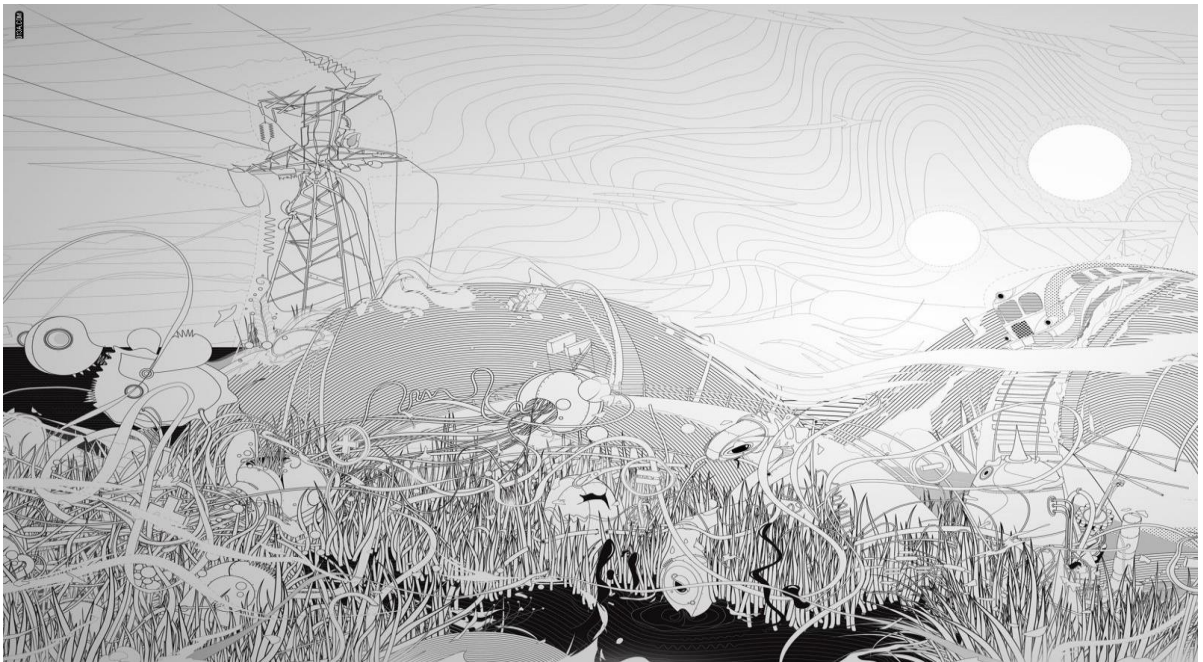


НАЦІОНАЛЬНИЙ ТЕХНІЧНИЙ УНІВЕРСИТЕТ УКРАЇНИ
“КИЇВСЬКИЙ ПОЛІТЕХНІЧНИЙ ІНСТИТУТ”

КАРАЄВА Н. В., ВАРАВА І. А.

ЕКОЛОГО-ЕКОНОМІЧНА ОПТИМІЗАЦІЯ ВИРОБНИЦТВА:
методи та засоби статистичного прогнозування

КОНСПЕКТ ЛЕКЦІЙ



Київ – 2016

УДК 66.012: 662.76:3

ББК 65.28я73

С.80

Рецензенти:

В. Г. Писаренко – доктор фізико-математичних наук, професор, завідувач відділу математичних проблем прикладної інформатики Інституту кібернетики ім. В. М. Глушкова НАН України;

Б. Я. Корнієнко – доктор технічних наук, доцент, завідувач кафедри комп'ютеризованих систем захисту інформації Національного авіаційного університету.

Рекомендовано до друку Вченою радою теплоенергетичного факультету НТУУ «КПІ», протокол № 8 від 28 березня 2016 р.

Електронне навчальне видання

Карасва Н. В., Варава І. А.

Еколого-економічна оптимізація виробництва: методи та засоби статистичного прогнозування : конспект лекцій. – К. : НТУУ «КПІ», 2016. – 80 с.

У навчальному виданні наведено лекційний матеріал за наступними напрямками: аналіз підходів та методів прогнозування стану територіально-виробничих систем; методи та засоби кореляційного аналізу; алгоритм розрахунку наявності мультиколінеарності в Excel; теоретико-методичні основи побудови множинної лінійної регресійної моделі в Excel; розрахунок прогнозів за допомогою надбудов ковзкого середнього; розрахунок прогнозів за допомогою надбудов експоненційного згладжування.

Розраховано на студентів та слухачів для самостійного ознайомлення з інформаційно-методичними положеннями статистичного прогнозування.

© Карасва Н.В., Варава І.А, 2016

НАЗВА ЛЕКЦІЇ

Вступ	4
1. Аналіз підходів та методів прогнозування стану територіально-виробничих систем	6
2. Кореляційний аналіз: теоретичні відомості	21
3. Обчислення кореляційної матриці в Excel	33
4. Усунення мультиколінеарності: теоретичні відомості	42
5. Алгоритм розрахунку наявності мультиколінеарності в Excel	48
6. Регресійний аналіз: теоретичні відомості	60
7. Побудова множинної лінійної регресійної моделі в Excel	66
8. Розрахунок прогнозів за допомогою надбудов ковзкого середнього ..	72
9. Розрахунок прогнозів за допомогою надбудов експоненційного згладжування	76
Список використаної літератури	79

ВСТУП

Еколого-економічна оптимізація спрямована на дотримання наукового обґрунтованого співвідношення між: матеріальними потребами розвитку суспільства та можливостями навколишнього природного середовища; мінімізацією антропогенного навантаження та раціоналізації суспільних потреб. Невіддільність завдань еколого-економічної оптимізації виробництва була однією з ключових проблем Конференції ООН з навколишнього середовища та сталого розвитку територій (червень 1992, Ріо-де-Жанейро). В її документах, зокрема, сформульовано такі положення: економічний розвиток у відриві від екології призводить до перетворення Землі на пустелю; екологія без економічного розвитку закріплює зубожіння та несправедливість; екологія без права на дію є часткою системи закріпачення; право на дію без екології торує шлях до колективного знищення, яке однаковою мірою стосується всіх. Для керування сталим розвитком цивілізації необхідно приймати рішення, які забезпечують мінімізацію навантаження на навколишнє середовище за наявності обмежених ресурсів. Саме у цьому полягає сутність задачі еколого-економічної оптимізації виробництва або у більш загальному розумінні мова має йти про територіально-виробничі системи. *Територіально-виробнича система (ТВС)* – це ділянка земної поверхні, що має окремі межі і просторовий базис діяльності, яка включає, крім виробництва, виробничу і соціальну інфраструктуру, розглянутих як полігон життєдіяльності, як фактор і ресурс розвитку території.

Головним завданням еколого-економічної оптимізації ТВС є виважене поєднання виробничих, природо відновних і соціальних функцій геосистем в інтересах досягнення належних просторово-екологічних умов життєдіяльності населення. Загалом, методи оптимізації або, як їх ще називають, методи теорії прийняття рішень є складовими математичних методів, які широко використовуються в екологічних і економічних дослідженнях.

Основними методами оптимізації є: аналіз, прогнозування і моделювання. Тому важливою складовою методології розробки стратегії

планування вирішення еколого-економічних проблем України є прогнозування стану ТВС на всіх рівнях керування.

Планування заходів й прогнозування результатів еколого-економічної оптимізації має базуватися на точних розрахунках потреб та їх ресурсному забезпеченні. Еколого-економічна оптимізація характеризується динамікою зміни початкових (вхідних) і похідних (розрахункових) показників ефективності функціонування ТВС за рівнем використання наявних виробничих потужностей та природних ресурсів, конкурентоспроможності продукції й обсягу її виробництва, рівня прибутку й платоспроможності, ефективності праці та рівня екологічної безпеки.

На підставі аналізу й оцінювання еколого-економічних показників стану ТВС прогнозують можливі зміни їх з огляду на науково-технічні та екологічні чинники у стратегічному й тактичному розумінні залежно від усталених тенденцій зміни ринкових ситуацій, зумовлених об'єктивними соціально-економічними процесами.

Наведені теоретичні знання в конспекті лекцій та методи розрахунку дають студенту достатню інформацію, щоб самостійно ознайомитися з теоретично-методичними положеннями і засобами статистичного прогнозування.



1. АНАЛІЗ ПІДХОДІВ ТА МЕТОДІВ ПРОГНОЗУВАННЯ СТАНУ ТЕРИТОРІАЛЬНО-ВИРОБНИЧИХ СИСТЕМ

План лекції:

- 1. Дефініція поняття "прогноз"*
- 2. Мета і завдання еколого-економічного прогнозування*
- 3. Суб'єкт, об'єкт, предмет еколого-економічного прогнозування*
- 4. Вимоги до організації прогнозування*
- 5. Генетичний і нормативний підходи в прогнозуванні*
- 6. Класифікація методів прогнозування*
- 7. Інтуїтивні методи прогнозування*
- 8. Формалізовані методи прогнозування*
- 9. Основні етапи побудови прогнозної економетричної моделі*

Прогноз – науково обґрунтоване судження стосовно можливих станів об'єкта в майбутньому, альтернативні шляхи й терміни їх здійснення. Прогноз має випадковий характер, та оскільки він будується на підставі аргументованих наукових уявлень про стан і розвиток об'єкта, здійснення його є доволі ймовірним. Процес розроблення прогнозів називають прогнозуванням.

Постійний еколого-соціально-економічний розвиток суспільства потребує невпинного ускладнення процедури управління. За ринкових умов прогнозування стає одним із вирішальних наукових чинників формулювання стратегії й тактики суспільного розвитку. Еколого-соціально-економічні (подалі – еколого-економічні) прогнози необхідні для визначення можливих цілей розвитку суспільства, забезпечення досягнення їх, сприяють зростанню економічних ресурсів для здійснення найімовірніших та економічно ефективних варіантів довготермінових, середньотермінових і поточних програм, обґрунтуванню основних напрямів економічної й технічної політики, вможливають передбачення наслідків рішень і заходів, що їх вживають у кожен поточний момент [1, 2].

У наш час прогнозування еколого-економічних процесів сформувалося у спеціальність, виникла нова галузь науки зі своєю специфічною методологією. Розроблення теорії та методів прогнозування еколого-економічної сфери дає змогу уникати експериментів над довкіллям і живими людьми, застосування нових екологобезпечних технологій і сучасної обчислювальної техніки в цій царині, забезпечує вибір раціонального способу управління суспільством.

Прогнозування еколого-економічних процесів – це наукова дисципліна, яка вивчає розроблення прогнозів розвитку національної економіки та екологічної сфери в майбутньому, ґрунтується на науковому пізнанні еколого-економічних явищ і використанні всієї сукупності методів, засобів і можливостей прогностики.

Метою еколого-економічного прогнозування є створення наукових передумов для прийняття ефективних рішень органами законодавчої та виконавчої влади, органами місцевого самоврядування. Ці передумови передбачають:

- науковий аналіз тенденцій зміни еколого-економічних процесів;
- варіантне передбачення розвитку еколого-економічних процесів з огляду на наявні тенденції й окреслену мету;
- оцінювання ймовірних наслідків рішень, які ухвалюються;
- обґрунтування напрямів еколого-економічного розвитку.

Завданням еколого-економічного прогнозування, з одного боку, є з'ясування перспективи найближчого або віддаленого майбутнього, вважаючи на реальні процеси сьогодення, а з іншого – сприяння розробленню оптимальних програм і планів розвитку ТВС, що має ґрунтуватися на пропонованому прогнозі й враховувати оцінку прийнятого рішення з позицій його наслідків у прогнозованому періоді.

Суб'єктами прогнозування еколого-економічного розвитку є органи державної влади й місцевого самоврядування, корпорації й підприємства,

також науково-дослідні й консалтингові організації, окремі експерти, яких залучають для розроблення й упровадження прогнозів.

Об'єктом еколого-економічного прогнозування є еколого-економічні процеси – тобто сукупність екологічних і економічних процесів формування та функціонування еколого-економічної системи, які характеризують динаміку зміни її параметрів на певному рівні господарювання.

Предметом еколого-економічного прогнозування є пізнання закономірностей соціально-економічних процесів в умовах екологічних обмежень у майбутньому, дослідження способів розроблення прогнозів.

Істотне значення має виконання *вимог*, пропонованих до *методики й організації еколого-економічного прогнозування*. Серед них виділяють наступні:

- варіантність розрахунків, в основі яких знаходиться облік розмаїтості умов розвитку і суспільних потреб;
- наукова обґрунтованість результатів прогнозування, що базується на об'єктивній і представницькій вихідній інформації;
- системність у розгляді об'єктів прогнозування, що припускає добір: факторів і умові відтворення, що найбільше істотно впливають на процеси розвитку ТВС; параметрів, що щонайкраще піддаються регулюючому впливу з боку суб'єктів прийняття рішень і ринку.

Однією з важливіших теоретичних проблем прогнозування є побудова типології прогнозів. Останню можна будувати залежно від різних критеріїв і ознак – цілей, завдань, об'єктів, методів організації прогнозування тощо [3-6]. До найважливіших із них належать:

- масштаб прогнозування (тобто мова йде про прогнозування ефективності діяльності суб'єктів господарювання на різних рівнях ієрархії управління) ;
- масштаб дії;
- термін прогнозування;

- тенденції зміни;
- функція прогнозу;
- характеру досліджуваних процесів.

За масштабом прогнозування розрізняють:

- макроекономічний (народногосподарський) та структурний (міжгалузевий і міжрегіональний) прогнози;
- прогнози розвитку народногосподарських комплексів (паливно-енергетичного, агропромислового, інвестиційного, виробничої інфраструктури, сфери обслуговування населення тощо);
- прогнози галузеві та регіональні;
- прогнози первинних ланок народногосподарської системи – підприємств, виробничих об'єднань;
- прогнози окремих виробництв і продуктів.

За масштабом дії прогноз може бути точковим або інтервальним. Точковий прогноз показника отримаємо підстановкою відомих значень фактора X_0 в отримане рівняння економетричної моделі $\hat{Y}_0 = X_0 \hat{A}$. При цьому дійсне значення показника Y для прогнозованого періоду дорівнює $Y_0 = X_0 A + u$, де u - вектор залишків. Помилка прогнозу обчислюється за виразом $v = Y_0 - \hat{Y}_0$. Дисперсія прогнозу дорівнює $\sigma_n^2 = \sigma_u^2 X_0^T (X^T X)^{-1} X_0$.

Інтервальний прогноз отримаємо для математичного сподівання залежної змінної Y та для індивідуального значення Y .

Довірчий інтервал для математичного сподівання залежної змінної Y :

$$\hat{Y}_0 - t_\alpha \hat{\sigma}_u \sqrt{X_0^T (X^T X)^{-1} X_0} \leq M(Y_0) \leq \hat{Y}_0 + t_\alpha \hat{\sigma}_u \sqrt{X_0^T (X^T X)^{-1} X_0},$$

де t_α - критичне значення t -критерію при $n-m-1$ ступенях вільності та рівні значущості α ; $\hat{\sigma}_u^2$ - скоригована дисперсія залишків, для якої маємо формулу

$$\hat{\sigma}_u^2 = \frac{\sum_{i=1}^n (Y - \hat{Y})^2}{n - m - 1} = \frac{u^T u}{n - m - 1}.$$

Інтервальний прогноз індивідуального значення має вигляд

$$\hat{Y}_0 - t_\alpha \hat{\sigma}_u \sqrt{1 + X_0^T (X^T X)^{-1} X_0} \leq Y_0 \leq \hat{Y}_0 + t_\alpha \hat{\sigma}_u \sqrt{1 + X_0^T (X^T X)^{-1} X_0} .$$

За терміном дії прогнози поділяють на:

- короткотермінові (від місяця до року);
- середньо термінові (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 припущенні, що у прогнозованому періоді не станеться ані кількісних, ані якісних суттєвих змін досліджуваного об'єкта. Середньо- та довготермінові 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. Форма оцінювання прогнозованих подій – якісна [4].

За способом опису тенденцій зміни чи об'єкта процесу прийнято виділяти *трендовий* і *факторний* прогнози. Зміст *трендового* підходу полягає в екстраполяції виправлених значень динамічного тимчасового ряду прогнозованого показника. Такий підхід використовується, як правило, для прогнозу показників, що характеризуються або монотонним зростанням, або монотонним убаванням. *Факторний* підхід припускає визначення кола, що виливають на прогнозований об'єкт (процес, показник) факторів і форм їхнього взаємозв'язку. Сутність факторного прогнозування полягає в тому, що перспективні оцінки прогнозованого показника одержують на основі заміни його динаміки динамікою значимих, щодо стійких тенденцій,

факторів і параметрів.

За способу визначення функціональних параметрів (тобто функціональною ознакою), що характеризують прогнозований об'єкт чи процес, прийнято виділяти *генетичний (пошуковий, ресурсний) і нормативний (цільовий)* підходи.

Генетичний підхід у прогнозуванні, що відбиває інерційний характер розвитку ТВС, ґрунтується на прогнозуванні стійких тенденцій у розвитку економіки і зводиться до перенесення залежностей, характерних для минулого і сьогодення, на майбутнє. *Пошуковий* прогноз ґрунтований на умовному продовженні в майбутнє тенденцій розвитку об'єкта, що вивчається в минулому й сучасному, абстрагуючись від можливих рішень, здатних радикально змінити ці тенденції (планів, програм тощо). Його завдання – з'ясувати, як розвиватиметься досліджуваний об'єкт за умов збереження наявних тенденцій.

Нормативний прогноз, на відміну від пошукового, розробляють на базі заздалегідь визначених цілей. Його завдання – визначити шляхи й терміни досягнення можливих станів об'єкта прогнозування в майбутньому, які визнано за мету. Іншими словами, *нормативний* підхід, що зв'язується з регулюванням стану ТВС, виходить з цілей розвитку, яких бажано досягти в прогнозованому періоді. Зміст нормативного підходу складається у визначенні параметрів відтворення, необхідних для досягнення заданих цілей розвитку ТВС. Разом з тим при даному підході враховується взаємозв'язок минулого, сьогодення і майбутнього. Це знаходить прояв у тім, що використані нормативні показники формуються в термінах минулого і сьогодення. Тобто, якщо пошуковий прогноз у визначенні майбутнього стану об'єкта відштовхується від минулого та нинішнього, то нормативний прогноз здійснюється у зворотному порядку: від заданого стану в майбутньому до наявних тенденцій та зміни їх задля поставленої мети.

