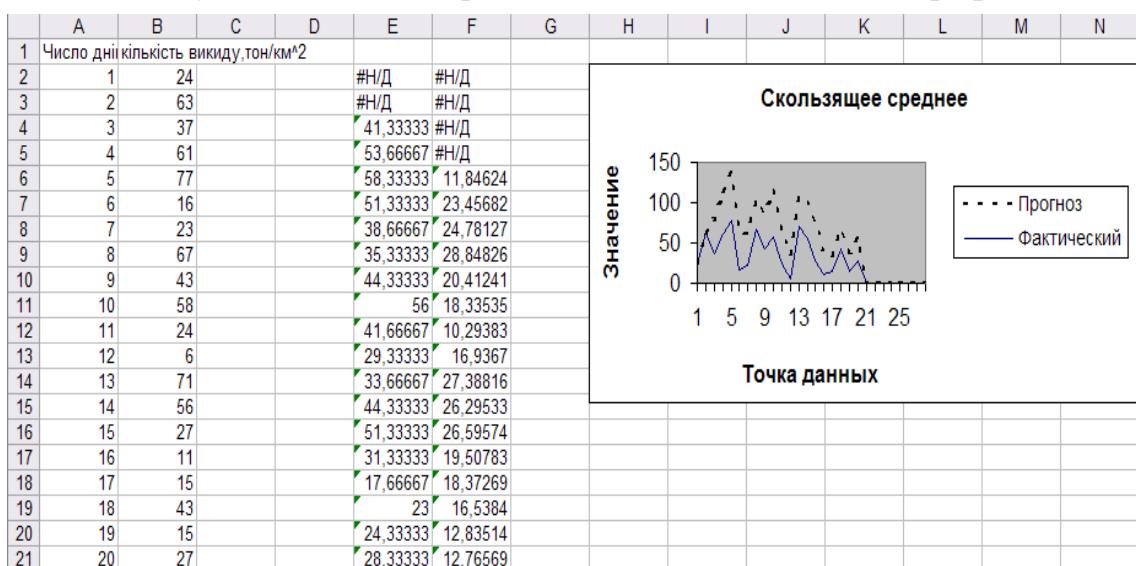


Рис. 32. Прогноз викидів забруднюючих речовин в атмосферне повітря за допомогою надбудов ковзкого середнього

На рис. 32 перші декілька показників ковзкого середнього відсутні внаслідок того, що надбудова "Скользящее среднее" повертає замість цих показників #N/D. Річ у тому, що ковзке середнє, включає дані трьох попередніх періодів, і не може бути обчислене до тих пір, поки не буде закінчений нагляд за третім періодом.

Алгоритм складання прогнозів ковзкого середнього за допомогою діаграм: виділіть дані своєї базової лінії \Rightarrow кладніть на кнопці "Мастер діаграм", розташованій на стандартній панелі інструментів, або виберіть команду "Вставка – Диаграмма" \Rightarrow на першому кроці виберіть тип діаграми "График" і кладніть на кнопці "Далее" \Rightarrow на другому кроці роботи засобу "Мастер діаграм" перевірте правильність посилань на комірку базової лінії та кладніть на кнопку "Далее" \Rightarrow на третьому кроці виберіть

параметри графіка, що включає як лінії, так і маркери. Клацніть на кнопці "Далее" ⇒ на останньому, четвертому, кроці роботи "Мастера діаграмм" визначите місце положення діаграми. Клацніть на кнопці "Готово".

В діалоговому вікні «Лінія тренда», що з'явилося, клацніть на вкладці «Type», виберіть лінію тренда. Потім клацніть на вкладниці «Параметри» і вкажіть, що необхідно показати рівняння, коефіцієнт достовірності, а також можна зробити прогноз на декілька періодів уперед. Перші декілька показників ковзкого середнього відсутні з тієї ж причини, що надбудова «Скользящее среднее» повертає замість цих показників #Н/Д.

На рис. 33 показаний прогноз викидів забруднюючих речовин в атмосферне повітря внаслідок діяльності модельного підприємства за допомогою діаграм надбудови ковзкого середнього.

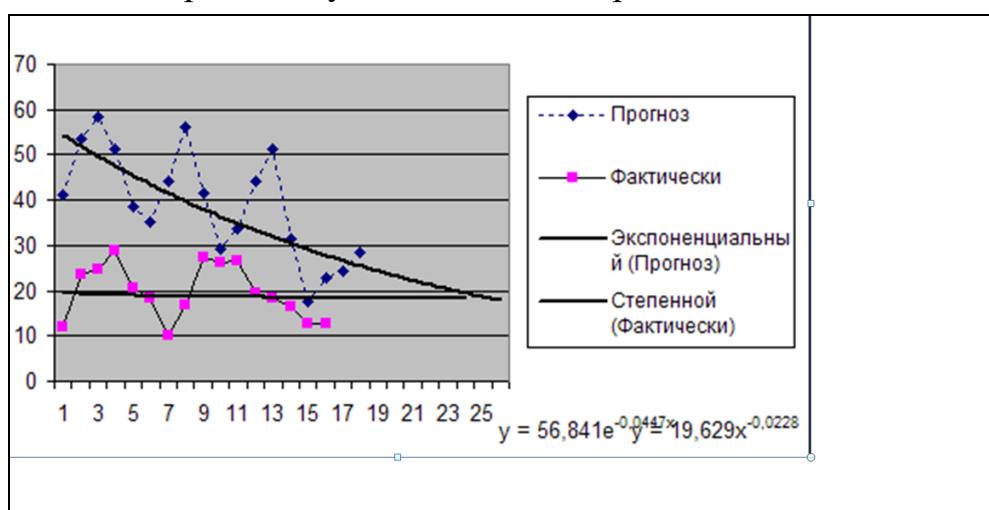


Рис. 33. Прогноз викидів забруднюючих речовин в атмосферне повітря за допомогою діаграм надбудови ковзкого середнього

Приклад 5. Розрахунок прогнозів за допомогою надбудов експоненційного згладжування. Експоненційне згладжування – спосіб згладжування часових рядів, обчислювальна процедура якого включає обробку всіх попередніх спостережень, при цьому враховується старіння інформації по мірі віддалення від прогнозного періоду. Інакше кажучи, чим "старше" спостереження, тим менше воно повинно впливати на величину прогнозної оцінки. Основою метода експоненційного згладжування є вибір параметра згладжування (згладжує константи) і початкових умов. Найбільше застосування метод знайшов для реалізації *середньострокових прогнозів*. Основна ідея цього методу полягає в тому, що кожен новий прогноз отримується шляхом зсування попереднього

прогнозу в напрямку, який би давав кращі результати порівняно зі старим прогнозом.

