ЛЬВІВСЬКИЙ ДЕРЖАВНИЙ УНІВЕРСИТЕТ

БЕЗПЕКИ ЖИТТЄДІЯЛЬНОСТІ

О.І. Полотай
Р.Л. Ткачук
В.І. Ящук
Т.Б. Брич

МОДЕЛЮВАННЯ ТА ПРОГНОЗУВАННЯ В СОЦІАЛЬНІЙ СФЕРІ

Навчальний посібник

Львів - 2025

УДК 316.4:519.86

Рецензенти: **Олег ГАРАСИМЧУК**, кандидат технічних наук, доцент кафедри захисту інформації Національного університету “Львівська політехніка”

Євген МАРТИН, кандидат технічних наук, доцент, начальник кафедри інформаційних технологій та систем електронних комунікацій Львівського державного університету безпеки життєдіяльності.

Полотай О.І., Ткачук Р.Л., Ящук В.І., Брич Т.Б. *Моделювання та прогнозування в соціальній сфері: навчальний посібник* / Львівський державний університет безпеки життєдіяльності. Львів : ЛДУБЖД, 2025. 347 с.

Метою навчального посібника є формування у здобувачів освіти цілісного уявлення про інструментарій моделювання та прогнозування, ознайомлення з класичними й сучасними підходами, а також розвиток навичок застосування цих методів у реальних умовах соціальної практики та сфери кібезбезпеки і захисту інформації.

Матеріал подано в доступній формі, з використанням прикладів та рисунків. Лабораторні роботи, які наведено у посібнику, дадуть змогу здобувачам освіти закріпити теоретичні знання на практиці, оволодіти сучасними методами обробки соціальних даних, опанувати сучасні методи аналізу кіберзагроз і даних безпеки, навчитись будувати математичні моделі, здійснювати моделювання і прогнозування в соціальній сфері та кіберпросторі з використанням спеціалізованого програмного забезпечення.

Посібник стане корисним для здобувачів вищої освіти спеціальності F5 «Кібербезпека та захист інформації», оскільки дозволяє сформувати аналітичне мислення, навички системного підходу до аналізу соціальних процесів, а також засвоїти методи прогнозування, що є важливими для прийняття обґрунтованих рішень у сфері інформаційної безпеки.

Рекомендовано рішенням Вченої ради Львівського
державного університету безпеки життєдіяльності
(Протокол № 4 від 26.11.2025 р.)

© Полотай О.І, 2025
© Ткачук Р. Л., 2025
© Ящук В. І., 2025
© Брич Т.Б., 2025
© ЛДУ БЖД, 2025

ЗМІСТ

Вступ. 4
Тема 1. Методологічні основи прогнозування. 6
Тема 2. Основні методи прогнозування: класифікація прогнозів. 23
Тема 3. Основи кореляційно-регресійного аналізу. 56
Тема 4. Евристичні методи обґрунтування й прийняття рішень в прогнозуванні. 73
Тема 5. Непараметричні та параметричні методи прогнозування. 91
Тема 6. Теоретичні основи економічної динаміки. 115
Тема 7. Аналітичне вирівнювання тренда (криві зростання) як метод середньо-строкового прогнозування. 123
Тема 8. Методи екстраполяції тенденції та побудови прогнозу по одному часовому ряду. 134
Тема 9. Моделі прогнозування та соціального-економічного демографічного розвитку. 150
Тема 10. Метод згладжування і сезонне прогнозування. 161
Тема 11. Використання ковзного середнього при прогнозуванні. 178
Тема 12. Експертні методи прогнозування. 193
Тема 13. Критерії визначення якісного прогнозу. 214
Лабораторні роботи 223
Лабораторна робота 1. Приклад обробки реальних даних в системі Statistica 229
Лабораторна робота 2. Прогнозування на основі залежності між двома змінними. Кореляційний аналіз. 237
Лабораторна робота 3. Побудова прогнозів за допомогою функцій регресії Excel. Регресійний аналіз. 250
Лабораторна робота 4. Побудова прогнозів із застосуванням методів екстраполяції тенденції. 265
Лабораторна робота 5. Побудова регресійних моделей для прогнозування багатовимірних процесів. 268
Лабораторна робота 6. Аналіз часових рядів та побудова прогнозу за допомогою непараметричних методів. 275
Лабораторна робота 7. Експертні методи прогнозування. 285
Лабораторна робота 8. Розрахунок прогнозів із застосуванням методу ковзного середнього. 292
Лабораторна робота 9. Розроблення бази економетричних моделей для оцінювання та прогнозування рівня кібербезпеки в Україні. 300
Тестові завдання 313
Рекомендована література 343

Вступ

У сучасному суспільстві процеси цифровізації охопили практично всі сфери життєдіяльності людини – від економіки та управління до освіти й соціальної взаємодії. Зростання обсягів інформації, швидкість її поширення та глибина впливу на соціальні процеси зумовлюють потребу у використанні інструментів моделювання та прогнозування. Саме вони дозволяють не лише аналізувати сучасні тенденції, а й передбачати можливі сценарії розвитку подій, що є важливим для своєчасного прийняття управлінських рішень.

Особливе значення ці підходи мають у сфері кібербезпеки та захисту інформації. Кіберзагрози змінюються динамічно, стають дедалі складнішими та масштабнішими, а наслідки їх реалізації впливають не лише на окремі організації, але й на функціонування суспільства в цілому. Тому для фахівців з кібербезпеки критично важливо володіти методами побудови моделей ризиків, прогнозування ймовірності реалізації атак, аналізу соціальних факторів, що впливають на рівень інформаційної безпеки.

Дисципліна «Моделювання та прогнозування в соціальній сфері» спрямована на формування у здобувачів вищої освіти цілісного уявлення про закономірності розвитку соціальних процесів, а також практичних навичок застосування математичних, статистичних і комп’ютерних методів для аналізу та прогнозування. Для спеціальності «Кібербезпека та захист інформації» цей курс є особливо актуальним, оскільки дозволяє інтегрувати знання з соціальної аналітики та інженерії безпеки.

З огляду на все вищевказане, метою даного посібника «Моделювання та прогнозування в соціальній сфері» є засвоєння знань про етапи, методи та моделі прогнозування соціально-економічних процесів. Він присвячений висвітленню теоретичних засад і практичних підходів до моделювання та прогнозування в соціальній сфері. Він орієнтований на здобувачів вищої освіти, які вивчають відповідну дисципліну.

Посібник структуровано відповідно до основних напрямів сучасного прогнозування: від методологічних основ і класифікації прогнозів до аналізу часових рядів, експертних методів та моделей соціально-економічної динаміки. Значна увага приділяється практичним навичкам роботи з даними — від використання програмних засобів (Statsoft Statistica, MS Excel, SPSS Statistics) до створення економетричних моделей для оцінювання стану кібербезпеки в Україні. Такий підхід забезпечує поєднання фундаментальної підготовки з орієнтацією на реальні завдання професійної діяльності.

Особливу цінність становлять лабораторні роботи, які спрямовані на формування прикладних компетентностей здобувачів освіти:

– обробка реальних соціально-економічних даних;

– використання кореляційно-регресійного аналізу для виявлення залежностей;

– прогнозування тенденцій на основі часових рядів і методів згладжування;

– застосування експертних підходів у ситуаціях високої невизначеності;

– побудова економетричних моделей для оцінки та прогнозування рівня кібербезпеки.

Таким чином, вивчення дисципліни сприятиме підготовці фахівців нового покоління, здатних не лише реагувати на наявні виклики у сфері інформаційної безпеки, але й проактивно передбачати ризики, формувати ефективні стратегії захисту та робити внесок у зміцнення стійкості цифрового суспільства.

Оволодіння методами моделювання та прогнозування у соціальній сфері є не лише актуальним, але й необхідним для прийняття зважених управлінських рішень. Цей посібник покликаний стати надійним супутником у процесі професійного розвитку та глибшого розуміння складних соціальних явищ, сприяючи підвищенню ефективності роботи фахівців різних рівнів.

Тема 1. Методологічні основи прогнозування.

1.1. Поняття про економічну кібернетику та її методи.

1.2. Економічне прогнозування, його суть, предмет та об’єкт.

1.3. Моніторинг та альтернативи прогнозування.

1.4. Основні принципи і суть прогнозування.

1.1. Поняття про економічну кібернетику та її методи

Кібернетика (від грець. Κυβερνήτik – мистецтво управління, Κυβερνήτηαο – правлю кермом, управляю, Κυβερνήτης – керманич) – це наука, що вивчає загальні принципи управління та передачі інформації в різних системах, зокрема в комп'ютерних, біологічних і соціальних. Вперше термін «кібернетика» ввів американський учений Норберт Вінер, один із творців першої електронно-обчислювальної машини. В сучасному контексті кібернетика охоплює не лише теоретичні основи управління, але й принципи захисту інформації, що дозволяє мінімізувати ризики в процесах обробки та передачі даних.

Економічна кібернетика, як підгалузь цієї науки, зосереджується на застосуванні методів кібернетики для управління інформацією в економічних системах. Вона поєднує математику, кібернетику та економіку, і включає такі важливі аспекти, як математичне програмування, економетрика, дослідження операцій, а також моделювання економічних процесів. В умовах швидкого розвитку цифрових технологій, економічна кібернетика особливо орієнтована на питання безпеки інформації, стандартизації даних, а також їх надійної обробки та захисту під час передачі.

Одним із ключових аспектів економічної кібернетики є захист інформаційних потоків в економічних системах. Оскільки процеси управління і регулювання економічною діяльністю неможливі без обробки та передачі великої кількості чутливої інформації, питання її безпеки набувають особливої ваги. Методи економічної кібернетики дозволяють знизити вразливість

інформаційних каналів, раціоналізувати обробку та захист даних, а також розробити надійні технічні засоби для забезпечення конфіденційності та цілісності інформації.

Захист економічної інформації включає розробку ефективних механізмів для захисту від кіберзагроз, зокрема, від несанкціонованого доступу, втручання та модифікацій даних. Це дозволяє не лише підвищити рівень безпеки, але й забезпечити стабільність і надійність функціонування соціально-економічних систем в умовах швидко змінюваного цифрового середовища.

Основні напрямки розвитку економічної кібернетики включають:

Теорія економічних систем і моделей: Вивчає методи моделювання економічних процесів з урахуванням їх інформаційних і кібербезпекових аспектів, включаючи класифікацію економічних моделей і побудову комплексних інформаційних систем.

Теорія економічної інформації: Оцінює потоки інформації в економічних системах як ключовий елемент для забезпечення ефективного управління та захисту даних на всіх етапах – від збору і обробки до передачі і зберігання.

В умовах глобальних кіберзагроз, захист інформації стає критичним фактором для успішного функціонування сучасних економічних систем. Впровадження передових методів кібербезпеки дозволяє мінімізувати ризики, підвищуючи ефективність і стабільність економічного управління в цифрову епоху.

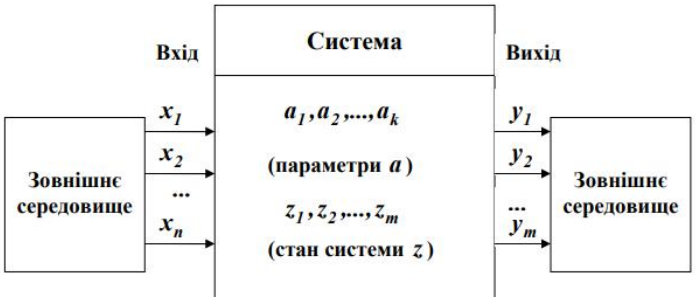


Рисунок 1.1 – Модель чорної скриньки

Соціально-економічна система – це складна інтегрована структура, що включає в себе взаємопов'язані соціальні та економічні інститути (суб'єкти) та

відносини, які регулюють процеси розподілу і споживання матеріальних і нематеріальних ресурсів. Вона охоплює всі етапи економічного циклу: виробництво, розподіл, обмін і споживання товарів і послуг, а також соціальні механізми, що визначають ці процеси.

Важливо, що в умовах цифрової трансформації та сучасних загроз кібербезпеки соціально-економічна система включає не лише традиційні економічні аспекти, а й інформаційні потоки, що з'єднують різні елементи цієї системи. Залучення інформаційних технологій, зокрема в сфері управління економікою, значно підвищує ефективність обробки даних і прийняття рішень, але й створює нові виклики у плані захисту цієї інформації від кіберзагроз.

Таким чином, соціально-економічна система стає не лише полем для аналізу економічних процесів, але й важливим об'єктом для забезпечення кібербезпеки, що включає захист інформації, стабільність функціонування її інститутів та збереження конфіденційності та цілісності даних, які циркулюють у системі.

1.2. Економічне прогнозування, його суть, предмет та об’єкт

Прогнозування є однією з ключових функцій управління поряд з аналізом, плануванням, організацією та мотивацією, а в умовах цифровізації воно набуває особливого значення для забезпечення кібербезпеки та захисту інформації. Активними користувачами прогнозних результатів виступають суб’єкти ринку, домашні господарства, державні й регіональні органи управління, які дедалі більше залежать від інформаційних систем і цифрових технологій, а отже – від рівня їх захищеності.

Один із засновників сучасного менеджменту Анрі Файоль слушно зауважив: «Управляти – означає передбачати, а передбачати – це вже майже діяти». У сфері кібербезпеки це означає своєчасне прогнозування кіберзагроз, уразливостей і можливих наслідків кібератак. Саме тому розробка прогнозів і побудова на їх основі планів захисту інформаційних ресурсів є необхідною

умовою стабільного функціонування підприємств і державних інституцій у динамічному цифровому середовищі.

Прогнозування в контексті кібербезпеки – це науково обґрунтоване передбачення ймовірного стану інформаційних систем, напрямів розвитку кіберзагроз і сценаріїв порушення безпеки в майбутньому. Воно базується на аналізі стійких взаємозв'язків між минулими інцидентами, поточним станом захисту та потенційними ризиками, що можуть виникнути надалі. Особливістю такого прогнозування є здатність виявляти приховані або ще не проявлені загрози, які неможливо безпосередньо спостерігати на поточний момент.

Прогнозування (від грец. prognosis – знання наперед) у сфері кібербезпеки є процесом формування прогнозів на основі аналізу тенденцій розвитку інформаційних технологій, методів атак і засобів захисту. Воно має дати відповіді на два принципові питання:

а) які кіберзагрози з найбільшою ймовірністю можуть виникнути в майбутньому;

б) якими способами слід впливати на умови функціонування інформаційних систем, щоб мінімізувати ризики та забезпечити заданий рівень захисту.

Застосування прогнозування дозволяє виявляти сталі закономірності або суттєві зміни в розвитку кіберзагроз, оцінювати ймовірність їх реалізації, аналізувати альтернативні сценарії кібератак і накопичувати наукову та практичну інформацію для обґрунтованого вибору стратегії кіберзахисту та управлінських рішень у сфері інформаційної безпеки.

Процес прогнозування є особливим видом наукового дослідження, спрямованого на вивчення перспектив розвитку загроз і засобів захисту в кіберпросторі. Він тісно пов'язаний із поняттям передбачення, яке полягає у випереджальному відображенні можливих подій на основі знання закономірностей розвитку цифрових технологій, суспільства та інформаційних процесів.

Важливим етапом кіберпрогнозування є ретроспекція – аналіз минулих кіберінцидентів, атак і збоїв у роботі інформаційних систем. Вона дозволяє впорядкувати накопичений досвід, виявити слабкі місця в системах захисту та сформувати основу для подальшого прогнозного аналізу.

Соціально-економічне прогнозування з урахуванням кібербезпеки передбачає розробку прогнозів розвитку економіки та суспільства з позицій інформаційних ризиків і цифрових загроз. Воно спирається на наукові методи пізнання соціально-економічних процесів і використовує інструменти економічної прогностики для оцінювання впливу кіберзагроз на економічну стабільність і суспільний розвиток.

Об’єктами такого прогнозування є економічні, соціальні та науково-технічні процеси на рівні держави, галузей і регіонів, що функціонують в умовах активного використання інформаційних систем і потребують надійного кіберзахисту.

Залежно від рівня конкретизації та характеру впливу на перебіг досліджуваних процесів у сфері кібербезпеки розрізняють три основні форми передбачення: гіпотезу, прогноз і план (програму). Вони відрізняються глибиною опрацювання, ступенем визначеності та практичною спрямованістю щодо забезпечення захисту інформації.

Гіпотеза є формою наукового передбачення на теоретичному рівні й у контексті кібербезпеки відображає загальні закономірності розвитку кіберзагроз, інформаційних технологій і засобів захисту. Вона ґрунтується на існуючих теоріях, виявлених тенденціях, причинно-наслідкових зв’язках між цифровізацією, уразливостями систем та активністю зловмисників. На рівні гіпотези формується узагальнене, переважно якісне уявлення про можливі напрями еволюції загроз і поведінку об’єктів кіберзахисту в майбутньому.

Прогноз у сфері кібербезпеки є науково обґрунтованим припущенням щодо ймовірного стану інформаційних систем, рівня їх захищеності або можливих сценаріїв реалізації кіберзагроз у визначеному майбутньому. Він

може також описувати альтернативні шляхи й строки досягнення необхідного рівня інформаційної безпеки. Прогнозування виступає ключовою ланкою між теоретичними уявленнями та практичними рішеннями, оскільки дозволяє обрати найбільш реалістичні й економічно доцільні варіанти кіберзахисту.

На відміну від гіпотези, прогноз має вищий рівень визначеності, адже базується не лише на якісних оцінках, а й на кількісних показниках: статистиці кіберінцидентів, імовірності атак, рівні ризиків та ефективності захисних механізмів. Водночас прогнози у сфері кібербезпеки мають імовірнісний характер і допускають кілька можливих сценаріїв розвитку подій, що відповідає динамічній природі кіберзагроз.

План є офіційним управлінським документом, у якому в контексті кібербезпеки визначаються конкретні цілі захисту інформації, терміни їх досягнення та засоби реалізації. Він передбачає постановку чітких завдань щодо впровадження технічних, організаційних і правових заходів безпеки, розподіл ресурсів і відповідальності, а також обґрунтування управлінських рішень. План має найвищий рівень конкретності та чіткості, хоча й спирається на результати прогнозування та досягнення прикладної теорії.

Програма у сфері кібербезпеки є рішенням щодо комплексу заходів, спрямованих на розв’язання науково-технічних, соціально-економічних або організаційних проблем захисту інформації. Вона може мати характер передпланового документа або деталізувати окремі напрями плану, наприклад розвиток систем кіберзахисту, підготовку фахівців чи реагування на інциденти.

Прогноз може існувати автономно, як інструмент аналізу майбутніх кіберризиків. Водночас план без прогнозу втрачає свою якісну й кількісну обґрунтованість, оскільки не враховує реальні тенденції розвитку загроз. Саме прогноз слугує основою для розроблення планів кібербезпеки, з яких обирається найбільш раціональний варіант. У плані вже закріплюється й реалізується конкретне управлінське рішення щодо забезпечення інформаційної безпеки, прийняте на підставі результатів прогнозування.

1.3. Моніторинг та альтернативи прогнозування

Соціально-економічний моніторинг у сучасних умовах цифровізації та зростання кіберзагроз постає як організований і систематичний процес спостереження не лише за динамікою економічних показників, а й за станом інформаційних процесів, що забезпечують функціонування економіки. Він охоплює аналіз змін у економічному середовищі з урахуванням надійності, цілісності та безпеки даних, які використовуються для управління та прийняття рішень.

Сьогодні здійснення соціально-економічного моніторингу є необхідним на державному, галузевому та регіональному рівнях, а також безпосередньо на рівні окремого підприємства, де економічна діяльність тісно пов’язана з використанням інформаційних систем і цифрових платформ. У цьому контексті моніторинг включає контроль ключових економічних показників разом із відстеженням ризиків порушення інформаційної безпеки, кіберінцидентів та уразливостей, що можуть впливати на стабільність господарської діяльності.

Основною метою такого моніторингу є збирання, фіксація та захист достовірних даних, які характеризують стан ринку, економічну ситуацію та ефективність функціонування суб’єктів господарювання. Забезпечення кібербезпеки на цьому етапі є критично важливим, оскільки спотворення, втрата або несанкціонований доступ до інформації можуть призвести до помилкових управлінських рішень.

Виконуючи роль підготовчого етапу перед оцінюванням і прогнозуванням, соціально-економічний моніторинг формує надійну інформаційну базу для подальших аналітичних і прогностичних досліджень. На основі захищених і перевірених даних шляхом порівняння альтернативних сценаріїв розвитку, з урахуванням можливих кіберризиків, приймаються обґрунтовані управлінські рішення, спрямовані на забезпечення економічної стабільності та інформаційної безпеки.

У сфері кібербезпеки відмова від системного прогнозування або використання спрощених підходів до передбачення майбутніх подій може мати критичні наслідки. Існує низка альтернативних підходів до прогнозування, які на практиці часто виявляються неефективними або навіть небезпечними для стійкості інформаційних систем і економічної діяльності загалом.

Принцип «може трапитися будь-що»

Цей підхід ґрунтується на уявленні про повну непередбачуваність подій у кіберпросторі та фатальність кіберзагроз. Вважається, що жодні превентивні заходи не здатні суттєво вплинути на рівень кіберризиків, а отже прогнозування атак, інцидентів і вразливостей є марним. Підприємства, які дотримуються такої позиції, фактично відмовляються від проактивного управління кібербезпекою, що робить їх надзвичайно вразливими до кібератак і інформаційних втрат.

Принцип «славного минулого»

Цей консервативний підхід полягає в переконанні, що методи кіберзахисту, які раніше забезпечували надійність інформаційних систем, залишатимуться ефективними і в майбутньому. Орієнтація виключно на перевірені часом технології та практики без урахування еволюції кіберзагроз і цифрових технологій призводить до застарілості систем захисту. Ігнорування динамічного розвитку кіберсередовища рано чи пізно створює критичні прогалини в безпеці.

Принцип «прогнозування крізь шори»

Цей підхід проявляється у прагненні до постійного нарощування існуючих засобів кіберзахисту за принципом «більше, швидше, потужніше» без переосмислення самих підходів до безпеки. Хоча він виглядає прогресивнішим, ніж повна відмова від прогнозування, його суттєвим недоліком є обмеженість бачення. Такий підхід не враховує появу нових векторів атак, альтернативних технологій захисту або необхідність зміни архітектури інформаційних систем, що може поставити організацію у вразливе становище.

Принцип «рішучих дій»

У межах цього підходу прогнозування кіберзагроз вважається зайвим, а рішення у сфері кібербезпеки ухвалюються лише після виникнення інцидентів. Управління безпекою має реактивний характер, без чітко сформованої стратегії та довгострокового бачення. Заходи вживаються в умовах кризи, коли бракує часу для глибокого аналізу й якісної реалізації рішень. Такий підхід не забезпечує ані стабільного рівня кіберзахисту, ані здатності ефективно адаптуватися до нових загроз.

Принцип прогнозування «за допомогою генія»

Цей підхід не заперечує необхідності прогнозування, однак зводить його до інтуїтивних оцінок окремого експерта або «видатної особистості» у сфері кібербезпеки. При цьому ігноруються системні, науково обґрунтовані методи аналізу даних, моделювання загроз і оцінювання ризиків. Хоча експертний досвід є важливим, відмова від формалізованих методів прогнозування знижує точність і надійність рішень, що особливо небезпечно в умовах складних і швидкозмінних кіберзагроз.

Отже, в умовах зростаючої залежності економіки та суспільства від цифрових технологій ефективне прогнозування кіберзагроз має базуватися на поєднанні наукових методів, аналітики даних і експертних оцінок, а не на спрощених або альтернативних підходах, які знижують рівень інформаційної безпеки.

1.4. Основні поняття, цілі та завдання прогнозування соціально-економічних процесів. Етапи та принципи прогнозування

Головним завданням функціонування сучасної держави є забезпечення сталого розвитку національної економіки в поєднанні з надійним соціальним захистом населення, що в умовах цифровізації неможливо без належного рівня кібербезпеки та захисту інформаційних ресурсів. Управлінські рішення, які ухвалюються органами державної влади, дедалі частіше спираються на цифрові

дані та інформаційні системи, а тому можуть призводити не лише до економічних, а й до кіберризиків і небажаних інформаційних наслідків. У зв'язку з цим важливим завданням державного управління стає завчасне прогнозування можливих соціально-економічних і кібернетичних наслідків прийнятих рішень та вибір найбільш обґрунтованих шляхів їх нейтралізації. Саме тому роль науково обґрунтованих прогнозів, зокрема з урахуванням кіберзагроз, суттєво зростає.

Потреба у прогнозуванні зумовлюється двома ключовими чинниками. По-перше, невизначеністю майбутнього, яка в цифровому середовищі посилюється швидким розвитком інформаційних технологій і еволюцією кіберзагроз. По-друге, тим, що наслідки управлінських рішень, у тому числі у сфері цифрової трансформації та кібербезпеки, часто проявляються не одразу, а через певний проміжок часу, коли виправити помилки значно складніше й дорожче.

Закономірності прогнозування почали активно досліджувати лише в останні десятиліття, що пояснюється зростанням складності соціально-економічних систем і їх залежністю від інформаційних потоків. У минулі періоди практична економіка мала обмежену потребу в прогнозах, здатних адекватно враховувати реакцію об'єкта управління на управлінські та технологічні впливи. Проте безперервний розвиток суспільства, цифровізація економіки та зростання ролі кіберпростору істотно ускладнили управлінські процеси. В умовах ринкової економіки прогнозування стало одним із ключових наукових інструментів формування стратегії та тактики розвитку держави з урахуванням економічних і кібербезпекових чинників.

Сьогодні прогнозування соціально-економічних процесів, зокрема з позицій інформаційної та кібербезпеки, сформувалося в окремий науковий напрям із власною методологією, методами аналізу та інструментами оцінювання ризиків. У розвинених країнах функціонують спеціалізовані науково-дослідні установи, які займаються прогнозуванням суспільного

розвитку, економічної динаміки та потенційних кіберзагроз, що дозволяє підвищити ефективність державного управління й стійкість національних соціально-економічних систем.

Базовими поняттями прогнозування є такі:

Прогнозування соціально-економічних процесів (СЕП) – це наукова дисципліна, що досліджує методи та засоби розроблення прогнозів майбутнього розвитку національної економіки та соціальної сфери. Вона ґрунтується на пізнанні закономірностей функціонування соціально-економічних явищ та застосуванні комплексу наукових методів і підходів.

Прогноз – це науково обґрунтоване судження щодо можливих майбутніх станів об’єкта, шляхів їх досягнення та термінів реалізації. Хоча прогноз має імовірнісний характер, його достовірність підвищується завдяки науковим аргументам і аналізу об’єктивних тенденцій розвитку. Самі розробники розглядають прогноз як найбільш очікуваний та вірогідний варіант майбутнього стану об’єкта.

Прогнозування – процес підготовки прогнозів, який, подібно до будь-якої форми трудової чи творчої діяльності, визначається суб’єктом та об’єктом дослідження, вибраними методами й інструментами, а також впливом зовнішнього середовища.

Варіант прогнозу – один з прогнозів, що становлять групу можливих прогнозів об'єкта прогнозування.

Метод прогнозування – спосіб дослідження об'єкта прогнозування, спрямований на розробку прогнозу.

Методика прогнозування – сукупність методів і правил розробки прогнозів конкретних об'єктів.

Система прогнозування – система методів прогнозування і засобів їх реалізації, що функціонує відповідно до основних принципів прогнозування.

Споживач прогнозу – організація, підприємство, установа або окрема особа, що використовує результати прогнозів, а також у ряді випадків формулює завдання на прогноз.

Суб'єктами прогнозування соціально-економічного розвитку в умовах цифровізації та зростання ролі кібербезпеки є органи державної влади й місцевого самоврядування, корпорації та підприємства, а також науково-дослідні й консалтингові установи та окремі експерти, яких залучають до розроблення й упровадження прогнозів. Усі ці суб'єкти у своїй діяльності активно використовують інформаційні системи та цифрові дані, тому якість прогнозування безпосередньо залежить від рівня захищеності інформації, достовірності даних і стійкості інформаційної інфраструктури до кіберзагроз.