Генетичний і нормативний підходи в прогнозуванні є альтернативними

способами виявлення параметрів, що характеризують прогнозований об'єкт чи процес. При генетичному підході взаємозв'язок процесів, що формують темпи і пропорції розвитку ТВС, розвертається від сьогодення до майбутнього, а при нормативному підході - від майбутнього до сьогодення.

Головною задачею прогнозування стає узгодження результатів прогнозних розрахунків, одержуваних на основі генетичного і нормативного підходів. Генетичний підхід покликаний забезпечити відповідність прогнозованих процесів реальним можливостям розвитку ТВС, а нормативний підхід повинний уможливити перехід від сформованих тенденцій до оптимального.

В процесі прогнозування виявляються існуючі проблеми розвитку ТВС й улаштовуються можливі способи їхнього рішення. Перелік таких проблем і їхня значимість можуть бути розкриті в процесі аналізу розбіжностей результатів перспективних розрахунків на основі генетичного і нормативного підходів. Величина таких розбіжностей характеризує міру складності назрілих питань розвитку ТВС.

Варіанти розвитку ТВС, які можуть розроблятися на основі спільного використання генетичного і нормативного підходів, при проведенні прогнозних розрахунків, наведено в табл. 1.

Таблиця 1

Методичні підходи при формуванні сценаріїв розвитку ТВС

Сценарії розвитку	Методи прогнозування
Песимістичний	Екстраполяція сформованих тенденцій на перспективу
Альтернативний	Сполучення генетичного і нормативного підходів
Оптимістичний	Нормативний

Розподіл прогнозів залежно від характеру досліджуваних процесів пов'язаний із різними аспектами процесу відтворення. За цією ознакою виокремлюють такі народногосподарські прогнози [2]:

- розвитку виробничих відносин;
- соціально-економічних передумов і наслідків науково-технічного прогресу;
- динаміки народного господарства (його темпів, чинників та структури);
- відтворення трудових ресурсів, зайнятості й підготовки кадрів;
- економічного використання природних ресурсів;
- відтворення основних фондів і капітальних вкладень;
- рівня життя населення;
- фінансових відносин;
- доходів і цін;
- зовнішніх економічних зв'язків тощо.

Кожен із названих напрямів має самостійне значення і може розроблятися на власній основі.

Для вибору методу прогнозування слід визначити мету і задачі прогнозу і період, на який він формується, врахувати специфіку об'єкта прогнозування, види, повноту і вірогідність вхідної інформації, а також ряд інших факторів.

Виходячи з цілей прогнозу, здійснюється вибір його виду. Тепер, за оцінками вчених, нараховується понад 150 різних методів прогнозування, але на практиці використовується тільки два-три десятки.

Класифікація методів прогнозування за загальним принципом дій і способом отримання інформації наведено на рис. 1.

Інтуїтивні методи використовуються у випадку неможливості кількісної оцінки окремих явищ (процесів) або неможливості врахування багатьох факторів через складність об'єктів. Ці методи базуються на використанні експертних оцінок. З їх допомогою встановлюється ступінь складності і актуальності проблеми, визначаються основні цілі та критерії, виявляються найбільш важливі фактори і взаємозв'язок між ними,

вибираються альтернативи, яким віддають найбільші переваги.

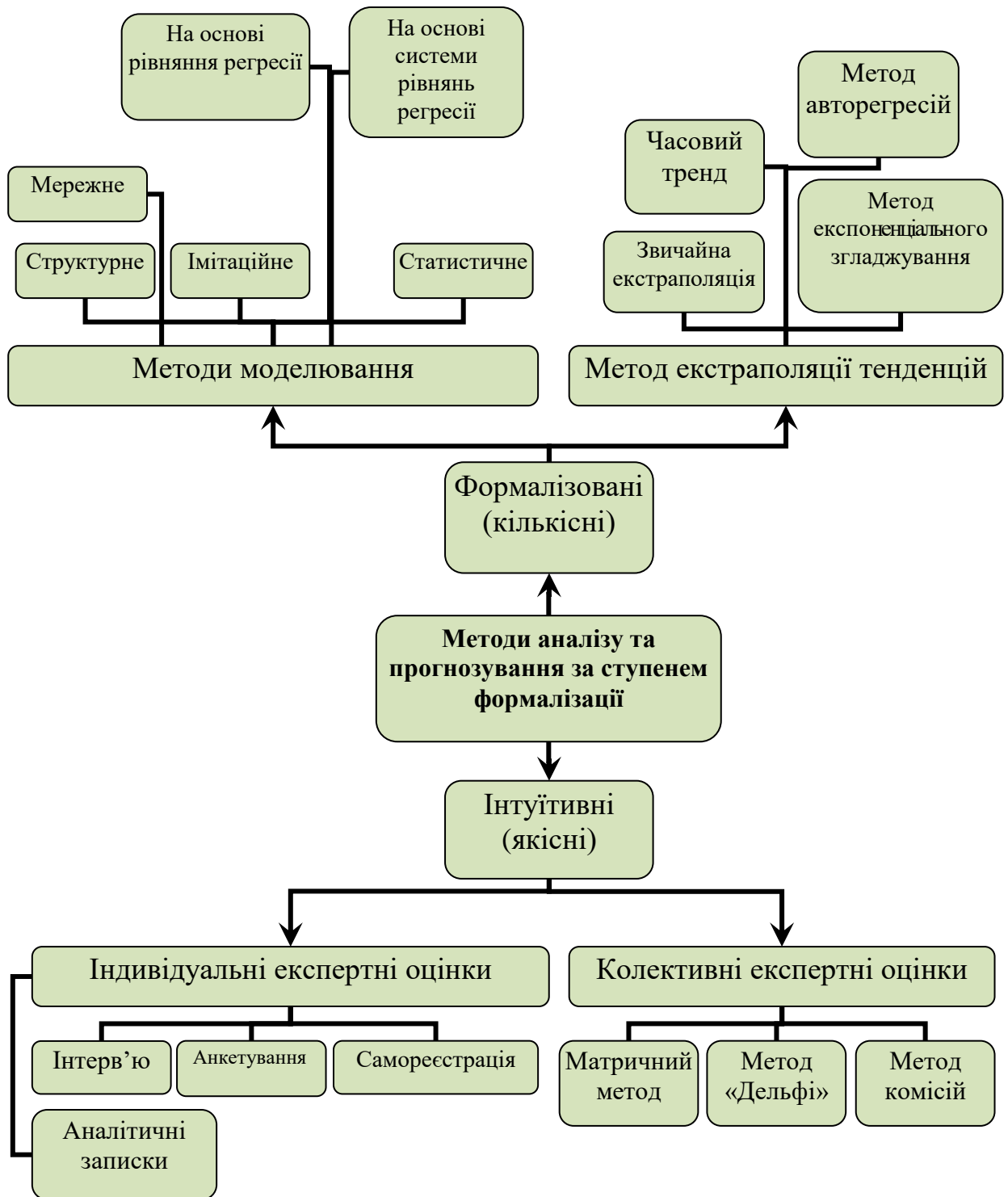


Рис. 1. Класифікація методів прогнозування

Серед *експертних* методів важливе місце відводиться методу експертних оцінок, в основі якого лежить використання оцінок визначеної групи людей – висококваліфікованих фахівців (експертів). Для одержання таких оцінок можуть використовуватися анкети, опитування, таблиці та інші

документи, за допомогою яких здійснюється збирання необхідної інформації. Існують очні, заочні, відкриті і закриті види опитувань. При цьому розрізняють *індивідуальні й колективні* експертні оцінки.

До складу *індивідуальних* експертних оцінок входять: метод "інтерв'ю", при якому здійснюється безпосередній контакт експерта з фахівцем зі схеми "питання-відповідь"; аналітичний метод, при якому створюється визначена прогнозована ситуація, складаються аналітичні доповідні записки; метод написання сценарію, що базується на визначенні логіки чи процесу явища в часі за різних умов.

Колективні експертні оцінки містять у собі такі методи: комісій, колективної генерації ідей ("мозкової атаки"), "Дельфі", матричний. Ця група методів базується на тому, що, по-перше, при колективному мисленні вища точність результату і, по-друге, при обробці індивідуальних незалежних оцінок, що дають експерти, можуть відбутися раціональні ідеї. Суть методу колективної експертної оцінки для розробки прогнозів складається у визначенні погодженості думок експертів з напрямків розвитку об'єкта прогнозування, що сформульовано раніше окремими фахівцями, а також в оцінці розвитку об'єкта, що не може бути визначена іншими методами (наприклад, аналітичним розрахунком, експериментом і ін.).

Метод *колективної генерації ідей* ("мозкова атака") дає можливість одержати продуктивні результати за короткий період часу і залучити всіх експертів до активного творчого процесу.

Формалізовані методи прогнозування базуються на аналітичних сітках, що містять рівняння, які репрезентують і сукупний попит, і сукупну пропозицію. До групи формалізованих методів входять *методи екстраполяції і моделювання*.

Метод *екстраполяції* – один з основних у прогнозуванні економіки. Загалом методи екстраполяції – це прийоми найменших квадратів, рухомих середніх, експоненційного згладжування. Під час формування прогнозів за

допомогою екстраполяції звичайно спираються на статистично обґрунтовані тенденції зміни тих чи інших кількісних характеристик об'єкта. Екстраполяційні методи є одними з найбільш розповсюджених і розроблених серед усіх способів економічного прогнозування. Ці методи передбачають, що на основі статистичних даних досліджуються закономірності і тенденції економічних явищ. Теоретичні знання про взаємозв'язок між економічними показниками мають підказати його конкретну аналітичну форму. Але оскільки одні й ті самі економічні процеси можуть бути описані різними функціями, то потрібно звернутися до статистичного аналізу і з його допомогою зробити вибір серед можливих альтернативних варіантів.

Найпростішою вважається лінійна форма зв'язку між двома змінними:

$$y = a_0 + a_1x,$$

де a_0 , a_1 – невідомі параметри.

Цей метод ґрунтується на припущенні, що незмінні фактори при розвитку даною явища в минулому будуть діяти й у майбутньому. При формуванні прогнозу за допомогою екстраполяції виходять з тенденцій зміни тих чи інших кількісних характеристик об'єкта. Екстраполюються оцінні, функціональні, системні і структурні характеристики.

Екстраполяційні методи є найбільш розповсюдженими і розробленими. Їх можна умовно поділити на прості та складні. Перші базуються на використанні аналітичних показників динамічних рядів, другі – на використанні статистичних формул. Складні методи екстраполяції тенденції у свою чергу можна поділити на трендові (аналітичні криві росту) і адаптивні. Адаптивні методи прогнозування основані на тому, що у процесі реалізації їх здійснюється обчислення послідовних у часі значень прогнозованого показника з врахуванням ступенів впливу попередніх рівнів. До них відносяться: методи експоненціального згладжування, метод гармонійних вагів, метод авторегресивних перетворень.

В основу аналітичних методів (кривих росту) прогнозування

закладений принцип отримання за допомогою методу найменших квадратів оцінки детермінованої компоненти, яка характеризує основну тенденцію.

Особливе місце у сучасному прогнозуванні займає статистичне моделювання – однофакторне і багатфакторне.

Крім цього методи прогнозування на основі статистичного моделювання можуть бути поділені на дві великі групи: прогнозування на основі поодиноких рівнянь регресії і прогнозування на основі системи рівнянь взаємопов'язаних рядів динаміки.

Регресійний аналіз – це математичний метод прогнозування, результатом якого є рівняння з однією або більшою кількістю незалежних змінних, яке використовується для визначення залежної змінної. Зміст регресійного аналізу полягає у дослідженні того, як зміна незалежних змінних впливає на залежну змінну. Один раз визначені взаємозв'язки вважаються ustalеними (у вигляді рівняння регресії), а майбутні значення залежної змінної прогноуються шляхом підстановки у рівняння певних значень незалежних змінних. Регресійний аналіз є відносно дорогим, але комплексним і надійним прийомом.

Найскладнішими серед методів кількісного прогнозування є комплексні методи *економетричного моделювання*. Цей метод вважається досить ефективним способом прогнозування можливого явища, нових чи майбутніх економічних і технічних способів і рішень. Вперше для цілей прогнозування складання моделей було розпочато в економіці з метою рішення задач організаційного керування, що характеризуються великою розмірністю і складністю і які неможливо вирішити за допомогою математичного програмування й аналізу в рамках теорії ймовірності.

Переважно, економетричні моделі "прив'язуються" до математичної моделі цілої економіки. Складні економетричні моделі базуються на численних рівняннях регресії, які кількісно описують взаємозв'язки між різними секторами економіки. Слід зазначити, що навіть складні

економетричні моделі не можуть забезпечити стовідсоткову точність прогнозів.

Модель є одним з важливих інструментів економічного прогнозування, наукового пізнання досліджуваного процесу. Конструювання моделі на основі попереднього вивчення об'єкта і визначення його істотних характеристик, експериментальний і теоретичний аналіз моделей, зіставлення результатів з даними об'єкта, коригування моделі складають зміст методу моделювання.

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.

Незалежні змінні моделі називаються факторами, наперед заданими змінними. Залежні змінні називаються показниками або результуючими ознаками. Інколи вживаються і інші назви "регресор", "регресант".

Економетрична модель, що будується на основі системи рівнянь, крім регресійних функцій, може включати алгебраїчні рівняння.

Побудова будь-якої економетричної моделі, незалежно від того, на якому рівні і для яких показників вона будується, здійснюється як послідовність п'яти кроків [5-7].

Крок 1. Встановлення причинно-наслідкового зв'язку між досліджуваними економічними показниками.

Крок 2. Вибір найбільш істотних ознак і встановлення вигляду і форми

функції зв'язку.

Крок 3. Знаходження параметрів функції зв'язку.

Крок 4. Оцінка достовірності отриманих результатів.

Крок 5. Прогнозування значень результуючої ознаки та їх аналіз.

Дослідження об'єкту моделювання і складання його економіко-математичного опису полягають у встановленні зв'язків між характеристиками процесу, виявленні його граничних і початкових умов та формалізації процесу у вигляді системи математичних співвідношень.

Процес побудови економіко-математичної моделі можна представити у вигляді алгоритму побудови, зображеного на рис. 2 [6].

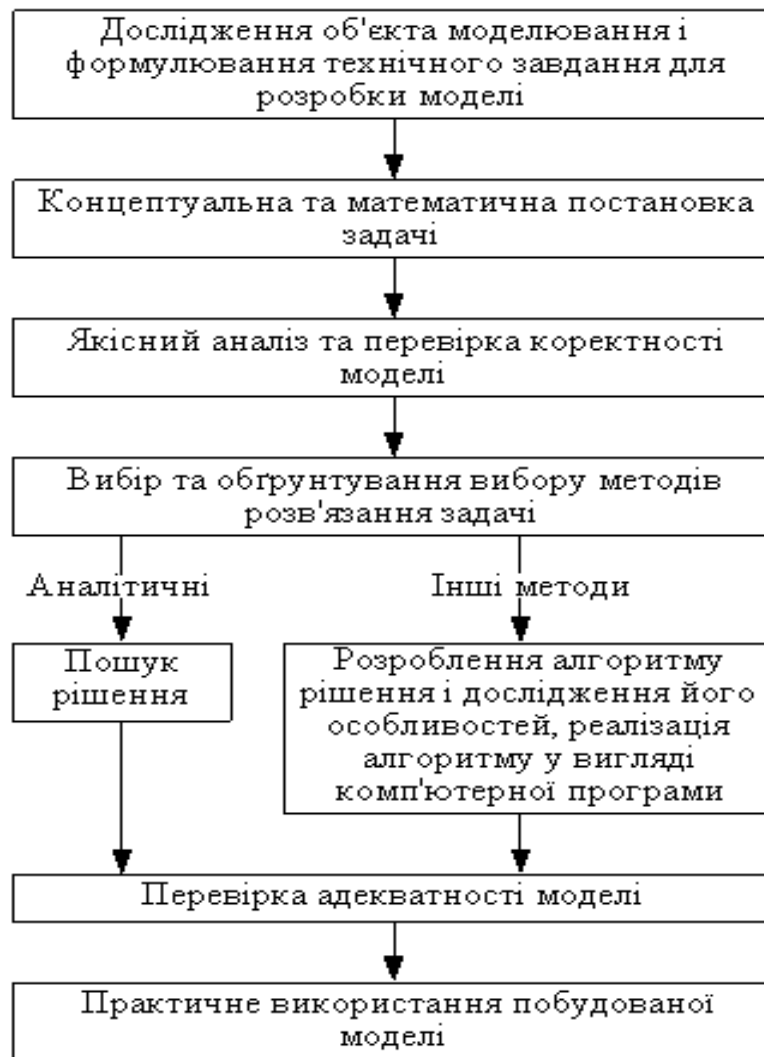


Рис. 2. Етапи побудови економіко-математичної моделі

Розробка економіко-математичних моделей розвитку соціального розвитку і сценаріїв розвитку досліджуваної системи дає можливість:

- зменшити невизначеність задачі;
- сформулювати обмежену множину рішень;
- визначити варіанти прийняття остаточного рішення відповідно до обумовлених критеріїв.

У загальному вигляді рішення задачі моделювання, з використанням прогнозування, можна схематично представити у вигляді усіченого «конуса рішень», представленого на рис. 3 [6].

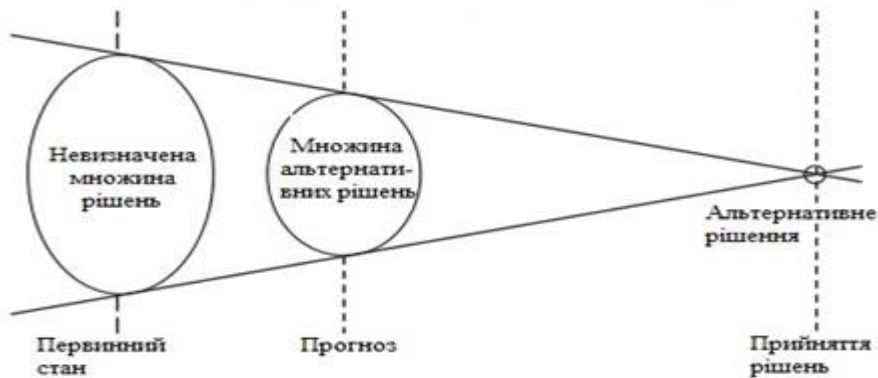


Рис. 3. Усічений конус рішення економіко-соціальних задач

Економетрична модель базується на єдності двох аспектів – теоретичного, якісного аналізу взаємозв'язків та емпіричної інформації.