Базове рівняння має такий вигляд:

$$F_{t+1} = \alpha \times D_t + (1 - \alpha) F_t,$$

де F_{t+1} – прогноз для часового періоду $t+1$;

D_t – фактичне значення показника у момент часу t ;

F_t - прогноз, зроблений в момент часу t ; α – константа згладжування ($0 < \alpha < 1$).

Константа згладжування є самокоригованою величиною. Константа згладжування – це величина між 0 і 1, яку вибирає укладач прогнозу в залежності від специфіки конкретного вживання. Якщо величина константи α вибирається рівною 0, то прогноз на наступний період буде рівний прогнозу на поточний період. З другого боку, якщо константа α приймається рівною 1, то даним минулих періодів не приділяється ніякого значення, і прогноз повністю залежить від фактичного попиту на поточний період.

На рис. 34 наведено приклад прогнозу викидів забруднюючих речовин в атмосферне повітря внаслідок діяльності модельного підприємства за допомогою надбудови експоненційного згладжування.

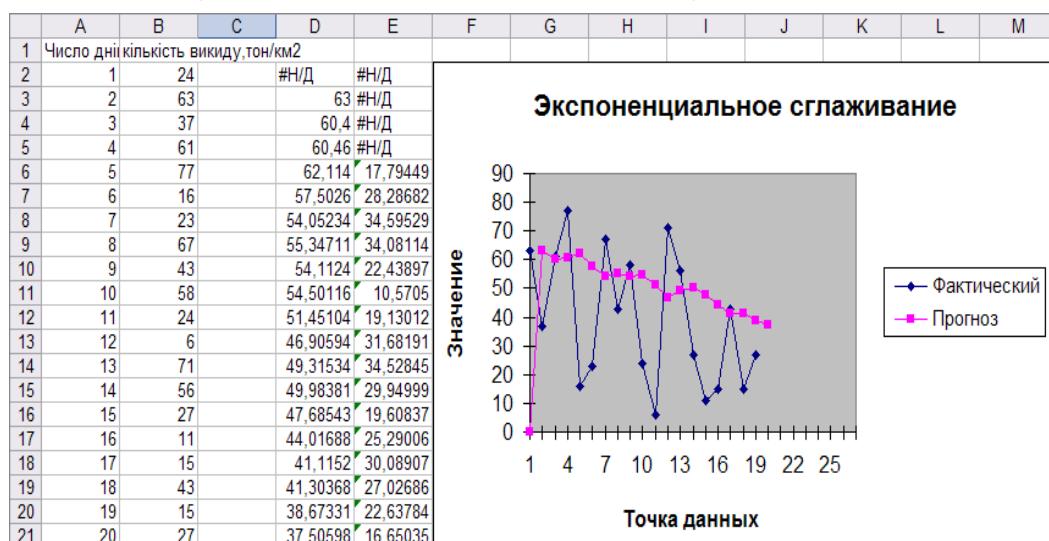


Рис. 34. Прогноз викидів забруднюючих речовин в атмосферне повітря за допомогою надбудови експоненційного згладжування

5. Кластерний аналіз інструментами Data Mining

Кластерний аналіз є набором різноманітних методів і алгоритмів класифікації (тобто методів об'єднання кластерів). Загалом, наявні методи побудови кластерних моделей за способами обробки даних утворюють два основні типи (*ієрархічний* і *неієрархічний*) та сім груп методів:

- 1) ієрархічні агломеративні;
- 2) ієрархічні дивізивні;
- 3) ітеративні методи групування;
- 4) факторні;
- 5) методи згущень;
- 6) методи пошуку модальних значень щільності;
- 7) методи, що використовують теорію графів.

Кожний тип включає безліч підходів і алгоритмів.

Вищезазначені положення свідчать, що для розбивки об'єктів на кластери використовуються дві процедури, що використовуються на різних етапах кластерного аналізу.

Перша – ієрархічна процедура дозволяє поєднувати елементи кластерів на базі понять відстані чи подібності між точками в багатомірному просторі ознак. Результатом такої розбивки є дендрограма (*дерево рішення*), що показує етапи об'єднання об'єктів дослідження в групи за характеристиками (показниками) соціально-економічного потенціалу.

Другий підхід для процедури розбивки на кластери – ітеративні (неієрархічні) методи угруповання. Для застосування цього підходу необхідно є попередня розбивка даних на задане число кластерів і наступна робота з первинними даними. Ітеративна процедура починається з розбивки даних на задане число кластерів. Потім обчислюються центри ваги цих кластерів. Далі кожен об'єкт поміщають у той кластер, центр ваги якого є найближчим. Обчислюються нові центри ваги кластерів. Обчислення повторюються доти, поки кластери не стають стійкими, тобто перестануть змінюватися.

Значне місце в прикладних статистичних дослідженнях займає *деревоподібна кластеризація*, що полягає в об'єднанні об'єктів у досить

великі кластери, використовуючи деяку міру подібності чи відстань між об'єктами. Типовим результатом цього виду кластеризації є ієрархічне дерево. Процедура починається з розгляду кожного об'єкта в класі. Далі поступово знижується поріг, що відноситься до рішення про об'єднання двох чи більше об'єктів в один кластер. Результатом обробки даних є з'язування разом усе більшого числа об'єктів, що складалися з елементів, що дуже відрізняються. На останньому кроці обчислення всі об'єкти об'єднуються разом.

Приклад класифікації методів кластерного аналізу наведена на рис. 35 [26].