Об'єктом соціально-економічного прогнозування виступають соціально-економічні процеси (СЕП), тобто сукупність економічних і соціальних явищ, що забезпечують формування та функціонування соціально-економічної системи та відображають динаміку зміни її параметрів на різних рівнях господарювання. У сучасних умовах ці процеси нерозривно пов'язані з інформаційними потоками та цифровими технологіями, що робить їх чутливими до кіберризиків. Завдання суб'єкта прогнозування полягає в тому, щоб за допомогою спеціальних методів і аналітичного інструментарію зібрати, захистити та проаналізувати інформацію про об'єкт або систему, а згодом трансформувати її у науково обґрунтовані уявлення про майбутній розвиток соціально-економічних і інформаційних процесів.

Економічні (природні) процеси являють собою взаємодію людини з природою, що реалізується через використання засобів праці з метою створення матеріальних продуктів виробничої діяльності, а також нематеріальних результатів — знань, інформаційних ресурсів та інноваційних продуктів. У цифровій економіці ці процеси значною мірою залежать від інформаційних технологій, автоматизованих систем управління та рівня їх кіберзахисту.

Соціальні (суспільні) процеси відображають взаємовідносини між людьми у сферах виробництва, розподілу, обміну та споживання створених благ. Вони формують простір соціальної економіки, що включає соціальні технології, цифрові платформи, а також пов’язані з ними політичні, управлінські й організаційні процеси, які потребують захисту від інформаційних впливів і маніпуляцій.

Взаємодія різних видів процесів зумовлює виникнення комбінованих форм. У межах економічних процесів це інформаційно-економічні, виробничо-економічні та інноваційно-економічні процеси, що безпосередньо залежать від безпеки даних і надійності цифрових систем. У сфері суспільних процесів формуються політико-економічні, соціально-економічні та організаційно-економічні процеси, стабільність яких значною мірою визначається рівнем інформаційної та кібербезпеки. Взаємозв’язок між природними і суспільними процесами забезпечується через механізми ринкового та інституціонального регулювання (рис. 1.2).

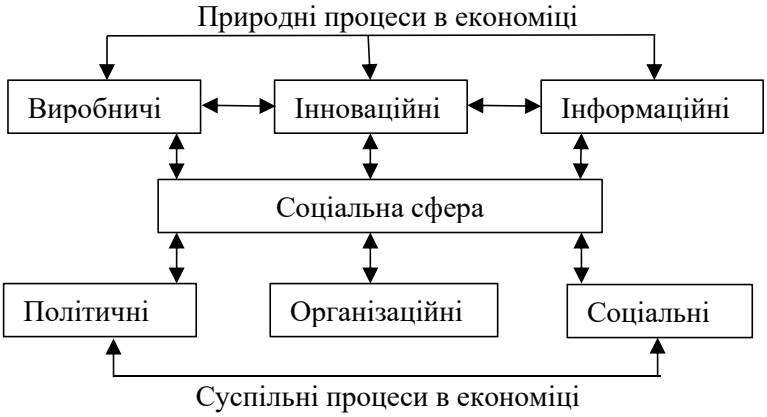


Рисунок 1.2 – Соціально сфера та її взаємозв’язок із іншими процесами

Предметом прогнозування у сфері кібербезпеки є виявлення закономірностей розвитку кіберзагроз і захисних технологій у майбутньому, а також дослідження методів і підходів до створення прогнозів у цій галузі.

Метою кіберпрогнозування є формування науково обґрунтованих передумов для прийняття управлінських рішень у сфері інформаційної безпеки

як на рівні державних структур, так і в межах корпоративного управління. Ці передумови передбачають:

- науковий аналіз тенденцій зміни соціально-економічних процесів;
- варіантне передбачення розвитку соціально-економічних процесів з огляду на наявні тенденції й окреслену мету;
- оцінювання ймовірних наслідків ухвалених рішень;
- обґрунтування напрямів соціального та науково-технічного розвитку.

Основним завданням прогнозування у сфері кібербезпеки є, з одного боку, визначення можливих сценаріїв розвитку кіберзагроз у ближчій та віддаленій перспективі з урахуванням реальних процесів, що відбуваються у сучасному кіберпросторі. З іншого боку, прогнозування має забезпечити основу для створення оптимальних програм і планів захисту інформаційних систем та інфраструктури конкретної організації. Такі програми повинні спиратися на результати прогнозу й включати оцінку управлінських рішень у сфері безпеки з погляду їхніх наслідків у прогнозованому періоді.

Етап прогнозування – частина процесу розробки прогнозів, що характеризується своїми задачами, методами і результатами.

Процес прогнозування в найзагальнішому вигляді включає 8 етапів:

1) передпрогнозна орієнтація;
2) розробка завдання на прогноз;
3) прогнозна ретроспекція;
4) прогнозний діагноз;
5) прогнозна проспекція;
6) верифікація прогнозу;
7) коректування прогнозу;
8) синтез прогнозів.

Основними функціями прогнозування є:

– науковий аналіз економічних, соціальних, науково-технічних процесів і тенденцій;

– дослідження об'єктивних зв'язків соціально-економічних явищ господарського розвитку у конкретних умовах;

– оцінка сформованого рівня розвитку, конкретної ситуації і виявлення тенденцій, які можуть скластися у майбутньому, передбачення нових ситуацій та їх оцінка;

– виявлення можливих альтернатив розвитку економіки у перспективі, нагромадження наукового матеріалу для обґрунтованого вибору певних рішень.

Принципи прогнозування:

1) варіантність – розробка декількох варіантів прогнозу, виходячи з особливостей робочої гіпотези, постановки мети (у нормативному прогнозуванні) і варіантів прогнозного фону;

2) верифікованість – перевірка достовірності, точності і обґрунтованості прогнозів;

3) безперервність – принцип прогнозування, що потребує коригування прогнозів в міру необхідності під час надходження нових даних про об'єкт прогнозування;

4) рентабельність – перевищення економічного ефекту від використання прогнозу над витратами на його розробку;

5) системність – взаємна ув'язка всіх прогнозованих показників, а також параметрів прогнозів;

6) узгодженість – принцип прогнозування, що потребує узгодження нормативних і пошукових прогнозів різної природи і різного періоду попередження.

Найтісніше прогнозування пов’язане із плануванням.

План та прогноз у сфері кібербезпеки є взаємопов’язаними й взаємодоповнювальними етапами управління захистом інформаційних систем. Провідна роль належить саме плану, який визначає конкретні заходи безпеки та виступає ключовою ланкою організаційного управління ризиками. Прогноз у

ції системі виконує функцію орієнтира: він допомагає врахувати можливі сценарії розвитку кіберзагроз та адаптувати сучасні практики під потенційні майбутні виклики. Прогнозування, у свою чергу, є інструментом підготовки планів кіберзахисту. Його застосування може мати різні форми: від попереднього моделювання ризиків перед створенням плану до аналізу наслідків після впровадження заходів безпеки, або ж паралельного розроблення разом із самим планом.

Разом з тим між прогнозом і планом існують суттєві відмінності. План у сфері кібербезпеки завжди має директивний характер і виступає як однозначне управлінське рішення, навіть якщо він включає кілька варіантів реалізації. Натомість прогноз є імовірнісним і багатоваріантним, оскільки базується на оцінці можливих сценаріїв розвитку кіберзагроз. Таким чином, прогнозування формує дослідницьку базу для планування, але ґрунтується на власних методах: наприклад, застосуванні систем моніторингу загроз, аналізі уразливостей чи використанні моделей штучного інтелекту для виявлення потенційних атак. Планування ж спирається на більш конкретні інструменти управління – від політик доступу до балансування ресурсів кіберзахисту.

Крім того, прогнозування та планування відрізняються рівнем пізнання об’єкта: прогноз відображає імовірні майбутні сценарії розвитку кіберзагроз, тоді як план конкретизує чіткі дії для їх запобігання, створюючи комплексну систему протидії.

Перелік запитань для самопідготовки:

1. Що таке економічна кібернетика?
2. Які основні завдання економічної кібернетики в системі управління?
3. Які методи використовуються в економічній кібернетиці?
4. Як економічна кібернетика взаємодіє з іншими науками?
5. У чому полягає відмінність між кількісними та якісними методами в кібернетиці?
6. Що таке економічне прогнозування?

7. У чому полягає суть, предмет і об’єкт економічного прогнозування?

8. Яку роль відіграє прогнозування в системі управління економікою?

9. Як пов’язані планування й прогнозування?

10. Які види економічного прогнозування ви знаєте?

11. Що таке моніторинг економічних процесів?

12. Яке значення має моніторинг для прогнозування?

13. Які альтернативи існують для традиційного прогнозування?

14. У чому відмінність між моніторингом і прогнозуванням?

15. Як поєднуються моніторинг і моделювання у прогнозуванні?

16. Які основні принципи економічного прогнозування?

17. У чому полягає суть прогнозного процесу?

18. Як забезпечується наукова обґрунтованість прогнозів?

19. Що таке сценарний підхід у прогнозуванні?

20. Як оцінюється точність та ефективність прогнозів?

Тема 2. Основні методи прогнозування: класифікація прогнозів.

2.1. Поняття прогнозування, його суть, предмет та об’єкт.

2.2. Типологія прогнозів.

2.3. Методи прогнозування та їх класифікація.

2.4. Динамічний (часовий) ряд і тренд.

2.1. Поняття прогнозування, його суть, предмет та об’єкт

Прогнозування є важливою єднальною ланкою між теорією та практикою в усіх царинах життя суспільства, і має дві різні площини конкретизації: власне передбачальну (дескриптивну, описову) і вказівну (прескриптивну, визначальну).

- Під передбаченням розуміють опис можливих чи бажаних перспектив, станів, розв’язань проблем майбутнього.
- Вказівка (рекомендація) дає спосіб розв’язання цих проблем, використання інформації про майбутнє у цілеспрямованій діяльності. Отже, в проблемі прогнозування розрізняють два аспекти: теоретико-пізнавальний та управлінський, пов’язаний із можливістю прийняття на підставі набутого знання управлінських рішень.

Прогнозування в кібербезпеці сьогодні мають надзвичайну актуальність, адже кількість кібератак щороку зростає, а їхні методи стають дедалі складнішими. Використовуючи математичні моделі та алгоритми прогнозування, організації можуть заздалегідь оцінювати ризики, визначати потенційні вразливості та будувати сценарії можливих атак.

Прогнозування дозволяє компаніям раціональніше розподіляти бюджет на кіберзахист: вкладати більше в ті ділянки, де ризик зламу найвищий. Наприклад, за допомогою моделей машинного навчання можна передбачити ймовірність фішингових атак чи зростання активності бот-мереж у певний період. Це знижує як прямі фінансові збитки від атак, так і витрати на ліквідацію їхніх наслідків.

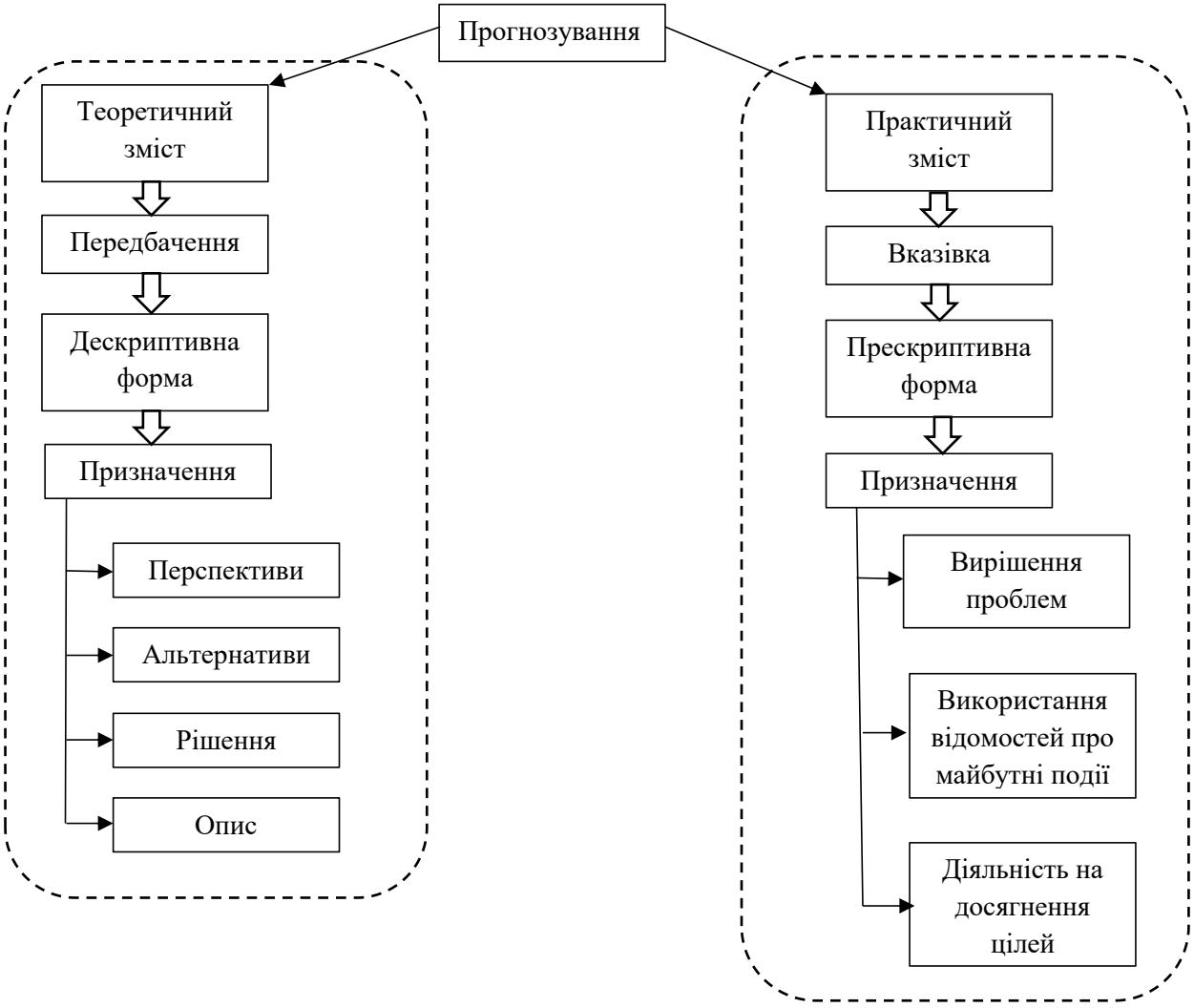


Рисунок 2.1 – Сутність прогнозування

Виходячи із загальноприйнятих понять теорії управління суспільними процесами, прогнозування відіграє провідну роль у системі управління процесами соціально-економічного розвитку як єднальна ланка між об'єктивно необхідними функціями підготовки й реалізації відповідних рішень (рис. 2.2).

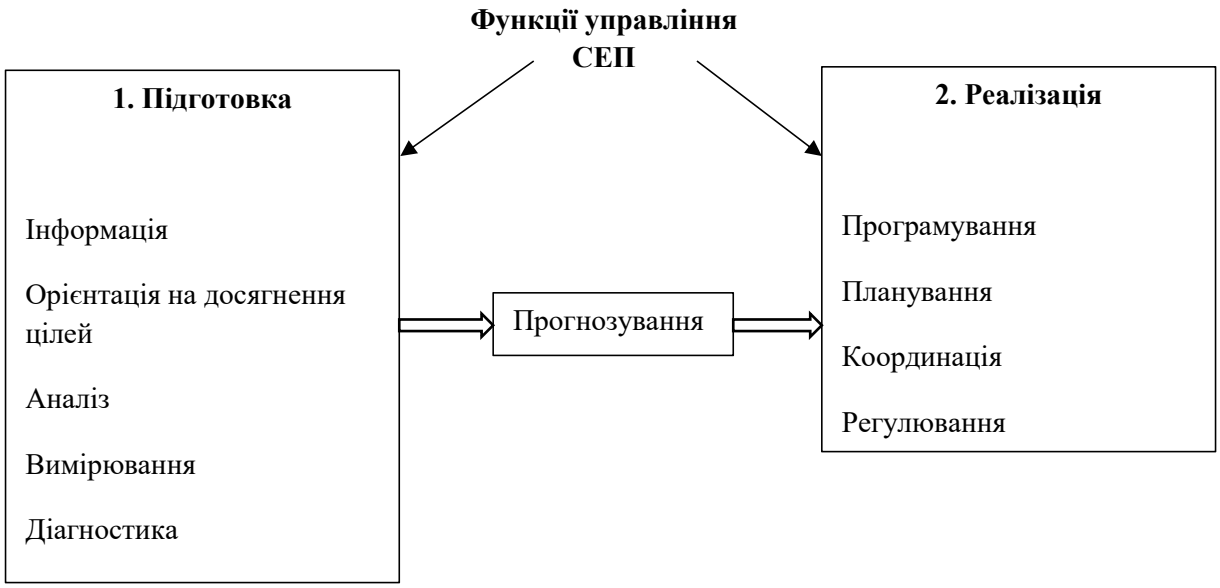


Рисунок 2.2 – Місце прогнозування серед функцій управління в соціальній сфері та кібербезпеці

Функції підготовчого блоку в умовах цифровізації та зростання значення кібербезпеки мають виражену дослідницько-аналітичну спрямованість і охоплюють п’ять ключових напрямів, кожен з яких пов’язаний із обробкою, аналізом і захистом інформації.

Інформаційна функція передбачає збирання, перевірку та захищене зберігання актуальних статистичних і аналітичних даних щодо динаміки соціально-економічних процесів на відповідному рівні моніторингу. Вона також охоплює аналіз сучасних тенденцій, кращих практик і нового досвіду, зокрема у сфері цифрової економіки та кібербезпеки, з метою запобігання викривленню або втраті інформації.

Функція цілеспрямування полягає у формуванні системи конкретних цілей соціально-економічного розвитку на визначений період із урахуванням національних стратегій підвищення якості життя, добробуту населення, гарантування економічної, соціальної та інформаційної безпеки. При цьому враховуються кіберризики, здатні вплинути на досягнення поставлених цілей.

Аналітична функція (аналіз) включає обробку, узагальнення й інтерпретацію зібраної інформації про перебіг соціально-економічних процесів із порівнянням її з цільовими індикаторами. У цьому контексті особлива увага приділяється достовірності даних, виявленню аномалій і потенційних інформаційних загроз, що можуть спотворити результати аналізу.

Функція вимірювання пов’язана зі створенням і вдосконаленням системи соціально-економічних показників, а також методів їх кількісної оцінки з використанням цифрових інструментів. Важливим аспектом є забезпечення цілісності показників, коректності обчислень і захисту вимірювальних даних від несанкціонованого втручання.

Функція діагностики полягає у формуванні узагальнених висновків щодо поточного стану й рівня соціально-економічного розвитку за аналогією з аудитом. Вона передбачає виявлення позитивних і негативних чинників, у тому числі пов’язаних з рівнем інформаційної та кібербезпеки, які впливають на стабільність і розвиток системи.

Завершальним етапом підготовчого блоку є прогнозування, що полягає в обґрунтуванні можливих сценаріїв розвитку соціально-економічної ситуації з урахуванням внутрішніх і зовнішніх чинників, включаючи технологічні зміни та кіберзагрози. Прогнозування інтегрує результати всіх попередніх функцій у єдиний аналітичний комплекс і реалізується за допомогою спеціалізованих методів і інформаційних технологій.

Водночас прогнозування формує підґрунтя для реалізації функцій виконавчого блоку, які мають програмно-управлінський характер:

1. Програмування – розроблення комплексних програм соціально-економічного розвитку з урахуванням вимог інформаційної безпеки;
2. Планування – складання детальних планів реалізації заходів, визначених у програмах, із чітким визначенням ресурсів, строків і відповідальних осіб;

3. Регулювання – внесення оперативних змін у процесі виконання програм на основі моніторингу результатів і реагування на нові соціально-економічні та кіберризики.

Таким чином, усі функції управління соціально-економічними процесами утворюють цілісну систему, що забезпечує узгодженість дій органів управління на різних рівнях, чіткий функціональний розподіл повноважень і підвищення стійкості розвитку за умов зростаючих інформаційних та кіберзагроз.

2.2. Типологія прогнозів

Однією з ключових теоретичних проблем прогнозування виступає розроблення типології прогнозів. Її можна формувати, спираючись на різні критерії та ознаки, такі як цілі і завдання прогнозування, специфіка об’єкта, застосовані методи організації досліджень тощо. Серед найбільш важливих критеріїв, за якими здійснюється класифікація прогнозів, виділяють: масштаб прогнозування, характер об’єкта, функції прогнозу, зміст прогнозних матеріалів, дискретність представлення результатів та період попередження.

Види прогнозів та їх призначення наведені в таблиці 2.1.

Таблиця 2.1

Види прогнозів та їх призначення

№ з/п	Класифікаційна ознака	Вид прогнозу	Коментар (призначення прогнозу, його зміст)
1	Масштаб прогнозу	Макроекономічний	Народногосподарський: економіка держави вцілому.
		Структурний	- міжгалузевий; - міжрегіональний.
		Розвитку народно-господарських комплексів	- паливно-енергетичного; - агропромислового; - інвестиційного; - виробничої інфраструктури; - сфери обслуговування населення, тощо.
		Галузеві	Окрема галузь народного господарства.
		Регіональні	У розрізі областей, міст, селищ, сіл.

		Первинних ланок народно-господарських систем	- підприємств; - виробничих об’єднань; - окремих виробництв і продуктів.
2	Характер досліджуваних процесів	Розвиток виробничих відносин	
		Соціально-економічних умов і наслідків науково-технічного прогресу	
		Динаміки народного господарства	- темпи; - чинники; - структура.
		Відтворення трудових ресурсів, зайнятості і підготовки кадрів	
		Економічного використання природних ресурсів	
		Відтворення основних фондів і капітальних вкладень	
		Рівня життя населення	
		Фінансових відносин, доходів і цін	
		Зовнішніх економічних зв’язків	
3	Зміст прогнозу	Пошуковий	Прогноз, змістом якого є визначення можливих станів об'єкту прогнозування в майбутньому.
		Нормативний	Прогноз, змістом якого є визначення шляхів і термінів досягнення можливих станів (що приймаються як заданих)
		Комплексний	Прогноз, що містить елементи пошукового і нормативного прогнозів.
4	Способи прогнозування	Експертний	Оснований на мобілізації професійного досвіду й інтуїції експертів (не достатньо статистичних даних)
		Екстраполяція	Перенесення закономірностей минулого і сучасного розвитку на майбутнє.
		Моделювання	Дослідження пошукових і нормативних моделей.
5	Характер відбиваних властивостей	Кількісний	Прогноз, який базується на: - кількісних показниках.
		Якісний	- якісних показниках.
		Системний	- системному представленні об'єкта прогнозування.

6	Дискретність представлення результату	Інтервальний	Прогноз, результат якого представлений у вигляді: - довірчого інтервалу характеристики об'єкта прогнозування для заданої вірогідності здійснення прогнозу.
		Точковий	- єдиного значення характеристики об'єкта прогнозування без вказівки довірчого інтервалу.
7	Період попередження	Оперативний	Прогноз з періодом попередження для об'єктів прогнозування:- до 1 місяця.
		Короткостроковий -	- від 1 місяця до 1 року.
		Середньостроковий	- від 1 понад 3 років.
		Довгостроковий	- від 5 до 15 років.
		Далекостроковий	- понад 15 років.
8	Ступінь детермінованості	Детерміновані (визначені або передбачувані)	Опис об’єкта без втрат інформації
		Стохастичні (імовірнісні)	Врахування випадкових складових для задоволення вимог точності та достовірності прогнозу
		Змішані	частково в детермінованому, частково в стохастичному вигляді
9	Характер розвитку об’єкта в часі	Дискретні (переривані)	Регулярна складова (тренд) яких зміюється стрибками у фіксовані моменти часу
		Аперіодичні	Опис регулярної складової у вигляді безперервної функції часу
		Циклічні	Опис регулярної складової у вигляді періодичної функції часу

За масштабом прогнозування в умовах цифровізації та підвищеної уваги до кібербезпеки виділяють такі рівні:

Макроекономічний (народногосподарський) прогноз – охоплює економіку країни в цілому, враховує соціально-економічні тенденції, державну політику та потенційні кіберзагрози на національному рівні, що можуть впливати на стабільність економічних процесів.

Структурний прогноз – міжгалузевий та міжрегіональний, спрямований на аналіз взаємозв’язків між різними секторами та регіонами країни,

включаючи оцінку ризиків витоку або порушення інформації між окремими економічними ланками.

Прогнози розвитку народногосподарських комплексів – зокрема паливно-енергетичного, агропромислового, інвестиційного, виробничої інфраструктури, сфери обслуговування населення тощо. В умовах цифровізації ці прогнози передбачають оцінку не лише економічних показників, а й вразливостей інформаційних систем, автоматизованих виробничих процесів та кіберзахисту критичної інфраструктури.

Галузеві та регіональні прогнози – передбачають деталізоване оцінювання економічного розвитку конкретних галузей або регіонів із врахуванням локальних особливостей і ризиків цифрових та інформаційних атак.

Прогнози первинних ланок народногосподарської системи – підприємств, виробничих об’єднань, окремих виробництв і продуктів. Вони включають аналіз економічної ефективності та стійкості до кібератак, захищеність корпоративних інформаційних систем, а також готовність до відновлення бізнес-процесів у разі порушень безпеки даних.

Таким чином, кожен рівень прогнозування у сучасних умовах інтегрує оцінку соціально-економічних параметрів і потенційних кіберризиків, забезпечуючи комплексне бачення майбутнього розвитку економічних систем. (рис. 2.3).

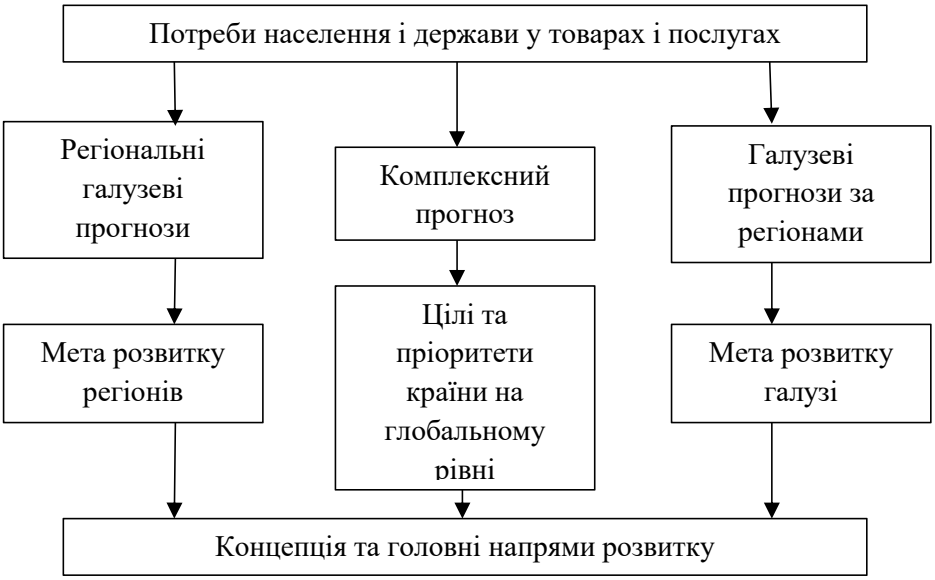


Рисунок 2.3 – Головні напрями розвитку країни

За періодом упередження прогнози в сфері кібербезпеки поділяють на:

Короткотермінові – від одного місяця до року; спрямовані на оцінку поточних кіберзагроз, вразливостей та можливих інцидентів у короткий період. Вони базуються на припущенні, що у прогнозованому періоді не відбудеться радикальних змін у кіберінфраструктурі або тактиках атак.

Середньотермінові – від одного до п’яти років; враховують як кількісні, так і якісні зміни в кіберсередовищі, наприклад, розвиток нових методів атак, впровадження сучасних систем захисту та зміни в політиках безпеки. У середньотермінових прогнозах переважає кількісна оцінка очікуваних подій.