Моделювання носить комплексний характер і охоплює всі рівні й аспекти розширеного відтворення. Ступінь впливу на динаміку соціального розвитку реалізується через систему кількісних показників – загроз, значення та кількість яких змінюються в міру зрушень у еколого-економічному стані суспільства, пріоритетах його розвитку. Загалом будь-яке еколого-економічне дослідження проходить такі етапи аналізу:

- формування теоретичної проблеми або гіпотези (теоретичний опис проблематичних явищ, визначення цілей дослідження, відбір економічних змінних тощо);

- вибір типу економіко-математичної моделі;
- розрахунок та оцінка параметрів моделі;
- інтерпретація отриманих результатів;
- застосування моделі (для прогнозування, для моделювання тощо).



2. КОРЕЛЯЦІЙНИЙ АНАЛІЗ: ТЕОРЕТИЧНІ ВІДОМОСТІ

План лекції

- 1. Дефініція поняття "кореляція" та "кореляційний зв'язок"*
- 2. Об'єкти кореляційних зв'язків для побудови економетричних прогнозних моделей*
- 3. Види кореляційних зв'язків*
- 4. Основні задачі кореляційного аналізу*
- 5. Однофакторна та багатофакторна кореляція*
- 6. Коефіцієнт детермінації*
- 7. Множинний коефіцієнт кореляції*
- 8. Основи побудови кореляційної матриці*
- 9. Алгоритм групування даних для кореляційного аналізу*

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 ймовірностний розподіл. Тому в еколого-економічних дослідженнях спостерігаються і розглядаються так звані статистичні (або кореляційні) залежності.

Також усі еколого-соціально-економічні явища взаємозв'язані та

взаємозумовлені і зв'язок (залежність) між ними носить причинно-наслідковий характер. Суть причинного зв'язку полягає в тому, що при необхідних умовах одне явище зумовлює інше і в результаті такої взаємодії виникає наслідок. Кореляційна залежність виникає тоді, коли одна з величин залежить не тільки від заданої другої, а й від деяких випадкових факторів; або, коли серед умов, від яких залежать обидві величини, є загальні для них обох.

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.

Кореляційний зв'язок може мати різну степінь – від повної незалежності до функціональної залежності. Крім того, характер зв'язку між різними величинами може бути різний. Тому виникає необхідність визначити форму, напрям і ступінь кореляційних зв'язків.

В якості *об'єктів кореляційних зв'язків* для побудови економетричних прогнозних моделей можуть розглядатися правові, економічні та організаційні засади формування цілісної системи прогнозних і програмних документів економічного і соціального розвитку України, окремих галузей економіки та окремих адміністративно-територіальних одиниць як складової частини загальної системи державного регулювання економічного і соціального розвитку держави.

Особливе місце в системі еколого-економічного прогнозування (моделювання) займають об'єкти, пов'язані з формуванням соціальних, економічних та екологічних ризиків та небезпек. При обговоренні проблеми об'єктів еколого-економічного прогнозування (моделювання) доцільно прийняти до уваги дослідження цього питання в різних наукових сферах та

адміністративній діяльності. Класифікація можливих кореляційно пов'язаних науково-дослідних сфер (об'єктів) прогнозування приведено на рис. 4 [8].

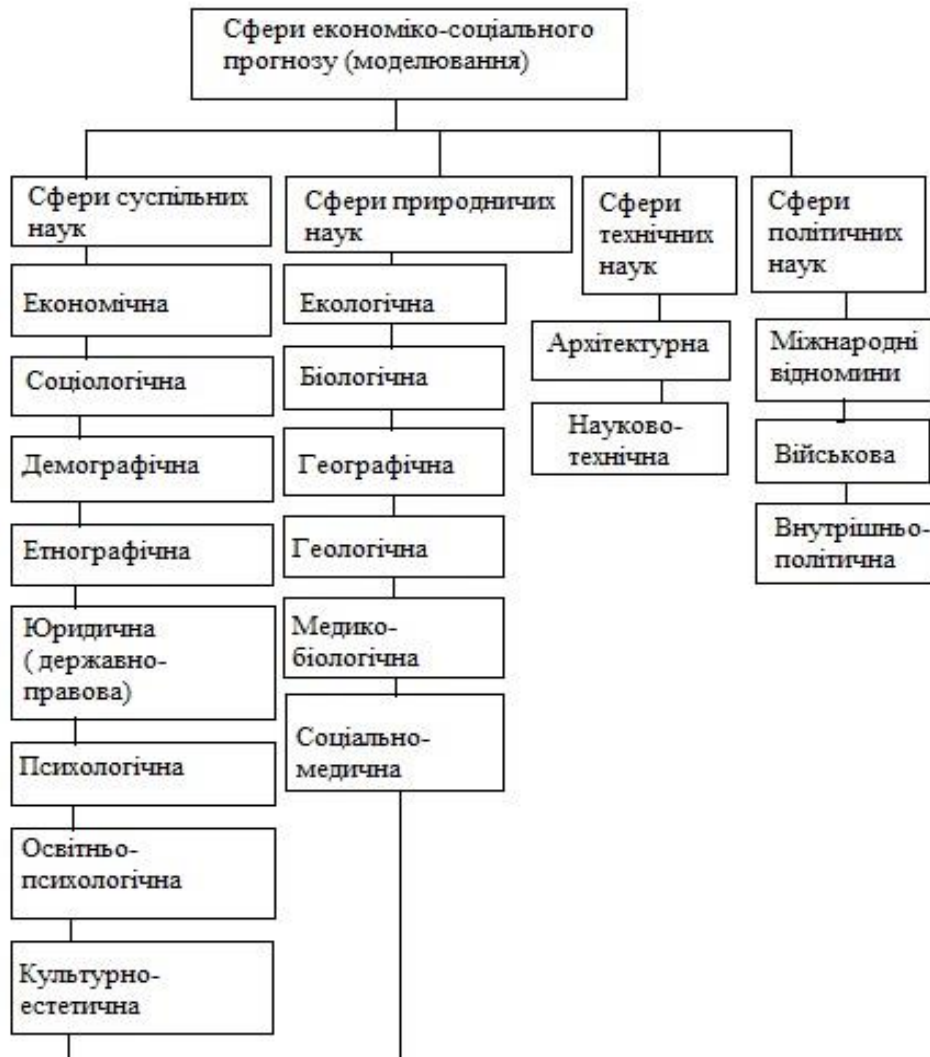


Рис. 4. Класифікації об'єктів еколого-економічного прогнозування

Найпростішим видом кореляційного зв'язку є зв'язок між двома ознаками: результативною і факторною. Такий зв'язок називають парною кореляцією або простою кореляцією. В економічних дослідженнях взаємозв'язку двох факторів серед множини функцій часто розглядається прямолінійна форма зв'язку, яка виражається рівнянням прямої лінії.

Щодо типізації кореляційного зв'язку, за *формою кореляція може бути прямолінійною і нелінійною (криволінійною)*, за *напрямком – прямою і оберненою*. При прямій 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.

Нелінійні форми зв'язку притаманні багатьом процесам у сільському господарстві. Так, ріст і розвиток рослин, накопичення ними продуктивної маси, як правило, в часі розвивається нелінійно.

При дослідженні криволінійних зв'язків, так само як і при вивченні лінійних зв'язків, принципове значення має вибір форми і рівняння зв'язку, яке найточніше відобразить наявний зв'язок. Для розв'язання цього завдання використовуються ті самі прийоми, що й при обґрунтуванні лінійного зв'язку. При цьому особлива увага належить графічному методу. Криволінійні форми зв'язку досить різноманітні. В статистичному аналізі найчастіше використовують параболу другого порядку, гіперболу і степеневу функцію.

В багатьох прикладних задачах необхідно виявити залежність між двома властивостями (ознаками) X і Y одного і того ж економічного об'єкту, або між певними ознаками різних об'єктів. Якщо вказані ознаки допускають кількісне вимірювання, і, з погляду економічної теорії, виходячи з економічної або екологічної характеристики об'єкту, ознака Y залежить від ознаки X , тоді X можна назвати незалежною змінною, або *факторною ознакою*, або просто фактором, а Y – залежною змінною або *результативною ознакою*.

Якщо кожному значенню факторної ознаки X відповідає одне і тільки одне значення результативної ознаки Y , то говорять, що між цими ознаками існує *функціональний зв'язок*: $Y=f(X)$.

Якщо кожному значенню факторної ознаки X відповідає безліч значень результативної ознаки Y , то говорять, що між цими ознаками існує *статистичний зв'язок*.

Наприклад, якщо X приймає l значень $X = \{x_1, x_2, \dots, x_l\}$ і кожному її значенню x_i відповідає множина значень Y , тобто:

значенню x_1 відповідає множина $\{y_{11}, y_{12}, \dots, y_{1m_1}\}$;

значенню x_2 відповідає множина $\{y_{21}, y_{22}, \dots, y_{2m_2}\}$;

...

значенню x_l відповідає множина $\{y_{l1}, y_{l2}, \dots, y_{lm_l}\}$;

то між X та Y існує статистичний зв'язок.

Слід відмітити, що вивчення статистичного зв'язку дуже складний і трудомісткий процес, у якому потрібно аналізувати багатомірні таблиці даних. Тому зазвичай вивчається не статистичний, а кореляційний зв'язок між X та Y . Тобто, статистичний зв'язок між змінними можна вивчати методом кореляційного аналізу.

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.

Розглянемо суть кореляційного аналізу. Нехай дано систему m нормально розподілених випадкових величин. Для випадкового вектора 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 щільності 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 вибіркового даних є коефіцієнт багатофакторної кореляції, який має вигляд

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

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

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

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:

$$R^2 = 1 - \frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{\sum_{i=1}^n (y_i - \bar{y})^2} \cdot \frac{n-1}{n-m}. \quad (3)$$

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

У випадку однофакторної лінійної регресії коефіцієнт кореляції можна обчислити за формулою

$$R_{xy} = \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^n (x_i - \bar{x})^2 \sum_{i=1}^n (y_i - \bar{y})^2}}. \quad (4)$$

Коефіцієнт кореляції R_{xy} – це відносна міра зв'язку між двома факторами. Тому значення коефіцієнта кореляції завжди розташовані між -1 та +1, тобто $-1 < R_{xy} < 1$.

Додатне значення коефіцієнта кореляції свідчить про прямий зв'язок між показниками, а від'ємне – про зворотній зв'язок.

В матричній формі коефіцієнт множинної кореляції має вигляд:

$$R = \sqrt{\frac{\hat{A}^T X^T Y}{Y^T Y} \cdot \frac{n-1}{m-1}}. \quad (5)$$

Статистичну значущість коефіцієнта детермінації можна перевірити за допомогою F - критерію:

$$F_R = \frac{R^2}{1-R^2} \cdot \frac{n-m-1}{m}. \quad (6)$$

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 \wedge 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 :

$$A = \begin{pmatrix} 1 & r_{12} & \dots & r_{1k} \\ r_{21} & 1 & \dots & r_{2k} \\ \dots & \dots & \dots & \dots \\ r_{k1} & r_{k2} & \dots & r_{kk} \end{pmatrix}; \quad (7)$$

– знайти визначник $|A|$ матриці A та алгебраїчне доповнення A_{ii} елемента r_{ii} цієї матриці;

– розрахувати множинний коефіцієнт кореляції за формулою:

$$R_i = \sqrt{1 - \frac{|A|}{A_{ii}}}. \quad (8)$$

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

$$t = \frac{R^2(n-k)}{(1-R^2)(k-1)}, \quad (9)$$

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

FRASПОБР ($\alpha; l_1; l_2$), 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 .

Частинний коефіцієнт кореляції характеризують тісний зв'язок між двома змінними при фіксованих значеннях інших незалежних змінних. Для розрахунку частинного коефіцієнта кореляції необхідно:

- побудувати матрицю парних коефіцієнтів кореляції A ;
- знайти алгебраїчні доповнення A_{ii}, A_{jj}, A_{ij} елементів r_{ii}, r_{jj}, r_{ij} відповідно;
- розрахувати частинний коефіцієнт кореляції за формулою

$$r_{ij,12\dots m} = \frac{A_{ij}}{\sqrt{A_{ii}A_{jj}}} \quad (10)$$

де A_{ij}, A_{ii}, A_{jj} – алгебраїчні доповнення, відповідно до елементів r_{ij}, r_{ii}, r_{jj} матриці парних коефіцієнтів кореляції між незалежними змінними.

Частинний коефіцієнт кореляції між змінними X_i і X_j , можна подати в такому вигляді:

$$r_{ij,12\dots m} = \frac{z_{ij}}{\sqrt{z_{ii}z_{jj}}} \quad (11)$$

де z_{ij}, z_{ii}, z_{jj} – елементи матриці $z = r^{-1}$.

Після визначення частинних коефіцієнтів кореляції необхідно перевірити гіпотезу про значущість кореляційного зв'язку між певними ознаками. Вводиться нульова гіпотеза H_0 про відсутність кореляційного зв'язку між ознаками і гіпотеза H_1 про наявність зв'язку між цими ознаками. Ці гіпотези перевіряють за допомогою t -критерію:

$$t_{ij} = r_{ij,12\dots m} \sqrt{\frac{n-m-1}{1-r_{ij,12\dots m}^2}}, \quad (12)$$

яка має t - розподіл Стюдента з $k = n - m - 1$ ступенями вільності, де n – загальна кількість спостережень, m – порядок матриці парних коефіцієнтів кореляції між незалежними змінними.

За таблицею розподілу Стюдента при рівні значущості α і числу ступенів вільності k знаходимо критичне значення $t_{\alpha,k}$. Якщо $|t_{ij}| \geq t_{\alpha,k}$, то нульову гіпотезу про відсутність кореляційного зв'язку між ознаками X_i і X_j слід відкинути; тобто при рівні значущості α можна вважати, що кореляційний зв'язок між ознаками X_i і X_j існує.

Коефіцієнт кореляції приймає значення на проміжку $[-1;1]$, тобто $-1 \leq r \leq 1$. Якщо $r > 0$, то зв'язок називається додатнім, тобто зі збільшенням значень X значення Y також збільшуються. Якщо $r < 0$, 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

Групування даних для кореляційного аналізу

Вибіркові 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 (позначення: k – кількість інтервалів для X ; m – кількість інтервалів для Y).

Крок 2. Зобразити дані графічно: побудувати на площині точки з координатами $(x_i; y_j)$. В результаті отримується площина, розбита на прямокутники, в кожному з яких може бути множина точок (рис. 5). Вказане графічне зображення вибірових даних називається *полем кореляції*.

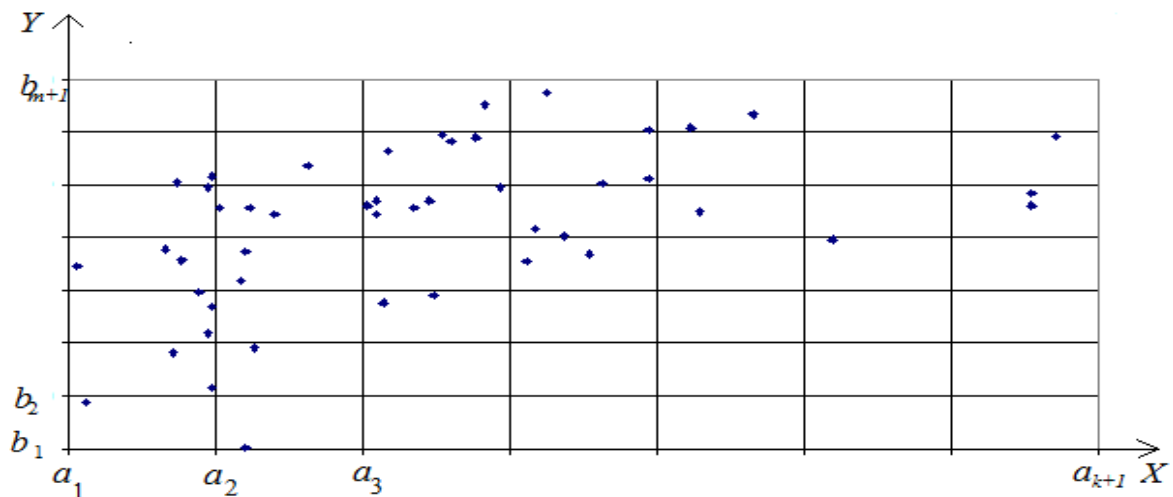


Рис. 5. Поле кореляції

Крок 3. Побудувати таблицю кореляційної матриці (табл. 4).

Шаблон кореляційної матриці

X (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			
Y (інтервали та їх середини)		$[b_1; b_2)$	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 потрапили 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 потрапили 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 потрапили 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 .

Кореляційну таблицю 4 можна розглядати як своєрідний подвійний статистичний ряд.

Крок 4. За даними таб. 4 будується ряд, що відображає залежність середнього значення Y від X (табл. 5).

Шаблон побудови ряду, що відображає залежність середнього значення
Y від X

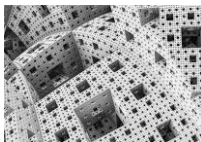
x_i	x_1	x_2	...	x_k
\bar{y}_{x_i}	\bar{y}_{x_1}	\bar{y}_{x_2}	...	\bar{y}_{x_k}
n_i	n_1	n_2	...	n_k

У першому рядку табл. 5 записуються середини інтервалів x_i . У другому – відповідні середні значення \bar{y}_{x_i} , що знаходяться за формулами:

$$\bar{y}_{x_1} = \frac{y_1 n_{11} + y_2 n_{12} + \dots + y_m n_{1m}}{n_1}; \quad \bar{y}_{x_2} = \frac{y_1 n_{21} + y_2 n_{22} + \dots + y_m n_{2m}}{n_2}; \quad \dots;$$

$$\bar{y}_{x_k} = \frac{y_1 n_{k1} + y_2 n_{k2} + \dots + y_m n_{km}}{n_k}.$$

У результаті отримується статистичний ряд, що містить значення X, відповідні середні значення Y та частоти. За даними такого ряду проводиться кореляційний аналіз.