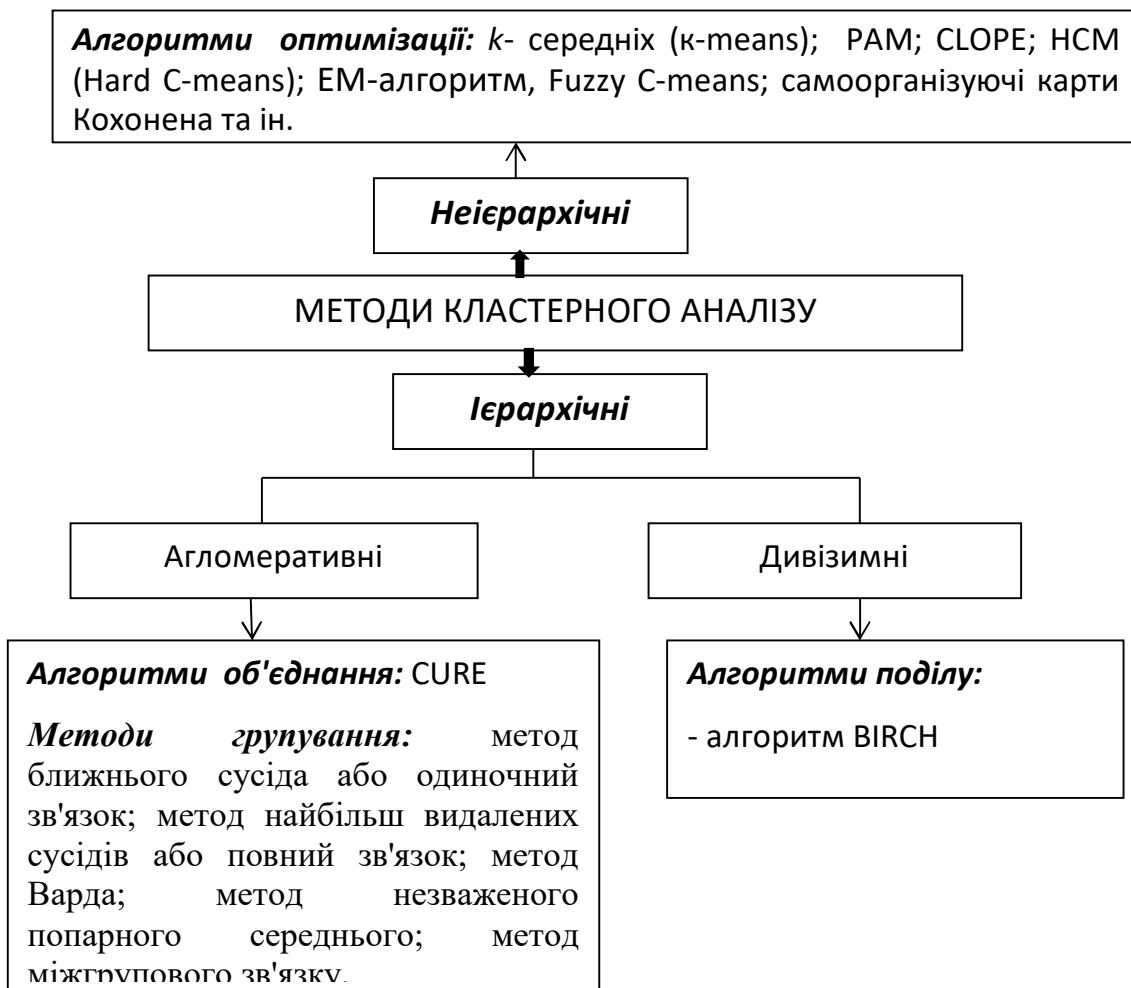


Рис. 35. Класифікація методів кластерного аналізу

Ієрархічні методи характеризуються побудовою ієрархічної чи деревоподібної структури, коли відбувається послідовне угруповання чи поділ об'єктів щодо інших об'єктів. При ієрархічній кластеризації виконується послідовне об'єднання менших кластерів у великі чи поділ великих кластерів на менші. Іншими словами, ієрархічні методи кластеризації розрізняються правилами побудови кластерів. У якості

правила виступають критерії, що використовуються при вирішенні питання про "схожість" об'єктів при їх об'єднанні в групу (агломеративні методи) або розділення на групи (дивізимні методи) (див. рис. 35).

Група ієрархічних агломеративних методів (*Agglomerative Nesting, AGNES*) характеризується послідовним об'єднанням вихідних елементів і відповідним зменшенням числа кластерів. На початку роботи алгоритму всі об'єкти є окремими кластерами. На першому кроці найбільш схожі об'єкти об'єднуються в кластер. На подальших кроках роботи алгоритму об'єднання продовжується до тих пір, поки всі об'єкти не складатимуть один кластер. Послідовне об'єднання вихідних елементів і зменшенням числа кластерів здійснюється на основі алгоритму *CURE (Clustering Using Representatives)*. Цей алгоритм виконує ієрархічну кластеризацію з використанням набору визначальних точок для визначення об'єкта в кластері.

Ієрархічні агломеративні методи розрізняються переважно за правилами групування кластерів. Існує багато різних правил групування, кожне з яких породжує специфічний ієрархічний метод. Найпоширенішими з них є:

- метод близького сусіда або одиночний зв'язок;
- метод найбільш відалених сусідів або повний зв'язок;
- метод Уорда;
- метод незваженого попарного середнього;
- метод між групового зв'язку.

Ієрархічні дивізимні (ділімі) методи (*Divisive Analysis, DIANA*) становлять логічну протилежність агломеративним методам. На початку роботи алгоритму всі об'єкти належать одному кластеру, який на подальших кроках ділиться на менші кластери, в результаті утворюється послідовність розщеплюючих груп.

Програмна реалізація алгоритмів кластерного аналізу широко представлена в різних інструментах *Data Mining*, які дозволяють вирішувати завдання досить великої розмірності. Наприклад, агломеративні методи можуть бути реалізовані програмними засобами «SPSS», дивізимні методи – програмними засобами «*Statgraphics*».

Приклад 6. Проведення кластерного аналізу програмними засобами

«SPSS». Алгоритм проведення кластерного аналізу стану регіонів України для вихідної статистичної інформації наведено нижче.

Крок 1. Завантаження даних. Для завантаження даних із книги Excel в програму SPSS необхідно вибрати в меню "Файл" пункт "Открыть", а потім підпункт "Данные...".

У вікні "Открыть данные" (рис. 36) вибрати тип файлу Excel (*.xls, *.xlsx, *.xlsm) та виділити необхідну книгу і натиснути кнопку "Открыть".