Довготермінові – від п’яти до 15–20 років; орієнтовані на стратегічні зміни в кіберпросторі, включно з технологічними інноваціями, змінами законодавства та глобальними тенденціями кіберзагроз. У таких прогнозах головний акцент робиться на якісних змінах, а кількісні показники служать допоміжними.

Далекоглядні – понад 20 років; спрямовані на виявлення загальних закономірностей розвитку кіберзагроз та інформаційних технологій у майбутньому. Оцінка таких прогнозів має переважно якісний характер і передбачає стратегічне планування розвитку кібербезпеки.

Класифікація прогнозів у кібербезпеці також здійснюється залежно від специфіки досліджуваних процесів, що охоплюють різні аспекти функціонування та захисту інформаційних систем. У цьому контексті виокремлюють такі напрями прогнозів:

Розвиток кіберзагроз і атак – прогнозування появи нових типів шкідливого ПЗ, методів атак, цільових груп та кіберзлочинних схем.

Технологічні інновації в кібербезпеці – оцінка впливу нових технологій, алгоритмів шифрування, систем виявлення вторгнень і захисту даних.

Динаміка стану кіберінфраструктури – аналіз темпів модернізації IT-інфраструктури, змін у структурі інформаційних систем та їхньому захисті.

Підготовка та розвиток кадрів – прогноз потреби в фахівцях з кібербезпеки, навчання персоналу та формування компетенцій у галузі захисту інформації.

Ефективність використання ресурсів – аналіз використання апаратних, програмних та людських ресурсів для забезпечення кіберзахисту.

Модернізація захисних систем і інвестиції – оцінка оновлення програмного забезпечення, апаратних засобів та фінансових вкладень у кіберзахист.

Безпека користувачів і рівень кіберстійкості організацій – прогнозування загального рівня захищеності користувачів та організацій від кіберзагроз.

Фінансові аспекти кібербезпеки – оцінка витрат на кіберзахист, економічних втрат від інцидентів та фінансових ризиків.

Міжнародна кібербезпека та співпраця – прогнозування розвитку зовнішніх кіберзагроз, міжнародних стандартів і взаємодії між державами та організаціями.

Кожен із цих напрямів має самостійне значення і може розроблятися окремо, водночас між ними зберігається методологічна єдність. Це забезпечує комплексний та науково обґрунтований підхід до прогнозування розвитку кіберзахисту та безпеки інформаційних систем.

Прогнозування кіберзахисту здійснюють у поєднанні з іншими видами прогнозів: технологічним, соціальним, політичним, демографічним, науково-технічним, прогнозуванням природних ресурсів та інфраструктури тощо. Результати цих прогнозів враховуються при формуванні стратегій кібербезпеки та плануванні розвитку інформаційних систем.

Зв'язок різних видів прогнозів проявляється у послідовності їх розроблення: наприклад, прогнози кіберзагроз і атак формуються з урахуванням тенденцій науково-технічного прогресу, зміни користувацьких патернів, демографічних процесів та стану технологічної інфраструктури. Економічні та ресурсні аспекти кібербезпеки також враховують ці дані для забезпечення комплексного підходу до захисту інформаційних систем.

За функціональною ознакою прогнозування кіберзахисту розрізняють два типи прогнозів: пошуковий і нормативний.

Пошуковий прогноз базується на умовному продовженні в майбутнє тенденцій розвитку кіберзагроз, вразливостей та атак, що спостерігалися в минулому і сучасному, абстрагуючись від можливих запобіжних заходів чи змін у політиці безпеки. Його завдання – визначити, як розвиватиметься кіберпростір та загрози у разі збереження існуючих тенденцій, наприклад, поширення певних типів шкідливого програмного забезпечення або методів соціальної інженерії.

Нормативний прогноз, навпаки, розробляють на основі заздалегідь визначених цілей кібербезпеки, таких як досягнення певного рівня захищеності критичної інфраструктури або інформаційних систем підприємства. Його завдання – визначити шляхи та терміни реалізації заходів, здатних забезпечити бажаний стан безпеки в майбутньому. Якщо пошуковий прогноз виходить із минулого і теперішнього стану, нормативний прогноз будується у зворотному порядку: від бажаного майбутнього стану системи до необхідних змін у політиках, технологіях та процедурах.

На практиці обидва типи прогнозів застосовуються спільно: пошукові прогнози допомагають оцінити потенційні ризики, а нормативні – визначити конкретні заходи для їх нейтралізації.

Прийняті рішення щодо кіберзахисту, у свою чергу, впливають на результати прогнозів. Тому для забезпечення актуальності і точності прогнозів необхідно їх систематично коригувати, наприклад, порівнюючи очікуваний стан системи з фактичним розвитком загроз і можливістю реалізації запланованих заходів.

За ступенем детермінованості об’єкти прогнозування кібербезпеки можна класифікувати так:

Детерміновані (визначені або передбачувані) – об’єкти, поведінка яких у майбутньому відносно прогнозована і описується чіткими правилами або алгоритмами без істотних втрат інформації. Наприклад, кількість користувачів у внутрішній мережі, структуровані оновлення систем.

Стохастичні (імовірнісні) – об’єкти, поведінка яких включає випадкові складові, що необхідно враховувати для точності прогнозів. Наприклад, ймовірність атак хакерів або поширення вірусів у глобальній мережі.

Змішані – об’єкти, які частково піддаються прогнозуванню у детермінованій формі, а частково мають випадкові (імовірнісні) характеристики. Наприклад, прогноз навантаження на сервери з урахуванням випадкових сплесків трафіку та планових процесів.

За характером розвитку в часі об’єкти прогнозування діляться на:

Дискретні (переривані) – об’єкти, чий стан змінюється стрибкоподібно у визначені моменти часу (наприклад, планові оновлення програмного забезпечення або проведення аудиту безпеки).

Аперіодичні (безперервні) – об’єкти, що змінюються плавно і безперервно з часом (наприклад, постійне накопичення даних або поведінка користувачів у системі).

Циклічні – об’єкти з періодичною регулярною зміною стану (наприклад, сезонні сплески спроб фішингових атак).

За ступенем інформаційної забезпеченості об’єкти прогнозування можна розділити на:

З повним кількісним забезпеченням – коли є достатня кількість ретроспективних даних для застосування методів статистичного прогнозування або екстраполяції. Наприклад, історичні дані про кількість виявлених атак за останні роки.

З неповним кількісним забезпеченням – коли дані частково відсутні або неповні, і точність прогнозу знижується.

З наявністю лише якісної інформації – коли відомі лише загальні тенденції чи факти, але відсутні числові дані.

З повною відсутністю ретроспективної інформації – нові системи, ще не введені в експлуатацію, або нові види кіберзагроз, для яких немає історичних даних.

Із типологією прогнозів у сфері кібербезпеки тісно пов’язане питання джерел інформації щодо майбутніх загроз та способів їх прогнозування. В науковій практиці розрізняють три основні джерела прогнозової інформації. Перше джерело – набутий досвід, який ґрунтується на знанні закономірностей розвитку кіберінцидентів, поведінки зловмисників та ефективності застосованих заходів захисту. Друге джерело – екстраполяція наявних тенденцій, коли закономірності розвитку кіберзагроз, зафіксовані у минулому та сьогоденні,

переносяться на майбутнє для оцінки потенційних ризиків. Третє джерело – побудова моделей об’єктів кіберзахисту та прогнозованих атак стосовно очікуваних або визначених умов.

Кожному із джерел відповідають певні способи прогнозування. Перший спосіб – експертний, заснований на залученні висококваліфікованих фахівців із кібербезпеки, які використовують свій досвід та інтуїцію для оцінки потенційних загроз. Цей підхід застосовується, коли статистичних даних про кіберінциденти недостатньо або їхня природа надто непередбачувана. Другий спосіб – екстраполяція, що передбачає аналіз минулих та поточних атак для виявлення закономірностей їх розвитку й перенесення цих закономірностей на майбутнє. Третій спосіб – моделювання, яке полягає у створенні пошукових і нормативних моделей кіберсистем та потенційних загроз, що дозволяє оцінити ефективність запобіжних заходів і підготувати сценарії реагування на атаки.

У науковій практиці моделі кіберзагроз та систем захисту класифікують залежно від математичних та аналітичних методів, що застосовуються для їх побудови. Формалізоване представлення вихідних умов та взаємозв’язків у таких моделях прийнято називати економіко-математичними або кібер-математичними моделями. Вони можуть бути статичними або динамічними, мати детермінований чи стохастичний характер, будуватися за лінійними чи нелінійними залежностями, а також відображати кіберпроцеси в межах сталої або змінної інфраструктури захисту.

Таким чином, інтеграція методів прогнозування у сфері кібербезпеки дозволяє систематично оцінювати ризики, планувати ефективні заходи протидії та підвищувати загальну кіберстійкість організацій.

Під методами прогнозування розуміють комплекс інтелектуальних прийомів і аналітичних підходів, що забезпечують можливість, спираючись на ретроспективні дані про внутрішні та зовнішні взаємозв’язки прогнозованого об’єкта, а також на кількісні вимірювання його характеристик у межах певного явища чи процесу, сформулювати науково

обґрунтовані й достовірні висновки щодо ймовірних напрямів розвитку та майбутнього стану цього об’єкта.

В даний час налічується понад 150 різних методів прогнозування, з яких на практиці використовується 15-20.

У процесі прогнозування кібербезпеки, як і в економічному прогнозуванні, застосовуються як загальні наукові методи та підходи до дослідження, так і специфічні методи, властиві саме сфері інформаційної безпеки. До загальних методів належать системний аналіз, статистична обробка даних, моделювання та порівняльний аналіз, що дозволяють виявляти закономірності розвитку загроз, оцінювати потенційні ризики та ефективність заходів захисту.

Специфічні методи прогнозування кібербезпеки враховують особливості інформаційних систем, поведінку зловмисників, технологічні вразливості та швидкі зміни у сфері кіберзагроз. Це, зокрема, методи імітаційного моделювання атак і захисних заходів, експертна оцінка ризиків, сценарне прогнозування кібератак, а також аналіз даних із систем моніторингу кіберінцидентів. Комбінація загальних і спеціалізованих методів дозволяє створювати комплексні та науково обґрунтовані прогнози, що слугують основою для планування превентивних заходів та підвищення кіберстійкості організацій.

У числі загальних методів можна виділити такі:

- історичний метод полягає в розгляді кожного явища у взаємозв'язку його історичних форм;

- комплексний метод полягає в розгляді явищ у їх взаємозалежності, використовуючи для цього методи дослідження не тільки цієї, але й інших наук, які вивчають ці явища;

- системний метод передбачає дослідження кількісних і якісних закономірностей протікання імовірнісних процесів в складних економічних системах;

- структурний метод дозволяє встановити причини досліджуваного явища, пояснити його структуру;

- системно-структурний метод передбачає, з одного боку, розгляд системи як динамічно розвивається цілого, а з іншого - розчленування системи на складові структурні елементи і розгляд їх у взаємодії.

Специфічні методи прогнозування в сфері кібербезпеки тісно пов'язані з кіберпрогностикою як науковою дисципліною. Серед основних інструментів цієї галузі провідне місце займають економіко-математичні та інформаційно-математичні підходи, методи побудови моделей розвитку кіберзагроз і вразливостей, а також прийоми статистичної екстраполяції та аналізу даних із систем моніторингу інцидентів і кібератак.

Важливим аспектом прогнозування кіберзагроз є питання його об'єктивної істинності. Під цим розуміють відповідність форми та параметрів прогнозу реальним можливостям розвитку кіберінцидентів і тих тенденцій, які мають шанс проявитися у майбутньому, але вже нині спостерігаються у вигляді перших ознак загроз, таких як нові види шкідливого ПЗ, атаки на критичну інфраструктуру чи вразливості в програмному забезпеченні. Такий підхід дозволяє своєчасно визначати пріоритети захисних заходів і підвищувати стійкість інформаційних систем.

Розгляд проблеми істинності прогнозу безпосередньо пов'язаний із визначенням критеріїв перевірки достовірності. Їх умовно поділяють на дві групи:

– практичні критерії, де ключову роль відіграє практика як універсальний спосіб перевірки істини на всіх етапах прогнозування;

– логічні або непрямі критерії, що включають можливість перевірки прогнозних висновків, їхню відповідність реальним умовам, логічну узгодженість та відсутність суперечностей.

Основними параметрами прогнозів є такі:

1) достовірність прогнозу;

2) джерело помилки прогнозу;

3) обґрунтованість прогнозу;

4) помилка прогнозу;

5) період підстави прогнозу;

6) період попередження;

7) прогнозний горизонт;

8) точність прогнозу.

Зважаючи на сучасні вимоги до прогнозування соціально-економічного розвитку, доцільно розглядати низку ключових аспектів технології формування прогнозів у взаємозв'язку. Сюди відносяться динаміка основних макроекономічних показників, демографічні процеси, рівень зайнятості населення, доходи громадян та система оплати праці, а також проблеми соціально-економічної безпеки, до якої сьогодні належить і кібербезпека. Адже розвиток цифрової економіки неможливий без забезпечення безпечного функціонування інформаційної інфраструктури, захисту персональних даних та стабільності критичних державних і корпоративних систем.

Розроблення прогнозів розвитку національної економіки ґрунтується на створенні комплексної системи прогнозування, яка включає як базові, так і соціально-економічні прогнози. Базові прогнози мають самостійні цілі та завдання, слугують фундаментом для побудови соціально-економічних прогнозів і забезпечують їх наукову обґрунтованість. У контексті соціально-економічного прогнозування базові прогнози умовно поділяють на два взаємопов'язані блоки:

Ресурсний блок, який надає інформацію про обсяги та перспективи розвитку ресурсної бази країни, зокрема трудових, фінансових, природних і технологічних ресурсів.

Блок зовнішніх впливів, що визначає чинники, сформовані поза межами національної економіки, проте впливають на соціально-економічні параметри

держави, включно з глобальними тенденціями у цифровій та кібербезпековій сферах.

У межах ресурсного блоку здійснюється демографічне прогнозування, прогноз природних ресурсів і науково-технічний прогноз.

Демографічний прогноз спрямований на оцінку перспективної зміни чисельності населення та його структури. Серед ключових показників – чисельність дітей дошкільного віку для планування розвитку мережі дитячих садків і шкіл; чисельність осіб пенсійного віку для органів соціального забезпечення; дані про статевовікову структуру населення, необхідні для розвитку медичних закладів і підготовки медичних кадрів; чисельність і склад сімей, що впливає на планування виробництва товарів широкого вжитку, будівництво житла та формування соціальних послуг.

Демографічне прогнозування передбачає врахування міграційних процесів, оскільки вони безпосередньо впливають на вікову структуру населення, а зміни у віковому та статевому складі віддзеркалюються у рівнях народжуваності та смертності. Така взаємозалежність є методологічною основою прогнозування трудових ресурсів, зайнятості населення та професійно-кваліфікаційної структури робочої сили. Крім того, прогноз дозволяє оцінювати середню тривалість життя, кількість домогосподарств і їхній розмір, напрямки та інтенсивність міграційних потоків, а також структурні зміни в суспільстві, що мають соціально-економічне значення.

Прогноз природних ресурсів охоплює оцінку обсягів використання вже освоєних і перспективних родовищ, а також потенціалу нових запасів, які ще не введені в промислову експлуатацію. Аналізуються паливно-енергетичні та мінеральні ресурси, лісові та гідроенергетичні ресурси, а також альтернативні джерела енергії, включно із сонячною, вітровою та припливно-відпливною енергетикою. Основною функцією прогнозу є визначення майбутніх обсягів ресурсів, умов їх залягання та просторового розподілу, що безпосередньо

впливає на формування структури національної економіки, регіональний розвиток, зовнішньоекономічні зв’язки та ефективність виробництва.

Прогноз природних ресурсів також враховує можливості переходу до використання ресурсів нижчої якості, освоєння родовищ у важкодоступних районах, застосування вторинних і супутніх ресурсів та розвиток альтернативних джерел енергії. Інформація з прогнозу ресурсів інтегрується до соціально-економічного прогнозування, включаючи цифрову економіку та кіберінфраструктуру, адже енергетична база, матеріали та технології впливають на можливості розвитку інформаційних систем і безпеки.

Прогноз науково-технічного прогресу передбачає оцінку розвитку фундаментальної науки, прикладних досліджень, галузевої науки, техніки та технологій, а також впровадження результатів науково-дослідних і конструкторських робіт у виробничу і невиробничу сферу. Його результати впливають на соціально-економічний розвиток, оптимізацію використання природних ресурсів та підвищення продуктивності праці. Наприклад, розвиток атомної енергетики або синтетичних матеріалів зменшує залежність економіки від традиційних ресурсів, а інновації в цифрових технологіях і кіберзахисті забезпечують стабільність критичних систем і розвиток нових ринків.

Від темпів науково-технічного прогресу залежить економічне зростання, формування технологічних укладів, а також здатність держави задовольняти суспільні потреби у товарах, послугах та безпечному цифровому середовищі. Впровадження нових технологій скорочує життєвий цикл інвестицій, стимулює розвиток інновацій та потребує адаптації економіки до нових умов.

Соціально-економічні прогнози поділяються на економічний та соціальний блоки. Економічний блок включає прогноз сукупної пропозиції, сукупного попиту, галузевої структури економіки та рівня інфляції. Соціальний блок охоплює прогноз рівня життя населення, зайнятості, житлово-комунального господарства, охорони здоров’я, освіти, культури і мистецтва.

Прогноз зайнятості населення є ключовим, оскільки інтегрує демографічні дані та слугує зв’язком між економічними і соціальними прогнозами.

Особливу увагу приділяють кібербезпеці, яка стала невід’ємною частиною соціально-економічного прогнозування. Прогнозування кіберзагроз і захисних заходів забезпечує стабільну роботу критичних систем, захист персональних даних, формування цифрової культури населення та розвиток IT-інфраструктури. Кібербезпека є одночасно первинним фактором розвитку цифрової економіки, оскільки задає вимоги до безпеки, і залежним від економічних та технологічних ресурсів, адже без належного фінансування та технологічного прогресу неможливе забезпечення захисту інформаційних систем.

На етапі формування інформаційної бази прогнозування визначають тип і склад вихідних даних, які повинні бути достовірними, точними, співставними та охоплювати всі часові проміжки. Початкову інформацію поділяють на фактографічну (бухгалтерська звітність, випереджальні дані, патенти, винаходи) та експертну (професійні оцінки фахівців). Сукупність цих даних формує інформаційний масив прогнозування, який характеризується повнотою початкової інформації та інформативністю змінних об’єкта прогнозування. Зі збільшенням часу від моменту спостереження зменшується інформативність ретроспективних даних, що відображає закономірність дисконтування інформації.

Інтеграція соціально-економічних, науково-технічних та кібербезпекових прогнозів дозволяє формувати надійну базу для стратегічного розвитку держави, забезпечувати стабільність цифрової та реальної економіки, мінімізувати ризики та створювати передумови для інноваційного зростання в усіх сферах.

2.3. Методи прогнозування та їх класифікація

Основною метою прогнозування є формування науково обґрунтованих варіантів можливих тенденцій розвитку чи змін керованого об’єкта, а також показників його стану у часовому та просторовому вимірах.

Джерелами інформації для здійснення прогнозних розрахунків виступають вербальні та письмові матеріали, отримані в процесі комунікацій між людьми або з відкритих друкованих джерел.

Для обробки інформації з відкритих джерел застосовуються такі методи:

– структурно-морфологічний аналіз, що дозволяє систематизувати та узагальнювати інформацію;

– метод визначення публічної активності, який базується на відстеженні рівня суспільного інтересу до певних тем;

– метод групування патентних документів, що дає змогу оцінити інноваційну активність у конкретних сферах;

– аналіз статистичних показників, які відображають динаміку об’єкта;

– термінологічний та лексичний аналіз, що використовується для виявлення змістових особливостей текстової інформації.

Для прогнозування в практичній діяльності застосовуються різні кількісні і якісні методи.

Кількісні методи базуються на інформації, яку можна отримати, на основі вивчення тенденцій зміни параметрів або маючи статистично достовірні залежності, що характеризують виробничу діяльність об'єкта управління. Як приклади таких методів можуть виступати: аналіз тимчасових рядів, каузальне (причинно-наслідкове) моделювання.

В основу якісних методів покладені експертні оцінки фахівців в сфері прийняття рішень: наприклад, методи експертних оцінок, висновок, моделі очікування споживача (опитування клієнтів).

Складні об’єкти підлягають прогнозуванню з використанням як кількісних, так і якісних методів. Наприклад, оцінювання економічної

кон’юнктури (сукупності показників, що відображають стан економіки у певний період часу) здійснюється на основі прогнозів у суміжних сферах: обмежень, пов’язаних із захистом довкілля, розвитку міжнародної торгівлі, динаміки попиту та пропозиції на товари, а також співвідношення між ними. Кожен із таких прогнозів, у свою чергу, спирається на проміжні оцінки окремих процесів.

До найбільш відпрацьованих методів прогнозування економічної кон’юнктури належать: метод «мозкової атаки», техніка Дельфі, екстраполяція тенденцій, морфологічний аналіз, імітаційне динамічне моделювання, структурний аналіз та інші. Водночас існують й інші підходи до класифікації методів прогнозування, які залежать від специфіки та особливостей самих прогнозів. (табл. 2.2).

У сфері кібербезпеки цілі, часові горизонти, умови здійснення прогнозу та специфіка його формування визначають вибір комплексу методів і підходів прогнозування. При цьому для різних типів загроз або сценаріїв розвитку подій можуть застосовуватися різні методи – від статистичного аналізу та машинного навчання до експертних оцінок чи імітаційного моделювання.

Приведемо коротку характеристику методів прогнозування, які використовуються частіше за інших, із вказівкою джерел, в яких їх суть викладена в розгорненому вигляді (табл. 2.3).

Таблиця 2.2

Класифікація прогнозів

Класифікаційний признак	Роль і місце прогнозу в управлінському рішенні
Цільовий	Визначення можливості реалізації цілей управління в кібербезпеці дозволяє уточнити стратегічні завдання організації та сформулювати її місію у цифровому середовищі. При цьому визначаються конкретні критерії досягнення мети – рівень захисту інформаційних ресурсів, мінімізація ризиків витоку даних, зниження кількості успішних атак чи скорочення часу реагування на інциденти. Чітке формулювання таких орієнтирів забезпечує узгодженість дій та підвищує ефективність системи кіберзахисту.

Пошуковий	Виявлення закономірних тенденцій у розвитку загроз та засобів кіберзахисту дає змогу визначити поточний стан інформаційної безпеки та спрогнозувати її майбутній рівень. Такий аналіз враховується під час розробки стратегічних рішень, що стосуються побудови системи захисту, розподілу ресурсів та вибору пріоритетних напрямів інвестування у безпеку.
Нормативний	Визначення шляху та етапів реалізації цілей у сфері кібербезпеки дає змогу структурувати процес прийняття управлінських рішень. На цій основі доцільно застосовувати відповідні методи управління – планові (розробка стратегій кіберзахисту), програмні (впровадження комплексних рішень безпеки) та інші інструменти, що забезпечують узгодженість дій і ефективний захист інформаційних ресурсів.
Програмний	Дослідження впливу різних факторів на окремих етапах досягнення цілей у сфері кібербезпеки дозволяє глибше зрозуміти механізми формування ризиків і ефективності захисних заходів. Формулюється гіпотеза щодо взаємовпливу технічних, організаційних та людських чинників на об’єкт прогнозування – наприклад, рівень стійкості інформаційної системи до кібератак. На цій основі визначаються ймовірнісні терміни досягнення проміжних цілей (зниження кількості інцидентів, скорочення часу реагування, підвищення рівня кіберграмотності персоналу) у процесі руху до головної мети – комплексного забезпечення інформаційної безпеки.
Проектний	Отримання аналітичного матеріалу, що забезпечує цільову спрямованість концепцій проєктів у сфері кібербезпеки, визначення їхнього життєвого циклу та критеріїв оцінки ефективності інвестицій у захист інформаційних систем має ключове значення. Результати проєктного прогнозування використовуються під час розробки інвестиційних та фінансових рішень, зокрема для вибору оптимальних технологій захисту, планування бюджету на впровадження систем безпеки та оцінки доцільності довгострокових проєктів у цифровій інфраструктурі.
Час (період) упередження	У сфері кібербезпеки застосовуються оперативні, коротко-, середньо- та довгострокові прогнози, які спрямовані на забезпечення, розробку, прийняття та реалізацію управлінських рішень різного рівня. Це стосується як тактичних рішень (реагування на інциденти в реальному часі), так і оперативних (планування заходів протидії актуальним загрозам), а також стратегічних (формування довгострокової політики кіберзахисту та інвестицій у розвиток захисних технологій).
Умови взаємозв'язку факторів	Прогнози у сфері кібербезпеки формуються з урахуванням детермінованих та ймовірнісних взаємозв’язків між факторами ризику та об’єктами захисту, а також під впливом невизначених загроз і умов. Саме ці умови визначають вибір методів прогнозування інцидентів та розробку стратегій прийняття рішень щодо захисту інформаційних систем.
Специфіка обробки інформації особою, що приймає рішення	Моделі прогнозування в кібербезпеці можуть бути представлені як математичними залежностями (формалізовані моделі, що дозволяють здійснювати прогнозування та прийняття

	рішень за допомогою комп'ютерних систем), так і у вигляді письмових або усних описів. Інформація, яка обробляється систематично або інтуїтивно, визначає тип прогнозів та рішень, наприклад, інтуїтивні оцінки загроз і потенційних інцидентів.

При вирішенні задач прогнозування і прийняття рішень суттєвою проблемою є кількість і якість обробки необхідної інформації.

Експертні методи застосовуються для прогнозування економічної кон'юнктури, розв'язання проблем науково-технічного прогресу та розвитку складних об'єктів. Вони використовуються у випадках, коли розвиток об'єкта не піддається предметному опису чи математичній формалізації, а також за відсутності достовірної статистики, в умовах великої невизначеності, без використання обчислювальної техніки або в екстремальних ситуаціях. Особливість таких методів полягає у формуванні колективної думки групи експертів (зазвичай 7–9 фахівців), проте вони потребують значних витрат часу на опитування та обробку даних.

Метод евристичного прогнозування застосовується переважно до науково-технічних об'єктів і проблем, розвиток яких складно формалізувати. Він дозволяє знаходити оптимальні способи створення або модернізації систем, прогнозувати розвиток великих і складних структур, а також формувати об'єктивізоване уявлення про перспективи розвитку вузької сфери. У цьому випадку математичний апарат є непридатним, натомість використовуються спеціально оброблені експертні оцінки, отримані шляхом систематизованих опитувань. Інформаційна база створюється на основі таблиць, заповнених фахівцями у конкретній галузі.

Колективна генерація ідей спрямована на формування блока ідей для прогнозування та прийняття рішень. Її застосування дозволяє визначити весь спектр можливих варіантів розвитку об'єкта управління, виявити альтернативні фактори, що впливають на прогноз, та розробити сценарії майбутнього розвитку. Основними особливостями цього методу є синтез об'єкта прогнозу, багатофакторний аналіз подій та оцінка їх з позиції детермінуючих чинників.

Морфологічний аналіз використовується у випадках, коли наявний обсяг інформації є обмеженим, але необхідно отримати систематизовані відомості щодо всіх можливих варіантів вирішення проблеми. Він дає змогу прогнозувати результати фундаментальних досліджень, оцінювати перспективи нових ринків чи формування нових потреб. Метод ґрунтується на виявленні структурних зв'язків між об'єктами, явищами та концепціями і вимагає відсутності упереджених оцінок. Процес передбачає кілька етапів: формулювання проблеми, аналіз параметрів, побудову так званого «морфологічного ящика», що включає всі можливі рішення, а також їх подальше дослідження.

Прогнозні графи та метод «дерева рішень» належать до структурних методів прогнозування, спрямованих на пошук рішень за умов збереження функцій об'єкта, але зі зміною його структури. Вони використовуються для формулювання сценаріїв досягнення прогнозованої мети та визначення відповідних критеріїв. Особливістю є необхідність правильного вибору графа, що відображає сутність відносин у межах об'єкта прогнозування.