3. ОБЧИСЛЕННЯ КОРЕЛЯЦІЙНОЇ МАТРИЦІ В EXCEL

План лекції

- 1. Алгоритм обчислення матриці кореляції**
- 2. Використання функції КОРРЕЛ**
- 3. Використання надбудови "Пакету аналізу"**

За даними табл. 6, наведемо приклад побудови кореляційної матриці та визначення ступеню взаємозв'язку між факторами (показниками, індикаторами) загроз енергетики в регіональному розрізі.

Таблиця 6

Значення регіональних показників еколого-економічних загроз енергетичної безпеки України за 2012 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.

1 спосіб (використання функції КОРРЕЛ)

Крок 1. Транспортувати табл. 6 в книгу Excel на Лист 1 (рис. 7).

	A	B	C	D	E	F	G
1	Регион	Частка газу природного	Частка ветхих та аварійних	Частка втрат теплоенергії при транспорт уванні	Частка викидів від енергетики	Середньод ушові обсяги викидів від енергетики	Викинутю 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-Львівська	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	н/д
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

Рис. 7. Скриншот 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).

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q
1	Регион	Частка газу природного	Частка ветхих та аварійних	Частка втрат теплоенергії при транспорт уванні	Частка викидів від енергетики	Середньод ушові обсяги викидів від енергетики	Викинутю 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-Львівська	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. Скриншот із виділеним діапазоном

Крок 3. Поставити курсор у верхню ліву комірку виділеного діапазону (див. рис. 8 – комірка I3) для розрахунку коефіцієнта кореляції (див. формулу 1) першого показника самого із собою.

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

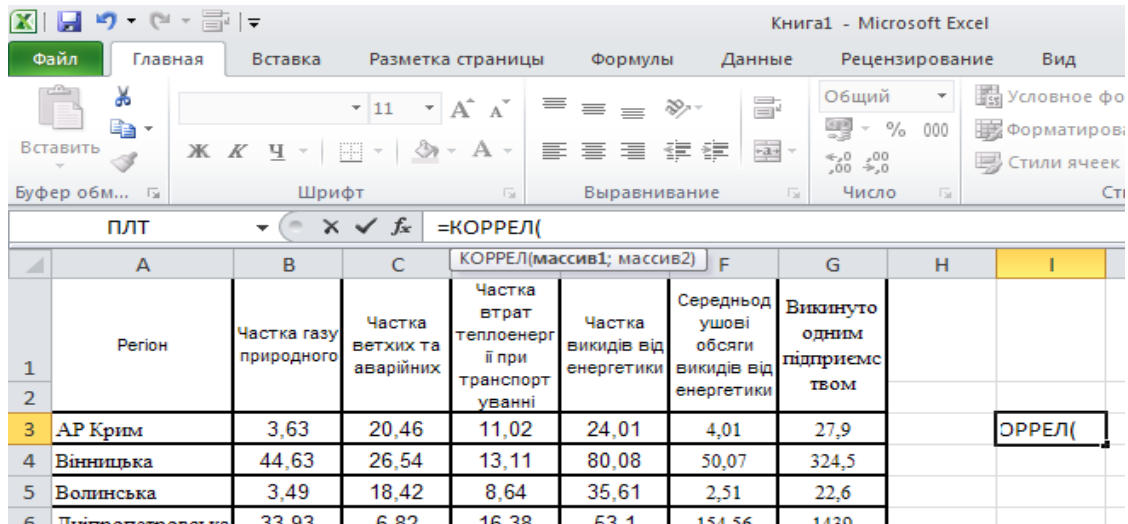


Рис. 9. Скриншот вводу функції «=КОРРЕЛ(»

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

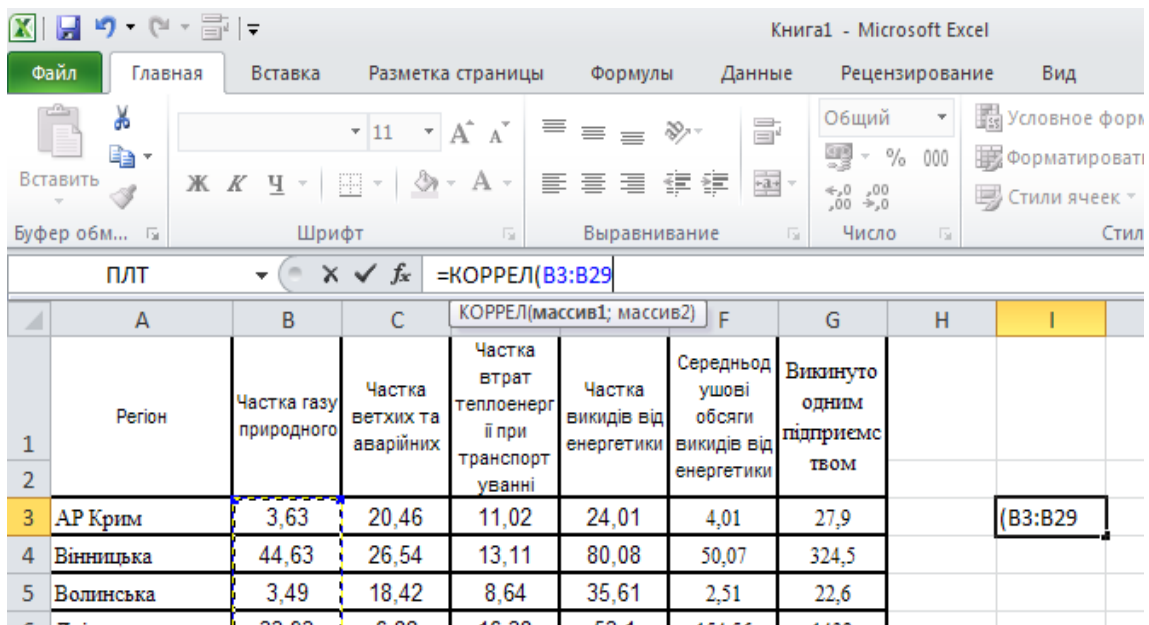


Рис. 10. Скриншот виділення даних

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

The screenshot shows the Microsoft Excel interface with the following data in the spreadsheet:

	A	B	C	D	E	F	G	H	I	J
1	Реґіон	Частка газу природного	Частка ветхих та аварійних	Частка втрат теплоенерґії при транспорт уванні	Частка викидів від енерґетики	Середньод ушові обсяги викидів від енерґетики	Викинуто одним підприємством			
2										
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			
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	н/д			
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			

Рис. 11. Скриншот розрахунку першого коефіцієнту 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).

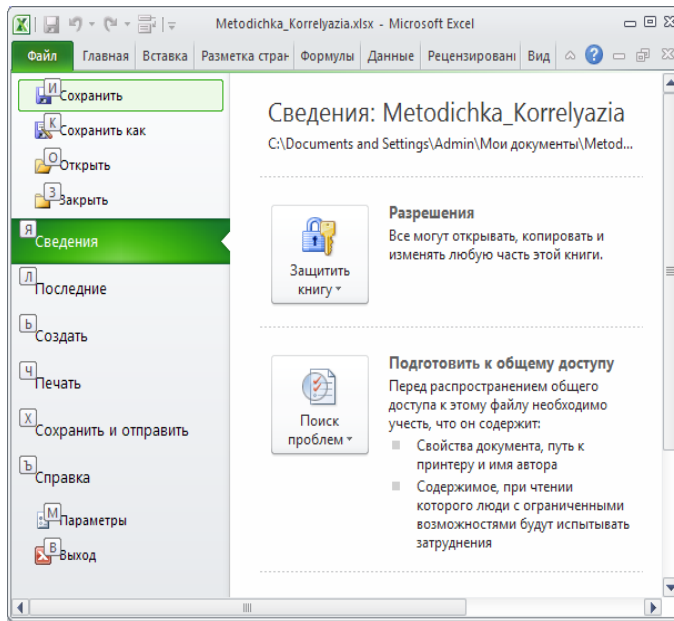
Region	Частка газу природного	Частка ветхих та аварійних	Частка втрат теплоенергії при транспортуванні	Частка викидів від енергетики	Середньодушові обсяги викидів від енергетики	Виклинуто одним підприємством								
АР Крим	3,63	20,46	11,02	24,01	4,01	27,9	1,000	-0,088	0,156	0,565	0,785	0,743		
Вінницька	44,63	26,54	13,11	80,08	50,07	324,5	-0,088	1,000	0,422	-0,017	-0,253	-0,192		
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		
Дніпропетровська	33,93	6,82	16,38	53,1	154,56	1439	0,565	-0,017	0,256	1,000	0,473	0,539		
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		
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		
Закарпатська	1,16	9,9	1,62	22,37	1,45	31,2								
Запорізька	79,8	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	9,9	20,5	3,48	0								
Луганська	40,48	7,26	5,29	38,94	77,4	274,1								
Львівська	21,17	28,7	14,11	45,88	23,77	251,9								
Миколаївська	6,15	4,43	10,42	26,42	5,7	54,9								
Одеська	1,24	40,22	13,91	22,61	2,67	31,9								

Рис. 12. Скриншот побудованої матриці кореляції

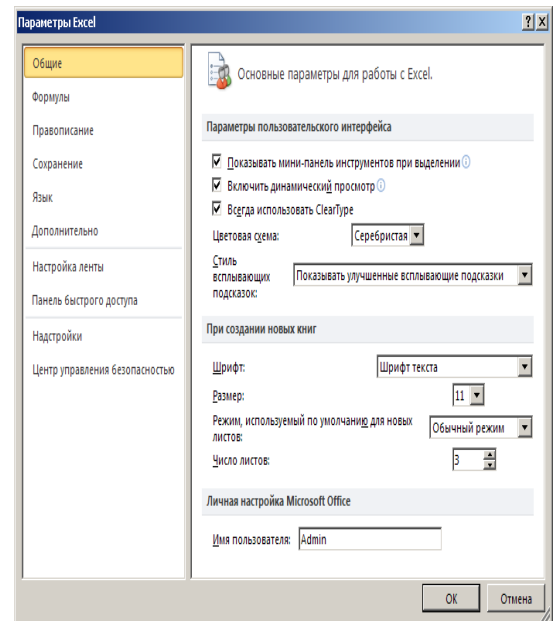
2 спосіб (використання надбудови "Пакету аналізу")

Крок 1. У випадку відсутності кнопки "Пакет аналіза" на закладці "Данные" в панелі інструментів в Excel необхідно підключити надбудову "Пакет аналіза" (рис. 13). Для цього необхідно перейти на закладку "Файл" і вибрати меню "Параметри" (див. рис. 13 – а). Активізувати вікно "Параметры Excel".

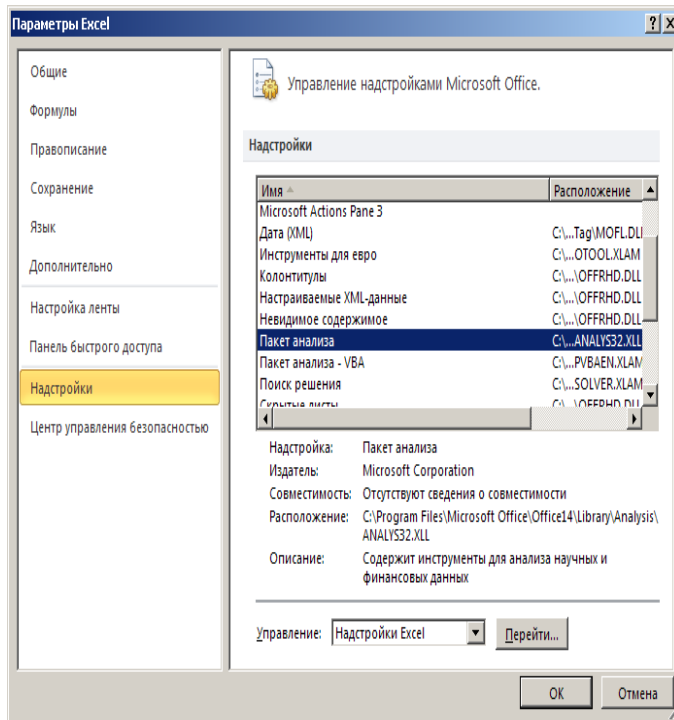
У вікні "Параметры Excel" (див. рис. 13 – б) вибираємо меню "Надстройки". У вікні меню "Надстройки" вибираємо "Пакет аналіза" (див. рис. 13 – в). Активізуємо надбудову "Пакет аналіза" (див. рис. 13– г).



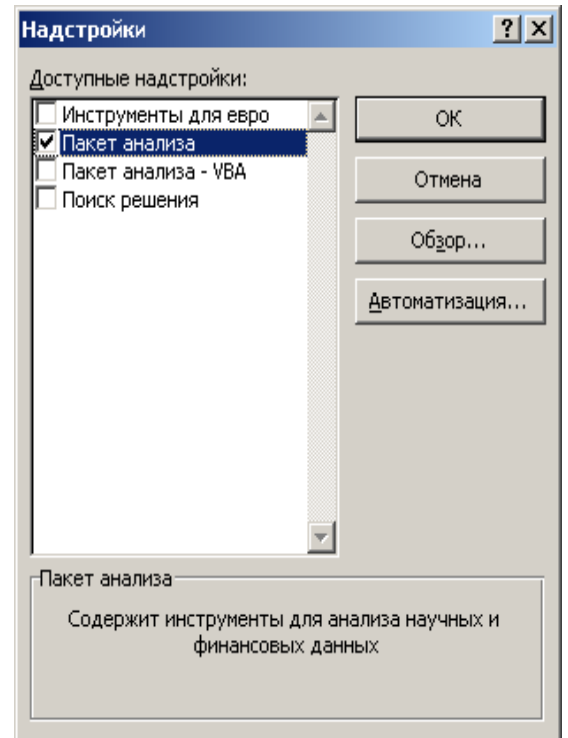
а



б



в



г

Рис.13. Скриншоты активации надбудови "Пакет анализа" в Excel

В іншому випадку на закладці "Данные" в панелі інструментів в Excel необхідно натиснути на кнопку "Анализ данных" (рис. 14).

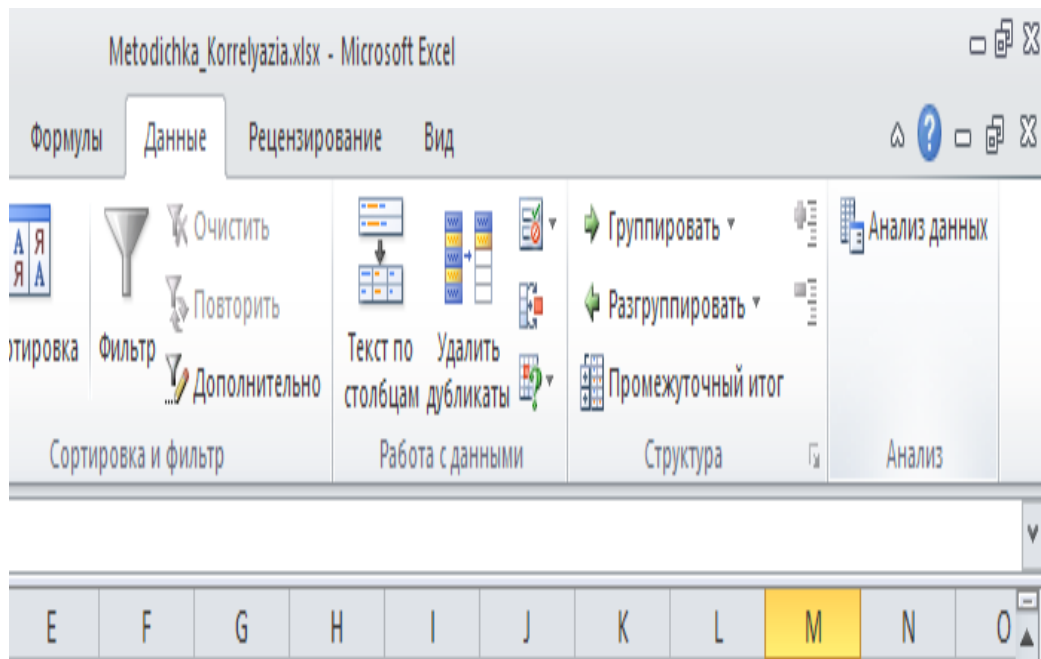


Рис. 14. Скриншот виклику надбудови "Пакет анализа"

У вікні "Анализ данных" для розрахунку матриці кореляції необхідно виділити пункт "Корреляция" та натиснути на кнопки "ОК" (рис. 15).

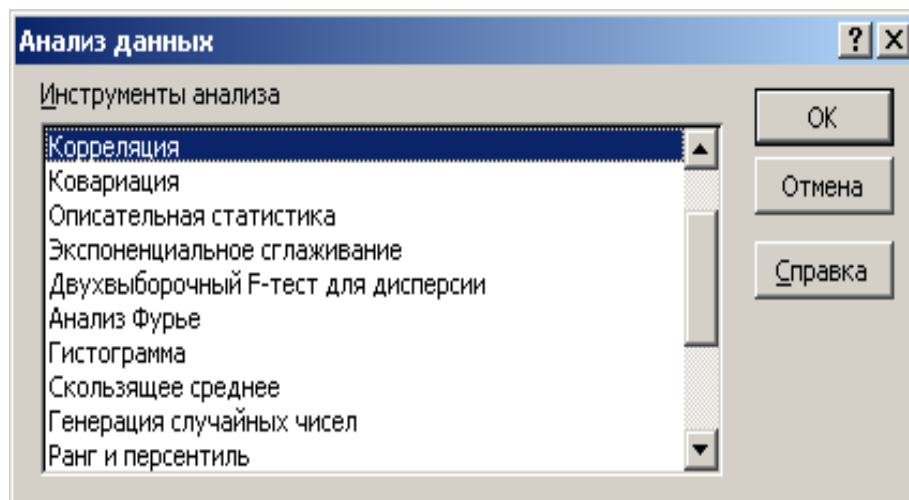




Рис. 15. Скриншот виклику інструменту "Корреляция"

У вікні "Корреляция" необхідно в елементі управління "Входной интервал:" натиснути кнопку .

При цьому відбудеться відображення робочого листа.

За допомогою мишки необхідно виділити масив комірок з даними та натиснути кнопку  для повернення до вікна «Корреляция» (рис. 16).