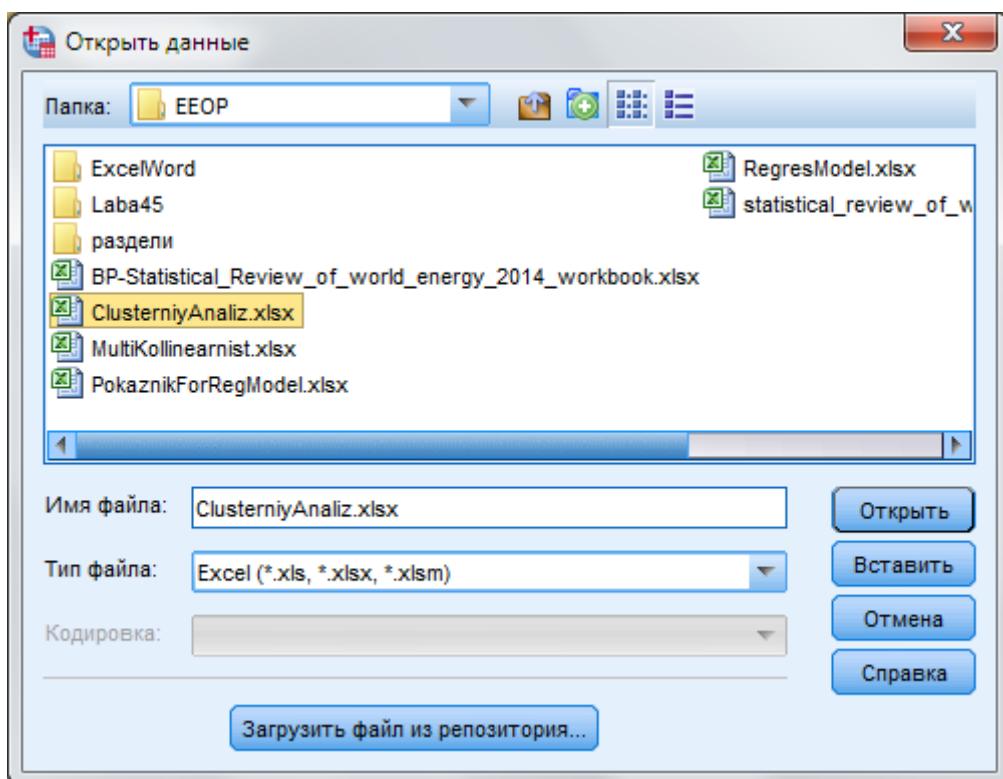


Рис. 36. Вибір книги Excel з даними

У вікні "Открытие файлов Excel" (рис. 37) виставити прапорець "Читать имена переменных из первой строки данных" та натиснути кнопку "OK".

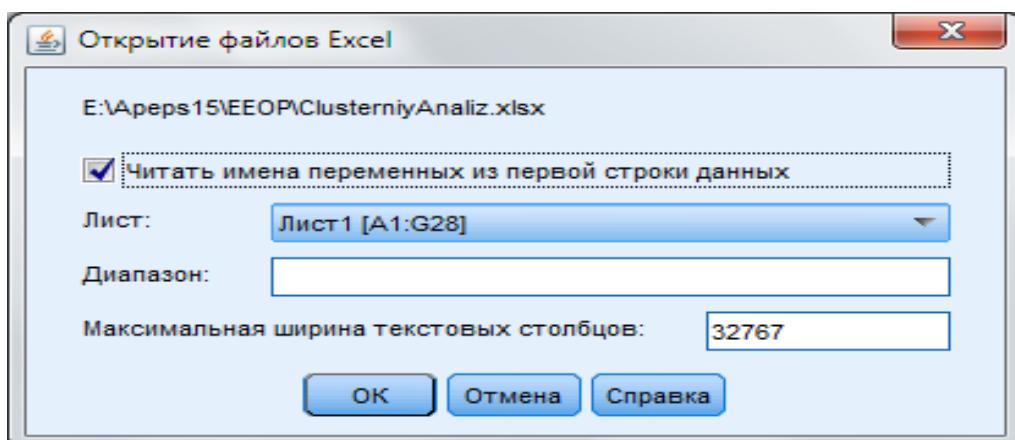


Рис. 37. Вибір назв факторів

Після цього дані завантажаться до закладки "Представление Данные" (рис. 38).

The screenshot shows the 'Data View' window in IBM SPSS Statistics. The title bar reads '*Без имени2 [Набор данных1] - Редактор данных IBM SPSS Statistics'. The menu bar includes 'Файл', 'Правка', 'Вид', 'Данные', 'Преобразование', 'Анализ', 'Прямой маркетинг', 'Графика', 'Сервис', 'Окно', and 'Справка'. Below the menu is a toolbar with icons for file operations. The main area displays a table with 11 rows and 8 columns. The columns are labeled: Регион, Часткагазу природного, Часткаветхи та аварійних, Часткавтра теплоенергопрітран..., Часткавики дівідендеретики, Середньодушові обсяги викидів..., Викинуто однім підприємством, and a blank column. The first row contains the header values. The data rows show values for 11 regions: АР Крим, Вінницька, Волинська, Дніпропетровська, Донецька, Житомирська, Закарпатська, Запорізька, Івано-Франківська, Київська, and Кіровоградська. The last column is empty. The status bar at the bottom indicates 'Процессор IBM SPSS Statistics готов' and 'Unicode:OFF'.

Рис. 38. Завантажені дані для кластеризації

Крок 2. Формування набору даних для кластеризації. Вибір даних для кластеризації полягає у визначенні факторів, що описують об'єкти в багатовимірному просторі ознак. Для цього необхідно за допомогою меню "Аналіз" вибрати пункт "Класифікація", а потім – підпункт "Ієрархіческая кластеризация" (рис. 39).

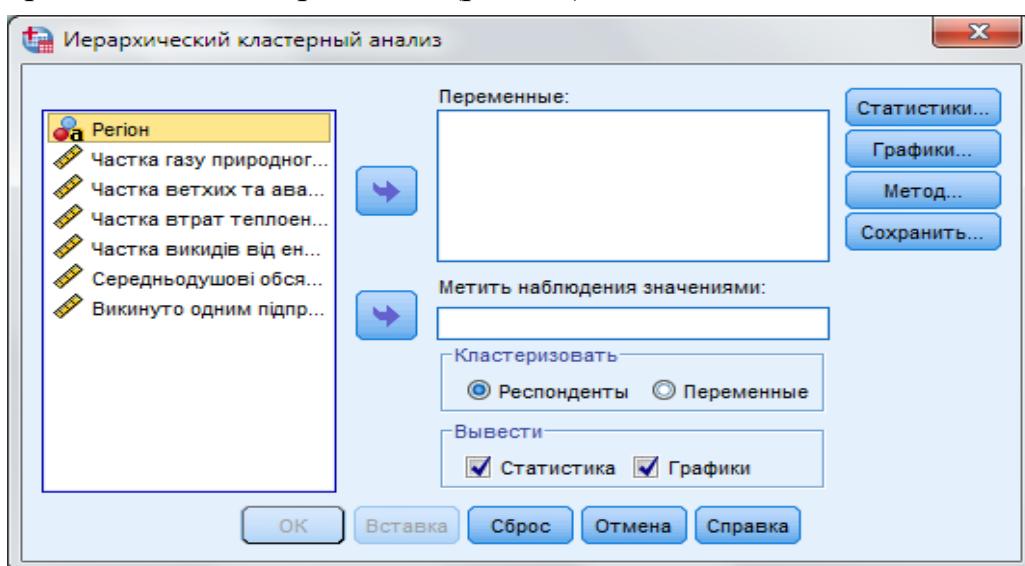


Рис. 39. Вікно "Иерархический кластерный анализ"