Математичні методи параметричного прогнозування застосовуються для виявлення тенденцій розвитку об'єктів, які мають достатні кількісні статистичні дані про їхній минулий і теперішній стан. Вони дозволяють розв'язувати задачі прогнозної екстраполяції за умов відповідності статистичних даних вимогам обраних математичних методів, наявності кількісної інформації та обмеженої глибини прогнозного горизонту в межах одного циклу розвитку об'єкта. Зокрема, метод найменших квадратів ефективний у випадках, коли протягом періоду прогнозування функціональна структура об'єкта не зазнає змін, змінюються лише значення параметрів.

Для прогнозування процесів, динаміка яких має коливальний або гармонійний характер, застосовується спектральний аналіз. Такий підхід використовується для оцінки сезонних коливань попиту, макроекономічних процесів та енергоспоживання.

Факторний аналіз передбачає максимально повний облік змінних, що характеризують об’єкт, та їх взаємозв’язків. Він належить до математичної статистики та включає широкий спектр методів, які дають змогу комплексно оцінювати вплив різних факторів.

Прогнозування за аналогією орієнтоване на розв’язання управлінських ситуацій, знайомих для осіб, що приймають рішення. Воно ґрунтується на використанні аналогів об’єктів, процесів чи подій, які вже мали місце раніше. Ефективність цього методу значною мірою залежить від наявності релевантних аналогів і потребує спеціальних навичок для коректного застосування.

Точність і перевірка прогнозів – важливий аспект прогнозування. Прогнозування майбутніх значень на регулярній основі ускладнюється важкою природою змінних. Саме тому необхідно включити в прогноз показник можливого відхилення значення змінної.

Вибір методу прогнозування значно покращує якість прогнозу, оскільки:

• забезпечує функціональну повноту, достовірність і точність прогнозу;

• зменшує тимчасові і матеріальні витрати на прогнозування. На вибір методу прогнозування впливають:

сутність практичної проблеми, що підлягає розв'язанню;

• динамічні характеристики об'єкта прогнозування у ринковому середовищі;

• вид і характер наявної інформації, типове уявлення про об'єкт прогнозування;

• комбінація фаз життєвого, ринкового циклу товару або послуги;

• період попередження і його співвідношення із передбачуваною тривалістю ринкового, життєвого циклу, циклу розробки або модифікації товару, послуги;

Усі перелічені фактори необхідно розглядати у взаємозв’язку як єдину систему, допускаючи виключення лише несуттєвих елементів. Особа, яка здійснює прогнозування, на етапі передпрогнозних досліджень повинна

впорядкувати та проаналізувати інформацію про об’єкт прогнозу, а далі залишити лише ті дані, що відповідають умовам завдання.

Під час вибору методу прогнозування кіберзагроз слід виокремлювати підходи, які ефективно застосовуються у конкретних умовах інформаційної системи, відкидаючи непридатні. Практично при цьому варто зосереджуватися на двох ключових критеріях – витратах на підготовку прогнозу та точності результатів. Необхідно оцінити бюджет на прогнозування, потенційні втрати від помилок та очікуваний ефект. Зазвичай оптимальним є прогноз, який забезпечує високу точність при прийнятній вартості реалізації.

При виборі методу прогнозування потрібно враховувати:

– наявність статистичних даних за необхідний період;

– компетентність прогнозиста;

– час, що необхідний для збору і аналізу інформації.

У ряді випадків для отримання незалежних прогнозів використовуються одночасно декілька методів.

Вимоги до прогнозів:

– своєчасність, певною мірою точності і визначеності інших показників;

– надійність, що виражено в знакових одиницях (доларах, одиницях продукції, обладнанні, кваліфікації персоналу і т. п.) і зафіксована на папері;

– простота методики прогнозування для розуміння і використання.

2.4. Динамічний (часовий) ряд і тренд

Динамічним рядом (рядом динаміки, часовим рядом) називається послідовність показників, які характеризують зміну явища (процесу, об'єкта) у часі. Окремі спостереження динамічного ряду називаються рівнями.

За часом, відображеним у динамічних рядах, вони поділяються на моментні й інтервальні.

В моментних рядах динаміки рівні виражають величину явища на відповідну дату, наприклад, залишки готової продукції на перше число кожного місяця, вартість основних фондів на початок чи кінець року та ін.

В інтервальних рядах рівні виражають розміри явищ за проміжок часу, наприклад, випуск продукції за місяць, квартал, рік.

При побудові динамічних рядів слід в першу чергу приділити увагу на порівнянність рівнів ряду. Це значить, що усі рівні повинні виражатися в однакових одиницях виміру, розраховуватися по єдиній методології, включати єдине коло об'єктів.

При аналізі тимчасового ряду великої тривалості, наприклад, за ряд років, можна помітити циклічні коливання. Це так званий цикл ділової активності, або економічний цикл, що складається з економічного підйому (буму), спаду, депресії і пожвавлення. Цей цикл повторюється регулярно. Важливо зрозуміти, що будь-який бізнес випробовує вплив процесів, що відбуваються в економіці в цілому, і тому їх потрібно враховувати при складанні довгострокового прогнозу. Проте при короткостроковому прогнозуванні можна нехтувати циклічною складовою ряду.

Важливим компонентом тимчасового ряду є залишкова варіація, яка залишається після того, як інші складові ряду відокремлені. Залишкова варіація може бути двох видів:

- аномальна варіація розглядає неприродне велике відхилення тимчасового ряду, яке надає дію на одиничний нагляд (наприклад, побоювання з приводу зараженості продукту харчування можуть привести до тимчасового скорочення обсягу продажів). Випадкові явища неможливо передбачати наперед, і ми повинні мати на уяві, що подібні події можуть мати місце, і ми повинні також враховувати їх при аналізі діяльності компанії і мати нагоду відповідним чином скоректувати наш прогноз у разі потреби;

- випадкова варіація враховує малі відхилення, які неможливо передбачати і які в довгостроковій перспективі з рівною вірогідністю можуть як

понизити, так і збільшити обсяг продажів. Проте ми не можемо пояснити абсолютно всі варіації обсягу продажів в кожному кварталі, в деяких кварталах обсяг продажів може бути трохи вище, в інших – трохи нижче, ніж очікувалося, причому без видимих причин.

У процесі господарської діяльності промислові підприємства, торговельні структури та сфера послуг постійно стикаються з циклічними коливаннями, обумовленими сезонністю виробництва та попиту на товари і послуги. Такі сезонні коливання можна розглядати як систематичні зміни в динамічному ряді, які виникають через специфічні умови виробництва, кліматичні фактори, культурні та соціальні традиції споживання, а також інші регулярні чинники.

Для підприємств, діяльність яких значною мірою залежить від сезонності, надзвичайно важливим є вивчення сформованих закономірностей цих коливань і побудова прогнозів на найближчу перспективу, передусім на наступний календарний рік. Такий прогноз дозволяє планувати обсяги виробництва, запаси, логістику та маркетингові кампанії, а також ефективно управляти ресурсами і персоналом. У сучасних умовах прогнозування сезонних коливань важливе не лише для економічної ефективності, а й для забезпечення безперервності цифрових та інформаційних процесів, що має значення для кібербезпеки підприємства. Наприклад, прогноз пікових періодів навантаження на онлайн-платформи чи корпоративні системи дозволяє планувати резерви інформаційної інфраструктури та заходи щодо захисту від кібератак під час пікових навантажень.

Аналіз сезонних коливань здійснюється за допомогою спеціальних показників – індексів сезонності, які формують так звану «сезонну хвилю». Використання цих індексів дозволяє виявляти закономірності зміни попиту чи виробництва, а також будувати прогнози на основі рядів динаміки з урахуванням виявлених тенденцій. Для цього застосовуються рівняння тренду, що відображають основну тенденцію зміни часового ряду.

Тренд є узагальненою характеристикою середніх тенденцій зміни явища в часі. Він допомагає визначити загальний напрямок розвитку процесу, не заглиблюючись у деталізацію впливу окремих факторів. Час у цьому випадку виступає як узагальнений показник, через який проявляється дія основних факторів, хоча сам по собі він не є механізмом формування закономірностей. Іншими словами, рівняння тренду відображає накопичену дію головних чинників на динамічний ряд, але без прямого врахування їхніх індивідуальних ефектів.

У практичних дослідженнях переважно розглядаються ймовірнісні (стохастичні) залежності. На відміну від функціональної залежності, де кожному значенню аргументу відповідає єдине значення функції, у стохастичних закономірностях одному й тому ж значенню аргументу може відповідати декілька можливих результатів, що формують певний розподіл ймовірностей. Такі залежності більш характерні для реальних процесів, оскільки неможливо врахувати всі чинники, що впливають на явище.

Рівняння, побудоване на основі стохастичних залежностей, складається з двох складових: детермінованої частини, що відображає вплив відомих і врахованих факторів, та випадкової частини, яка формується під дією неконтрольованих або неврахованих факторів. Наприклад, у сфері кібербезпеки до детермінованих факторів можна віднести заплановані технічні оновлення або впровадження систем контролю доступу, а до випадкових – непередбачувані кібератаки чи збої в роботі мережі через сторонні чинники.

Таким чином, тренд і сезонні індекси дозволяють підприємствам і державним органам не лише планувати виробництво та послуги, але й прогнозувати навантаження на цифрову інфраструктуру та системи кіберзахисту, що особливо важливо для забезпечення стабільної роботи критично важливих ресурсів у будь-який сезон.

Аналітичне вирівнювання тренда є одним із найпоширеніших і водночас найбільш інформативних методів прогнозування, що застосовується для аналізу

динамічних рядів у різних сферах, зокрема в економіці, промисловості та сфері послуг. Суть методу полягає в побудові рівняння, яке відображає основну тенденцію зміни явища у часі, абстрагуючись від випадкових коливань та сезонних відхилень.

Екстраполяція тренда дозволяє робити прогнози лише за умови, що поведінка явища у минулому достатньо точно описується побудованим рівнянням, а фактори, що визначали його динаміку, залишаються відносно стабільними й у майбутньому. Іншими словами, прогноз є обґрунтованим тільки тоді, коли немає очікуваних радикальних змін зовнішніх або внутрішніх умов, які могли б серйозно змінити тенденцію. У випадку виконання цих умов прогноз формується шляхом підстановки у рівняння тренда значень незалежної змінної t, що відповідають обраному горизонту прогнозування, наприклад, наступного року, кварталу або місяця.

Вибір конкретної форми рівняння тренда здійснюється на основі графічного аналізу динамічного ряду. За зовнішнім виглядом графіка можна визначити характер змін показника: чи він демонструє монотонне зростання або спад, наявність точки перегину, циклічність, тренд насичення чи інші закономірності. Цей підхід дозволяє візуально оцінити, яка математична форма рівняння (лінійна, поліноміальна, експоненційна або логістична) найбільш адекватно описує динаміку явища.

При побудові рівняння тренда необхідно вирішити два ключові завдання. По-перше, слід перевірити, чи адекватно воно відображає реальні процеси, включаючи часовий тренд та основні закономірності, що склалися в минулому. По-друге, важливо оцінити відповідність рівняння статистичним критеріям достовірності, таким як коефіцієнт детермінації або середня квадратична помилка. Комплексне рішення цих питань дозволяє переконатися, що рівняння є водночас логічно обґрунтованим і статистично достовірним, а отже – придатним для прогнозування.

У сучасному контексті цей підхід можна ефективно застосовувати й у сфері кібербезпеки. Наприклад, прогнозування трендів у кількості кібератак, обсягах трафіку шкідливого ПЗ або зростанні вразливостей інформаційних систем потребує точного визначення тенденцій минулого. Якщо зміни в цифровій інфраструктурі відбуваються плавно та передбачувано, екстраполяція тренда дозволяє планувати ресурси для захисних заходів, визначати необхідну пропускну здатність мереж, а також розробляти стратегії превентивного реагування на потенційні загрози. Водночас у разі можливих радикальних змін, наприклад, появи нового типу атак або технологічних проривів у сфері штучного інтелекту та криптографії, прогнозування на основі старих трендів може бути недостатньо точним і потребуватиме корекції шляхом моделювання сценаріїв.

Таким чином, аналітичне вирівнювання тренда є не лише інструментом економічного та соціального прогнозування, але й потужним засобом планування ресурсів і протидії ризикам у кіберпросторі, оскільки дозволяє систематизувати закономірності минулого і на їх основі приймати стратегічні рішення щодо майбутнього розвитку як економіки, так і цифрової безпеки.

Перелік запитань для самопідготовки:

1. Що таке прогнозування в економіці?
2. У чому полягає суть прогнозного процесу?
3. Хто або що є об’єктом прогнозування?
4. Який предмет дослідження у прогнозуванні?
5. Чим відрізняється прогноз від плану?
6. Які основні типи прогнозів ви знаєте?
7. У чому різниця між короткостроковими, середньостроковими та довгостроковими прогнозами?
8. Які існують прогнози за рівнем деталізації та сферами застосування?
9. Як класифікуються прогнози за об’єктом і метою?

10. Що таке умовно-нормативні прогнози?
11. Які основні групи методів прогнозування існують?
12. У чому суть експертних методів прогнозування?
13. Як працюють статистичні методи прогнозування?
14. Що таке моделювання і яке місце воно займає у прогнозуванні?
15. Які переваги та недоліки мають різні методи прогнозування?
16. Що таке динамічний (часовий) ряд у прогнозуванні?
17. Як визначається тренд у часовому ряді?
18. Які види трендів розрізняють (лінійний, експоненційний тощо)?
19. У чому суть декомпозиції часових рядів?
20. Як часові ряди використовуються для побудови прогнозів?

Тема 3. Основи кореляційно-регресійного аналізу

3.1. Короткі теоретичні відомості.

3.2. Види залежностей: регресійна залежність.

3.3. Лінійна регресія і метод найменших квадратів.

3.4. Множинна регресія і етапи її побудови.

3.5. Сутність кореляції.

3.6. Лінійна кореляція.

3.7. Коефіцієнт лінійної кореляції.

3.1. Короткі теоретичні відомості

Вивчення кореляційного зв'язку між ознаками починається з регресійного аналізу, який вирішує проблему встановлення форми зв'язку, або виду рівняння регресії, та визначення параметрів рівняння регресії.

В регресійному аналізі розрізняють рівняння парної (простої) та множинної (багатофакторної) регресії. Коли зв'язок із результативною ознакою у здійснюється з одним видом факторної ознаки ***x***, то рівняння регресії має назву рівняння парної регресії. Якщо результативна ознака ***y*** пов'язана з декількома видами факторних ознак x_j (j=1-т), то така залежність має назву рівняння множинної регресії. Обмежимось розглядом рівнянь парної регресії, як найбільш простим випадком зв'язку між ознаками, що достатньо широко використовується в статистичній практиці обстеження економічних явищ.

Найбільш часто для характеристики кореляційного зв'язку між ознаками використовують такі види рівнянь парної регресії, або кореляційних рівнянь:

а) лінійний $Y = a_0 + a_1x$;

б) параболічний $Y = a_0 + a_1x^2$;

в) гіперболічний $Y = a_0 + a_1\ 1/x$;

г) степеневий $Y= a_o x^{a1}$,

де a_0, a_1 - параметри рівнянь регресії, які підлягають визначенню.

Параметри a_j (j=1-m) в рівняннях регресії визначаються методом найменших квадратів, який запропоновано в XVIII ст. французьким математиком Лежандром. Цей метод найкращим чином відповідає кореляційній таблиці і припускає знаходження таких значень параметрів рівняння регресії, при яких сума квадратів відхилень табличних (фактичних) значень результативної ознаки у від теоретичних Y за лінією регресії була б мінімальною:

$$S = \sum(y - Y)^2 = \min \quad (3.1)$$

Функція S параметрів рівняння регресії a_j буде мінімальною тоді, коли виконуються необхідні умови знаходження екстремуму цієї функції - дорівнювання нулю перших похідних функції за параметрами:

$$\partial S : \partial a_0 = 0; \partial S : \partial a_1 = 0 \quad (3.2)$$

Із цих умов визначається система нормальних рівнянь для знаходження параметрів a_0 та a_1.

У випадку лінійного виду рівняння регресій система нормальних рівнянь записується у вигляді:

$$\sum y = na_0 + a_1\sum x$$
$$\sum xy = a_0\sum x + a_1\sum x^2 \quad (3.3)$$

де n – кількість одиниць сукупності (тобто заданих пар значень ***x*** і ***y***).

Розв'язавши цю систему, знаходимо такі значення параметрів:

$$a_0 = (\sum y\sum x^2 - \sum xy\sum x) : (n\sum x - \sum x\sum x) \quad (3.4)$$

$$a_1 = (n\sum xy - \sum x\sum y) : (n\sum x^2 - \sum x\sum x) \quad (3.5)$$

або

$$a_0 = y_{сер} - a_1\, y_{сер};\ a_1 = (x\, y_{сер} - x_{сер}\, y_{сер}) : (x^2{}_{сер} - x_{сер}{}^2) \quad (3.6)$$

де $\boldsymbol{xy_{сер}}$ – середня із добутку факторної ознаки на результативну.

Використавши рівняння регресії, можна знайти теоретичне значення Y для будь-якого значення факторної ознаки ***x***.

У рівнянні параметр a_0 економічного змісту не має, а геометрично він відповідає значенню ординати ліній регресії Y при x=0. Параметр a_1 називається

коефіцієнтом регресії і показує зміну результативної ознаки Y при зміні факторної ознаки *x* на одиницю; геометрично параметр a_1 відповідає куту нахилу (в радіанах) прямої лінії регресії до горизонтальної осі.

Для оцінки впливу факторної ознаки на результативну може розраховуватись *коефіцієнт еластичності* в середньому для усієї сукупності:

$$K_e = a_1 x_{сер} / y_{сер} \quad (3.7)$$

де $x_{сер}$, $y_{сер}$ – середні величини фактичних даних відповідно за факторною та результативною ознаками в цілому для сукупності.

Коефіцієнт еластичності показує, на скільки процентів у середньому зміниться результативна ознака при зміні факторної ознаки на 1%.

Прикладами використання лінійного рівняння регресії є такі залежності: між електроозброєністю праці (х) на 1 робітника і випуском готової продукції (у) для однорідних підприємств; між стажем роботи (х) та виробкою 1 робітника за зміну (у) тощо.

Для вибору виду рівняння регресії необхідно побудувати графік залежності фактичних даних *y=f(x)* і за групуванням точок на графіку встановити візуально, до якого виду (лінійного чи нелінійного) можна віднести лінію регресії.

3.2. Види залежностей: регресійна залежність

Основною метою будь-якого сучасного дослідження є використання його результатів для майбутніх потреб, тобто прогнозування стану об’єкта чи явища, що вивчається. Досліджуючи явище у взаємозв’язку з іншими факторами чи процесами, необхідно виокремити ті з них, що чинять найбільший вплив, а також оцінити його силу та характер. Це означає визначення якісних і кількісних характеристик зв’язку між об’єктом дослідження та чинниками, які впливають на його розвиток.

Надалі «основну» величину, яка досліджується, називатимемо *залежною змінною* і позначатимемо літерою *Y*, інші, впливаючі на *Y*, величини

називатимемо *незалежними змінними* і позначатимемо літерами *x1, x2, ..., xk*. Як *Y*, так і *x1, x2, ..., xk*, вважатимемо числовими.

Розрізняють два види зв'язків.

Якщо значення залежної змінної стає відомим, як тільки відомі значення незалежних змінних, то зв'язок є *динамічним* або функціональним, оскільки в цьому випадку існує закон, за яким обчислюється *Y* залежно від *x1, x2, ..., xk*, *Y = f* (*x1, x2, ..., xk*). Приклади таких зв'язків: закон вільного падіння тіла; закон Ома; закон Бойля - Маріотта; зв'язок між вартістю одиниці товару і ціною, сплаченою за його партію; залежність продуктивності праці і витрат робочого часу.

Інша ситуація спостерігається тоді, коли на основі значень незалежних змінних можна встановити лише певну усереднену тенденцію у зміні залежної змінної. Наприклад, відомо, що між типами кібератак і вразливостями систем існує статистичний зв’язок: для цього створюються спеціальні таблиці та бази даних, які враховують тип системи та рівень захисту, проте вони дають лише орієнтовні середні оцінки ймовірності інцидентів. Подібні залежності прийнято називати кореляційними (від лат. correlatio – співвідношення).

Встановлення математичної форми кореляційного зв’язку здійснюється за допомогою регресійного аналізу. У цьому випадку залежна змінна розглядається як випадкова величина, тоді як незалежні змінні підлягають прямому чи опосередкованому контролю. Кореляційний аналіз досліджує спільний розподіл усіх вимірюваних змінних, а також забезпечує оцінку точності прогнозування одних показників на основі інших.

На відміну від функціонального зв'язку в регресійному аналізі йдеться про встановлення функції регресії *M(y/ x1, x2, ..., xk) = f(x1, x2, ..., xk)* де символ *M (·/·)* позначає математичне сподівання випадкової величини при заданих значеннях незалежних змінних.

Оскільки незалежні змінні *x1, x2, ..., xk* є контрольованими та керованими, а *Y* – випадковою величиною, то за даними експерименту, в

якому *x1, x2, ..., xk* набули конкретних значень, можна судити лише про оцінку параметра, пов'язаного з розподілом *Y*. А оцінок же, як правило, можна побудувати багато.

З точки зору подальших використань бажано мати оцінку якомога простішого вигляду і яка задовольняла б деякий критерій оптимальності (подібний до незміщеності, наприклад, для оцінок параметрів).

Із всіх елементарних функцій (виключаючи константу) найбільш простою є лінійна, цей випадок розглянемо надалі детально, як найбільш прозорий з ідейної точки зору і такий, що в той же час дає можливість для подальших узагальнень.

3.3. Лінійна регресія і метод найменших квадратів

Опишемо спочатку математичну постановку завдання, вважаючи, що вивчається одна залежна змінна *y*, що залежить від однієї незалежної змінної *x* (так зване завдання парної регресії).

Нехай залежність між *x* і *y* має вигляд:

$$y = a_0 + a_1 x_1 + \varepsilon, \tag{3.8}$$

де a_0, a_1 – постійні коефіцієнти, що називаються *параметрами моделі; e* – випадкова величина з математичним сподіванням 0 і дисперсією *s2* .

В цьому випадку рівняння регресії перетворюється на рівняння прямої

$$\bar{y}(x) = M\left(\frac{y}{x}\right) = a_0 + a_1 x \tag{3.9}$$

Передбачимо, що незалежна змінна набула значень *x1, x2, ..., xn*, внаслідок чого залежна змінна набула значень y*1, y2, ..., yn*. У припущенні лінійної залежності отримуємо *n* рівностей:

$$y_i = a_0 + a_1 x_1 + \varepsilon, i = 1...n \tag{3.10}$$

Потрібно за значеннями пар *(xi, yi)* оцінити невідомі a_0, a_1 . Як ми вже знаємо, кожне завдання оцінювання пов'язане з деяким критерієм якості. У теорії, що викладається нами, таким критерієм є критерій найменших квадратів:

$$\sum_{i=1}^{n} \varepsilon_i^2 - min \tag{3.11}$$

Запишемо цю суму інакше, так, щоб була помітна залежність від a_0, a_1:

$$\sum_{i=1}^{n} \varepsilon_i^2 = \sum_{i=1}^{n} [\bar{y}(x) - y_i]^2 = \sum_{i=1}^{n} [y_i - a_0 - a_1 x_i]^2 \tag{3.12}$$

Тепер остаточно приходимо до такої задачі: знайти такі значення невідомих параметрів a_0, a_1 щоб функція

$$Q(a_0, a_1) = \sum_{i=1}^{n} [y_i - a_0 - a_1 x_1]^2 \tag{3.13}$$

набула найменшого значення.

Метод розв'язання цієї задачі відомий з курсу вищої математики. Знаходимо часткові похідні функції Q і прирівнюємо їх до нуля, внаслідок чого приходимо до системи лінійних рівнянь:

$$\begin{cases} \frac{\partial Q}{\partial a_0} = -2 \sum_{i=1}^{n} [y_i - a_0 - a_1 x_i] = 0 \\ \frac{\partial Q}{\partial a_1} = -2 \sum_{i=1}^{n} [y_i - a_0 - a_1 x_i] x_i = 0 \end{cases} \tag{3.14}$$

Після очевидних перетворень отримуємо систему:

$$\begin{cases} n a_0 + a_1 \sum_{i=1}^{n} x_i = \sum_{i=1}^{n} y_i \\ a_0 \sum_{i=1}^{n} x_i + a_1 \sum_{i=1}^{n} x_i^2 = \sum_{i=1}^{n} x_i y_i \end{cases} \tag{3.15}$$

Позначимо вибіркові середні:

$$\bar{x} = \frac{1}{n} \sum_{i=1}^{n} x_i,\ \bar{y} = \frac{1}{n} \sum_{i=1}^{n} y_i,\ \overline{xy} = \frac{1}{n} \sum_{i=1}^{n} x_i y_i,\ \bar{x}^2 = \frac{1}{n} \sum_{i=1}^{n} x_i^2 \tag{3.16}$$

У цих позначеннях після ділення кожного рівняння системи на n вона набуде вигляду:

$$\begin{cases} a_0 + a_1 \bar{x} = \bar{y} \\ a_1 \bar{x} + a_1 \bar{x}^2 = \bar{x}\bar{y} \end{cases}, \tag{3.17}$$

а шукані оцінки коефіцієнтів рівняння регресії будуть такими:

$$\begin{cases} \bar{a}_0 = \frac{\overline{x^2}\bar{y} - \bar{x}\overline{yx}}{\overline{x^2} - (\bar{x})^2} \\ \bar{a}_0 = \frac{\bar{x}\bar{y} - \bar{y}\bar{x}}{\overline{x^2} - (\bar{x})^2} \end{cases}. \tag{3.18}$$

Якщо ввести ще позначення $S_x^2 = \overline{x^2 - \overline{(x)^2}}$ і перетворити вираз $\hat{a}_0$

$$\hat{a}_0 = \frac{\overline{x^2}\bar{y} - \bar{x}\overline{yx} \pm \bar{y}\overline{(x)^2}}{S_x^2} = \frac{S_x^2 \bar{y} - \bar{x}(\bar{x}\bar{y} - \overline{yx})}{S_x^2} = \bar{y} - \hat{a}_1 \bar{x}, \tag{3.19}$$

то оцінка функції регресії набуде такого вигляду:

$$\tilde{\bar{y}}(x) = \widetilde{a_0} + \widetilde{a_1}x = \bar{y} - \widetilde{a_1}\bar{x} + \tilde{a}x = \bar{y} + \widetilde{a_1}(x - \bar{x}). \qquad (3.20)$$

3.4. Множинна регресія і етапи її побудови

Регресія називається множинною, якщо вона відображає залежність результативної ознаки від декількох чинників.

Побудова моделей множинної регресії включає декілька етапів:

– вибір форми зв'язку (рівняння регресії);

– відбір факторних ознак;

– забезпечення достатнього об'єму сукупності для отримання незміщених оцінок.

Розглянемо детальніше кожен із зазначених аспектів. Дані, що збираються для проведення регресійного аналізу, здебільшого мають характер “історичних” спостережень, тобто відображають значення певних чинників у попередні періоди часу або в різних географічних регіонах. Саме ці дані слугують основою для розрахунку оцінок коефіцієнтів регресії та для визначення рівня адекватності побудованої моделі фактичним змінам результативної ознаки.

Задля досягнення поставлених цілей можуть застосовуватися різні методики, проте найбільш універсальним і водночас найпоширенішим підходом є метод найменших квадратів.

Вибір форми залежності між змінними часто становить складність, оскільки математичний апарат дозволяє виражати зв’язки між ознаками великою кількістю різних функцій. Для будь-якого з можливих варіантів зв’язку може бути запропоновано низку рівнянь, які в тій чи іншій мірі описуватимуть наявні співвідношення.