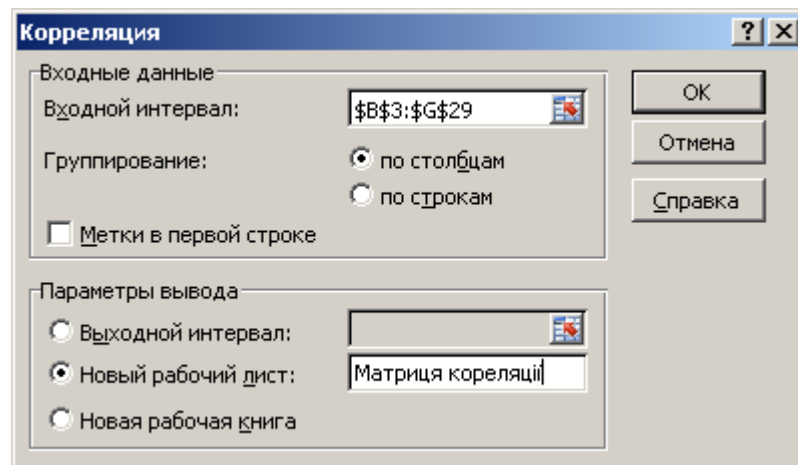


Рис. 16. Скриншот вікна інструменту "Корреляция"

В області "Параметры вывода" необхідно перемикнути радіокнопку у положення "Новый рабочий лист" та ввести назву листа у текстове поле. При натисканні на кнопку "ОК" з'явиться новий робочий лист із трикутною матрицею кореляції (рис. 17).

	Столбец 1	Столбец 2	Столбец 3	Столбец 4	Столбец 5	Столбец 6
Столбец 1	1					
Столбец 2	-0,088272	1				
Столбец 3	0,1558293	0,4219607	1			
Столбец 4	0,5650226	-0,01714	0,25577835	1		
Столбец 5	0,7851412	-0,252555	0,21361612	0,47327355	1	
Столбец 6	0,7430597	-0,19238	0,26573895	0,53934917	0,940766	1

Рис. 17. Побудована за допомогою інструменту "Корреляция" матриця коефіцієнтів кореляції



4. УСУНЕННЯ МУЛЬТИКОЛІНЕАРНОСТІ: ТЕОРЕТИЧНІ ВІДОМОСТІ

План лекції

- 1. Дефініція поняття "мультиколінеарність"*
- 2. Основні наслідки мультиколінеарності*
- 3. Методи дослідження мультиколінеарності*
- 4. Сутність та алгоритм методу Фаррара-Глобера*
- 5. Методи уникнення явища мультиколінеарності*
- 6. Метод головних компонентів*

Мультиколінеарність – це явище, що використовується для опису проблеми, коли нестрога лінійна залежність між пояснювальними змінними призводить до отримання ненадійних оцінок регресії. Проте така залежність, зовсім необов'язково дає незадовільні оцінки. Якщо всі інші умови задовільні, тобто якщо кількість спостережень і вибіркові дисперсії пояснювальних змінних великі, а дисперсія випадкового члену – мала, то в результаті можна отримати досить позитивні оцінки.

Мультиколінеарність виникає за рахунок отримання нестрогої залежності одній (або більше) незадовільних умов, це питання ступеня визначеності явища, а не його виду [9-12]. Оцінки регресії будуть не задовільні від неї у відповідному ступені, коли тільки всі незалежні змінні будуть абсолютно некорельовані. Розгляд цієї проблеми починається тільки тоді, коли вона суттєво впливає на результати оцінки регресії.

Таким чином, 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 ($r_{x_i x_j} \rightarrow 1, i \neq j$). 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 відхилятися від реальних.

Основні 3 наслідки мультиколінеарності.

1. Падає точність оцінювання, яка виявляється так:

- а) помилки деяких конкретних оцінок стають занадто великими;
- б) помилки досить корельовано одна з одною;
- в) дисперсії оцінок параметрів різко збільшуються.

2. Оцінки параметрів деяких змінних моделі можуть бути незначущими через наявність їх взаємозв'язку з іншими змінними, а не і тому, що вони не впливають на залежну змінну. У такому разі множина вибіркового даних не дає змоги цей вплив виявити.

3. Оцінки параметрів стають досить чутливими до обсягів сукупності спостережень. Збільшення сукупності спостережень іноді може і причинитися до істотних змін в оцінках параметрів. Тому в економетричних дослідженнях важливо з'ясувати, чи існують між факторними змінними взаємозв'язки і наскільки вони суттєві.

Існує багато методів дослідження мультиколінеарності. Основна частина цих методів полягає в дослідженні матриці коефіцієнтів парної кореляції:

$$r = \begin{pmatrix} r_{x_1x_1} & r_{x_1x_2} & r_{x_1x_3} & r_{x_1x_4} & \dots & r_{x_1x_m} \\ r_{x_2x_1} & r_{x_2x_2} & r_{x_2x_3} & r_{x_2x_4} & \dots & r_{x_2x_m} \\ \dots & \dots & \dots & \dots & \dots & \dots \\ r_{x_mx_1} & r_{x_mx_2} & r_{x_mx_3} & r_{x_mx_4} & \dots & r_{x_mx_m} \end{pmatrix}. \quad (13)$$

Проте, коли до моделі входять більш як дві пояснювальні змінні, вивчення питання про мультиколінеарність не може обмежуватись

інформацією, що її містять елементи кореляційної матриці. Явище мультиколінеарності в жодному разі не зводиться лише до існування парної кореляції між незалежними змінними.

Більш загальна перевірка мультиколінеарності передбачає знаходження визначника (детермінанта) кореляційної матриці r , який ще називається *детермінантом кореляції* і позначається $|r|$. Числові значення детермінанта кореляційної матриці задовольняють умову: $|r| \in [0,1]$.

Якщо $|r| = 0$, то існує повна мультиколінеарність, а коли $|r| = 1$, мультиколінеарність відсутня. Чим ближче $|r|$ до нуля, тим певніше можна стверджувати, що між пояснювальними змінними існує мультиколінеарність. Незважаючи на те, що на числове значення $|r|$ впливає дисперсія факторних змінних, цей показник можна вважати точковою мірою рівня мультиколінеарності.

Якщо в економетричній моделі знайдено мале значення параметра A_k при високому рівні *частинного коефіцієнта детермінації* R_j^2 і при цьому F -критерій істотно відрізняється від нуля, то це також свідчить про наявність мультиколінеарності.

Коли коефіцієнт частинної детермінації R_j^2 , який обчислено для регресійних залежностей між однією пояснювальною змінною та іншими, має значення, яке близьке до одиниці, то можна говорити про наявність мультиколінеарності.

Нехай при побудові економетричної моделі на основі покрокової регресії введення нової пояснювальної змінної істотно змінює оцінку параметрів моделі при незначному підвищенні (або зниженні) коефіцієнтів кореляції чи детермінації. Тоді ця змінна перебуває, очевидно, у лінійній залежності від інших, які було введено до моделі раніше.

Усі ці ознаки мультиколінеарності мають один спільний недолік: жодна з них чітко не розмежовує випадки, коли мультиколінеарність істотна і коли нею можна знехтувати.

Для дослідження наявності мультиколінеарності широко застосовується метод **Фаррара-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) кожної 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 Фаррара-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:

$$1) x_{ik}^* = \frac{x_{ik} - \bar{x}_k}{\sigma_{x_k}}, \quad 2) x_{ik}^* = \frac{x_{ik} - \bar{x}_k}{\sqrt{n\sigma_{x_k}^2}}, \quad (14)$$

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:

$$1) r = \frac{1}{n} X^{*T} X^*, \quad 2) r = X^{*T} X^*, \quad (15)$$

де X^* – матриця стандартизованих незалежних (пояснювальних) змінних;

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

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

$$\chi^2 = - \left[n - 1 - \frac{1}{6}(2m + 5) \right] \ln |r|, \quad (16)$$

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

Значення цього критерію порівнюється 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:

$$r^{-1} = \left(\frac{1}{n} X^{*T} X^* \right)^{-1}. \quad (17)$$

Крок 5. Обчислення F - критерію:

$$F_k = (z_{kk} - 1) \frac{n-m}{m-1}, \quad (18)$$

де z_{kk} – діагональні елементи матриці r^{-1} . Фактичні значення критерію 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:

$$R_{x_k}^2 = 1 - \frac{1}{z_{kk}}. \quad (19)$$

Крок 6. Знаходження частинних коефіцієнтів кореляції:

$$r_{kj,1,2,\dots,m} = - \frac{z_{kj}}{\sqrt{z_{kk} \cdot z_{jj}}}, \quad (20)$$

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 - критерію:

$$t_{kj} = \frac{r_{kj,1,2,\dots,m} \sqrt{n-m-1}}{1 - r_{kj,1,2,\dots,m}}. \quad (21)$$

Фактичні значення критерію 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.

Найпростіше позбутися мультиколінеарності в економетричній моделі можна, відкинувши одну зі змінних мультиколінеарної пари. Але на практиці вилучення якогось чинника часто суперечить логіці економічних зв'язків. Тоді можна перетворити певним чином незалежні змінні моделі:

- а) узяти відхилення від середньої;
- б) замість абсолютних значень узяти відносні;
- в) замінити фактор X_j^* , який має тісний зв'язок з X_i на фактор $X_j^* = X_i - X_j$, після чого перевірити наявність мультиколінеарності між факторами X_i і X_j^* ;
- г) стандартизувати пояснювальні змінні тощо.

За наявності мультиколінеарності змінних потрібно звертати увагу й на специфікацію моделі. Іноді заміна однієї функції іншою, якщо це не суперечить апріорній інформації, дає змогу уникнути явища мультиколінеарності.

Іноді просте збільшення спостережень у моделі пом'якшує проблему мультиколінеарності. Проте отримати додаткові дані не завжди легко, оскільки на практиці це вимагає значних витрат. Коли жодний з розглянутих способів не дає змоги позбутися мультиколінеарності, то для її виправлення використовують статистичні методи, зокрема: факторний аналіз, гребенева регресія та метод головних компонентів.

Подамо *метод головних компонентів* у вигляді покрокового алгоритму:

Крок 1. Нормалізація всіх пояснювальних змінних:

$$X_{ij}^* = \frac{X_{ij} - \bar{X}_j}{\sigma_{x_j}}, \quad i = 1, 2, \dots, n; \quad j = 1, 2, \dots, m. \quad (22)$$

Крок 2. Обчислення кореляційної матриці

$$r = X^{*T} X^*. \quad (23)$$

Крок 3. Знаходження характеристичних чисел матриці r з рівняння

$$|r - \lambda E| = 0, \quad (24)$$

де E – одинична матриця розміром $m \times m$.

Крок 4. Власні значення λ_k упорядковуються за абсолютним рівнем

вкладу кожної головної компоненти до загальної дисперсії.

Крок 5. Обчислення власних векторів s_k розв'язуванням системи рівнянь

$$(r - \lambda E)S = 0, \quad (25)$$

за таких умов:

$$s_j^T s_k = \begin{cases} 0, & k \neq j \\ 1, & k = j \end{cases}.$$

Крок 6. Знаходження головних компонентів – векторів

$$Z = XS, \quad (26)$$

де $S = (s_1, s_2, \dots, s_m)$ - матриця власних векторів.

Головні компоненти мають задовольняти умови:

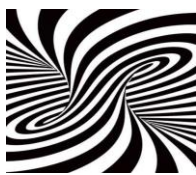
$$\sum_{i=1}^n Z_{ki} = 0, \quad \frac{1}{n} Z_k^T Z_k = \lambda_k, \quad k = 1, 2, \dots, m, \quad s_j^T s_k = 0, \quad k \neq j.$$

Крок 7. Визначення параметрів моделі з головними компонентами $Y = Z \cdot B$:

$$\hat{B} = Z^{-1} Y. \quad (27)$$

Крок 8. Знаходження параметрів вихідної моделі $\hat{Y} = X \cdot \hat{A}$:

$$\hat{A} = S \cdot \hat{B}. \quad (28)$$



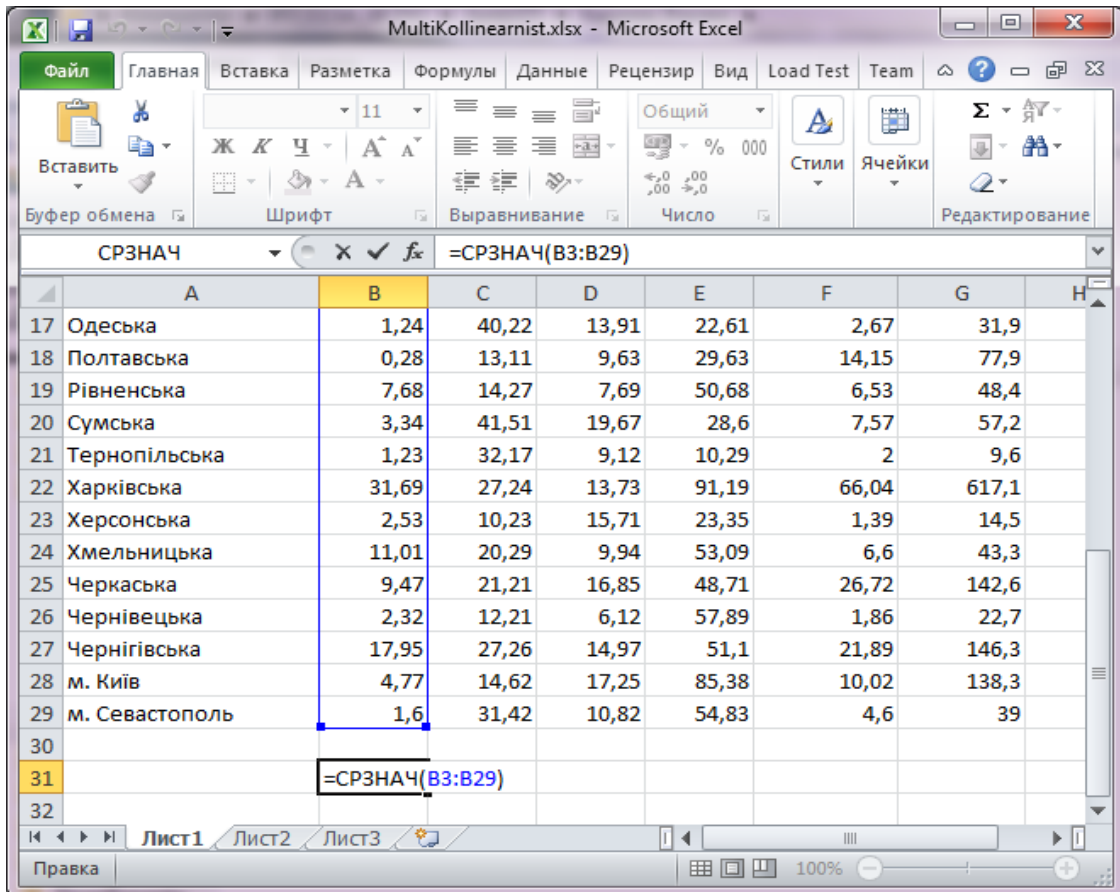
5. АЛГОРИТМ РОЗРАХУНКУ НАЯВНОСТІ МУЛЬТИКОЛІНЕАРНОСТІ В EXCEL

План лекції

- 1. Обчислення рядка середніх значень і дисперсії*
- 2. Формування нормалізованої матриці*
- 3. Розрахунок детермінанта кореляційної матриці*
- 4. Обчислення F-критерію (Фішера)*
- 5. Обчислення t-критеріїв Стьюдента*
- 6. Вибір показників для регресійної моделі*

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 В31 та ввести формулу «=СРЗНАЧ(В3:В29)», де В3:В29 – діапазон значень першого індикатора (рис. 18).



	A	B	C	D	E	F	G	H
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	
23	Херсонська	2,53	10,23	15,71	23,35	1,39	14,5	
24	Хмельницька	11,01	20,29	9,94	53,09	6,6	43,3	
25	Черкаська	9,47	21,21	16,85	48,71	26,72	142,6	
26	Чернівецька	2,32	12,21	6,12	57,89	1,86	22,7	
27	Чернігівська	17,95	27,26	14,97	51,1	21,89	146,3	
28	м. Київ	4,77	14,62	17,25	85,38	10,02	138,3	
29	м. Севастополь	1,6	31,42	10,82	54,83	4,6	39	
30								
31		=СРЗНАЧ(В3:В29)						
32								

Рис. 18. Скриншот розрахунку середнього значення індикатора
Виділивши комірку В31 і затиснувши ліву кнопку миші, перетягнути курсор вправо для вибору шести клітинок (відповідно до кількості індикаторів) і натиснути комбінацію клавіш «Ctrl»+«R» для копіювання формули (рис. 19).

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)».

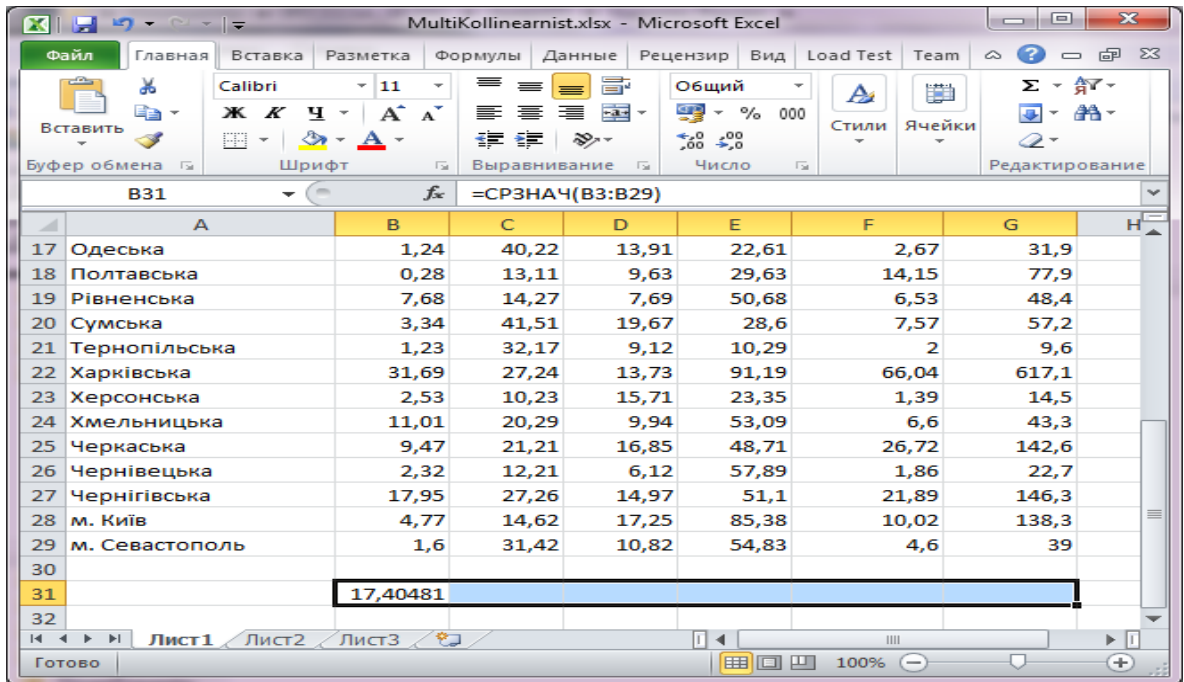


Рис. 19. Скриншот копіювання формули

Аналогічно копіюємо формулу в наступні п'ять правих комірок (рис. 20).