З метою підготовки даних до кластерізації необхідно в поле "Метить наблюдения значениями:" перенести з лівого списку пункт "Регіон", а в

спісок "Переменные" – всі інші фактори (рис. 40).

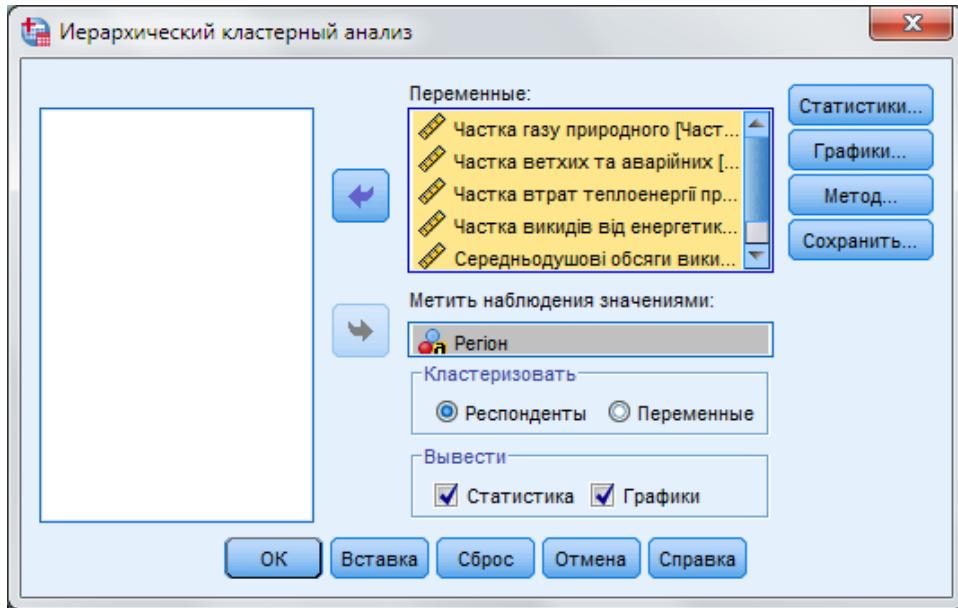


Рис. 40. Підготовка даних до кластеризації

Крок 3. Вибір методу кластеризації. При натисканні кнопки "Метод..." з'являється вікно "Иерархический кластерный анализ: Метод" (рис. 41). В ньому необхідно вибрати у випадаючому списку "Метод:" один з трьох методів: "Близкий сосед", "Самый дальний сосед" або "Центроидная кластеризация".

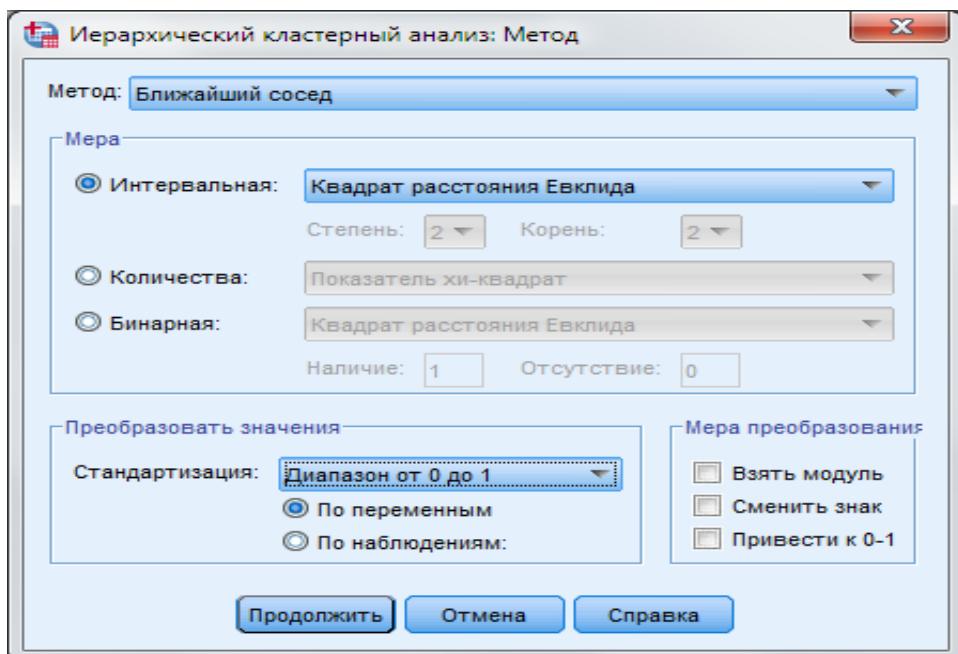


Рис. 41. Вибір методу кластеризації

У випадаючому списку "Стандартизация" залишаємо значення "Нет" (кластеризація по фактичним даним) або обираємо пункт "Диапазон от 0 до 1" (кластеризація по стандартизованим даним). Далі натискаємо кнопку

"Продолжить" і повертаємося до попереднього вікна (див. рис. 41).

Крок 4. Визначення кількості результуючих кластерів. При натисканні на кнопку "Статистики" з'являється вікно "Ієрархіческий кластерний аналіз: Статистики". В цьому вікні в області "Принадлежность к кластерам" вибираємо пункт "Одно решение" та задаємо кількість кластерів (рис. 42).