Найбільш практичним підходом до визначення оптимальної форми вихідного рівняння регресії є метод послідовного перебору різних рівнянь із подальшою оцінкою їх відповідності фактичним даним

Суть даного методу полягає в тому, що велике число рівнянь (моделей) регресії, відібраних для описування зв'язків якого-небудь соціально-економічного явища або процесу, реалізується на ЕОМ за допомогою спеціально розробленого алгоритму перебору з подальшою статистичною перевіркою, головним чином на основі *t*-критерію Стьюдента і *F*-критерію Фішера. Хоча з іншого боку, спосіб перебору є досить трудомістким і пов'язаний з великим об'ємом обчислювальних робіт.

Важливим етапом у процесі побудови рівняння множинної регресії після його визначення є відбір та включення факторних змінних. Основна складність полягає в тому, що більшість потенційних факторів виявляють взаємозалежність між собою. Одним із головних завдань при формуванні такої моделі є визначення оптимальної кількості змінних, тобто встановлення раціональної розмірності рівняння. Збільшення кількості факторних ознак, з одного боку, дає змогу точніше відобразити досліджуване явище, а з іншого – робить рівняння занадто громіздким, що ускладнює його практичне використання та потребує значних обчислювальних ресурсів. Тому актуальним підходом є скорочення розмірності шляхом виключення другорядних, економічно та статистично несуттєвих факторів. Це дозволяє спростити модель і зробити її зручнішою у застосуванні. Водночас надмірне зменшення кількості змінних може призвести до надто спрощеної моделі, яка вже неадекватно відображатиме досліджувані процеси.

Одним із найбільш поширених методів відбору факторних ознак є крокова регресія (кроковий регресійний аналіз). Її суть полягає у поступовому введенні факторів до рівняння з подальшою перевіркою їхньої статистичної значущості. При використанні «прямого методу» змінні додаються послідовно, а після кожного кроку оцінюється, наскільки зменшується сума квадратів

залишків і підвищується значення множинного коефіцієнта кореляції. Паралельно застосовується і «зворотний метод», який передбачає поетапне виключення змінних, що втрачають значущість за результатами перевірки з використанням t-критерію Стьюдента. Фактор вважається незначущим у разі, якщо його включення лише впливає на значення коефіцієнтів регресії, але не сприяє зниженню суми квадратів залишків і не покращує якість моделі. Навпаки, якщо при введенні змінної коефіцієнт множинної кореляції зростає, а коефіцієнт регресії залишається стабільним або змінюється несуттєво, така ознака розглядається як вагома і обов'язково враховується у рівнянні.

У тих випадках, коли включення факторної змінної призводить не лише до зміни значень, а й до зміни знаків коефіцієнтів регресії, при цьому множинний коефіцієнт кореляції не зростає, такий фактор визнається недоцільним для подальшого використання у моделі.

Розглянемо процес побудови моделей множинної регресії на прикладі оцінки ризиків кіберзагроз. Відомо, що ймовірність успішної атаки на систему залежить не лише від одного фактора, наприклад, наявності вразливостей, а й від багатьох інших чинників, таких як конфігурація мережі, рівень підготовки персоналу та використання захисного ПЗ. Оскільки більшість завдань кібербезпеки потребує одночасного врахування декількох факторів, у таких випадках доцільно застосовувати моделі множинної регресії для прогнозування ризиків.

Методологічні підходи, що використовуються у множинному регресійному аналізі, загалом ґрунтуються на тих самих принципах, що й простий регресійний аналіз. Проте їхня відмінність полягає в модифікаціях та доповненнях, які зумовлені необхідністю врахування взаємодії між кількома факторами одночасно.

Разом із тим побудова множинної регресії супроводжується низкою додаткових труднощів. Найсуттєвіша з них полягає в тому, що для виявлення взаємозв'язків між змінними вже недостатньо використати просту діаграму

розсіювання. Вона придатна лише для наочного відображення впливу кожного незалежного чинника окремо на результативну змінну. Однак така інформація є обмеженою, адже незалежні змінні здатні суттєво взаємодіяти між собою і впливати одна на одну майже так само, як і на залежну змінну. Розглянемо етапи проведення регресійного аналізу:

– побудова системи факторів, що істотно впливають на результативну ознаку;

– розроблення моделі, яка відображає загальний зміст взаємозв'язків, що вивчаються, і кількісне оцінення її параметрів;

– перевірка якості моделі;

– оцінення впливу окремих факторів.

На початковому етапі побудови моделі проводиться відбір факторів, що справляють істотний вплив на результативну змінну. Такий відбір зазвичай ґрунтується на змістовному аналізі. Водночас, аби забезпечити надійність отриманих оцінок, не рекомендується включати занадто велику кількість змінних: їх число має становити не більше третини від загального обсягу використаних даних.

Однак на початку дослідження дослідник зазвичай не має чіткої впевненості щодо того, які саме фактори є найвагомішими. Тому при використанні обчислювальних методів добір змінних часто здійснюється безпосередньо в процесі формування моделі за допомогою методу послідовної регресії. Його сутність полягає у поетапному введенні нових факторів та оцінюванні внеску кожного з них.

Існує й альтернативний підхід: на стартовій стадії до попередньої моделі включається широкий перелік потенційних факторів без істотних обмежень, а вже надалі здійснюється їх ретельне оцінювання та остаточний відбір. Другий етап починається з розробки моделі, що відображає загальний зміст взаємозв'язків, що вивчаються.

Регресійна модель – це рівняння (або система рівнянь), що показує, які фактори, на думку дослідника, підлягають аналізу з точки зору взаємозв'язків.

Регресійне рівняння дає також уявлення про форму зв'язку. Класифікація таких рівнянь подана на рис. 3.1.

Регресія називається парною, якщо вона відображає залежність між результативною ознакою й одним фактором.

Регресія називається множинною, якщо вона відображає залежність результативної ознаки від декількох факторів.

Якщо залежності лінійні щодо параметрів (але не обов'язково лінійні щодо незалежних змінних), то регресія називається лінійною. У іншому випадку регресію називають нелінійною.

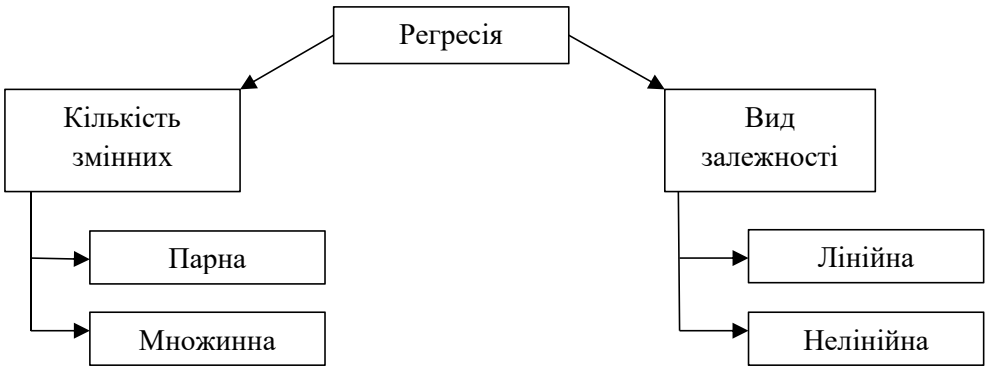


Рисунок 3.1 – Класифікація регресійної залежності

Після відбору змінних і визначення способу їх подання наступним кроком є встановлення форми рівняння регресії. Вибір аналітичної форми залежності здійснюється на основі певних математичних функцій, серед яких найчастіше застосовуються лінійна, логарифмічна, степенева, експоненціальна та поліноміальна.

Для обчислення параметрів цих функцій використовуються спеціалізовані програмні засоби. Найбільш зручними для інтерпретації результатів є програми, орієнтовані на лінійний регресійний аналіз, тому на практиці більшість дослідників надає перевагу саме цьому варіанту.

Дані, що застосовуються у регресійному аналізі, зазвичай мають форму «історичних зведень». Вони подаються у вигляді числових значень факторів за

попередні періоди часу або для різних територіальних одиниць. На основі цих даних визначаються оцінки коефіцієнтів регресії та перевіряється, наскільки побудована модель адекватно відображає зміни результативної ознаки.

Для досягнення поставлених цілей можуть використовуватися різні методи оцінювання параметрів, однак найбільш універсальним і водночас найчастіше застосовуваним є метод найменших квадратів. Оцінки за методом найменших квадратів – це ті величини коефіцієнтів регресійного рівняння, що мінімізують суму квадратів відхилень значень результативної ознаки, що спостерігаються дійсно, (*Yi*) від тих значень, що одержують з рівняння *Y(Xi*).

Для оцінювання якості побудованої моделі та достатності обраного набору пояснювальних факторів зазвичай використовується коефіцієнт детермінації *R2*– це відношення дисперсії, що пояснюються регресійним аналізом, до загальної дисперсії.

Коефіцієнт детермінації дає кількісну оцінку міри аналізованого зв'язку. Він показує частину варіації результативної ознаки, яка перебуває під впливом факторів, що вивчаються, тобто визначає, яка частка варіації ознаки *Y* враховується в моделі й обумовлена впливом на неї незалежних факторів.

Чим ближче *R2* до 1, тим точніше рівняння регресії пояснює фактор, що вивчається (при функціональному зв'язку *R2* дорівнює 1, а через відсутність зв'язку – 0).

3.5. Сутність кореляції

Важливим завданням описової статистики є виявлення зв'язку між змінними, оцінювання його напряму та інтенсивності. Порівнюючи різні види зв'язків, можна виділити три типи залежностей між змінними X і Y:

1) функціональна залежність визначає значення змінної Y від X однозначно;

2) кореляційна залежність визначає середнє значення змінної Y від X;

3) стохастична залежність визначає розподіл змінної Y від X.

Найбільш загальною вважається стохастична залежність. Кореляційна залежність є залежністю стохастичною, функціональна – розглядається як окремий випадок кореляційної залежності.

Кореляція (від лат. correlatio - співвідношення) – це статистична залежність між випадковими величинами, що носить ймовірнісний характер.

Кореляційна залежність між явищами чи величинами (параметрами) відрізняється від функціональної тим, що вона не має суворо визначеного характеру. У такій залежності зміна однієї або кількох змінних супроводжується систематичними змінами іншої чи інших змінних. Кількісним показником тісноти цього зв'язку виступає коефіцієнт кореляції.

Оцінити кореляційний зв'язок можна також якісно, використовуючи графічний підхід. Для цього будують кореляційне поле – множину точок, нанесених у систему координат, де по осі абсцис відкладаються значення факторної змінної x, а по осі ординат – результативної змінної y. Таке поле дає змогу візуально простежити наявність та характер (зокрема напрям) зв'язку між двома ознаками. Розташування точок на площині відображає ступінь і напрям залежності, що дозволяє зробити попередні висновки про взаємозв'язок між досліджуваними величинами.

Кореляційна залежність між двома випадковими величинами X та Y може бути:

1. Позитивна кореляція – при збільшені одного параметра (X) другий (Y) також збільшується.
2. Негативна кореляція – при збільшені одного параметра (X) другий (Y) зменшується.
3. Нульова кореляція – взаємозв'язок між випадковими величинами в даних умовах відсутній.
4. Лінійна кореляція – якщо обидві функції $M(Y)_x = f(x)$ та $M(X)_y = g(y)$ описуються прямими лініями регресії.

5. Нелінійна кореляція – якщо обидві функції $M(Y)_x = f(x)$ та $M(X)_y = g(y)$, або одна із них не описується прямими лініями регресії.

3.6. Лінійна кореляція

Кореляційний зв’язок між величинами X та Y необхідно оцінювати чисельно, оскільки за допомогою кореляційного поля неможливо чітко охарактеризувати (розрізнити) кореляцію (рис. 3.2 А і рис. 3.2 Б).

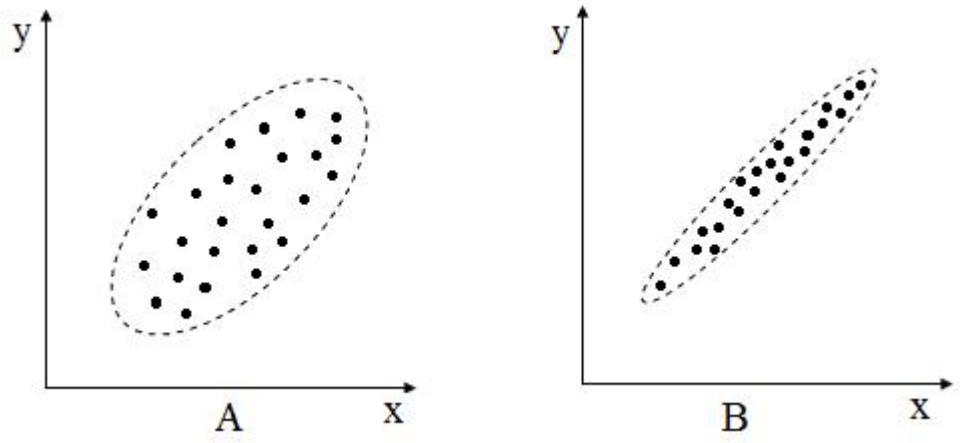


Рисунок 3.2 – Приклади кореляційного поля

Лінійний кореляційний зв'язок для емпіричних даних, виміряних за шкалою інтервалів або відношень, оцінюється за допомогою коефіцієнта кореляції Пірсона r_{xy}.

3.7. Коефіцієнт лінійної кореляції

Обчислюють коефіцієнт лінійної кореляції за формулою:

$$r_{xy} = \frac{\sum_{i=1}^{n}(x_i - \bar{x}) \cdot (y_i - \bar{y})}{\sqrt{\sum_{i=1}^{n}(x_i - \bar{x})^2 \cdot \sum_{i=1}^{n}(y_i - \bar{y})^2}} \qquad (3.21)$$

де n – об’єм вибірки; $\bar{x} = \frac{\sum_{i=1}^{n} x_i}{n}$, $\bar{y} = \frac{\sum_{i=1}^{n} y_i}{n}$ - середні значення параметрів X та Y.

Коефіцієнту кореляції притаманні такі властивості:

1. Коефіцієнт кореляції може набувати значення від -1 до +1, тобто $-1 \le r \le 1$.

2. Знак (+) вказує на позитивну кореляцію.

3. Знак (-) вказує на негативну кореляцію.

4. При $|r| = 1$ – маємо функціональний зв'язок між параметрами (ознаками).

5. При $r = 0$ – кореляція в даних умовах між ознаками відсутня.

6. Коефіцієнт кореляції вказує на ступінь спряженості (сили взаємозв'язку) між значеннями досліджуваних ознак:

$\Rightarrow$ $|r| < 0,|$ – зв'язок слабкий;

$\Rightarrow$ $0,3 \le |r| \le 0,5$ – зв'язок помірний;

$\Rightarrow$ $0,5 < |r| \le 0,7$ – зв'язок значний;

$\Rightarrow$ $0,7 < |r| \le 0,9$ – зв'язок сильний;

$\Rightarrow$ $|r| > 0,9$ – дуже сильний, наближається до функціонального зв'язку.

Коефіцієнт кореляції є випадковою величиною, оскільки вираховується із випадкових величин. Тому потрібно знати значущість (вірогідну оцінку) обчисленого коефіцієнта кореляції. Для вірогідної оцінки коефіцієнта кореляції найчастіше використовують метод – Z, який був запропонований Фішером. Емпіричним критерієм достовірності служить тестова статистика t_z, яка обчислюється за формулою:

$$t_z = \left[0,5 \cdot \ln\left(\frac{1+r}{1-r}\right) - \frac{|r|}{2(n-1)} \right] \cdot \sqrt{n-3} \qquad (3.22)$$

Отримане значення t_z порівнюють із стандартним значенням t_{st} (p = 0,05, k = n-2), яке знаходять за таблицею Стюдента. При $t_z > t_{st}$, нульова гіпотеза відхиляється, рівень значущості обчисленого r_{xy} є достатнім. При $t_z \le t_{st}$ нульова гіпотеза не відхиляється, рівень значущості обчисленого r_{xy} є недостатнім.

Статистична недостовірність емпіричного коефіцієнта кореляції ще не доказує, що зв'язок між варіюючими ознаками відсутній. При достатньому числі спостережень такий зв'язок може виявитись достовірним.

Розрахувати необхідний об'єм вибірки *n* для запланованої значущості коефіцієнта кореляції можна за формулою:

$$n = \frac{t_z^2}{Z^2} + 3 \qquad (3.23)$$

де $Z = 0{,}5 \ln \frac{1+r}{1-r}$; r – обчислений з недостатньою значущістю коефіцієнт кореляції; t_z – табличне значення коефіцієнта Стюдента при заданих *(p, k = ∞);* при $p = 0{,}095$ *і* $k = \infty$, $t_z = 1{,}96$.

Дослідження кореляційної залежності між ознаками X та Y може проводитись в різний час, різними лаборантами або в різних лабораторіях. У таких випадках ми будемо мати емпіричні коефіцієнти кореляції, обчислені на незалежних вибірках. Як випадкові величини, значення цих коефіцієнтів будуть відрізнятися $r_1 \neq r_2$. Важливим є знати наскільки ця різниця є достовірною.

Критерій значущості t_D для оцінки різниці між коефіцієнтами кореляції також обчислюють методом Z. Критерієм достовірності служить відношення різниці $Z_1 - Z_2$ до її похибки m_{D_z} :

$$t_D = \frac{Z_1 - Z_2}{m_{D_z}} \qquad (3.24)$$

де $Z_1 = 0{,}5 \ln \frac{1+r_1}{1-r_1}$; $Z_2 = 0{,}5 \ln \frac{1+r_2}{1-r_2}$; $m_{D_z} = \sqrt{\frac{1}{n_1 - 3} + \frac{1}{n_2 - 3}}$.

При $t_D < t_{st}$ $(p,\ k = \infty)$ нульова гіпотеза не відхиляється і різниця між коефіцієнтами кореляції є незначущою, тобто носить випадковий характер.

Перелік запитань для самопідготовки:

1. Що таке кореляційно-регресійний аналіз?
2. Які задачі вирішує кореляційно-регресійний аналіз в економіці?

3. Які статистичні поняття лежать в основі цього аналізу?
4. У чому суть регресійної залежності між змінними?
5. Які типи регресійних залежностей існують (лінійна, нелінійна)?
6. Чим регресійна залежність відрізняється від функціональної?
7. Що таке лінійна регресія?
8. Як формується рівняння лінійної регресії?
9. У чому полягає метод найменших квадратів (МНК)?
10. Що таке множинна регресія?
11. Які змінні використовуються у множинній регресії?
12. Які основні етапи побудови моделі множинної регресії?
13. Як оцінюється якість множинної регресійної моделі?
14. Що таке кореляція?
15. Як пов’язана кореляція з регресією?
16. Що означає сила та напрямок кореляційного зв’язку?
17. У чому полягає суть лінійної кореляції?
18. Як визначити наявність лінійної кореляції між змінними?
19. Які приклади економічних показників мають лінійну кореляцію?
20. Що таке коефіцієнт кореляції Пірсона?
21. Як інтерпретується значення коефіцієнта (від –1 до +1)?
22. Яке значення коефіцієнта свідчить про сильний, слабкий або відсутній зв’язок?
23. Як перевіряється статистична значущість коефіцієнта кореляції?
24. Які функції Excel використовуються для обчислення коефіцієнтів кореляції?
25. Як інтерпретувати результати кореляційного аналізу в Excel?

Тема 4. Евристичні методи обґрунтування й прийняття рішень в прогнозуванні

4.1. Індивідуальне і групове прийняття рішень.

4.2. Методи групової роботи.

4.1. Індивідуальне і групове прийняття рішень

Евристика у контексті кібербезпеки розглядається як сукупність спеціалізованих методів і прийомів аналізу, спрямованих на виявлення нових способів протидії кіберзагрозам та розв'язання нетипових інцидентів інформаційної безпеки. Вона охоплює як технології формування запитів і пошуку рішень у процесі аналізу загроз, так і навчальні підходи, засновані на використанні навідних запитань під час підготовки фахівців з кіберзахисту. У ширшому розумінні евристика виступає теоретичною основою методів, що сприяють розвитку аналітичного й критичного мислення спеціалістів у сфері кібербезпеки.

Основу евристичних підходів у кібербезпеці становить система принципів, вимог і правил, які визначають загальну стратегію та тактичні дії окремого експерта або команди під час реагування на складні, слабо формалізовані чи принципово нові кіберінциденти. Застосування таких підходів дозволяє активізувати творчий потенціал аналітиків, стимулює генерування інноваційних рішень і підвищує ефективність управлінських рішень у сфері інформаційної та кібернетичної безпеки.

Перш ніж перейти до аналізу конкретних евристичних методів, які найчастіше використовуються в управлінні кібербезпекою, доцільно розглянути особливості прийняття рішень як окремими фахівцями з безпеки, так і колективами спеціалізованих підрозділів.

Система управління кібербезпекою передбачає ухвалення рішень різного рівня складності та значущості. Вищий рівень керівництва організації приймає стратегічні рішення, пов'язані з формуванням політики інформаційної безпеки,

визначенням пріоритетних напрямів захисту інформаційних ресурсів, вибором архітектури систем безпеки та доцільністю впровадження нових захисних технологій. Менеджери середньої та операційної ланок зосереджуються на тактичних і поточних завданнях, зокрема на плануванні заходів реагування на інциденти, управлінні персоналом служби кібербезпеки та розподілі відповідальності між фахівцями. Співробітники, які не виконують управлінських функцій, приймають рішення в межах своїх професійних обов'язків, пов'язаних із моніторингом, аналізом та усуненням загроз.

Суб'єктами прийняття рішень у сфері кібербезпеки є як окремі експерти, так і колективні органи – команди реагування на інциденти, аналітичні центри або міждисциплінарні робочі групи. Відповідно, рішення можуть бути індивідуальними або груповими, кожен із цих підходів має свої переваги та обмеження.

Індивідуальні рішення в кібербезпеці часто вирізняються високим рівнем креативності та швидкістю ухвалення, що є критично важливим під час реагування на інциденти в реальному часі. Такі рішення можуть включати нестандартні методи виявлення атак або інноваційні способи нейтралізації загроз. Водночас вони пов'язані з підвищеним ризиком помилок, оскільки ґрунтуються на обмеженому досвіді та знаннях однієї особи, що є суттєвим недоліком в умовах складних багатовекторних кібератак.

Саме складність сучасних кіберзагроз, які потребують комплексного технічного, організаційного та аналітичного підходу, зумовлює поширення групових форм прийняття рішень. Для цього створюються спеціалізовані команди, що об'єднують експертів з різних напрямів – мережевої безпеки, криптографії, аналізу шкідливого програмного забезпечення та управління ризиками.

Колективне прийняття рішень у сфері кібербезпеки забезпечує ширшу інформаційну базу, оскільки учасники команди обмінюються досвідом, знаннями та аналітичними напрацюваннями. Це дозволяє сформувати

комплексне бачення кіберінциденту та розробити декілька альтернативних сценаріїв реагування. Крім того, групові рішення сприяють підвищенню відповідальності та довіри до обраних заходів захисту, а також краще сприймаються персоналом організації як результат спільної та узгодженої роботи. Завдяки механізмам колективного аналізу, критичного обговорення та взаємного коригування пропозицій такі рішення є більш обґрунтованими та менш вразливими до помилок.

Таким чином, як індивідуальні, так і групові рішення відіграють важливу роль у системі управління кібербезпекою. Однак сучасна практика дедалі частіше орієнтується на колективні підходи, які забезпечують більшу гнучкість, врахування різних професійних точок зору та ефективне реагування на динамічні й складні кіберзагрози.

Водночас колективне ухвалення рішень у сфері кібербезпеки має низку обмежень. Передусім воно характеризується підвищеними часовими витратами, оскільки потребує формування спеціалізованої команди, доведення до її членів інформації про інцидент або загрозу та створення умов для результативної координації дій між фахівцями. Процес узгодження різних технічних і організаційних підходів до аналізу кібератаки чи вибору механізмів захисту нерідко є тривалим, тоді як у кібербезпеці фактор часу має критичне значення, особливо під час реагування на інциденти в режимі реального часу. Зі збільшенням кількості учасників команди ускладнюється управління взаємодією, що додатково подовжує процес формування остаточного рішення. У зв’язку з цим у практиці управління кіберінцидентами оптимальною вважається команда чисельністю п’ять–сім осіб.

Ще одним недоліком групових рішень у сфері кібербезпеки є ризик домінування позиції більшості або впливу авторитетних експертів і керівників, що може призводити до придушення альтернативних, потенційно інноваційних ідей інших учасників. У таких умовах знижується ефективність евристичного пошуку та обмежується варіативність запропонованих сценаріїв реагування.

Додатковою проблемою є розмитість відповідальності, оскільки у випадку колективного ухвалення рішень складно однозначно визначити суб’єкта, відповідального за помилки або неефективність обраних заходів кіберзахисту.

4.2. Методи групової роботи

Метод «мозкової атаки». Іноді виникають ситуації, коли особам, що приймають рішення досить складно знайти будь-який варіант рішення проблеми. В цьому випадку можливе застосування методу "мозкової атаки", який дозволяє виявити і співставити індивідуальні судження, спектр ідей стосовно розв'язання проблем, а потім прийняття відповідного рішення.

Метод мозкової атаки у сфері кібербезпеки застосовується як колективний евристичний інструмент генерування ідей під час аналізу нових або складних кіберзагроз, розроблення сценаріїв реагування на інциденти та пошуку нетипових способів захисту інформаційних систем. Його сутність полягає у вільному висловленні учасниками команди максимального числа пропозицій щодо виявлення, локалізації чи нейтралізації кібератак без їх негайної критичної оцінки, що дозволяє активізувати творчий потенціал фахівців різних напрямів – аналітиків SOC, інженерів з безпеки, спеціалістів з ризик-менеджменту та управління. У подальшому згенеровані ідеї підлягають систематизації, відбору та техніко-економічному аналізу, в результаті чого формується обґрунтований набір рішень, здатних підвищити ефективність системи кіберзахисту та адаптивність організації до динамічних загроз.

Існує декілька різновидів методів "мозкових атак" (ММА): пряма "мозкова атака", зворотна, подвійна атака, "конкуренція ідей" та ін. Розглянемо найбільш поширені з них, тобто перший та другий різновиди.

Пряма «мозкова атака» – метод колективного генерування ідей стосовно рішення творчої задачі. Її мета – відбір ідей.

Правила для учасників:

1. Учасники сідають за стіл обличчям одне до одного.

2. Не дозволяються суперечки, критика, оцінки того, що пропонується.

3. Час виступу учасника – 1-2 хвилини.

4. Пропонуються будь-які ідеї, навіть марні.

5. Кількість ідей важливіша за їхню якість. Рекомендації:

• ідеї слід схвалювати;

• необхідно вірити в те, що проблему можна розв'язати;

* дозволяється ставити запитання, що мають розвивати ідею;

* слід прагнути вирішення висунутої проблеми;

* всі учасники є рівноправними;

• не слід думати про наслідки того, що буде сказано;

• групі не висувається якась конкретна задача, а характеризується проблема взагалі;

• не слід робити жестів, що відвертають увагу від розв'язання проблеми. Керівник повинен:

• спрямовувати хід дискусії, ставити запитання, що стимулюють до вирішення завдання;

• підказувати, використовувати жарти, репліки, які створюють неформальну обстановку. Обмеження та умови:

• кількість учасників – 4-15;

• бажаним є різний рівень їхньої освіти та різна спеціалізація;

• необхідне дотримання балансу рівня активності, темпераменту;

• час роботи – від 15 хвилин до 1 години. Ідеї відбирають фахівці – експерти, які формують оцінку в два етапи. Спочатку відбираються найбільш оригінальні й раціональні ідеї, а потім – оптимальні, з урахуванням завдань і цілей її вирішення.

Зворотна «мозкова атака» припускає замість генерації нових ідей критику наявних.