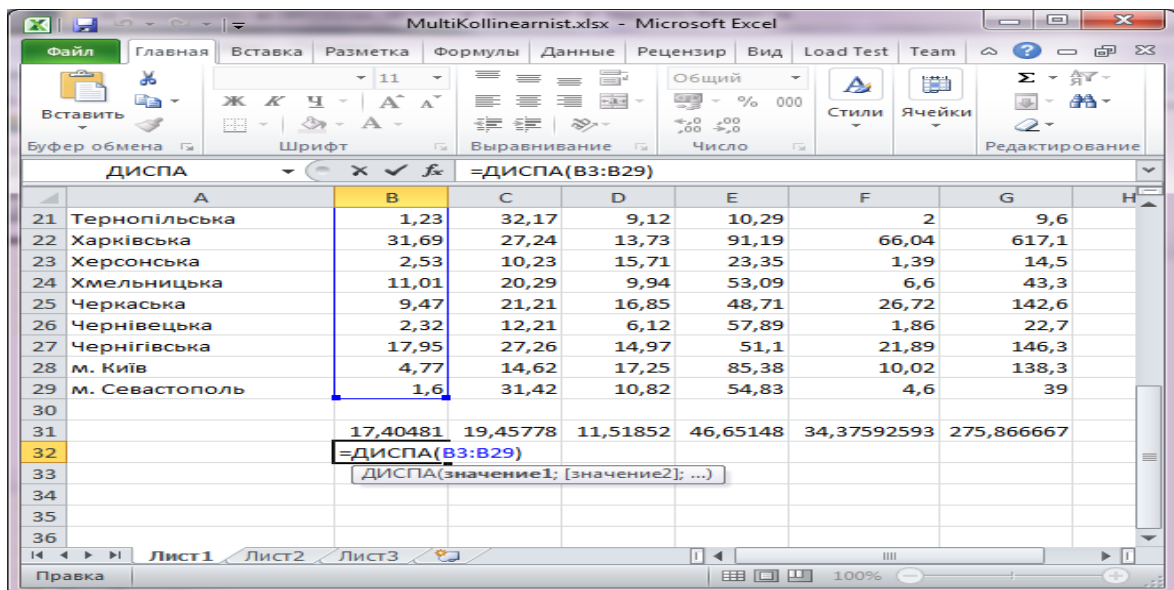


Рис. 20. Скриншот визначення дисперсії

Крок 3. Формування нормалізованої матриці (див. формулу (14)). 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 вибрати комірку I3 та задати формулу

« $=(B3-B\$31)/\text{КОРЕНЬ}(27*B\$32)$ », де B\$31 – зафіксоване середнє значення, B\$32 – зафіксована дисперсія, 27 – кількість об’єктів.

Крок 4. Розрахунок кореляційної матриці на основі нормалізованих даних (див. формулу (15)). 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 « $=\text{МУМНОЖ}(\text{ТРАНСП}(I3:N29);I3:N29)$ » (рис. 21 – а). Після цього необхідно одночасно натиснути комбінацію клавіш « $\text{Ctrl} + \text{Shift} + \text{Enter}$ » (див. рис. 21 – б).

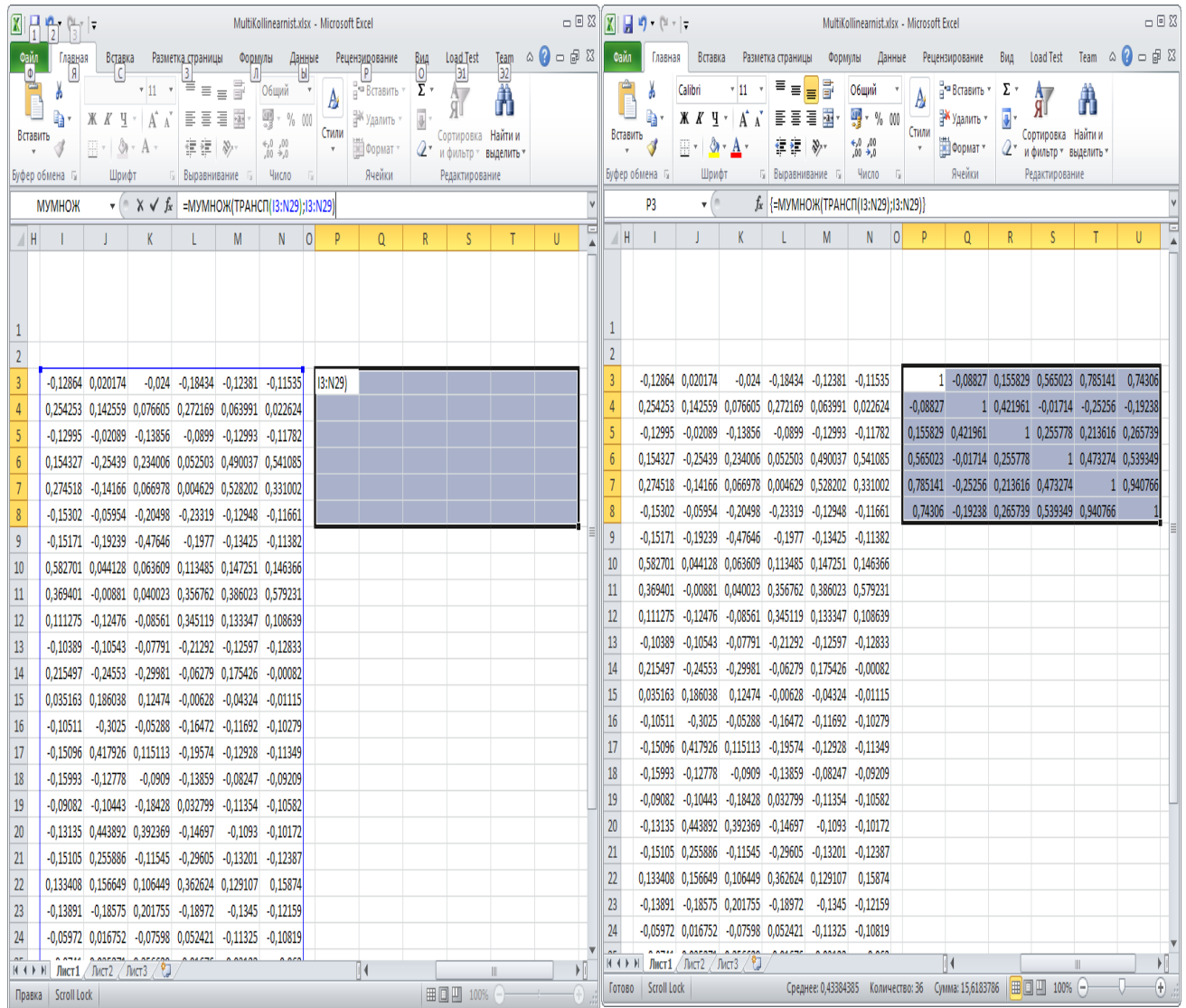


Рис.21. Скриншот зберігання значень коефіцієнтів кореляції

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. 22).

The screenshot shows an Excel spreadsheet with a correlation matrix in the range P3:U8. The formula bar shows the formula =МОПРЕД(P3:U8) entered in cell P10. The matrix data is as follows:

	P	Q	R	S	T	U
3	1	-0,08827	0,155829	0,565023	0,785141	0,74306
4	-0,08827	1	0,421961	-0,01714	-0,25256	-0,19238
5	0,155829	0,421961	1	0,255778	0,213616	0,265739
6	0,565023	-0,01714	0,255778	1	0,473274	0,539349
7	0,785141	-0,25256	0,213616	0,473274	1	0,940766
8	0,74306	-0,19238	0,265739	0,539349	0,940766	1

Cell P10 contains the formula =МОПРЕД(P3:U8) and the value 0,5. A tooltip for the function МОПРЕД(массив) is visible over the cell.

Рис. 22. Скриншот визначення детермінанта кореляційної матриці

Після натиснення клавіші «Enter» у комірці P10 з'явиться значення детермінанту кореляційної матриці.

Крок 6. Обчислення критерію χ^2 (див. формулу (16)).

Для розрахунку фактичного значення χ^2 - критерію у вибрану комірку, наприклад P12, вводять формулу « $=(27-1-(2*6+5)/6)*LN(P10)$ », де 27 – кількість об'єктів, 2*6 – подвоєна кількість факторів, P10 – значення детермінанту кореляційної матриці (рис. 23).

The screenshot shows an Excel spreadsheet with a correlation matrix in columns I through U and rows 3 through 24. The formula bar at the top displays the formula for cell P12: $=(27-1-(2*6+5)/6)*LN(P10)$. The value 0,015791 is shown in cell P12. The formula bar also shows the function name LN and the cell reference P10.

	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	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	$=(27-1-(2*6+5)/6)*LN(P10)$						
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							
15	0,035163	0,186038	0,12474	-0,00628	-0,04324	-0,01115							
16	-0,10511	-0,3025	-0,05288	-0,16472	-0,11692	-0,10279							
17	-0,15096	0,417926	0,115113	-0,19574	-0,12928	-0,11349							
18	-0,15993	-0,12778	-0,0909	-0,13859	-0,08247	-0,09209							
19	-0,09082	-0,10443	-0,18428	0,032799	-0,11354	-0,10582							
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							
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							

Рис. 23. Скриншот розрахунку фактичного χ^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 P12) з табличним значенням (комірка R12) χ^2 -критерію, приходимо до висновку, що між індикаторами існує мультиколінеарність.

Крок 7. Визначення оберненої матриці (див. формулу (17)). Для визначення оберненої матриці до відповідної матриці кореляції нормалізованого набору даних необхідно виділити квадратний діапазон комірок, наприклад, P14:U19. В рядку формул ввести вираз «=МОБР(P3:U8)» і одночасно натиснути комбінацію клавіш «Ctrl»+«Shift»+«Enter» (рис. 24).

The screenshot shows the Microsoft Excel interface with the following data in the spreadsheet:

	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. 24. Скриншот визначення оберненої матриці

Крок 8. Обчислення F-критерію (Фішера).

Фактичні значення F-критерію розраховується за формулою (18) і зберігається в масиві комірок P21:U21 (рис. 25).

The screenshot shows the Microsoft Excel interface with the following data in the worksheet:

	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	=(U19-1)*(27-6)/(6-1)		
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. 25. Скриншот 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. 25 наведено формулу для обчислення значення останнього F-критерію.

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

Крок 9. Розрахунок частинних коефіцієнтів кореляції (див. формулу (20)). 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 елемента; 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. 26).

	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	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								
26	-0,14088	-0,14589	-0,25986	0,091502	-0,13258	-0,11777		-1	0,231535	-0,18166	0,377493	0,464812	-0,10136	
27	0,005091	0,157052	0,166136	0,036219	-0,05091	-0,06027			-1	0,506765	-0,06725	=T15/КОРЕНЬ(\$Q\$15*T18)	-0,090687	
28	-0,118	-0,09738	0,275883	0,31532	-0,09931	-0,064				-1	0,193998	0,08933	0,090687	
29	-0,1476	0,240789	-0,03362	0,066588	-0,12141	-0,11019						-1	0,852851	
30														-1
31														

Рис. 26. Верхня права трикутна матриця частинних коефіцієнтів кореляції

Крок 10. Обчислення t -критеріїв Стьюдента (див. формулу (21)). 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. 27).

	L	M	N	O	P	Q	R	S	T	U	V	W	X
17	-0,19574	-0,12928	-0,11349		-0,88	0,107857	-0,25461	1,696587	1,149283	-1,25396			
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	-0,10136			
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										=U29*КОРЕНЬ(27-6)/КОРЕНЬ(1-U29^2)			
37													

Рис. 27. Скриншот розрахунку матриці t -критеріїв Стьюдента

Крок 11. Розрахунок табличного значення t -критерію Стьюдента. Табличне значення цього критерію в Excel обчислюється за допомогою статистичної формули **СТЮДЕНТ.ОБР.2X** (двосторонній обернений розподіл Стьюдента). В комірку Q38 введемо вираз «**=СТЮДЕНТ.ОБР.2X(0,05;27-6)**» (рис. 28).

	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. 28. Скриншот 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 парами 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. 29).

	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	-0,23448
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	-1,25396
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	-0,10136
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				

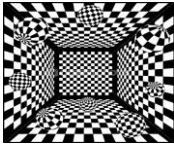
Рис. 29. Скриншот порівняння фактичних і табличного значень критеріїв Стюдента

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

- значення фактичного критерію Стюдента для пари регіональних індикаторів до якої входить показник F приймає найбільше значення, а саме – 7,48 (див. рис. 29 – комірka U36);

- показник F зустрічається найбільшу кількість разів у матриці критеріїв Стюдента (2 рази).

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



6. РЕГРЕСІЙНИЙ АНАЛІЗ: ТЕОРЕТИЧНІ ВІДОМОСТІ

План лекції

1. *Можливості математико-статистичного апарату аналізу еколого-економічної інформації*
2. *Основні методи регресійного аналізу*
3. *Сутність основних етапів регресійного аналізу*
4. *Основні завдання факторного аналізу*
5. *Види однофакторної нелінійної моделі*
6. *Види багатфакторних регресійних моделей*

Моделювання системи стану еколого-економічного розвитку територій України в умовах невизначеності інформаційного середовища та непередбачуваності зовнішнього впливу або невмотивованої поведінки внутрішніх структурних елементів керованих об'єктів тяжіє до використання стохастичних методів.

Стохастичні моделі реалізуються в межах простого динамічного та багатфакторного регресійного аналізу, аналізу за допомогою авторегресійних залежностей. Опис та умови застосування приведених стохастичних методів наведено в табл. 7.

Математико-статистичний апарат, що дозволяє встановити вид кореляційної залежності називається регресійним аналізом, а функція, що описує цю залежність, називається рівнянням регресії.

Основними методами регресійного аналізу є:

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

Регресійний аналіз проводиться за такими етапами [13]:

- встановлення виду кореляційної залежності результативної ознаки Y від факторної ознаки X ;
- побудова регресійної моделі;
- перевірка статистичної значущості побудованої моделі.

Таблиця 7

Опис методів стохастичного моделювання

Назва методу	Опис
Простий динамічний аналіз	Ґрунтується на передумові, що прогнозний показник Y змінюється прямо- (обернено-) пропорційно протягом певного періоду.
Багатофакторний регресійний аналіз	Базується на використанні більш ніж однієї незалежної змінної в рівнянні регресії. В результаті якісного аналізу виділяють множину X , які визначають, зміну прогнозного показника Y , і будують регресійну залежність.
Авторегресійні залежності	Специфіка процесів, які відбуваються у середовищі економіки, полягає в тому, що вони характеризуються, по-перше, взаємозалежністю і, по-друге, певною інерцією. Побудова авторегресійних залежностей направлена на визначення показника Y з урахуванням моменту часу t .

Перший етап регресійного аналізу є найважливішим, оскільки помилки у виборі виду залежності призводять до побудови регресійної моделі, що не відповідає емпіричним даним і не може використовуватися для прогнозування. Вибіркові дані для вивчення кореляційного зв'язку між ознаками X та Y зазвичай мають вигляд пар їх значень: $(x_1; y_1)$, $(x_2; y_2)$, ..., $(x_n; y_n)$, x_i – значення величини X , y_i – значення Y , n – кількість пар значень, $i = \overline{1, n}$. Якщо їх кількість достатньо велика, то для зручності розрахунків дані групуються і будується статистичний ряд, що містить значення X , відповідні середні значення Y та частоти (табл. 8).

Таблиця 8

Групування даних для вивчення кореляційного зв'язку

x_i	x_1	x_2	...	x_k
\bar{y}_{x_i}	\bar{y}_{x_1}	\bar{y}_{x_2}	...	\bar{y}_{x_k}
n_i	n_1	n_2	...	n_k

Згруповані дані табл. 8 зображуються графічно, що часто дозволяє визначити вид залежності Y від X .

Ламана лінія, що сполучає крапки з координатами $(X_i; \bar{Y}_{xi})$, називається *емпіричною лінією регресії*. Якщо емпірична лінія регресії значно наближається до прямої лінії, то висувається гіпотеза про наявність лінійного зв'язку між досліджуваними еколого-економічними ознаками (рис. 30).

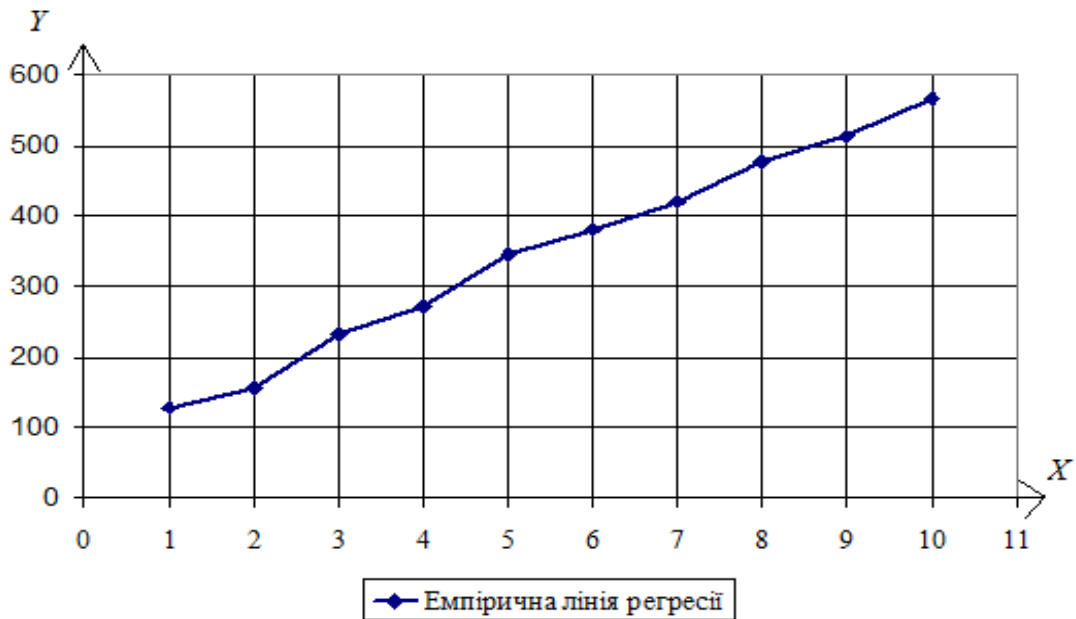


Рис. 30. Гіпотетична лінійна залежність

В іншому випадку висувається гіпотеза про наявність нелінійного зв'язку (рис. 31).