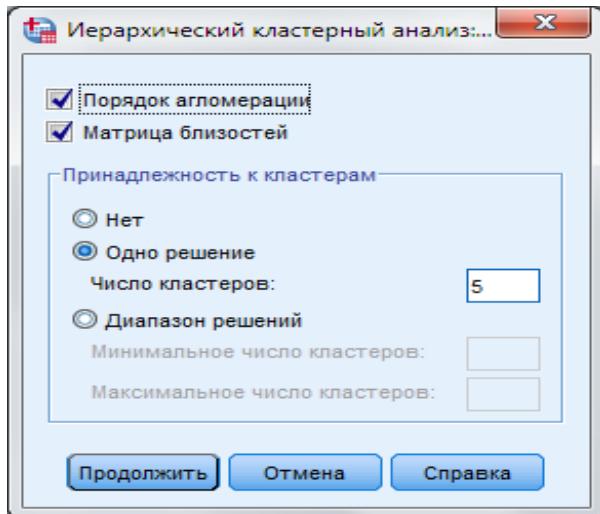


Рис. 42. Визначення кількості кластерів

Крок 5. Ініціалізація дендрограми. При натисканні на кнопку «Графики» з'являється вікно «Ієрархіческий кластерний аналіз: Графики». В цьому вікні встановлюємо прaporець «Дендрограмма». При натисканні на кнопку «Продолжить» дане вікно закриється (рис. 43).

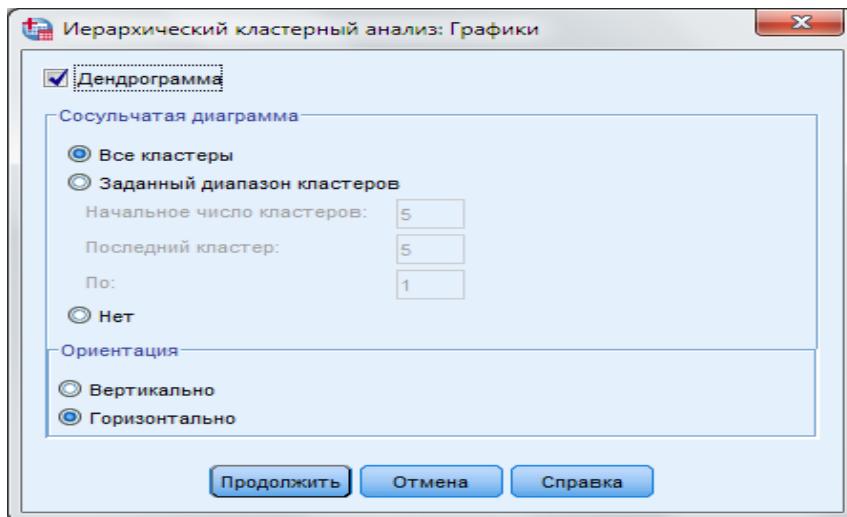


Рис. 43. Вікно ініціалізації дендрограми

Крок 6. Для виконання процедури кластеризації з обраними на попередніх кроках налаштуваннями у вікні натискають кнопку "OK". Результат кластеризації з'являється у новому вікні виводу (рис. 44).

```

* Вывод1 [Документ1] - Viewer IBM SPSS Statistics
Файл Правка Вид Данные Преобразование Вставка Формат Анализ Прямой маркетинг Графика Сервис Окно Справка
GET DATA /TYPE=XLSX
/FILE='E:\Арера15\EEOP\ClusterniyAnaliz.xlsx'
/SHEET=name 'Лист1'
/CELLRANGE=full
/READNAMES=on
/ASSUMEDSTRWIDTH=32767.
EXECUTE.
DATASET NAME Наборданных1 WINDOW=FRONT.
DATASET DECLARE D0.18831124713892555.
PROXIMITIES Часткагазуприродного Часткаветхихтааварійних Часткавтраптло
одушовюбсягивикидівденергетики Викинтооднимпідприємством
/MATRIX OUT(D0.18831124713892555)
/VIEW=CASE
/MEASURE=SEUCLID
/PRINT NONE
/ID=Perion
/STANDARDIZE=VARIABLE RESCALE.

Близости

[Наборданных1]

```

Рис. 44. Вікно виводу результату кластеризації

Крок 7. Аналіз результатів кластеризації. У лівій частині вікна в дереві виводу результати кластеризації можна перевірити в табличному вигляді, натиснувши пункт "Принадлежность к кластерам" (рис. 45). У таблиці перший стовпчик відповідає назві об'єкту, в другому вказується номер кластеру до якого належить об'єкт.

Наблюдение	Кластеры
1:АР Крим	1
2:Вінницька	2
3:Волинська	1
4:Дніпропетровська	3
5:Донецька	3
6:Житомирська	1
7:Закарпатська	1
8:Запорізька	4
9:Івано-Франківська	3
10:Київська	2
11:Кіровоградська	1
12:Луганська	5
13:Львівська	1
14:Миколаївська	1
15:Одеська	1
16:Полтавська	1
17:Рівненська	1
18:Сумська	1
19:Тернопільська	1
20:Харківська	2

Рис. 46. Табличний вигляд результату кластеризації

Також результати кластеризації можна перевірити в графічному

вигляді – вузол «Дендрограмма». Дендрограма показує порядок об'єднання об'єктів в кластери (рис. 47).