Правила учасників:

• критикується, обговорюється і оцінюється кожна ідея за такими критеріями: відповідності початковим вимогам, можливості її реалізації або відсутність такої; реалізації за затратами, застосування в іншій сфері;

• критика викладається лаконічно, позитивно. Ідеї, що потребують тривалого часу для обговорення, розглядаються пізніше;

• виступати кожному слід кілька разів, але краще – по колу;

• тривалість виступу – 15-20 хвилин;

• бажано проводити захід у першій половині дня, у спокійній обстановці. Зворотна "мозкова атака" може бути проведена одразу після прямої, коли після колективного генерування ідей формуються контрідеї. При цьому відбувається:

• систематизація і класифікація ідей та групування за ознаками, які виражають загальні підходи до вирішення проблеми;

оцінка ступеня реалізації ідей і пошук перешкод, що протидіють цьому;

• оцінка критичних зауважень і відбір ідеї та контрідеї, які витримують критику. Доцільно всіх учасників поділити на групи:

• генераторів ідей (коли проводиться пряма "мозкова атака");

• аналізу проблемної ситуації і оцінки ідей;

• генераторів контрідей.

Цей метод базується на принципах:

• співтворчості у процесі пошуку рішення;

• довіри до творчих можливостей і здібностей один одного;

• оптимального поєднання інтуїтивного та логічного. Переваги методу:

• можливість використання всіх фахівців, що є в апараті управління;

• удосконалення соціально-психологічних внутрішньогрупових процесів. Недоліки методу:

• дозволяє знайти ідею в загальному вигляді, при цьому не гарантує її ретельної розробки;

• не може застосовуватися при вирішенні проблеми, яка потребує великих розрахунків;

• потребує високої підготовленості керівника, який має володіти навичками організації мислетехнічних, психотехнічних процесів у групі;

• не завжди вдається подолати інерцію мислення (внаслідок закону інерції).

Метод ключових запитань доцільно застосовувати як для пошуку додаткової інформації в умовах виникнення проблемної ситуації, так і для впорядкування вже наявних даних, необхідних для розв'язання управлінського завдання. Суть підходу полягає у використанні системи запитань, які виступають своєрідними інтелектуальними «стимулами». Вони допомагають сформувати стратегію та тактику вирішення проблеми, активізують інтуїтивне мислення, сприяють побудові логічних алгоритмів та підводять людину до знаходження оптимального рішення.

Метод ключових запитань у сфері кібербезпеки використовується як структурований евристичний підхід, спрямований на глибокий аналіз кіберзагроз, інцидентів та вразливостей інформаційних систем шляхом послідовного формулювання спеціально підібраних запитань. Його застосування дозволяє системно дослідити причини виникнення кібератак, можливі вектори проникнення, потенційні наслідки для критичних активів і ефективність наявних заходів захисту. Завдяки зосередженню уваги на запитаннях типу «що сталося», «чому це стало можливим», «які компоненти є найбільш уразливими» та «які альтернативні дії можуть зменшити ризики», метод ключових запитань сприяє розвитку аналітичного мислення фахівців, підвищує обґрунтованість управлінських рішень і забезпечує комплексний підхід до управління кібербезпекою.

Постановка правильно сформульованих запитань здатна не лише структурувати процес мислення, а й скерувати його у напрямі пошуку нових ідей чи уточнення важливих деталей. Історично відомо, що ще в Стародавньому Римі державним діячам радили для отримання максимально повної та достовірної інформації про подію ставити собі низку базових

запитань і намагатися дати на них відповіді. Цей перелік охоплював такі ключові питання: хто? що? з якою метою? де? чому? яким чином? коли?

Таким чином, метод ключових запитань є універсальним інструментом, який допомагає як у зборі даних, так і в організації процесу мислення, роблячи його більш системним і результативним.

Принципові вимоги до використання методу:

• Проблемність і оптимальність. Вправно поставленими запитаннями необхідно знижувати проблемність задачі до оптимального рівня або знизити невизначеність проблеми.

• Дроблення інформації. За допомогою запитань намагатись розділити проблему на підпроблеми.

• Цілеполягання. Кожне нове запитання повинне формувати стратегію, модель розв'язання проблеми.

Рекомендації для керівника:

1) запитання повинно стимулювати думку, а не підказувати ідею вирішення проблеми;

2) запитання повинні містити мінімальну інформацію;

3) якщо ставиться серія запитань, то необхідно:

* знижувати рівень проблемності задачі;

• слід ставити запитання, які логічно пов'язані між собою, цікаві за формою, спонукаючи до виникнення несподіваних поглядів на проблему, що розв'язується;

4) необхідно стимулювати як емпіричне, аналітичне (аксіоматичне), так і діалектичне (творче) мислення;

5) проблему слід розділити на підпроблеми, етапи рішення. Рекомендації для членів групи (апарату управління):

1) фіксувати найбільш характерні питання й намагатись систематизувати їх;

2) ставити перед собою запитання:

• які спрощують проблему;

• які дозволяють осмислити її з нового несподіваного боку;

• які стимулюють використання засвоєних знань навичок, що вироблені;

• які розділяють проблему на підпроблеми;

• які спонукають до самоорганізації, самоконтролю. Рекомендується поетапна постановка питань, яка має стимулювати розв'язання проблеми, підвищуючи її визначеність.

Послідовність дій: що невідомо? що дано?

В чому полягають умови? Чи є можливим задовольнити умовам?

Чи є достатніми умови для визначення невідомого (або ж ні, або ж надмірно, або суперечливо)?

Зробити креслення (рисунок).

Ввести відповідні позначення.

Розділити умову на частини. Знаходження ідеї і розробка плану рішення: Як знайти зв'язок між тим, що дано і невідомими даними?. Чи відоме будь-яке споріднене завдання? Чи можна ним скористатися? Чи можна використати метод її вирішення?

Чи не слід ввести будь-який допоміжний елемент, щоб скористатися колишнім завданням?

Чи можливо сформулювати завдання інакше, простіше? Чи можливо придумати доступніше завдання? Більш загальне? Більш приватне? Аналогічне?

Чи не можна вирішити частину завдання, задовольнити частини умови? Чи не можна витягнути що-небудь корисне з даних? Чи всі дані і умови використані? Чи взяті до уваги всі поняття, що містяться у завданні?

Здійснення плану:

• Реалізовуючи план рішення, контролюйте свої кроки.

• Чи відомо вам, що сформований план правильний?

• Чи зумієте ви довести, що він правильний? Контроль і самоконтроль одержаного рішення:

• Чи можна перевірити результат?

• Чи можна перевірити хід рішення?

• Чи можна отримати той самий результат інакше?

• Чи можна перевірити правильність отриманого результату?

• Чи можна в будь-якому іншому завданні використати отриманий результат? Чи можна вирішити завдання, зворотне цьому?

Метод вільних асоціацій ґрунтується на спостереженні, що у процесі генерації ідей використання нових асоціативних зв'язків суттєво підвищує результативність творчої діяльності. Це пояснюється тим, що поява асоціацій створює можливість для встановлення несподіваних і нестандартних зв'язків між елементами проблеми, яку необхідно вирішити, та попереднім життєвим чи професійним досвідом учасників групи. Таким чином, метод спирається на природні особливості роботи людського мозку, який здатний продукувати оригінальні ідеї внаслідок утворення нових асоціативних комбінацій.

Метод вільних асоціацій у сфері кібербезпеки застосовується як евристичний підхід, що сприяє пошуку нестандартних рішень під час аналізу нових типів кіберзагроз, прогнозування можливих сценаріїв атак та розроблення інноваційних заходів захисту інформаційних систем. Його сутність полягає у вільному породженні асоціативних зв'язків між, на перший погляд, непов'язаними поняттями, технологіями або подіями, що дозволяє фахівцям виходити за межі стандартних шаблонів мислення. У процесі такого асоціативного аналізу експерти можуть знаходити неочевидні вектори атак, приховані взаємозв'язки між компонентами інфраструктури або альтернативні способи реагування на інциденти. Застосування методу вільних асоціацій активізує креативне мислення команди з кібербезпеки та підвищує її здатність адаптуватися до динамічних і непередбачуваних кіберзагроз.

Ефективність цього методу полягає в тому, що він дозволяє подолати інерційність мислення та вийти за межі стандартних логічних схем. Завдяки використанню вільних асоціацій відкривається широкий простір для творчих

пошуків, а отримані результати нерідко виявляються несподіваними й водночас надзвичайно корисними для розробки оптимального управлінського рішення.

Рекомендації для керівника:

• не поспішати вирішувати проблему, а спробувати неодноразово її переформулювати, поглянути на неї з іншого боку;

• пропонувати слово, поняття, спробувати "викликати" якийсь образ, який може стати стимулом для утворення несподіваних вільних асоціацій, що сприяють виникненню ідеї розв'язання проблеми;

• на початкових етапах генерування ідей на основі утворення асоціацій критика заборонена;

• зміна слів, понять, картин, що створюють асоціації, повинна відбуватися швидко.

Рекомендації для членів групи (апарату управління):

• висловлювати, демонструвати все, що прямо або побічно викликає виникнення асоціацій, ідей;

• необхідно фіксувати будь-які ідеї, що зароджуються у колег;

• потім напрацювання ідей слід систематизувати і класифікувати;

• за допомогою критичного аналізу вибираються найбільш суттєві ідеї.

Метод інверсії у сфері кібербезпеки використовується як евристичний підхід, що полягає у свідомій зміні напрямку мислення шляхом розгляду проблеми «від протилежного», зокрема через моделювання дій потенційного зловмисника. Застосування цього методу передбачає постановку запитань на кшталт «як можна обійти існуючі механізми захисту» або «які дії призвели б до максимального порушення безпеки», що дозволяє виявити приховані вразливості інформаційних систем і процесів. У результаті інверсійного аналізу формуються більш ефективні заходи протидії, удосконалюються політики безпеки та підвищується стійкість кіберінфраструктури до атак. Метод інверсії сприяє розвитку аналітичного та критичного мислення фахівців з кібербезпеки

й широко використовується під час проведення тестувань на проникнення та оцінювання ризиків

Часто ідею розв’язання проблеми вдається знайти саме тоді, коли пошук спрямовується у протилежний бік – туди, де погляди суперечать усталеним нормам, продиктованим логікою та здоровим глуздом. У випадках, коли звичні логічні прийоми та мисленнєві процедури виявляються непродуктивними, найефективнішим може виявитися вибір протилежної альтернативи.

Основою цього методу є принцип дуалізму, діалектична єдність і максимально доцільне використання протилежних (прямих і обернених) прийомів творчого мислення, що спираються на діалектичний підхід до аналізу досліджуваного об’єкта.

Приведемо рекомендації до використання методу.

Рекомендації для керівника:

• спонукати членів групи до неодноразового переформулювання проблеми з метою її осмислення;

• поряд з прямим завданням висувати зворотні;

• добиватися діалектики аналізу і синтезу у міркуванні;

• пропонувати членам групи у процесі розв'язання проблеми використовувати протилежні процедури.

Рекомендації для членів групи:

• пам'ятати, що інверсія – пошук ідей в напрямах, протилежних традиційним поглядам, переконанням, здоровому глузду, формальній логіці;

• починати вирішувати задачу із спроби її переформулювати;

• до будь-якої ідеї шукати контрідею;

• вирішуючи задачі, намагатися використовувати протилежні процедури, засоби.

Метод дозволяє:

• розвивати діалектику мислення;

• знаходити вихід з безвихідних ситуацій;

• відшукувати оригінальні рішення різного рівня складності. Застосування методу потребує розвинених творчих здібностей, базових знань, умінь і навичок.

Метод особистої аналогії у сфері кібербезпеки застосовується як евристичний підхід, що передбачає уявне ототожнення фахівця з окремим елементом інформаційної системи або з потенційним порушником з метою глибшого розуміння процесів атаки та захисту. У межах цього методу експерт намагається «поставити себе на місце» мережевого вузла, сервера, користувача чи зловмисника, аналізуючи, які дії, обмеження або загрози впливають на його функціонування. Такий підхід дозволяє виявити неочевидні вразливості, помилки конфігурації або поведінкові фактори, що можуть бути використані під час кібератаки. Застосування методу особистої аналогії сприяє розвитку інтуїтивного та системного мислення фахівців і підвищує ефективність прийняття рішень у процесі забезпечення кібербезпеки.

Під час розв’язання проблеми нерідко об’єкт, закони функціонування якого ще невідомі, замінюється іншим – подібним, але з уже дослідженими властивостями. Застосовуються різні типи аналогій: прямі, суб’єктивні, символічні та фантастичні. Для осіб, що ухвалюють рішення, особливо важливими є особисті аналогії – коли досліджуваному об’єкту приписуються власні почуття, емоції, наміри чи функції. Такий підхід дозволяє «злитися» з об’єктом, ніби відчути його, усвідомити та випробувати його переваги й недоліки на собі. В основі методу лежить заміщення досліджуваного об’єкта (або процесу) іншим – найчастіше самим собою.

Метод дозволяє:

• розвивати уяву, фантазію – образне мислення і на цій основі отримувати оригінальні рішення проблем;

• "досліджувати" об'єкти, що вивчаються, не витрачаючи кошти, ресурси. Завдяки використанню методу можна отримати тільки ідею вирішення проблеми.

Метод номінальної групи. Цей метод використовується для знаходження найкращих варіантів розв'язання проблем. Умови і вимоги:

• притягуються експерти, які добре вирішують проблеми, але тільки ті, що раніше разом не працювали;

• учасники в процесі спільної роботи можуть генерувати власні ідеї з урахуванням позиції колег, можуть їх переглядати;

• склад групи – не більше 12-15 чол. (мінімум – 6-8 чол.);

• експерти не повинні бути зв'язані службовими відносинами;

• тривалість роботи групи – 5 годин (з урахуванням освоєння методу);

• висловлені ідеї сприймаються не лише на базі особистого ставлення до них, але і з урахуванням обстановки в групі;

• відбувається взаємне підживлення експертів, у результаті чого ідея ніби відокремлюється від авторів.

Етапи роботи групи. Ведучий:

1) зачитує ретельно сформульоване завдання експертам. Експерти письмово пропонують ідеї вирішення. Процедура триває 20 хвилин;

2) пропонує кожному учаснику назвати одну ідею і записати їх на дошці (вивести на монітор комп'ютера). Кожен експерт називає одну ідею. По колу перераховуються ідеї, при цьому виключаються коментарі до них і їх оцінки. Відбувається індивідуальний аналіз кожної ідеї всіма учасниками;

3) зачитує першу ідею і пропонує кожному висловити свою думку по одному з чотирьох варіантів:

* залишити ідею в списку;

• уточнити ідею, переформулювати, скоротити її і т. ін.;

• об'єднати з іншою ідеєю;

* викреслити (при повній згоді всіх членів групи).

Обговорення ідей повинно відбуватися швидко, без повторів і дискусій. Це, з одного боку, поглиблює і розширює самі ідеї, а з іншого – навчає

учасників. Індивідуальні знання і досвід стають груповими. У результаті експерти залишають 12-15 ідей.

4) відбувається голосування і ранжування. Кожному експерту видають по 8 карток (6 карток, якщо група складається з 6-8 чол.) і пропонують відібрати 8 найбільш важливих ідей. Картки ранжуються: найважливіша отримує цифру 8, а найменше важлива – 1. 3 шести ідей, що залишилися, найбільш важливій привласнюється ранг 7, найменше важливій – 2 і т. д. Оцінка здійснюється таємно, що підсилює об'єктивність підходів до визначення пріоритетів, але оскільки раніше експерти були ознайомлені з думкою інших (на третьому етапі), це сприяє уніфікації позицій учасників.

5) підраховуються голоси і підводяться підсумки.

Метод синектики у сфері кібербезпеки використовується як комплексний евристичний підхід, спрямований на розв’язання складних і нестандартних проблем шляхом поєднання різнорідних ідей, знань та аналогій із різних галузей науки і практики. Його сутність полягає у свідомому використанні прямих, символічних, особистих та фантастичних аналогій для переосмислення кіберзагроз, архітектури захисту або процесів реагування на інциденти. Залучення фахівців з різним професійним досвідом – кібербезпеки, IT-інфраструктури, управління ризиками та навіть суміжних дисциплін – дозволяє сформувати нетривіальні підходи до виявлення вразливостей і розроблення інноваційних механізмів протидії кібератакам. Застосування методу синектики сприяє подоланню стереотипів мислення, підвищує креативність колективних рішень і зміцнює адаптивність системи кібербезпеки до нових та складних загроз.

Суть методу полягає в отриманні найбільш нестандартних ідей завдяки навчанню учасників застосовувати під час «мозкової атаки» прийоми аналогії, інтуїтивного мислення, абстрагування та вільних міркувань. Використання несподіваних метафор, ігрових елементів та асоціацій сприяє тому, що вирішення проблеми набуває оригінального й непередбачуваного характеру.

Умови і вимоги:

1) проблема формулюється в загальному (образному) вигляді;

2) обговорення слід починати не з проблеми, а з аналізу її загальних ознак, які характеризують ситуацію, що склалася;

3) не рекомендується зупинятися на одній, навіть оригінальній ідеї рішення;

4) при виникненні труднощів у процесі вирішення проблеми слід звернутися до аналізу ситуації, в якій виникла проблема;

5) висунення ідей та їх відбір залежать від майстерності, такту, винахідливості керівника, його уміння стимулювати творче мислення учасників;

6) критичний відбір і оцінку ідей розв'язання проблеми краще здійснювати поетапно:

* аналіз кожної висунутої ідеї;

* групування і класифікація ідей;

* критичний аналіз висунутих ідей;

* відбір оригінальних, оптимальних рішень. Приведемо рекомендації використання методу. Рекомендації для керівника;

• застосовуються всі рекомендації методів "мозкової атаки", аналогії, інверсії, вільних асоціацій та інші, що входять в метод синектики;

• оптимальний, результативний склад групи – 3-15 чол. із різними здібностями, професійними інтересами і підготовкою;

• необхідно спонукати членів групи до багаторазового переформулювання проблеми;

• не слід заспокоюватися при отриманні вдалої ідеї;

Рекомендації для членів групи:

• максимально використовувати особистий досвід, знання, уміння;

• не вдаватися до передчасних, поспішних формулювань;

• висуваючи ідеї, використовувати аналогії, метафори, інверсію, елементи гри, міркування вголос;

• об'єкт аналізувати з різних позицій: зовнішніх і внутрішніх, наукових і життєвих, в різних ситуаціях.

Метод 635**.** Метод 635 у сфері кібербезпеки застосовується як структурований колективний евристичний підхід до швидкого генерування ідей під час аналізу кіберзагроз, розроблення заходів захисту та сценаріїв реагування на інциденти. Суть методу полягає в тому, що шість учасників команди з кібербезпеки протягом визначеного часу пропонують по три ідеї щодо розв’язання конкретної проблеми, після чого передають результати наступному учаснику для подальшого розвитку й уточнення, і такий цикл повторюється п’ять разів. Це дозволяє за короткий проміжок часу отримати значну кількість різнопланових пропозицій, зменшуючи вплив авторитетів і забезпечуючи рівну участь усіх членів команди. Застосування методу 635 підвищує ефективність колективного мислення, сприяє виявленню нетипових підходів до кіберзахисту та покращує якість управлінських рішень у сфері кібербезпеки.Група з шести осіб отримує завдання проаналізувати та сформулювати проблему. Кожен учасник протягом п’яти хвилин у спеціальній формі записує три варіанти її розв’язання, після чого передає бланк сусідові. Наступний учасник, ознайомившись із попередніми ідеями, додає під ними ще три власні пропозиції. Вони можуть спиратися на вже висунуті рішення або бути абсолютно новими. Процес триває доти, доки всі учасники не заповнять кожен із формулярів.

Умови: рекомендована кількість учасників – 6. Час на ротаційну фазу може збільшуватися на наступних фазах. Метод дозволяє отримати до 108 (6х3х6) пропозицій.

Метод Дельфі**.** Цей метод здійснюється як багаторівневе анкетування з поетапною обробкою результатів і переданням їх експертам, які працюють незалежно один від одного. Кожному учаснику пропонуються питання та варіанти відповідей без необхідності їх аргументування. Наприклад, відповіді можуть містити числові оцінки певних параметрів. Отримані результати

опрацьовуються з метою визначення середніх та крайніх значень. Експертам повідомляють підсумки першого туру із зазначенням діапазону оцінок. Якщо думка експерта суттєво відхиляється від середнього показника, він повинен обґрунтувати свою позицію.

У наступному турі учасники можуть змінити власні оцінки, пояснивши причини корекції. Після повторної обробки результати знову доводяться до експертів. У разі відхилень від середнього значення вони надають додаткові коментарі. Опитування триває доти, доки оцінки не набудуть стабільного характеру.

Ітеративність процедури, поєднана з аналізом результатів і вимогою аргументування, стимулює експертів критично переосмислювати власні погляди. Водночас збереження анонімності відповідей унеможливлює вплив авторитету окремих учасників та прояви конформізму.

Перелік запитань для самопідготовки:

1. У чому полягає відмінність між індивідуальним і груповим прийняттям рішень?
2. Які переваги має індивідуальне прийняття рішень?
3. Які недоліки притаманні груповому прийняттю рішень?
4. У яких ситуаціях доцільно застосовувати індивідуальний підхід, а в яких — груповий?
5. Які основні евристичні методи групового прийняття рішень ви знаєте?
6. У чому суть методу "мозкового штурму" (brainstorming)?
7. Як працює метод Дельфі?
8. Які етапи проведення групової експертної оцінки?
9. У чому полягає метод номінальних груп і як він застосовується?
10. Які переваги та обмеження має метод синектики?
11. Як забезпечується об’єктивність у групових методах прогнозування?
12. Які чинники впливають на ефективність прийняття рішень?

Тема 5. Непараметричні та параметричні методи прогнозування

5.1. Суть і принципи непараметричних методів прогнозування.

5.2. Суть і принципи параметричних методів прогнозування.

5.1. Суть і принципи непараметричних методів прогнозування

Непараметрична статистика як галузь математичної статистики дає змогу вивчати випадки, які передбачають нерівномірний або невідомий розподіл показників тестів.

Непараметричні методи статистики – методи математичної статистики, які застосовують за відсутності функціонального виду генеральних розподілів.

Використання їх розпочалося із запровадженням критеріїв знаків у XVIII ст. У другій половині XIX ст. стали застосовувати ранги і коефіцієнти рангової кореляції. Рангові методи, які розробляв Ч.-Б. Спірмен, викликали інтерес у наукових колах. Унаслідок цього сформувалася непараметрична статистика як самостійна галузь математичної статистики.

Для визначення статистичних залежностей у непараметричній статистиці застосовують моду *Mo*, *U*-критерій Манна – Вітні, Вілкінсона, коефіцієнт Пірсона χ^2 та інші – для шкали найменувань; коефіцієнти асоціацій *Ф* і контингенції *Q*, перетворений коефіцієнт кореляції Пірсона – для дихотомічної шкали найменуваль; коефіцієнт спряженості Пірсона *C* (для великих вибірок) і Чупрова *K* – для *M*×*N* – клітинної спряженості; моду *Mo,* медіану *Me (Md)*, коефіцієнт рангової кореляції Спірмена *R* – для шкали рангів.

У профвідборі використовують такі коефіцієнти:

а) для шкали найменувань – коефіцієнт погодженості Пірсона χ^2, коефіцієнт контингенції *Q*, коефіцієнт асоціації *Ф* для чотириклітинної спряженості, для *M*×*N* – клітинної спряженості, коефіцієнт взаємної спряженості Пірсона *C* і коефіцієнт взаємної спряженості Чупрова *K*;

б) для шкали порядків – коефіцієнт рангової кореляції Спірмена *Rs*, призначений для кореляції двох випадкових змінних, заснованих на рангових вимірюваннях і не діючих значеннях.

Розглядають особливості розрахунків деяких із названих коефіцієнтів, які найчастіше використовують при обробці тестів, в межах непараметричних методів математичної статистики.

Одними із найпростіших математично-статистичних обчислювальних мір є міри центральної тенденції, що дають змогу вираховувати середні значення.

Середнє (центральне) значення – узагальнювальний показний місця і рівня центру розподілу, тобто значення ознаки, навколо якої концентруються всі інші варіювальні значення.

Він є дуже важливим і поширеним параметром вибірки. Математична статистика трактує багато видів середніх величин. Кожна з них – це специфічна абстракція, яка має певний зміст, специфічне смислове навантаження, що робить середню величину "сплавом" кількісної і якісної оцінок. Тому вибір форми середнього значення – це вибір способу опосередковування даних, наявних у повному обсязі початкової вибірки. На всіх подальших висновках позначається таке опосередковування, оскільки воно відображає специфіку вибраної середньої величини і в цьому значенні є відносним. Проблема вибору адекватної середньої величини пов'язана з формою статистичного розподілу і відповідного її аналізу.

Отже, визначальним для кожного виду середньої величини є її якісний зміст – знання того, в якому значенні вона середня, в яких межах відбувається усереднення. В цьому контексті розглядають основні види середніх значень.

Основними мірами центральної тенденції є: мода, медіана, середнє арифметичне. Одні з них застосовують у непараметричній, а інші – у параметричній статистиці.

Мода (*Mo*) – величина, що найчастіше спостерігається в числовому ряді.

Вона вказує найтиповіше значення статистичної ознаки і становить особливий інтерес в асиметричному розподілі, де навколо значення моди концентруються найбільші частоти вибірки. Мода є середнім скупченням варіант, позначає місце і величину дії найсильнішого фактора серед тих, що формують статистичну сукупність, виявляють досліджувану ознаку. Мода становить клас найбільшої властивості, віднесеної до конкретних умов вимірювання.

Параметр моди незамінний для шкали найменувань, тобто у разі суто якісної градації варіант – конкретних значень оцінюваного параметра. Однак мода придатна і для всіх інших шкал вимірювання, особливо для випадків великої варіативності показника, для багатовершинних розподілів, де вона є своєрідним вираженням типу.

Емпіричне визначення (обчислення) параметра моди подібне до обчислення медіани, ґрунтується на формулі найпростішої інтерполяції:

$$M_0 = x_{M_0} + \Delta x \frac{f_{M_0} - f_{M_0-1}}{\left(f_{M_0} - f_{M_0-1}\right) + \left(f_{M_0} - f_{M_0+1}\right)} \qquad (5.1)$$

де x_{M_0} – початок модального, тобто найчастішого, розряду (інтервалу); Δx – величина (ширина) модального розряду; f_{M_0} – частота модального розряду; f_{M_0-1} – частота розряду, суміжного модальному зліва, тобто меншого, ніж x_{M_0}; f_{M_0+1} - частота розряду, суміжного модальному справа.

Крім емпіричного (тобто заснованого на результатах конкретного вимірювання) обчислення, існує також обчислення, засноване на теоретичному визначенні моди як точки, що відповідає максимуму кривої розподілу. При цьому до емпіричного розподілу підбирають одну з найбільш відповідних теоретичних кривих розподілу (наприклад, набір кривих Пірсона), а потім параметр моди обчислюють аналітично, тобто на основі відомої формули тієї чи іншої підібраної теоретичної кривої. Цей шлях обчислення є математично громіздким і використовується лише в спеціальних дослідницьких задачах.

Медіана (*Me, Md*) – середнє, центральне (за місцем) значення змінної в загальному впорядкованому ряді варіант вибірки.

Медіана є своєрідною серединою, справа і зліва від якої порівну розташовується решта варіант, тоді як їх питома вага, тобто абсолютна величина, не береться до уваги. Отже, *Me* як середня, центральна величина є передусім якісною. Наприклад, для семи числових значень показника обсягу уваги (п = 7) (4, 6, 6, 8, 7, 6, 5) середнє арифметичне дорівнює 6. Для обчислення медіани (Мої) необхідно попередньо упорядкувати ряд числових значень показника від його мінімального значення до максимального, чи навпаки (для прикладу: 4, 5, 6, 6, 6, 7, 8 чи 8, 7, 6, 6, 6, 5, 4). Тоді медіану можна визначити як числове значення показника, що розташований в упорядкованому ряді даних посередині (медіана поділяє упорядкований ряд даних на дві однакові частини). За непарної кількості значень у ряду медіаною буде те числове значення показника, порядкове місце якого визначається виразом:

$$M_d = \frac{n-1}{2} \tag{5.2}$$

Коефіцієнт погодженості Пірсона χ^2 використовують для порівняння частот двох розподілів: двох емпіричних або емпіричного і теоретичного.