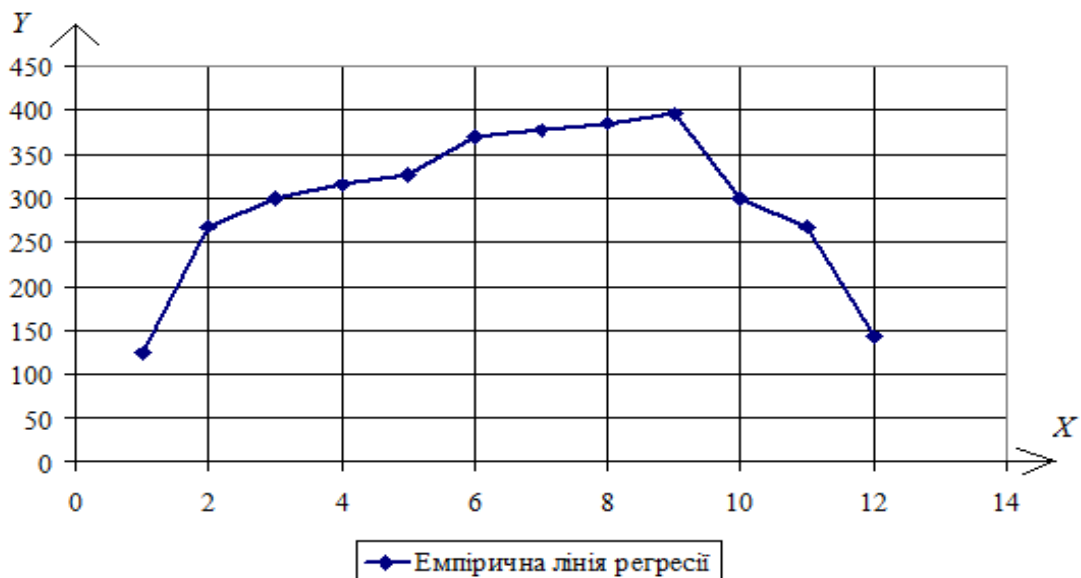


Рис. 31. Гіпотетична нелінійна залежність

Явище, яке залежить від, багатьох факторів, можна описати за допомогою множинної регресії. Дослідивши взаємозв'язок процесів у минулому і визначивши функціональний зв'язок між ними, можна з деякою вірогідністю планувати майбутнє. Найбільш досконалою і вивченою серед усіх багатовимірних регресійних моделей є лінійна. Лише деякі природні та економічні процеси можна моделювати за допомогою лінійної моделі. Її вибір залежить від процесу і часового проміжку його спостереження. Деякі процеси при малому проміжку їх спостереження можна з певним наближенням моделювати лінійною багатофакторною моделлю.

Основними завданнями факторного аналізу є:

- відбір факторів для аналізу певного результативного показника на основі принципу причинності та зв'язку;
- класифікація і систематизація факторів з метою забезпечення можливостей системного підходу, що включає їх розподіл за групами залежно від загальних ознак. Це дає змогу глибше розібратись у причинах змін явищ, точніше оцінити місце та роль кожного фактора у формуванні результативного показника. Системний підхід передбачає взаємопов'язане вивчення факторів з урахуванням їх зовнішніх і внутрішніх зв'язків, взаємодії та взаємопідпорядкованості;
- визначення форми залежності між факторами і результативним показником: функціональна вона чи стохастична, пряма чи обернена, прямолінійна чи криволінійна.

Регресійна залежність включає кілька етапів підбору рівняння та оцінювання тісноти кореляційного зв'язку:

- моделювання взаємозв'язків між факторами і результативним показником, що передбачає побудову конкретного математичного рівняння;
- розрахунок впливу факторів та оцінювання ролі кожного з них у змінах результативного показника, що включає використання різних способів;

– практичне застосування факторної моделі – останній етап регресійного аналізу, який полягає у визначенні резервів збільшення результативного показника, плануванні та прогнозуванні його значення при змінах ситуації.

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 еколого-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:

$$y = a_0 + a_1 z_1 + a_2 z_2 + \dots + a_m z_m,$$

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

– квадратична

$$y = a_0 + a_1x_1^2 + a_2x_2^2 + \dots + a_mx_m^2,$$

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 багатofакторних моделей регресійного аналізу:

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

Багатofакторні моделі *першого типу* мають вигляд

$$y = a_1F_1(x_1) + a_2F_2(x_2) + \dots + a_nF_n(x_n) + l, \quad (29)$$

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 багатofакторні нелінійні відносно параметрів моделі можна зводити до багатofакторних лінійних регресійних моделей. Прикладом таких багатofакторних моделей може бути модель

$$y = a_0x_1^{a_1} \cdot x_2^{a_2} \cdot \dots \cdot x_n^{a_n}l. \quad (30)$$

Такого вигляду регресіями 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:

$$\begin{aligned}
 y &= a_1 e^{a_2 x_1} e^{a_3 x_2} \dots e^{a_{n+1} x_n} l, \\
 y &= a_0 e^{a_1 x_1} \dots e^{a_n x_n} \dots x_{k+1}^{a_{k+1}} \dots x_{k+n}^{a_{k+n}} l, \\
 y &= a_0 a_1^{x_1} a_2^{x_2} \dots x_n^{a_n} e^{a_{n+1}} l.
 \end{aligned}
 \tag{31}$$

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.



7. ПОБУДОВА МНОЖИННОЇ ЛІНІЙНОЇ РЕГРЕСІЙНОЇ МОДЕЛІ В EXCEL

План лекції

1. *Способи побудови множинної лінійної регресійної моделі в Excel*
2. *Розрахунок транспонованої матриці в Excel*
3. *Розрахунок оберненої матриці*
4. *Розрахунок матриці коефіцієнтів рівняння регресії*
5. *Побудова регресійної моделі засобом "Регрессия" із надбудови "Пакет анализа"*

Побудувати множинну лінійну регресійну модель в Excel можна двома способами:

1. Засобами матричні функції МУМНОЖ та МОБР.
2. Засобом "Регрессия" із надбудови "Пакет анализа".

Перший спосіб. Використовуючи таблицю з вихідною інформацією (рис. 32) 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) фактора (див. рис. 32).

	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
27				

Рис. 32. 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 стовпців нової матриці). Для першої комірки J2 вводимо формулу «=МУМНОЖ(ТРАНСП(A2:D26);A2:D26)», де A2:D26 – діапазон значень матриці. Приклад розрахованої транспонованої матриці наведено на рис. 33.

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. 33. Скріншот розрахунку транспонованої матриці

Обернена матриця (рис. 34) 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. 34. 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. 35). Виділяємо першу комірку і вводимо значення «=МУМНОЖ(ТРАНСП(I3:N27);I3:N27)».

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.

	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. 35. Результат множення транспонованої матриці на матрицю залежної змінної

Результат побудови матриці коефіцієнтів рівняння регресії надано на рис. 36.

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. 36. Матриця коефіцієнтів рівняння регресії

Згідно даних рис. 36, рівняння регресії матиме вигляд:

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

Другий спосіб. Для побудови лінійної регресійної моделі необхідно:

1) Викликати засіб "Регресія" із надбудови "Пакет аналіза", натиснувши на кнопку "Анализ данных" на закладці "Данные" панелі інструментів Excel (рис. 37).

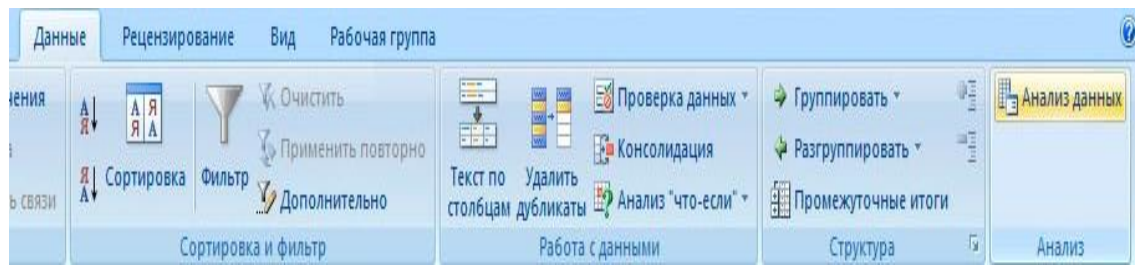


Рис. 37. Кнопка "Анализ данных" на панели інструментів

2) Викликати *Сервис – Анализ данных – Регрессия – ОК*. З'явиться вікно для надання вхідних даних (рис. 38). У вікні "Регрессия" вибрати 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 "ОК".

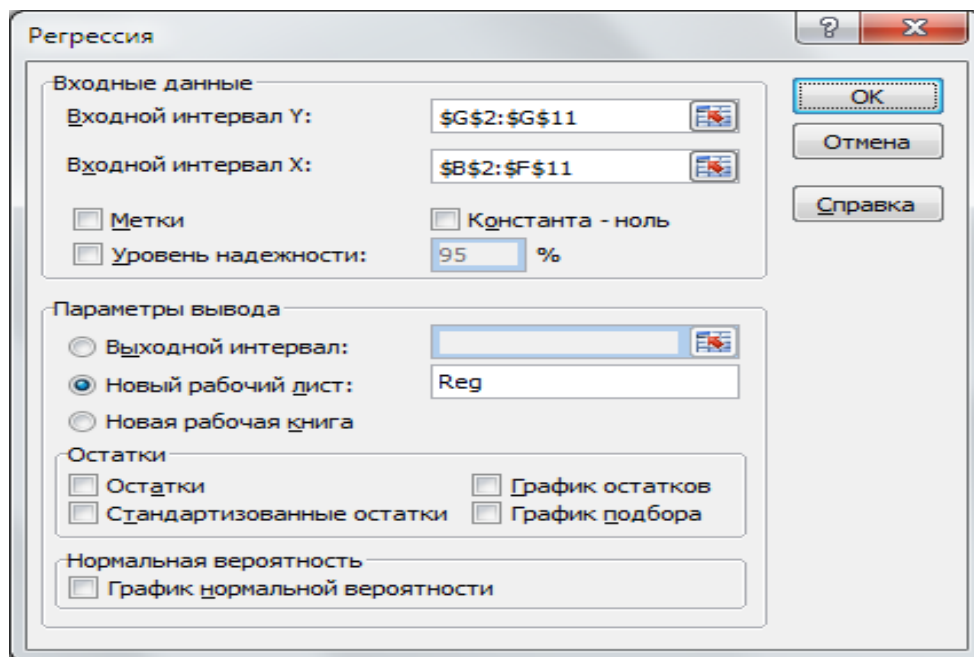


Рис. 38. Діалогове вікно функції "Регрессия"

Приклад і результати роботи функції "Регрессия" наведені на рис. 39.

Файл Правка Вид Вставка Формат Сервис Данные Окно Справка									
I23									
A	B	C	D	E	F	G	H	I	J
1	Регресійний аналіз								
2	Вхідні дані			Вихідні дані					
3	№	Значення		ВЫВОД ИТОГОВ					
4	i	X	Y						
5	1	1	328	Регрессионная статистика					
6	2	2	329	Множественный R	0,972633354				
7	3	3	329	R-квадрат	0,946015642				
8	4	4	345	Нормированный R-квадрат	0,937018249				
9	5	5	352	Стандартная ошибка	5,786032717				
10	6	6	370	Наблюдения	8				
11	7	7	377						
12	8	8	385	Дисперсионный анализ					
13					df	SS	MS	F	Значимость F
14	Регрессия				1	3520,005952	3520,005952	105,1433059	5,01935E-05
15	Остаток				6	200,8690476	33,4781746		
16	Итого				7	3720,875			
17									
18					Коэффициенты	Стандартная ошибка	t-статистика	P-Значение	
19	Y-пересечение				310,6785714	4,508440371	68,91043151	6,28282E-10	
20	Переменная X 1				9,154761905	0,892804231	10,25394099	5,01935E-05	

Рис. 39. Скріншот результатів регресійного аналізу

На рис. 39 у графі «Коэффициенты» 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*, $R^2 = 0,97$.

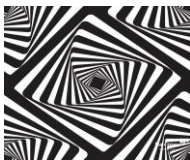
Крім того, може бути надано:

- графік підбору – порівняльна діаграма, що містить емпіричну і теоретичну лінії регресії;
- таблиця залишків – різниць емпіричних і теоретичних значень Y (рис. 40).

	A	B	C	D	E	F	G
1				Регресійний аналіз			
2	Вхідні дані			Вихідні дані			
3	№	Значення		ВЫВОД ОСТАТКА			
4	i	X	Y				
5	1	1	328				
6	2	2	329		Наблюдение	Предсказанное Y	Остатки
7	3	3	329		1	319,8333333	8,166666667
8	4	4	345		2	328,9880952	0,011904762
9	5	5	352		3	338,1428571	-9,142857143
10	6	6	370		4	347,297619	-2,297619048
11	7	7	377		5	356,452381	-4,452380952
12	8	8	385		6	365,6071429	4,392857143
13					7	374,7619048	2,238095238
14					8	383,9166667	1,083333333
15							
16							
17							
18							
19							
20							
21							
22							
23							
24							
25							
26							
27							
28							
29							

Графік підбору: Переменная X 1. Показує залежність Y від X з порівнянням фактичних даних (Y) та передбачених значень (Предсказанное Y). Ось X має значення від 0 до 10, а ось Y – від 0 до 500. Фактичні дані позначені синіми ромбами, а передбачені – рожевими квадратами.

Рис. 40. Додаткові результати регресійного аналізу



8. РОЗРАХУНОК ПРОГНОЗІВ ЗА ДОПОМОГОЮ НАДБУДОВ КОВЗКОГО СЕРЕДНЬОГО

План лекції

- 1. Передумови використання методу ковзкого середнього прогнозу*
- 2. Алгоритм складання прогнозів за допомогою надбудов ковзкого середнього*
- 3. Алгоритм складання прогнозів ковзкого середнього за допомогою діаграм*

Використання методу ковзкого середнього прогнозу будь-якого періоду є не що інше, як отримання середнього показника декількох результатів наглядів тимчасового ряду. 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. Іноді при складанні прогнозу ці методи більш ефективні, ніж методи, що засновані на довготривалих спостереженнях. 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.

Щоб визначити, скільки спостережень бажано включити в ковзке середнє, потрібно виходити з попереднього досвіду і наявної інформації про набір даних. Одне відхилення в наборі даних для, наприклад у нашому випадку, чотирьохкомпонентного середнього може змінити весь прогноз. Тобто, чим менше компонентів у прогнозі, тим менше ковзке середнє відгукується на зміну ситуації.

Наведемо приклади алгоритмів складання прогнозу в Excel на основі методу ковзкого середнього.

Приклад 1. Алгоритм складання прогнозів за допомогою надбудов ковзкого середнього.

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

Крок 2. У меню "Сервіс" ви знайдете нову команду "Аналіз даних". Виберіть команду "Сервіс – 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 клацніть на кнопці ОК. З'явиться діалогове вікно "Скользящее среднее".

Крок 4. У полі "Входной интервал" введіть дані про вашу базову лінію

або вкажіть діапазон у робочому листі, посилання на нього з'явиться в цьому полі.

Крок 5. У полі "Интервал" 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 скриншоті рис. 41 показаний прогноз викидів забруднюючих речовин в атмосферне повітря внаслідок діяльності модельного підприємства за допомогою надбудов ковзкого середнього та виведений його графік.

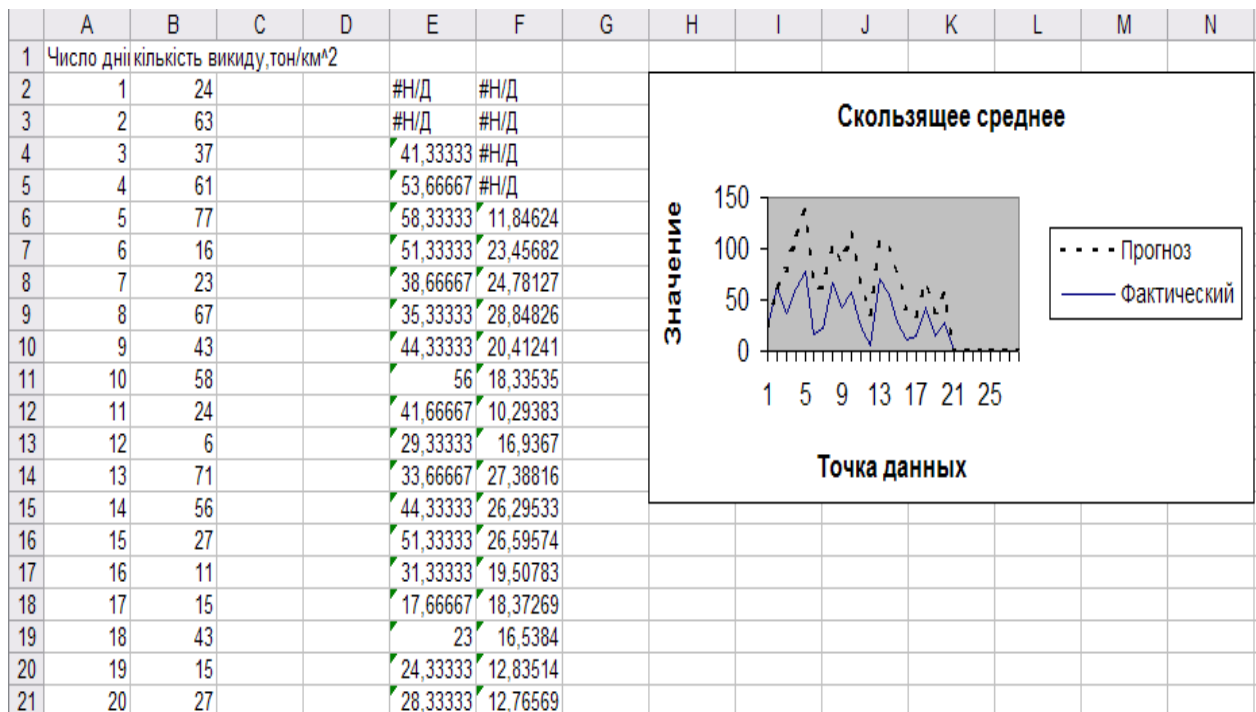


Рис. 41. Прогноз викидів забруднюючих речовин в атмосферне повітря за допомогою надбудов ковзкого середнього

На даному рисунку перші декілька показників ковзкого середнього відсутні внаслідок того, що надбудова "Скользящее среднее" повертає замість цих показників #Н/Д. 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.