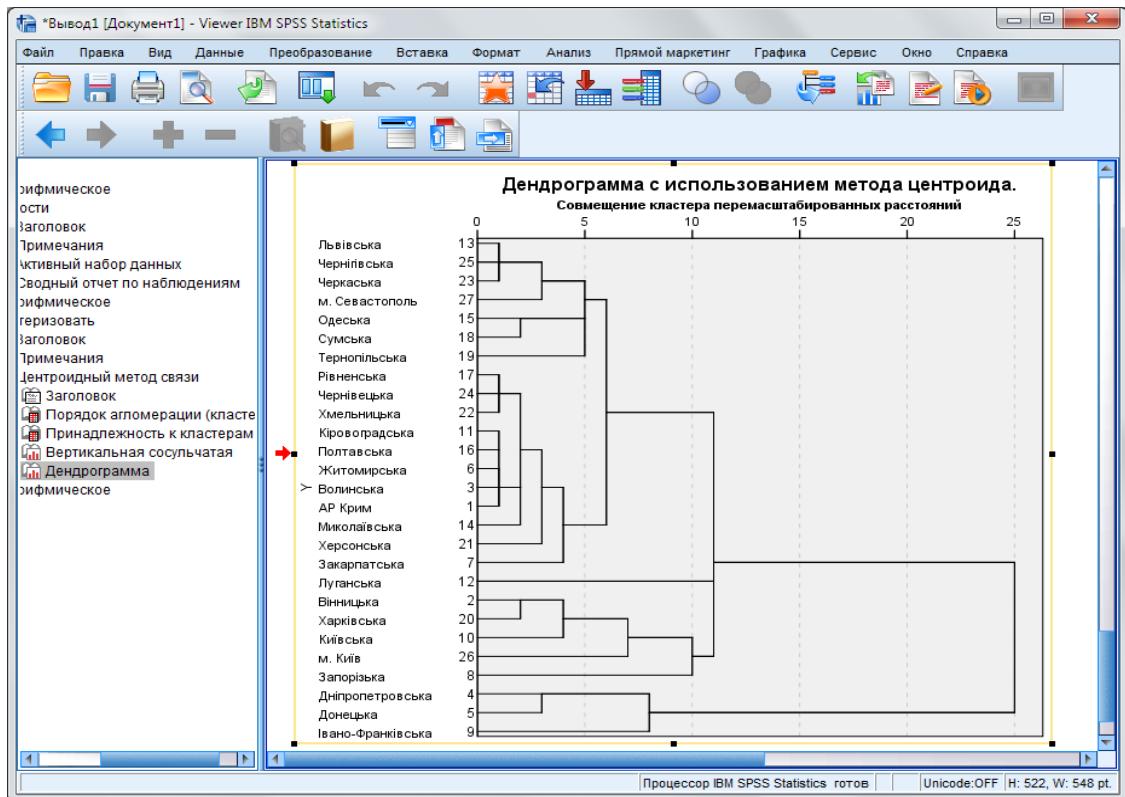


Рис. 47 – Дендрограмма кластеризації

Для того, щоб порівняти результати кластеризації різними методами (зокрема, з стандартизацією або без стандартизації), з різною кількістю результиуючих кластерів необхідно змінити відповідні налаштування для кроку 3 (метод кластеризації, стандартизація) та кроку 4 (кількість кластерів).

6. Представлення результатів комп’ютерних практикумів у протоколі

Вимоги до оформлення протоколів. Протокол комп’ютерного практикуму, охайно оформлюється у вигляді текстового файлу. На титульному аркуші зверху з вирівнюванням по центру вказують назву вузу, факультету та кафедри. Нижче під ними розміщують номер, назву комп’ютерного практикуму. Ще нижче, з вирівнюванням вправо, вказують

прізвище автора роботи, групу та прізвище викладача, який буде здійснювати перевірку роботи.

Структура протоколу містить дві частини: 1) детальний опис побудови багатофакторної регресійної моделі в Excel із усунення мультиколінеарності та власної розробки системи прогнозування ключових показників соціально-економічного потенціалу управління станом довкілля; 2) ранжування та виділення однорідної сукупності країн (або області України) на основі кластерного аналізу інструментами Data Mining (SPSS).

Основні завдання первого блоку комп'ютерного практикуму і рекомендації, щодо представлення результатів у протоколі.

Задачі:

- на основі аналізу екологічних і соціально-економічних загроз побудувати «дерево цілей» відтворення соціально-економічного потенціалу сталого розвитку об'єкту дослідження (це може бути будь-яка країна, група країн або регіони України);
- на основі побудованого дерева цілей і концептуально-теоретичних положень, що наведені у цьому навчальному посібнику сформувати систему показників (неменшу ніж 7) у табличному вигляді. Обов'язково надавати посилання на літературні джерела;
- сформувати мету, задачі і розробити алгоритми для кожного підблоку: кореляційний аналіз, усунення мультиколінеарності, регресійний аналіз;
- побудувати прогнозну економетричну модель із усунення мультиколінеарності засобами Excel, що описує кореляційно-регресійний зв'язок між соціальними, економічними, екологічними показниками. В протоколі мають бути представлені всі проміжні результати (відповідні скріншоти, таблиці) і висновки покрокової побудови моделі;
- розробити відповідний програмний продукт і надати кратку характеристику.

Основні завдання другого блоку комп'ютерного практикуму і рекомендації, щодо представлення результатів у протоколі.

Задачі:

- сформувати мету і задачі комп'ютерного практикуму;

– навести коротку характеристику, переваги і недоліки будь-яких 4 методів. Доцільно результати аналізу переваг і недоліків представити у вигляді таблиці (табл. 5);

Таблиця 5. Переваги і недоліки методів кластерного аналізу

Метод	Переваги	Недоліки
Метод близнього сусіда	<ul style="list-style-type: none"> – простота використання отриманих результатів; – рішення не унікальні для конкретної ситуації, можливе їх використання для інших випадків; – метою пошуку є не гарантовано вірне рішення, а найкраще з можливих. 	<ul style="list-style-type: none"> – при використанні методу виникає необхідність повного перебору навчальної вибірки при розпізнаванні, наслідок цього - обчислювальна трудомісткість; – типові завдання даного методу - це завдання невеликої розмірності за кількістю класів і змінних.
-----	-----	-----

– сформувати таблицю з вхідними показниками (або індексами);
 – навести результати (у вигляді відповідних скріншотів) кластерного аналізу інструментами *Data Mining* (SPSS) за вибраними методами;
 – проаналізувати і зробити висновок, щодо впливу застосування того чи іншого методу в рамках однієї групи на структуру кластерів. Щоб відповісти на це питання, доцільним є результати аналізу інструментами *Data Mining* представити у вигляді таблиці 6;

Таблиця 6. Результати кластерного розподілу регіонів з використання різних методів кластерного аналізу

№ кластеру	Методи із застосуванням нормалізації		Методи без застосування нормалізації	
	Метод близнього сусіда	Центроїдний метод	Метод дальнього сусіда	Метод Уорда
1	Інші (21)	Інші (17)	Інші (17)	Інші (8)
2	Луганська	Донецька	Донецька, Запорізька, Київська	Дніпропетровська
3	-----	-----	-----	-----

– обґрунтувати вибір методу кластерного аналізу за даними табл.6;
 – висновки представити у вигляді таблиці 7 і опису відмінностей кожного кластеру за факторами соціально-економічного потенціалу

управління станом довкілля.