Коефіцієнт погодженості Пірсона χ^2 (Критерій χ^2 (хі-квадрат)) – коефіцієнт, що ґрунтується на наближенні частоти прояву ознаки у різних вибірках, виміряної за номінативною шкалою.

Розрахунок здійснюють за формулою:

$$\chi 2 = \frac{\sum (n_{i1}-n_{i2})^2}{n_{i2}} \tag{5.3}$$

де n_{i1} – частота P^1 прояву властивості у першого досліджуваного; n_{i2} – частота P^2 прояву властивості у другого досліджуваного.

Застосування критерію вимагає, щоб обсяг розподілів, що зіставляються, мав не менше 20-30 варіант, а мінімальна їх частота – не менше п'яти (інакше слід укрупнити розряди).

Інші вчені пропонують таку формулу критерію $\chi^{2'}$:

$$\chi 2 = \frac{\sum (f_i - \overline{f}_i)^2}{\overline{f}_i} \quad (5.4)$$

де f_i – кожна частота двох вибірок, що зіставляються, відповідна єдиному аргументу; $\overline{f}_i$ – середнє значення цієї частоти за двома вибірками (або очікувана частота).

При зіставленні двох емпіричних вибірок обчислення спрощуються, якщо формулу χ^2 перетворити у такий спосіб:

$$\chi 2 = \sum \frac{(f_i^{'} - f_i^{''})^2}{f_i^{'} + f_i^{''}} \quad (5.5)$$

де $f_i^{'}$ і $f_i^{''}$ – частоти двох вибірок, що зіставляються. Отриману суму порівнюють з табличним значенням того або іншого рівня значущості.

Алгоритм обчислення вказаного коефіцієнта можна з'ясувати на прикладі. При виявленні реалізації кібезагрози у двох типах операційної системи отримано результати випадків одночасної реалізації різних видів загроз, розміщені в табл. 5.1.

Таблиця 5.1

Результати кількості одночасної реалізації різних видів кіберзагроз

Обсяг одночасного виявлення кіберзагрози	x_i	2	3	4	5	6
Кількість досліджуваних кіберзагроз з першої операційної системи	$f_i^{'}$	4	7	8	12	9
Кількість досліджуваних кіберзагроз з другої операційної системи	$f_i^{''}$	4	9	12	7	8

Слід з'ясувати, чи значущою є відмінність частот у цих двох групах. Кількість складених розрядів f = 5.

Обчислення χ^2 наведено в табл. 5.2 (за спрощеною формулою).

Таблиця 5.2

Обчислення коефіцієнта погодженості

x_i	f_i'	f_i''	$f_i' - f_i''$	$(f_i' - f_i'')^2$	$f_i' + f_i''$	$\frac{(f_i' - f_i'')^2}{f_i' + f_i''}$
2	4	4	0	0	8	0
3	7	9	-2	4	16	0,25
4	8	12	-4	16	20	0,8
5	12	7	5	25	19	1,32
6	9	8	1	1	17	0,06
						Σ=2,43 χ^2=2,43

Для прийняття або відхилення гіпотези про схожіть характеристик різних операційних систем, з точки зору реалізації кібератак всередині них, скористаємося табличним значенням критерії Пірсона і його порівняння з отриманим критерію. Для цього потрібно обчислити значення *к* – кількість ступенів свободи, яку визначають як зменшену на одиницю кількість розрядів f, що зіставляються. В нашому випадку $к = f - 1 = 5 - 1 = 4$.

Оскільки табличне значення $\chi_{0,5}(4)=9,49$ і обчислене емпіричне $2,43 < \chi_{0,5}$ означає відсутність відмінностей між частотами у двох типах досліджуваних операційних систем, а обидві емпіричні сукупності можна вважати вибірками з однієї генеральної сукупності (операційні системи з приблизно однаковими властивостями захисту, або операційні системи одного сімейства).

Для перевірки гіпотез використовують *U*-критерій Манна–Вітні. Розглянемо приклад його розрахунку. У процесі дослідження кількості реалізованих кібератак було отримано дані двох видів: одні – під час роботи з віртуальною машиною (В), а інші – в процесі експерименту з реальною операційною системою (О) (табл. 5.3).

Дані обох груп об'єднують в порядку зменшення числових значень, закодувавши належність даних до певного виду дослідження буквами "О" (операційна система) і "В" (віртуальна машина).

Кожному числовому значенню отриманого ряду надають ранг (порядковий номер). У разі збігу числових значень їм приписують середнє

значення тих рангів, які були б подані кожному з них за відсутності збігу. Наприклад, дві 6 ділять між собою ранги 1 і 2, тому кожній з них надають ранг, який дорівнює (1+2)/2=1,5

Таблиця 5.3

Таблиця значень для розрахунку коефіцієнта Манна – Вітні

Код	В	В	В	В	В	В	В	О	О	О	О	О	В	В	В	В	О	О	О	О	О
К-ть атак	6	6	5	5	5	5	5	5	4	4	4	4	4	4	4	3	3	3	3	3	2
Ранг	1,5	1,5	5,5	5,5	5,5	5,5	5,5	5,5	12	12	12	12	12	12	12	18	18	18	18	18	21

Знаходять суму рангів (R) для кожного ряду даних: для віртуальної машини – R_1 =84,5; для операційної системи – R_2 =146,5.

Для перевірки обчислень використовується вираз:

$$R_1 + R_2 = \frac{N}{2}(N+1), \tag{5.6}$$

де $N = n_1 + n_2$

Підставляючи значення, отримуємо:

$$R_1 + R_2 = 84{,}5 + 146{,}5 = 231,\ \frac{N}{2}(N+1) = \frac{21}{2} * 22 = 231$$

Отже, операції рангування і підсумовування рангів виконано правильно.

Обчислюють значення U-критерію Манна-Вітні за формулою:

$$U = n_1 * n_2 + \frac{n_1*(n_1+1)}{2} - R_1 \tag{5.7}$$

або

$$U' = n_1 * n_2 + \frac{n_2*(n_2+1)}{2} - R_2. \tag{5.8}$$

Підставляючи значення, отримується такий рзультат:

$$U = 11 * 10 + \frac{11 * 12}{2} - 84{,}5 = 110 + 66 - 84{,}5 = 91{,}5$$

За даними таблиці 5.4 знаходять пограничні значення *U*-критерію Манна – Вітні (U_{min} і U_{max}) для n_1=11 і n_2=10, U_{min}=26, U_{max}=84.

Таблиця 5.4

Пограничні значення U-критерію Манна – Вітні

(рівень достовірності – 95%)

	10	11	12	13	14	15	16	17	18	19	20
10	23 77	26 84	29 91	33 97	36 104	39 111	42 118	45 125	48 132	52 138	55 145
11	26 84	30 91	33 99	37 106	40 114	44 121	47 129	51 134	55 143	58 151	62 158
12	29 91	33 99	37 107	41 115	45 123	49 131	53 139	57 147	61 155	65 163	69 171
13	33 97	37 106	41 115	45 124	50 132	54 141	59 149	63 158	67 167	72 175	76 184
14	36 104	40 114	45 123	50 132	55 141	59 151	64 160	67 171	74 178	78 188	83 197
15	39 111	44 121	49 131	54 141	59 151	64 161	70 170	75 180	80 190	85 200	90 210
16	43 118	47 129	53 139	59 149	64 160	70 170	75 181	81 191	86 202	92 112	98 222
17	45 125	51 136	57 147	63 158	67 171	75 180	81 191	87 202	93 213	99 224	105 235
18	48 132	55 143	61 155	67 167	74 178	80 190	86 202	93 213	99 225	106 236	112 248
19	52 138	58 151	65 163	72 175	78 188	85 200	92 212	99 224	106 236	113 248	119 261
20	55 145	62 158	69 171	76 184	83 197	90 210	98 222	105 235	112 248	119 261	127 273

Гіпотеза H_0 відхиляється, якщо $U<U_{min}$ або $U>U_{max}$. Обчислене значення 7 = 91,5 > U_{max} =84? Отже, потрібно прийняти H_1 про достовірність відмінностей M_1 і M_2. Можна зробити такий статистично обгрунтований висновок: при роботі з віртуальною машиною реалізовується більша кількість кібератак ніж під час роботи в умовах реальної операційної системи (в дужках при цьому вказують величину статистичного критерію і рівень значущості: U=91,5; p<0,05). Якщо значення критерію обчислено за формулою U, то формулу U' зазвичай використовують для завершальної перевірки обчислень, використовуючи вираз:

$$U = n_1 - n_2 - U' \qquad (5.9)$$

Потім проводиться обчислення:

$$U^{'} = 11 - 10 + \frac{10 * 11}{2} - 146{,}5 = 18{,}5$$

$$U = 11 * 10 - 18{,}5 = 110 - 18{,}5 = 91{,}5$$

Отже, всі обчислення правильні.

Для визначення статистичного зв'язку змінних, виміряних за дихотомічною шкалою найменувань, використовують коефіцієнт контингенції Q і коефіцієнт асоціації $Ф$.

Приклад. Для визначення алгоритму обчислення коефіцієнтів контенгенції та асоціації необхідно розрахувати статистичний зв'язок між думками адміністраторів безпеки, щодо їх думки про ефективність використання віртуальної машини при дослідженні кібератак (В) чи реальної операційної системи (О), виміряною дихотомічною шкалою найменувань за даними ТАХО (тест адекватних характеристик об'єкта) і ДДО (диференційно-діагностичний опитувальник).

Спочатку необхідно визначити кількість співвідношень технічної і гуманітарної спрямованостей за тестом ТАХО і ДДО (табл. 5.5).

Таблиця 5.5

Таблиця результатів дослідження думок адміністраторів безпеки

№з/п	ТАХО	ДДО
1	В	В
2	О	В
3	О	О
4	В	О
5	О	О
6	О	О
7	В	В
8	В	О
9	В	В
10	О	О

Результати слід занести до таблиці квадратів (табл. 5.6). Це необхідно для розрахунку залежності за формулами:

а) для обчислення коефіцієнта контингенції:

$$Q = \frac{(ad-bc)}{(ad+bc)} \quad (5.10)$$

б) для обчислення коефіцієнта асоціації:

$$\Phi = \frac{(ad-bc)}{\sqrt{(a+b)(c+d)(a+c)(b+d)}} \quad (5.11)$$

Таблиця 5.6

Таблиця квадратів для розрахунку коефіцієнтів контингенції та асоціації

		ДДО	
		В	О
ТАХО	В	3 (a)	2 (b)
	О	1 (c)	4 (d)

Для вказаного прикладу це матиме такий результат:

а) обчислення коефіцієнта контингенції:

$$Q = \frac{(12-2)}{(12+2)} = 0{,}71$$

б) обчислення коефіцієнта асоціації:

$$\Phi = \frac{(12-2)}{\sqrt{(3+2)(1+4)(3+1)(2+4)}} = 0{,}41$$

Отримані результати інтерпретують на основі співвідношення у такий спосіб: якщо теоретичні Q і Φ більші ніж 0,5, це свідчить про сильний зв'язок; якщо теоретичні Q і Φ менші ніж 0,5 – про слабкий зв'язок.

Для визначення статистичного зв'язку змінних, виміряних за порядковою шкалою, використовують коефіцієнт рангової кореляції.

5.2. Суть і принципи параметричних методів прогнозування

Параметричні методи статистики − методи математичної статистики, які застосовують у випадках, коли тестові показники виміряно за інтервальною шкалою, шкалою відношень або абсолютною шкалою при дотриманні розподілу Гаусса.

Параметричну статистику застосовують для методів, які орієнтовані на статистичну норму. Використовують методи аналізу даних за допомогою: мод *Mo*, медіани *Me* (*Md*) і середнього арифметичного *Mx* (міри центральної тенденції), середнього квадратичного відхилення α, коефіцієнта варіації *V* (міри мінливості), коефіцієнта кореляції Пірсона *Rxy* (міри зв'язку), *t*-критерію Стьюдента, υ-вритерію Велша, *f*-критерію Фішера (статичтичний висновок) і прогнозування в соціальній сфері за допомогою лінійної і нелінійної регресії (моделі регресії).

Особливості розрахунків деяких із названих коефіцієнтів розглянуто в межах параметричних методів математичної статистики.

Середнє арифметичне – величина, сума негативних і позитивних відхилень від якої дорівнює нулю. У статистиці його позначають буквою "M" або "X". Щоб підрахувати середнє арифметичне, потрібно додати всі значення ряду і розділити суму на кількість значень, які додавалися.

Середнє арифметичне (*M*) у найпростішому випадку обчислюється за формулою:

$$M = \frac{\sum x_i}{n}, \tag{5.12}$$

де x_i – числові значення показника, n – їх кількість (або кількість досліджуваних).

У процесі дослідження експерт-аналітик може отримувати дані, розкидані в досить широкому діапазоні розкиду. Чим сильніше розкидані варіанти щодо середнього, тим більшим виявляється і середнє квадратичне відхилення.

Середнє квадратичне відхилення (σ) – міра різноманітності об'єктів, що входять до групи, яка показує, наскільки в середньому відхиляється кожна варіанта від середнього арифметичного.

Середнім квадратичним відхиленням (σ) випадкової величини X називають корінь квадратний із дисперсії, тобто:

$$\sigma(X) = \sqrt{D(X)}, \tag{5.13}$$

де $\sigma(X)$ – середнє квадратичне відхилення, $D(X)$ – дистперсія випадкової величини *X*.

Величину кореня квадратного з дисперсії ще називають стандартним відхиленням.

Обчислення дисперсії застосовують для виокремлення вибіркової сукупності, визначення помилки вибірки, однорідності досліджуваної сукупності за певною ознакою.

Дисперсія випадкової величини *X* – математичне сподівання квадрата відхилення випадкової величини від її математичного сподівання.

її обчислюють за формулою:

$$\sigma = \frac{\sum (x - x_{сер})^2}{n}, \tag{5.14}$$

де $\sum (x - x_{сер})^2$ – сума квадратів різниць між середнім та індивідуальним значенням ознаки, *n* – кількість варіант.

Дисперсією також називають варіації значення залежної змінної. Ті варіації, які можна пояснити дією незалежної змінної, називають первинною дисперсією; ті ж, що є результатом дії інших факторів, – вторинною дисперсією або дисперсією помилки. Контролюючи рівень інших потенційних змінних, експериментатор прагне максимізувати долю первинної дисперсії.

Розкиданість значень характеризує і розмах *W* – різниця між найбільшим і найменшим значеннями в ряду:

$$W = x_{max} - x_{min} \tag{5.15}$$

Розмах є найпростішою і приблизною мірою розкиданості значень, оскільки під час її обчислення використовують лише крайні елементи ряду.

Точнішою мірою розкиданості даних є стандартне відхилення (σ, S_x), яке обчислюють за формулами:

$$\sigma = \sqrt{\frac{n_i (x_i - M)^2}{n-1}}, \tag{5.16}$$

де x_i – числові значення показника; *n* – їх кількість (якщо *n*>30, то в знаменнику замість *n*-1 буде *n*), *M* – середнє арифметичне.

$$\sigma = \sqrt{\frac{\sum d^2}{n-1}} \tag{5.17}$$

де $\sum d^2$ – сума отриманих значень, відмінних від середнього арифметичного, *n* – кількість досліджуваних випадків.

Класичний спосіб обчислення σ досить об'ємний. Тому при n<20 можна можна використати експрес-метод визначення σ. Розрахунок здійснюють за формулою:

$$\sigma = \frac{W}{d_n}, \tag{5.18}$$

де W – розмах, d_n – табличний коефіцієнт.

Коефіцієнт варіації дає змогу порівняти в абсолютних одиницях варіативність вибірок незалежно від їх середнього значення.

Коефіцієнт варіації (*V*, *CV*) – виражене у відсотках відношення стандартного відхилення до середнього арифметичного значення.

Його обчислюють за формулою:

$$V = \frac{\sigma}{X} \tag{5.19}$$

$$CV = \frac{\sigma}{M} \tag{5.20}$$

де σ – стандартне відхилення, *CV* – коефіцієнт варіації, *X*, *M* – середнє арифметичне.

Сигма (σ) є величиною іменованою, тобто різною у різних одиницях вимірювання, і залежить не тільки від ступеня варіювання, а й від одиниць вимірювання. Тому за сигмою можна порівнювати мінливість лише одних і тих самих показників, а зіставляти сигми різних ознак за абсолютною величиною не можна. Щоб порівняти за рівнем мінливості ознаки будь-якої розмірності (виражені в різних одиницях вимірювання) і уникнути впливу масштабу вимірювань середнього арифметичного на величину сигми, застосовують коефіцієнт варіації – приведення даних до однакового масштабу величини σ.

Для нормального розподілу існує точна кількісна залежність частот і значень, що дає змогу прогнозувати появу нових варіант: ліворуч і праворуч від середнього арифметичного розміщено 50% варіант, в інтервалі від $M - 1\sigma$ до $M + 1\sigma$ – 68%, в інтервалі від $M - 1{,}96\sigma$ до $M + 1{,}96\sigma$ – 95% варіант. Орієнтуючись на ці характеристики нормального розподілу, можна оцінити ступінь близькості до нього досліджуваного розподілу психологічної ознаки.

Кореляційні залежності можуть бути таких видів: 1) коефіцієнт кореляції Пірсона *Rxy* що показує наявність статистичного зв'язку між змінними x і y, за якого кожній змінній x відповідає одне чи кілька визначених значень y, а розподіл y змінюється разом зі змінною x. Змінна може бути односпрямованою (+) і різноспрямованою (-). За певної кількості вимірів n-кореляційні зв'язки можуть бути значущими і незначущими. Це необхідно знати для достовірних висновків про причинно-наслідкові зв'язки змінних. Рівень значущості коефіцієнтів кореляції визначають за формулою розрахунку t-критерію за допомогою таблиць квантилей t-розподілу Стьюдента для довірчої ймовірності;

2) точково-бісеріальний коефіцієнт кореляції Пірсона Rpb, що є методом кореляційного аналізу відношення змінних. Змінні можуть вимірюватись за дихотомічною шкалою найменувань та за інтервальною шкалою відношень або порядку. Точково-бісеріальний коефіцієнт кореляції застосовують і для визначення дискримінативності завдань тестів.

Розрахунок кореляційних залежностей дає змогу здійснити факторний, кластерний і кореляційний аналіз емпіричних даних.

Факторний аналіз – розділ багатомірного статистичного аналізу, сутність якого полягає у виявленні безпосередньо невимірюваної ознаки, що є головним компонентом (похідною) групи вимірюваних тестових показників. Кластерним аналізом є сукупність статистичних і якісних методів, призначених для формування відносно віддалених одна від одної груп споріднених між собою об'єктів за інформацією про зв'язки (міри спорідненості) між ними. Факторний і кластерний аналіз практикуючі психологи використовують не часто.

Завдання психологічного дослідження полягає у з'ясуванні взаємозв'язків між двома або більше наборами даних. Однією з найпростіших форм виявлення такого зв'язку є кореляція. Саме кореляційний аналіз допомагає точно кількісно оцінити ступінь узгодженості змін (варіювання) двох і більше ознак.

Кореляційний аналіз – метод дослідження взаємозалежності ознак у генеральній сукупності, що є випадковими величинами і мають нормальний багатомірний розподіл.

Наочно інтеркореляційні показники подають як таблиці інтеркореляцій змінних, кореляційні матриці і графіки.

Ступінь узгодженості змін характеризує тіснота зв'язку – абсолютна величина коефіцієнта кореляції.

Наявність кореляції між двома результатами означає, що при зміні одного-результату інший також змінюється, тобто між результатами існує і виявляється зв'язок.

Якщо значення певної величини може змінюватися, її називають змінною. Кореляція між двома змінними може бути позитивною або негативною. Позитивною кореляцією вважають такий зв'язок між змінними, коли значення обох змінних збільшуються або зменшуються пропорційно: із зменшенням (збільшенням) однієї зменшується (збільшується) інша. Простим прикладом позитивної кореляції є зв'язок між зростом і вагою людини – із збільшенням зросту збільшується і вага. За негативної кореляційний зв'язок є обернено пропорційним: збільшення однієї змінної супроводжується зменшенням іншої (наприклад, температура повітря і кількість одягу: чим тепліше на вулиці, тим менше одягу вдягають).

Кореляція ще не означає наявності причинно-наслідкового зв'язку. Вона свідчить лише про наявність зв'язку, але не про те, що одна із змінних є причиною, а інша – наслідком. Існування причинно-наслідкового зв'язку встановлюють іншими методами.

Висновок про причинно-наслідкову залежність між явищами, що вивчаються, тільки на підставі статистичної значущості зв'язку між відповідними ознаками (на підставі коефіцієнта кореляції) робити не варто. Зазвичай статистичний зв'язок між ознаками – це необхідна, але не достатня умова причинно-наслідкового зв'язку між ними. Твердження про те, що явище А є причиною явища В, справедливе, якщо: одночасно явища А і В статистично пов'язані; А відбувається раніше В; відсутня альтернативна інтерпретація появи Б, крім А (іншими словами – відсутня загальна причина С сумісної мінливості А і В).

Застосування кореляційного методу дає змогу обґрунтувати наявність тільки статистичного зв'язку – однієї з трьох ознак причинно-наслідкового.

Наявність кореляції між двома подіями у сфері кібербезпеки не завжди означає, що одна є причиною іншої. Наприклад, якщо під час запуску шкідливого файлу користувач втратив доступ до даних, це не означає, що сам факт відкриття файлу автоматично «створив» атаку – причиною є прихований шкідливий код у ньому. Тобто зв'язок між подіями може бути одностороннім: зараження зумовлює втрату даних, але не навпаки. У багатьох випадках залежність між показниками пояснюється третьою змінною. Скажімо, можна виявити кореляцію між кількістю попереджень антивірусної системи та кількістю підозрілих листів у корпоративній пошті. Однак це не означає, що антивірус «створює» підозрілі листи або що збільшення кількості листів примушує антивірус частіше працювати. Обидва показники зростають під впливом третього чинника – зростання активності зловмисників у фішингових кампаніях.

Після виявлення позитивної або негативної кореляції необхідно встановити, наскільки вона тісна. На це вказує коефіцієнт кореляції, який позначають буквою *r*. Величина г варіює в діапазоні від -1 до +1. У разі прямо пропорційної залежності однієї ознаки від іншої коефіцієнт кореляції дорівнює одиниці (тобто ознака корелює сама із собою). Негативний коефіцієнт кореляції

свідчить про різну спрямованість варіювання ознак: при збільшенні одного інший зменшується, і навпаки.

Зазвичай виявляють кореляції з коефіцієнтами, що знаходяться в діапазоні між нулем (відсутність кореляції) і одиницею (ідеальна кореляція), і чим ближче значення *r* ±1, тим тіснішим є зв'язок. Значення *r* виражають в десяткових дробах (наприклад, -0,23; + 0,5). При низьких значеннях *r* (зазвичай низькими вважають значення, що не перевищують 0,2 при n < 30) кореляція, як правило, не є статистично значущою.

Нульова величина коефіцієнта кореляції свідчить про відсутність взаємозв'язку між ознаками, але це трапляється дуже рідко, бо у соціальній сфері всі явища взаємопов'язані (переважно опосередковано, ці зв'язки можуть виявлятися лише на рівні тенденцій).

Враження, що значення *r* є безпосереднім показником сили кореляції, не є правильним. Наприклад можна дійти висновку, що, оскільки при ідеальній позитивній кореляції *r* = +1, то *r* = 0,7 відповідає 70% ідеальної кореляції (або *r* = 0,4 відповідає 40% ідеальної негативної кореляції). Насправді коефіцієнт кореляції є оманливим числом. Щоб знайти, яку відсоткову частку від ідеальної кореляції становить дане значення *r*, необхідно піднести його до квадрата, а результат помножити на 100. Якщо *r* = 0,7, то така кореляція становить 49% від ідеальної (0,7 0,7 о 100 = 49). Так само негативна кореляція *r* в - 0,4 становить 16% від ідеальної негативної кореляції. Тому "ступінь ідеальності" кореляції може бути набагато меншим, ніж видається, якщо зважати на значення *r*.

Фахівці статистики вважають, що коефіцієнт кореляції вказує на частку змін однієї змінної, які можна передбачити за змінами іншої. Існує багато методів вимірювання кореляції, і вибір конкретного методу залежить від типу даних.

Алгоритм обчислення коефіцієнта кореляції Пірсона (*r*) є мірою кореляції між двома змінними, розподіленими за нормальним законом (наприклад, для виявлення взаємозв'язку рівня розвитку суспільства й адаптивних можливостей

особистості). Перевага цього методу полягає в тому, що на величину кореляції не впливає те, в яких одиницях вимірювання представлені ознаки. Недоліком методу є складність математичних обчислень, особливо для великих масивів даних. Його усувають застосуванням прикладних програм (наприклад, Excel).

Коефіцієнт кореляції Пірсона r_{xy} обчислюють за формулою:

$$r_{xy} = \frac{n\sum x_i y_i - (\sum x_i)(\sum y_i)}{\sqrt{[n\sum x_i^2 - (\sum x_i)^2][n\sum y_i^2 - (\sum y_i)^2]}}, \quad (5.21)$$

де x_i y_i – числові значення корелюючих ознак, n – кількість пар числових значень корелюючих ознак.

Розглянемо приклад визначення коефіцієнта кореляції ПІрсона. В таблиці 5.7 наведені деякі значення пар числових значень корелюючих ознак.

Таблиця 5.7

Таблиця визначення коефіцієнта кореляції Пірсона

	x_i	y_i	x_i-y_i	x_i^2	y_i^2
	10	2	20	100	4
	10	4	40	100	16
	9	3	27	81	9
	9	5	45	81	25
	8	3	24	64	9
	7	6	42	49	36
	7	9	63	49	81
	5	9	45	25	81
	5	7	35	25	49
	4	10	40	16	100
Сума	74	58	381	590	410

Отримані з обчислень дані перевіряють на значущість за таблицею критичних значень (табл. 5.8).

За неможливості дотримання цих вимог, а змінні є обчисленими за порядковою шкалою (чи переведені за нею), доцільніше використовувати (особливо за малого обсягу даних) як міру зв'язку непараметричну статистику.

Таблиця 5.8

Критичні значення rxy (для рівнів значущості 0,05 і 0,01)

n	0,05	0,01	*n*	0,05	0,01	*n*	0,05	0,01

10	0,63	0,77	21	0,43	0,55	32	0,35	0,45
11	0,60	0,74	22	0,42	0,54	33	0,34	0,44
12	0,58	0,71	23	0,41	0,53	34	0,34	0,44
13	0,55	0,68	24	0,40	0,52	35	0,33	0,43
14	0,53	0,66	25	0,40	0,51	36	0,33	0,42
15	0,51	0,64	26	0,39	0,50	37	0,32	0,42
16	0,50	0,62	27	0,38	0,49	38	0,32	0,41
17	0,48	0,61	28	0,37	0,48	39	0,31	0,41
18	0,47	0,59	29	0,37	0,47	40	0,31	0,40
19	0,46	0,58	30	0,36	0,46	41	0,31	0,40
20	0,44	0,56	31	0,36	0,46	42	0,30	0,39

Непараметричним еквівалентом цієї оцінки є коефіцієнт кореляції Снірмена (p) (наприклад, для порівняння зв'язку між знаннями з кібербезпеки персоналу підприємства та кількістю інцидентів інформаційної безпеки на цьому підприємстві), що обчислюється за формулою:

$$\rho = 1 - \frac{6d_i^2}{n(n^2-1)}, \tag{5.22}$$

де d_i^2 – різниця рангів пари даних, n – кількість пар даних.

Обчислення значень для цієї формули представлене у табл. 5.9.

Таблиця 5.9

Алгоритм розрахунку значень коефіцієнта кореляції Спірмена (р)

n	Ранг x_i	Ранг y_i	d_i	d_i^2
1	1	5	-4	16
2	2	3	-1	1
3	3	2	1	1
4	4	4	0	0
5	5	6	-1	1
6	6	1	5	25
7	7	10	-3	9
8	8	9	-1	1
9	9	7	2	4
10	10	8	2	4
Сума			0	62

Якщо обчислений $\rho > \rho_{табл.}$ Перебуває на певному рівні значущості, то він статистично достовірний (відмінний від нуля). З отриманих результатів та таблиці 5.10 видно, що $\rho = 0{,}63 > \rho = 0{,}564$ (але менше $\rho = 0{,}746$). Отже коефіцієнт кореляції значущий ($\rho < 0{,}63$) і між змінними у середньому існує позитивний зв’язок.