Приклад 2. Алгоритм складання прогнозів ковзкого середнього за

допомогою діаграм наведено на рис. 42.

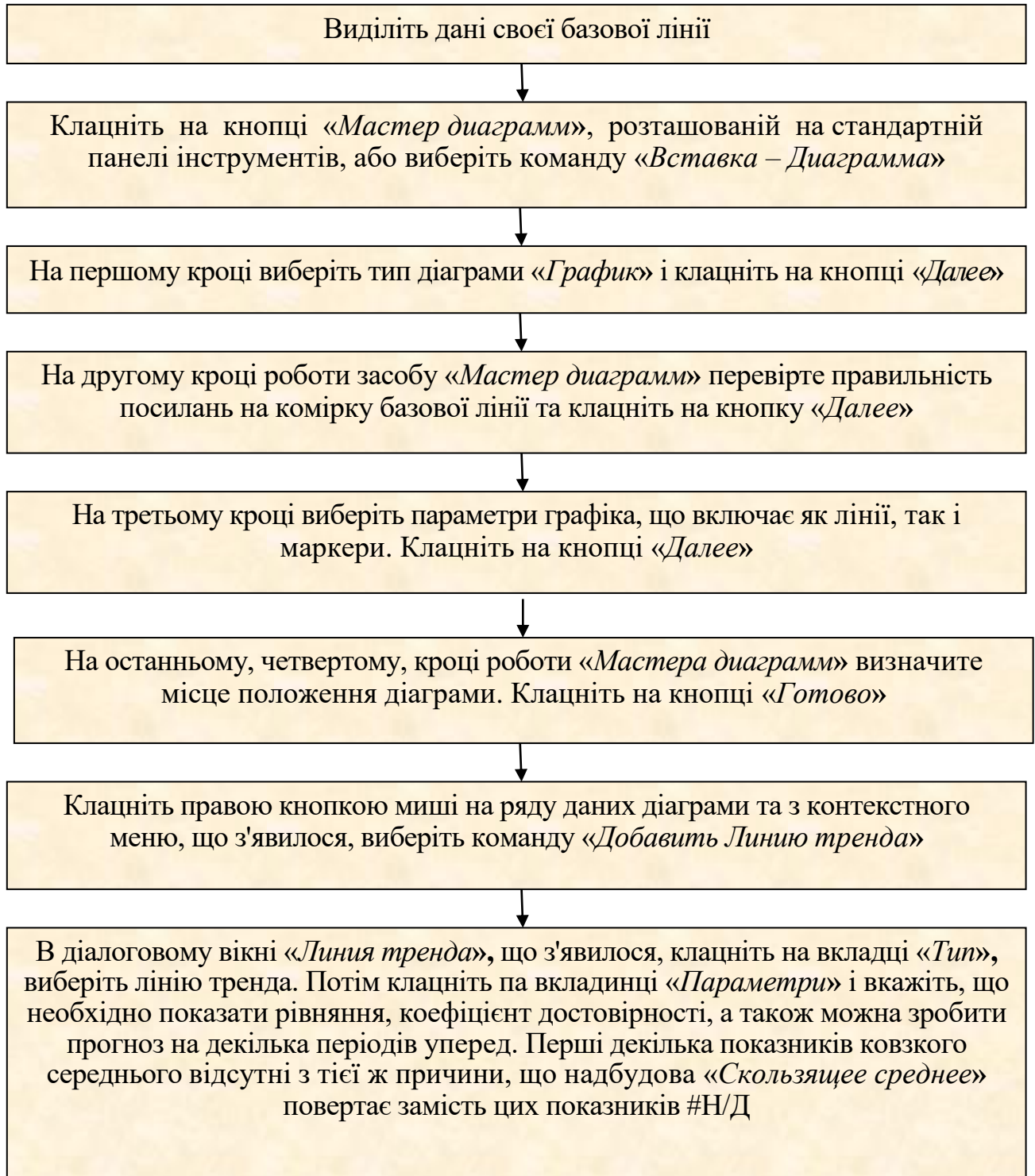


Рис. 42. Алгоритм складання прогнозів ковзкого середнього за допомогою діаграм

На скриншоті рис. 43 показаний прогноз викидів забруднюючих речовин в атмосферне повітря внаслідок діяльності модельного підприємства

за допомогою діаграм надбудови ковзкого середнього.

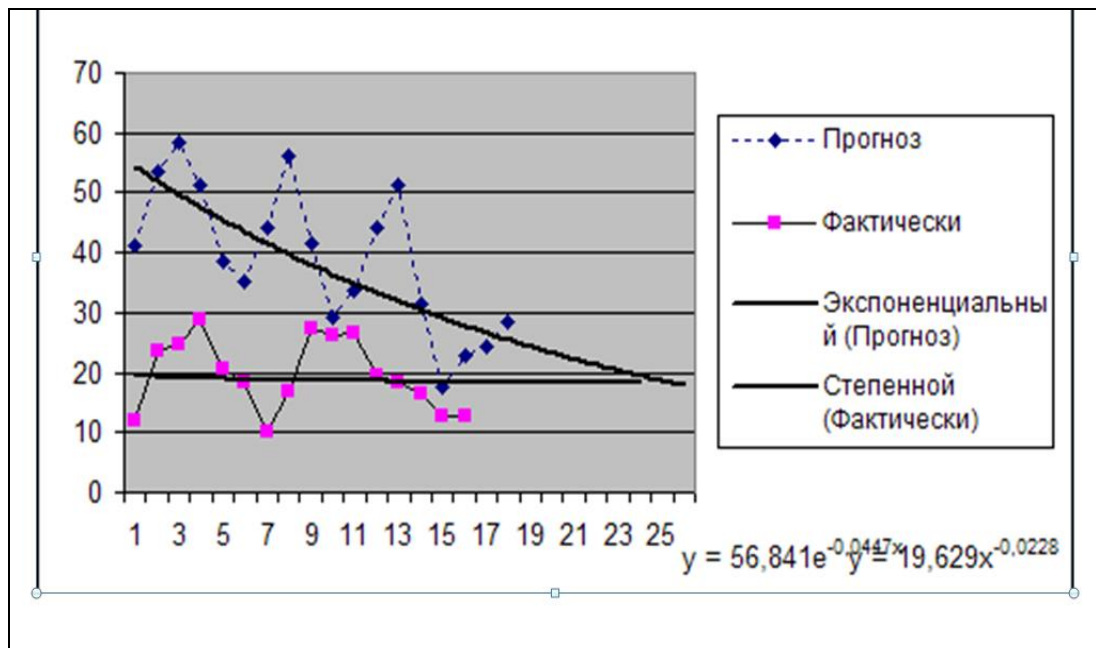


Рис. 43. Прогноз викидів забруднюючих речовин в атмосферне повітря за допомогою діаграм надбудови ковзкого середнього



9. РОЗРАХУНОК ПРОГНОЗІВ ЗА ДОПОМОГОЮ НАДБУДОВ ЕКСПОНЕНЦІЙНОГО ЗГЛАДЖУВАННЯ

План лекції

- 1. Експоненційне згладжування як метод середньострокового прогнозування*
- 2. Переваги методу експоненційного згладжування*
- 3. Константа згладжування*
- 4. Алгоритм складання прогнозів за допомогою надбудов експоненційного згладжування*

Методика, що використовується для подолання проблем, пов'язаних з короткостроковим прогнозуванням, носить назву "Експоненційне 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 прогнозової оцінки. Даний метод прогнозування вважається досить ефективним і відмінковим. *Основні переваги* методу полягають у можливості обліку ваг вихідної інформації, в простоті обчислювальних операцій, гнучкості опису різних динамік процесів. Метод експоненціального згладжування дає можливість отримати оцінку параметрів тренда, що характеризують не середній рівень процесу, а тенденцію, що склалася на момент останнього спостереження.

Для методу експоненційного згладжування основним моментом 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:

$$F_{t+1} = \alpha \times D_t + (1 - \alpha) F_t, \quad (32)$$

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 залежно 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.

Алгоритм обчислення прогнозів за допомогою надбудов

експоненційного згладжування наведено на рис. 44.

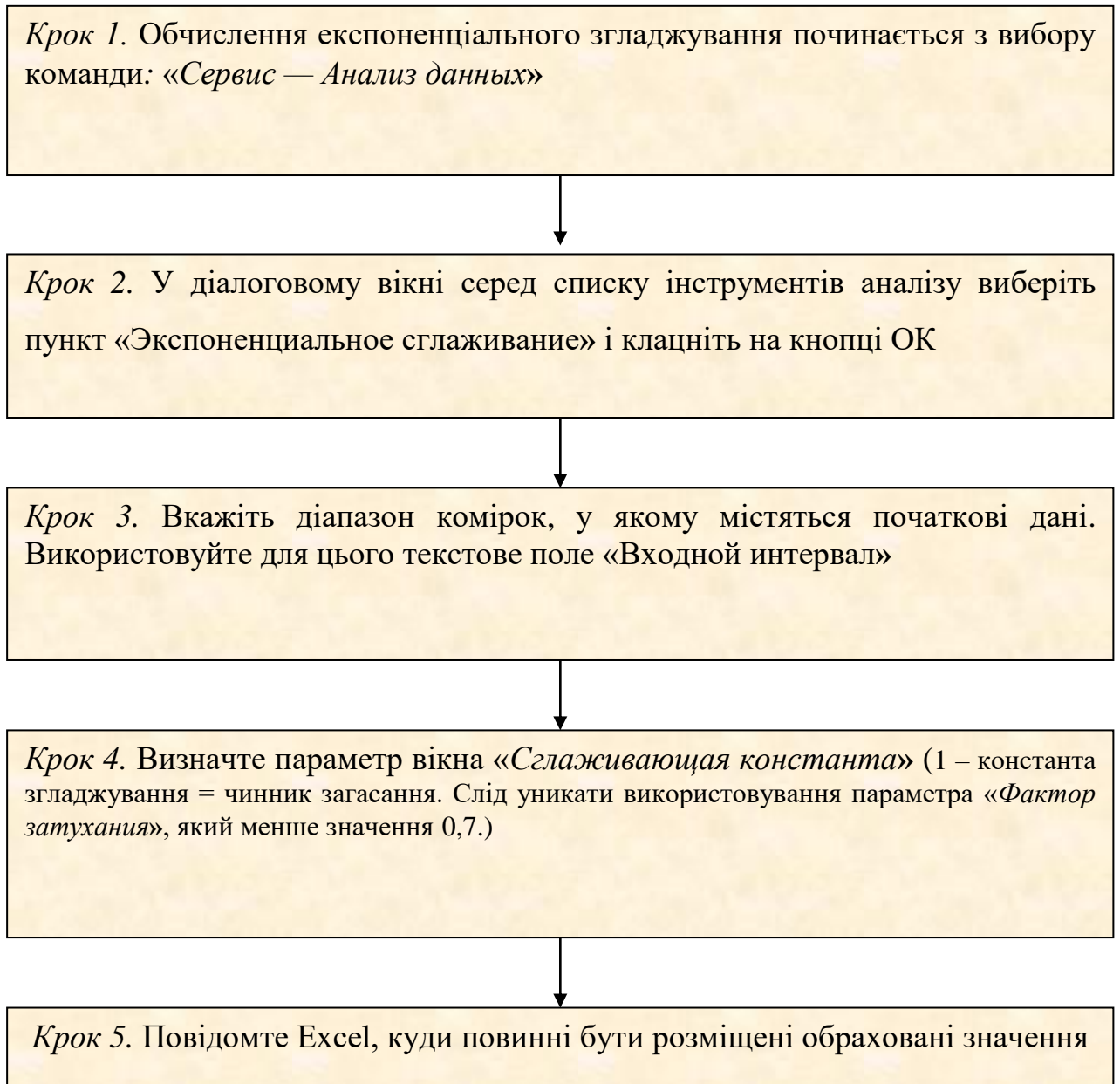


Рис. 44. Алгоритм складання прогнозів за допомогою надбудов експоненційного згладжування

На скриншоті рис. 45 наведено приклад прогнозу викидів забруднюючих речовин в атмосферне повітря внаслідок діяльності модельного підприємства за допомогою надбудови експоненційного згладжування.

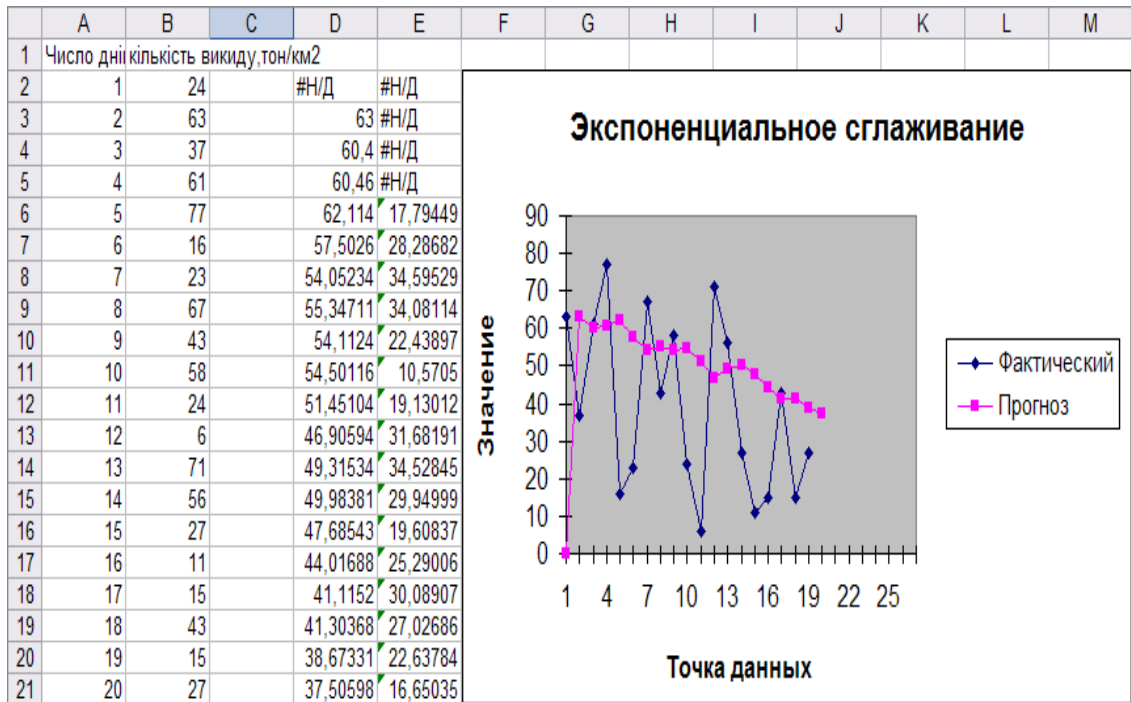


Рис. 45. Прогноз викидів забруднюючих речовин в атмосферне повітря за допомогою надбудови експоненційного згладжування

Побудована модель прогнозу має супроводжуватися додатковою інформацією стосовно її точності та адекватності.

Список використаної літератури:

1. Секторальні моделі прогнозування економіки України: научне издание / За ред. В. М. Гейця ; НАН України. Ін-т екон. прогнозування. – К. : Фенікс, 1999. – 304 с.
2. Присенко Г. В. Прогнозування соціально-економічних процесів : навч. посіб. / Присенко Г. В., Равікович Є. І. – К. : КНЕУ, 2005. – 378 с.
3. Басовский Л. Е. Прогнозирование и планирование в условиях рынка : Учебное пособие / Басовский Л. Е. – М. : ИНФРА-М, 2001. – 260 с.
4. Парсаданов Г. А. Прогнозирование национальной экономики (Теоретико-методологические аспекты) / Парсаданов Г.А. – М. : Высш. шк., 2002. – 222 с.
5. Прогнозування соціально-економічного розвитку регіону: питання теорії і методології / НАН України. Інститут регіональних досліджень / наук. ред. Є. І. Бойко. – Львів, 2005. – 234 с.
6. Сухоруков А. І. Моделювання та прогнозування соціально-економічного розвитку регіонів України: монографія / Сухоруков А. І., Харазішвілі Ю. М. –

- К. : НІСД, 2012. – 368 с.
7. Економіко-математичне моделювання : навч. посібник / За заг. ред. В. В. Вітлінського. – К. : КНЕУ, 2008. – 536 с.
 8. Экономико-экологическое прогнозирование (методология, методы, приложения) / Под ред. В. Н. Степанова – Одесса : ИПРЭЭИ НАН Украины, 2003. – 350 с.
 9. Магнус Я. Р. Эконометрика. Начальный курс: [учеб.] / Я. Р. Магнус, П. К. Катышев, А. А. Пересецкий. – М. : Дело, 2005.- 576 с.
 10. Лещинський О.Л. Економетрія / О.Л. Лещинський. – К. : МАУП, 2003. – 208 с.
 11. Кузьмичов А. І. Економетрія. Моделювання засобами MS Excel: [навчальний посібник] / А. І. Кузьмичов. – К. : ЦУЛ, 2011. – 214 с.
 12. Наконечний С. І. Економетрія. / С. І. Наконечний, Т. О. Терещенко. – К. : КНЕУ, 2006. – 528 с.
 13. Економіко-математичні моделі та методи проектного менеджменту: програма, тематичні плани і методичні рекомендації з навчально-методичного комплексу дисципліни для слухачів денної та заочної форми навчання спеціальності 8.000003 «Управління проектами» / Укладач І.А. Сенча. – Одеса: ОРІДУ НАДУ, 2009. – 101 с.