Таблиця 7. Кластерний розподіл регіонів України за методом Уорда

№ клас-теру	Область	КВ	ЗЗП	ДН	ЗН	ІПП	КІ
1	Вінницька	11,661	11,7	71930	640,9	108,2	11744,1
	Житомирська	15,1	5,5	52753	510,6	109,5	7722,0
	Івано-Франківська	13,719	5,5	55947	559,0	112,0	9707,8
	Миколаївська	10,095	84,9	51964	489,7	101,5	11178,0
	Полтавська	20,627	27,2	69116	575,0	98,9	15855,6
	Сумська	17,13	102,6	50410	481,4	101,7	6947,1
	Хмельницька	10,542	1,9	55848	516,0	101,6	10499,9
	<i>Середні значення</i>	<i>14,41975</i>	<i>32,5375</i>	<i>57409</i>	<i>536,375</i>	<i>104,0625</i>	<i>10224,84</i>
2	-----	-----	-----	-----	-----	-----	-----

Приклад висновку. Результати табл. 7 свідчать, що особливістю 1 кластеру є найменше значення заборгованості заробітної плати (ЗЗП, млн. грн.) серед інших кластерів (це значення менше у 1.6 рази у порівнянні з 2 кластером, 4.4 рази у порівнянні з 3 кластером, що має найбільше значення серед кластерів, 4.2 рази у порівнянні з 4 кластером та 2.3 рази у порівнянні з 5 кластером). Якщо порівнювати усі інші показники з середніми значенням в рамках України, то усі значення 1 кластеру крім індексу промислової продукції мають менші значення – кількість вибулих людей (КВ, тис. осіб) менше у 1.2 рази, дохід населення (ДН, млн. грн.) менше у 1.4 рази, зайнятість населення (ЗН, тис. осіб) менше у 1.2 рази, індекс промислової продукції (ІПП, %) більше усього у 1.01 рази, а капітальні інвестиції (КІ, млн. грн.) менше у 1.7 рази. Особливістю другого кластеру є.....

Список використаної літератури:

1. Руська Р. В. Економетрика : навчальний посібник. Тернопіль : Тайп, 2012. 224с.
2. Вернадский, В. Научная мысль как планетарное явление URL: <http://vernadsky.lib.ru/e-texts/archive/thought.pdf> (дата звернення: 24.12.2019).
3. Вернадский В.И. Биосфера и ноосфера – М. : Наука, 1989. 261 с. URL: <http://www.spsl.nsc.ru/win/nelbib/vernadsky.pdf> (дата звернення: 24.12.2019).
4. Вернадский В.И. Несколько слов о ноосфере, 1943-1944 – URL:

<http://vernadsky.name/wp-content/uploads/2013/01/neskolko-slov-o-oosfere.pdf>

(дата звернення: 24.12.2019).

5. Schultz, T. Capital Formation by Education. *Journal of Political Economy*. 1960. № 6. Р. 571-583.
6. Shultz, T. Investment in Human Capital. N.-Y., L., 1971. 272 р.
7. Беккер Г. С. Человеческое поведение. Экономический подход. Избранные труды по экономической теории / Пер. с англ. , сост., науч. ред-р послес-я Р.И. Капелюшников; предис-я – М.И. Левин. - М. : ГУВШЭ, 2003. 672 с.
8. Людський розвиток регіонів України: аналіз та прогноз (колективна монографія) / За ред. Е.М. Лібанової. – К. : Ін-т демографії та соціальних досліджень НАН України, 2007. – 367 с.
9. Сен, А. Развитие как свобода. Пер. с англ. Е. Полецкой ; Под ред. и с послеслов. Р. М. Нуриева. М. : Новое издательство, 2004. 432 с.
10. Human Development Report 1990. N. Y. : Oxford University Press, 1990. 116 р. URL: <http://hdr.undp.org> (дата звернення: 24.12.2019).
11. Шаульська, Л. В. Стратегія розвитку трудового потенціалу України : монографія. Донецьк : Ін-т економіки промисловості НАН України, 2005. 386 с.
12. Карава, Н. В. Войтко, С. В. Сорокина, Л. В. Ризик-менеджмент сталого розвитку енергетики: інформаційна підтримка прийняття рішень : навчальний посібник. Київ : Альфа Реклама, 2013. 308 с.
13. Національна доповідь «Цілі Сталого Розвитку: Україна» URL: http://un.org.ua/images/SDGs_NationalReportUA_Web_1.pdf (дата звернення: 20.11.2019).
14. Human Development Reports UNDP "Human Development Indices and Indicators: 2018 Statistical Update". URL: <http://hdr.undp.org/en/2018-update> (дата звернення: 20.11.2019).
15. Методика вимірювання регіонального людського розвитку; затверджена Рішенням Президії НАН України та колегії Державної служби статистики України від 13.06.2012 р. № 123-м. URL:
http://www.idss.org.ua/arhiv/Metodika_RLR.zip (дата звернення: 28.11.2019).
16. Климко С.Г., Пригода В.М., Сизоненко В.О. Людський капітал: світовий досвід і Україна. К. : Основа, 2006. 224 с.
17. Корчагин Ю.А. Российский человеческий капитал: фактор развития или деградации? : монография. Воронеж: ЦИРЭ, 2005. 252 с.
18. Romer P.M. Increasing returns and long-run growth / P.M. Romer. *Journal*

- of Political Economy*. October, 1986. V. 94. P. 1002–1037.
19. Lucas, R. On the Mechanism of Economic Development / R. Lucas. *Journal of Monetary Economics*. 1988, vol. 22, P. 3–421.
20. Barro R., Sala-i-Martin X. Economic growth. –2nd Edition. Cambridge, Mass.; The MIT Press, 2003. 428 p.
21. Grossman M. The Human Capital Model. *Handbook of Health Economics* / Ed. by A.J. Culver, J.P. Newhouse. Vol. 1A. Amsterdam: Elsevier, 2000. P. 347-408.
- Міжнародна мікроекономіка. Навч. посібник. Видання 3-те перероб. та доп. Київ. Центр навчальної літератури, 2011. 362 с.
22. Економіко-математичне моделювання : навч. посібник / За заг. ред. В. В. Вітлінського. К. : КНЕУ, 2008. 536 с.
23. Лещинський О. Л. Економетрія : навч. посібник. К. : МАУП, 2003. 208 с.
24. Наконечний С. І., Терещенко Т.О., Романюк Т. П. Економетрія: Підручник. вид. 3-те, доп. та перероб. К. : КНЕУ, 2004. 520 с.
25. Кузьмичов А. І. Економетрія. Моделювання засобами MS Excel: навч. посібник. К. : ЦУЛ, 2011. 214 с.
26. Караєва Н.В. Методологічні аспекти кластеризації регіонів України за рівнем викликів енергетичної безпеки / Н.В. Караєва //Науковий вісник Одеського національного економічного університету. – Науки: економіка, політологія, історія. – 2016. – № 1 (233). – С. 40-55.