Таблиця 5.10

Критичні значення r (для рівнів значущості 0,05 і 0,01)

n	0,05	0,01	*n*	0,05	0,01
5	0,900	1,000	16	0,425	0,601
6	0,829	0,943	18	0,399	0,564
7	0,714	0,893	20	0,377	0,534
8	0,643	0,833	22	0,359	0,508
9	0,600	0,783	24	0,343	0,485
10	0,564	0,746	26	0,329	0,465
12	0,506	0,712	28	0,317	0,448
14	0,456	0,645	30	0,306	0,432

Перевага цього методу полягає в можливості здійснення нескладних математичних обчислень з використанням калькулятора для невеликих за обсягом вибірок. Недоліком методу є обмеження, що спричинені складністю оброблення значних масивів даних і необхідністю рангування рядів значень.

Варіантом кореляції є також зв'язок, який описує рівняння регресії (параметрична оцінка). Виявляючи кореляцію, дослідник зацікавлений продемонструвати наявність зв'язку між двома змінними. При визначенні рівняння регресії робиться спроба не тільки продемонструвати наявність зв'язку, а й передбачити фактичне значення однієї змінної за значенням іншої. Наприклад, побудувавши рівняння регресії, яке описує залежність між кількістю спроб несанкціонованого входу та ймовірністю зламу системи, можна не лише показати, що між цими показниками існує зв’язок, а й отримати формулу для прогнозування ризику успішної атаки залежно від частоти таких спроб. За допомогою класичного рівняння регресії зазвичай аналізують відношення між двома змінними. Якщо ж застосувати багатовимірне рівняння регресії, то можна передбачити ймовірність інциденту безпеки вже за кількома показниками одночасно — наприклад, кількістю фішингових листів, рівнем завантаженості мережі та числом уразливих пристроїв у системі. У багатьох випадках такий підхід дає значно точніші результати. Скажімо, використовуючи просте рівняння регресії, можна спрогнозувати ризик

зараження шкідливим програмним забезпеченням лише на основі кількості підозрілих вкладень у пошті. А багатовимірна регресія дозволить врахувати ще й наявність застарілого програмного забезпечення та відсутність оновлень системи, що зробить прогноз набагато надійнішим. Однак застосування цього методу вимагає використання спеціалізованих програм статистичного аналізу та достатнього обсягу даних для коректного моделювання.

Статистичні методи застосовують у визначеному довірчому інтервалі, який задають відповідно до потреб точності вимірів.

Довірчий внтервал – інтервал $(X \pm \varepsilon)$, який охоплює невідомий параметр із заданою точністю. У соціальних дослідженнях максимальне значення ε задають у межах 5%, тобто $\varepsilon \leq 0{,}05$.

Довірчий інтервал пов'язаний з інтервальним оцінюванням. Інтервальна оцінка параметра (у т. ч. середнього арифметичного) – це інтервал числової осі, в межах якого лежить значення параметра. У цьому разі замість одного числа (точкової оцінки) вказують нижню і верхню межі інтервалу, тобто групу суміжних точок на числовій осі, одна з яких, імовірно, і є значенням параметра. Коли інтервал будують з урахуванням вірогідності потрапляння значення параметра в межі інтервалу, його називають довірчим інтервалом. Довірчий інтервал виражається за допомогою довірчого коефіцієнта. Він вказує, що будь-який статистичний висновок є висновком імовірнісним – таким, що містить ризик допустити помилку (наприклад, з незначною вірогідністю - 0,05, або 5%).

Приклад побудови 95% довірчих інтервалів. Припустимо, дослідники хочуть оцінити середній час реакції співробітників компанії на підозрілий лист електронної пошти. Для цього було випадково обрано 16 працівників, яким наприкінці робочого тижня надіслали тестові фішингові повідомлення. Показником виступав час (у секундах) від моменту отримання листа до його позначення як «підозрілого» в системі безпеки. Було зафіксовано такі результати: 30, 40, 40, 45, 45, 45, 50, 50, 50, 50, 55, 55, 60, 60, 60, 65.

Ці дані дозволяють побудувати 95-відсоткові довірчі інтервали для середнього часу реакції та медіани, що дає змогу зробити висновки про загальний рівень обізнаності персоналу у сфері кіберзагроз.

Отриманий ряд даних характеризується середнім арифметичним $M = 50$ і стандартним відхиленням а=9. При аналізі цих величин у вчителя-дослідника зазвичай виникають стандартні питання для всіх подібних ситуацій: чи можна очікувати, що при повторенні подібних дослідів у тих самих умовах і з тим самим матеріалом, але на інших вибірках будуть отримані подібні результати; яких висновків можна дійти про певні якості особистості, спираючись тільки на отримані дані і не проводячи ніяких додаткових дослідів. Ця група питань пов'язана з проблемою надійності статистичних характеристик, розрахованих за вибірковими даними.

Для з'ясування надійності можна скористатися експрес-методами оцінювання меж довірчих інтервалів (табл. 5.11).

Таблиця 5.11

Значення коефіцієнта k_n для розрахунку меж довірчих інтервалів середніх арифметичних (рівень достовірності - 95%)

n	k_n	n	k_n
5	0,51	13	0,18
6	0,40	14	0,17
7	0,33	15	0,16
8	0,29	16	0,15
9	0,25	17	0,14
10	0,23	18	0,14
11	0,21	19	0,13
12	0,19	20	0,13

Верхню і нижню межі довірчого інтервалу середнього арифметичного можна визначити за формулами:

верхня межа = $M + k_n*W$, нижня межа = $M - k_n*W$,

де W – розмах, k_n – табличний коефіцієнт.

Тоді при $M = 50$, $W = 35$, k_{16}=0,15 можна отримати:

верхня межа = 50+0,15*35=50+5=55, нижня межа = 50-0,15*35=50-5=45.

Отже, $M_{0,95} \approx 45 \div 55$.

Ще простіше оцінюють межі довірчого інтервалу медіани. Для цього на впорядкованій сукупності числових значень швидкості читання 30, 40, 40, 45, 45, 45, 50, 50, 50, 50, 55, 55, 60, 60, 60, 65 достатньо за табл. 5.12 знайти номери значень, які можуть бути прийняті на цьому рівні достовірності як нижня і верхня межі довірчого інтервалу M_d (у прикладі точкова оцінка медіани дорівнює 50). За табл. 5.12 значення 4 і 13 у ряду даних (подані жирним шрифтом).

Таблиця 5.12

Порядкові номери числових значень у впорядкованому за збільшенням ряді даних, які можуть бути прийняті як нижня і верхня межі довірчого інтервалу медіани (рівень достовірності - 95%)

Число значень в ряді	Нижня межа	Верхня межа	Число значень в ряді	Нижня межа	Верхня межа
11	2	10	31	10	22
12	3	10	32	10	23
13	3	11	33	11	23
14	3	12	34	11	24
15	4	12	35	12	24
16	**4**	**13**	36	12	25
17	5	13	37	13	25
18	5	14	38	13	26
19	5	15	39	13	27
20	6	15	40	13	27
21	6	16	41	14	28
22	6	17	42	15	28
23	7	17	43	15	29
24	7	18	44	16	29
25	8	18	45	16	30
26	8	19	46	16	31
27	8	20	47	17	31
28	9	20	48	17	32
29	9	21	49	18	32
30	10	21	50	18	33

Отже, нижня межа довірчого інтервалу M_d дорівнює 45, верхня – 60. $M_{d\ 0,95} = 45 \div 60$.

Отримавши інтервальну оцінку середнього або медіани, що відображає і розкиданість числових значень показника і їх кількість, можна зробити попередні діагностичні висновки.

Перелік запитань для самопідготовки:

1. Що таке непараметричні методи прогнозування?
2. У чому полягає головна відмінність непараметричних методів від параметричних?
3. Які переваги мають непараметричні методи у випадках економічного прогнозування?
4. Які приклади непараметричних методів найчастіше застосовуються?
5. У яких ситуаціях доцільно застосовувати непараметричні методи?
6. Що таке параметричні методи прогнозування?
7. Які припущення лежать в основі параметричних методів?
8. Як параметричні методи враховують залежності між змінними?
9. Які приклади параметричних моделей ви знаєте?
10. Коли краще використовувати параметричні, а коли непараметричні методи?

Тема 6. Теоретичні основи економічної динаміки.

6.1. Поняття моделі та суть методу моделювання.

6.2. Етапи моделювання соціально-економічних систем і процесів.

6.3. Модель типу «життєвий цикл».

6.1. Поняття моделі та суть методу моделювання

У вивченні кібернетичних систем важливе місце посідає метод моделювання.

Модель – це створений людиною штучний об’єкт будь-якої природи (уявний чи матеріально реалізований), що відтворює або замінює оригінал таким чином, аби дослідження моделі давало нову інформацію про сам об’єкт-оригінал.

Моделювання як спосіб пізнання ґрунтується на здатності людини виокремлювати спільні ознаки й властивості різних об’єктів та встановлювати між ними взаємозв’язки. Завдяки цьому з’являється можливість досліджувати характеристики певних об’єктів не безпосередньо, а опосередковано – через вивчення доступніших для аналізу моделей, подібних до оригіналів.

Важливою властивістю будь-якої моделі є її спрощеність порівняно з оригіналом. Це неминуче, адже під час моделювання відображається лише обмежена кількість характеристик об’єкта, а ресурси дослідження завжди мають межі.

Через це виникають певні відмінності між моделлю та об’єктом-оригіналом. Модель, яка дає змогу успішно досягати поставленої мети, вважається адекватною. Адекватність означає, що вимоги до точності, повноти й достовірності виконано не повністю, а лише настільки, наскільки потрібно для розв’язання завдання.

Основне призначення моделі полягає в отриманні нових знань про об’єкт, який вона відображає.

Створення моделей у сфері кібербезпеки значно спрощує аналіз інформаційних систем та процесів захисту, часто стаючи єдиним ефективним способом дослідження складних мережевих інфраструктур або сценаріїв кібератак. Моделювання дозволяє прогнозувати поведінку системи за різних умов, оцінювати наслідки потенційних загроз, тестувати стратегії реагування та оптимізувати заходи захисту без ризику для реальних ресурсів, що робить його одним із ключових методів пізнання та управління кібербезпекою.

Схему функціонування моделі показано на рис. 6.1. Щоб отримати працездатну модель, необхідно визначити числові значення змінних, які використовують під час розв'язування задачі. Коли схему описано математично і визначено необхідні числові значення, то для вирішення задачі необхідно використати відповідний метод.

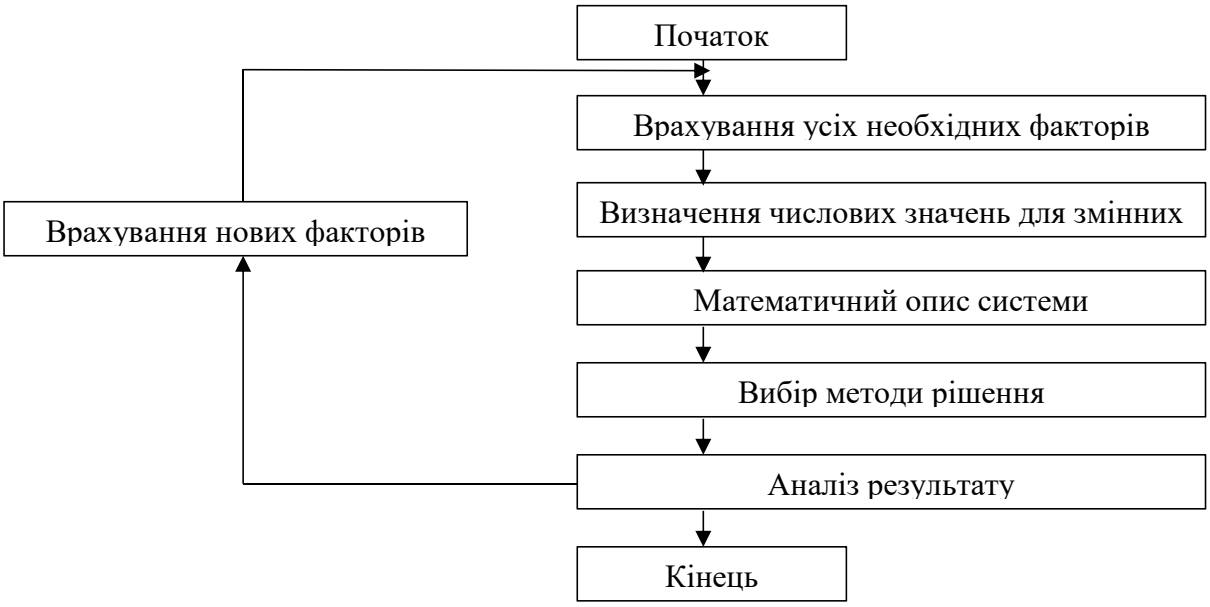


Рисунок 6.1 – Схема функціонування моделі

Під час проведення модельних експериментів варто зважати на кілька важливих моментів:

– використання готових моделей не потребує знання їхньої історії, однак для створення нових моделей необхідно розуміти закономірності їхнього розвитку;

– на стадії розробки моделей ключове значення мають інтелектуальні здібності людини, які неможливо повністю формалізувати;

– відсутність можливості цілковитої формалізації основних етапів моделювання робить цей процес подібним до мистецтва;

– сутність науки й мистецтва моделювання полягає у поділі процесу на окремі кроки, їх ретельному дослідженні та описі з максимальною мірою формалізації, а також у розгляді та оцінюванні альтернатив;

– загалом моделювання є нерозривним поєднанням наукового підходу, творчості та практичного досвіду.

6.2. Етапи моделювання соціально-економічних систем і процесів

Основними етапами дослідження кіберзагроз та інформаційних процесів за допомогою математичного моделювання є: визначення мети, складання плану дослідження, формулювання проблеми, побудова моделі кіберсистеми або сценарію атаки, організація числового або симуляційного розв’язання задачі, збирання даних про систему та загрози, перевірка адекватності моделі та впровадження отриманих результатів для підвищення ефективності захисту.

Першочерговим завданням дослідника у кібербезпеці є прогнозування потенційних наслідків атак або поведінки інформаційної системи, що базується на характеристиках об’єкта, експериментальних умовах та обраній методиці моделювання. Визначення мети дослідження має бути достатньо конкретним, щоб досягти практичних результатів, але водночас не надто вузьким, оскільки моделювання часто дозволяє отримати відповіді на ширший спектр питань щодо безпеки системи. Надмірно загальна мета може ускладнити процес дослідження через розпорошеність завдань. Тому завдання моделювання зазвичай визначають спільно фахівці з кібербезпеки та замовники дослідження.

Сам процес вивчення об’єкта або певної ситуації має творчий характер, тож жорстке визначення термінів виконання етапів не завжди є виправданим.

Проте суворий контроль за дотриманням графіка стимулює ініціативність дослідників і сприяє своєчасному завершенню експериментів.

Не менш важливим аспектом є забезпечення дослідження необхідними ресурсами, з урахуванням можливості отримання додаткових корисних результатів. Аналіз проміжних підсумків та оперативне інформування замовника допомагають підтримувати інтерес до дослідження й забезпечують активну участь усіх зацікавлених сторін.

Етап формулювання проблеми є ключовим у процесі моделювання кібербезпеки. Він потребує тісної взаємодії між замовником і виконавцем та включає збір і опрацювання інформації, необхідної для чіткого усвідомлення сутності проблеми: характеру потенційних загроз, історії інцидентів, прогнозування можливого розвитку атак, а також для аналізу взаємозв'язків між змінними чи параметрами інформаційної системи.

Під час побудови моделі кіберсистеми необхідно визначити перелік змінних, які слід враховувати, та встановити рівень їх деталізації або агрегування. Важливо розмежувати змінні на ті, що виступають вхідними даними, які можна коригувати в процесі моделювання, та ті, що залишаються незалежними від рішень керівних органів безпеки. Прикладами вхідних даних можуть бути параметри мережевого трафіку, рівень доступу користувачів, конфігурації систем безпеки або обсяги ресурсів для реагування на інциденти.

Керовані змінні у моделюванні кібербезпеки обираються під час експерименту або планування заходів захисту. До них належать, наприклад, рівень інтенсивності моніторингу мережевого трафіку, кількість ресурсів, виділених на реагування на інциденти, налаштування систем виявлення загроз або масштаби впровадження нових політик безпеки — усі вони спрямовані на підвищення ефективності захисту інформаційної системи та оптимізацію стратегій кібербезпеки.

Некеровані змінні, навпаки, не підлягають безпосередньому контролю під час експерименту, але істотно впливають на результати моделювання та

можливість досягнення поставлених цілей. Прикладами таких змінних у кібербезпеці є ймовірність виникнення нових вразливостей, поведінка користувачів, особливості дій зловмисників або затримки в оновленні програмного забезпечення.

У випадках, коли завдання в сфері кібербезпеки має повну визначеність і допускає значну кількість варіантів реагування на загрози, для пошуку оптимальних рішень застосовують методи дослідження операцій. Багато управлінських і технічних завдань у сфері інформаційної безпеки належать саме до такого типу, а найефективнішим інструментом їх вирішення виступають методи математичного моделювання.

Задачі з неповною визначеністю містять параметри з високим рівнем невизначеності, наприклад, потенційні дії зловмисників або швидкість поширення шкідливого ПЗ. Для їх аналізу та пошуку рішень використовується системний аналіз, який поєднує математичне моделювання з іншими аналітичними та евристичними методами.

Невизначені задачі, як правило, характеризуються браком достовірної інформації, тому їх розв’язання значною мірою залежить від експертних оцінок, евристичних підходів, а також досвіду й інтуїції кваліфікованих фахівців із кібербезпеки. Розвиток науково-технічного прогресу та вдосконалення методів моделювання поступово підвищує рівень визначеності таких задач, що сприяє їх кращій структуризації.

Етап збору даних у кібербезпеці спрямований на перевірку адекватності моделі та практичну застосовність отриманих результатів. Для цього обирають відповідну методику тестування, використовуючи як аналітичні, так і експериментальні підходи, що дозволяють оцінити чутливість, реалістичність і логічну узгодженість моделі.

Перевірка чутливості здійснюється шляхом аналізу зміни результатів моделювання при варіюванні вхідних параметрів — наприклад, інтенсивності атак, доступності ресурсів чи ефективності заходів захисту. Реалістичність

моделі визначається порівнянням її структури з реальною кіберінфраструктурою, організаційними процедурами та можливими сценаріями атак. Логічна узгодженість перевіряється особливо при екстремальних значеннях вхідних даних.

Важливим елементом тестування є використання контрольних даних із реальних інцидентів або відомих сценаріїв атак, що дозволяє об'єктивно оцінити властивості моделі.

Отже, застосування математичного моделювання та системного аналізу у сфері кібербезпеки спрямоване на отримання практично значущих результатів, які підвищують ефективність управління ризиками, прогнозування загроз і прийняття обґрунтованих рішень щодо захисту інформаційних систем.

6.3. Модель типу «життєвий цикл»

У процесі дослідження та моделювання динамічних систем, зміни в яких мають еволюційний характер, тобто проявляються як послідовність перетворень у функціонуванні системи, її структурі та взаємодії із зовнішнім середовищем, важливого значення набуває концепція життєвого циклу системи та різні її моделі.

Сам термін «життєвий цикл» виник у зв'язку з біологічними системами, зокрема живими організмами, життя яких чітко поділяється на певні стадії: зародження, розвиток, зрілість, старіння та відмирання.

Аналогічний підхід застосовують і під час системного аналізу штучних систем, де період існування також можна розділити на послідовні етапи — від створення до занепаду чи зникнення. Такий підхід дає можливість будувати динамічні моделі функціонування системи, базуючись на декомпозиції процесу за стадіями: від початкової до завершальної.

Подібний принцип можна простежити, наприклад, у житті людини, яке умовно поділяють на молодість, зрілість і старість; або у розвитку наукових

теорій чи процесі розв’язання певних проблем, які теж мають свій життєвий цикл.

У сучасній практиці управління кіберзахистом та планування заходів безпеки ґрунтується на концепції життєвого циклу інформаційних систем і програмних продуктів. Організація захисту, моніторинг і реагування значною мірою базуються на моделі життєвого циклу програмного забезпечення та цифрових сервісів. Особливий інтерес становлять моделі життєвого циклу проєктів у сфері кібербезпеки, оскільки вони дозволяють системно планувати як профілактичні, так і реактивні дії протягом усього періоду експлуатації інформаційних систем. (рис. 6.2).

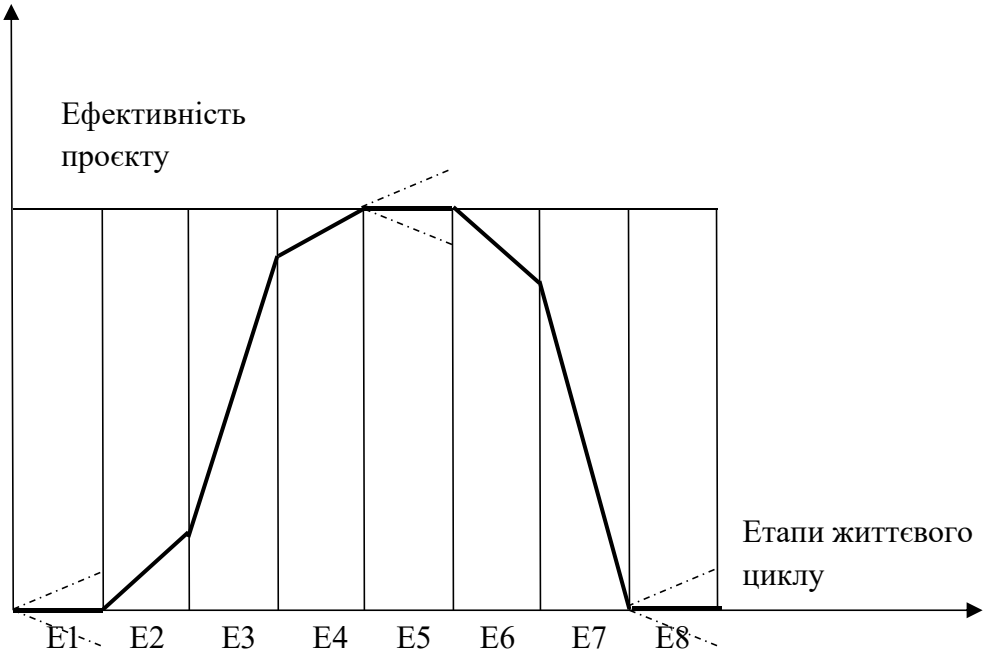


Рисунок 6.2 – Динаміка розвитку проєкту в сфері кібербезпеки

Розвиток будь-якого кіберпроєкту або системи захисту можна описати кривою життєвого циклу, що включає вісім етапів: планування й формування задумів (E1), впровадження засобів захисту (E2), зростання ефективності системи (E3), зрілість і стабільне функціонування (E4), насичення, коли ресурсність і потужності досягають максимуму (E5), поступовий занепад через появу нових загроз чи технологічну відсталість (E6), руйнування або критична втрата актуальності (E7), та остаточна ліквідація або заміна системи (E8).

Ці вісім етапів включають як прогресивний, так і регресивний розвиток. Позитивна динаміка (E2, E3, E4) відображає зростання ефективності й прогресивний розвиток системи безпеки. Негативна динаміка (E6, E7) вказує на регрес і втрату актуальності. Етапи E1, E5 та E8 можуть характеризувати як прогресивний розвиток (наприклад, стратегічне оновлення чи модернізація), так і регресивний (наприклад, втрата гнучкості чи затримка з оновленнями) – залежно від напряму розвитку, що позначається у моделі пунктирною лінією.

Перелік запитань для самопідготовки:

1. Що таке модель у контексті економічної динаміки?
2. Яке призначення моделей у дослідженні соціально-економічних процесів?
3. У чому полягає суть методу моделювання?
4. Які бувають типи економічних моделей?
5. Які вимоги висуваються до економічної моделі?
6. Які основні етапи процесу моделювання соціально-економічних систем?
7. Що таке формалізація, і яку роль вона відіграє у моделюванні?
8. Як відбувається перевірка й валідація моделі?
9. Чому важливо враховувати зворотний зв’язок у моделях соціально-економічних процесів?
10. Як відбувається адаптація моделі до зміни зовнішніх умов?
11. Що таке модель «життєвого циклу» в економічному прогнозуванні?
12. Які етапи включає модель життєвого циклу (виведення на ринок, зростання, зрілість, спад)?
13. Як змінюється економічна активність на різних етапах життєвого циклу?
14. Як модель «життєвого циклу» допомагає в прийнятті управлінських рішень?

Тема 7. Аналітичне вирівнювання тренда (криві зростання) як метод середньо-строкового прогнозування

7.1. Поняття про аналітичне вирівнювання тренду.

7.2. Методу перевірки різниць середніх рівнів.

7.3. Метод Форстера-Стюарта.

7.4. Метод Ірвіна.

7.1. Поняття про аналітичне вирівнювання тренда

Для аналізу тенденції на основі динамічних рядів і побудови прогнозу з врахуванням закономірностей, що склалися в «передісторії», широко застосовується залежність, яка має назву рівняння тренда:

$$y = f(t) + u_t, \tag{7.1}$$

де $f(t)$ – детермінована випадкова компонента процесу (явища),

u_t – стохастична випадкова компонента процесу.

У випадку функціональної залежності кожному значенню аргументу відповідає лише одне певне значення функції. Натомість у стохастичній закономірності одному значенню аргументу може відповідати кілька можливих значень функції, що утворюють певний розподіл.

Тренд відображає напрям і характер змін процесу з часом, тобто є усередненою оцінкою тенденцій його розвитку на основі попередніх даних.

Одним із поширених методів прогнозування у сфері кібербезпеки є аналітичне вирівнювання тренда. Його основна мета полягає у створенні середньострокового прогнозу розвитку кіберзагроз або ефективності систем захисту на майбутній період. Суть методу полягає у виборі математичної функції (моделі тренда), яка найбільш точно відображає динаміку атак чи змін у рівні безпеки, у визначенні її параметрів та використанні цієї функції для прогнозування подальшого розвитку ситуації. Екстраполяція тренда може застосовуватися лише тоді, коли розвиток процесу (наприклад, зростання

кількості атак певного типу) добре описується побудованим рівнянням, а умови, що визначали тенденцію в минулому, не зазнають істотних змін у майбутньому. При додержані цих умов екстраполяція здійснюється шляхом підстановки у рівнянні тренда $y = f(t) + u_t$, значення незалежної змінної t, яка відповідає величині горизонту прогнозування:

$$\hat{y}t + \mathrm{L} = f(t_n + L), \tag{7.2}$$

де L – величина горизонту прогнозування (період, на який складається прогноз).

Ці умови зумовлюють необхідність проведення попереднього аналізу часових рядів економічних показників, що передбачає виявлення та усереднення аномальних значень рівнів ряду, а також визначення трендової складової у вихідному часовому ряді.

7.2. Методу перевірки різниць середніх рівнів

Пошук оптимальної прогнозної моделі здійснюється в межах двох типів часових рядів – стаціонарних і нестаціонарних. Для цього проводять перевірку стаціонарності та оцінку наявності тренда (ідентифікацію тренда) за допомогою різних методів. Стаціонарні часові ряди характеризуються відсутністю тренда, а також незмінністю середнього значення та дисперсії в часі.

Початковий етап аналізу часових рядів передбачає виявлення й усунення аномальних спостережень, а також встановлення факту наявності або відсутності тренда у вихідних даних. Аномальним вважається таке значення ряду, яке не відповідає реальним можливостям досліджуваної економічної системи або істотно спотворює ключові характеристики ряду, зокрема параметри трендової моделі.

Причини появи аномалій можуть бути різними. Якщо вони зумовлені технічними похибками чи іншими помилками першого роду, такі спостереження необхідно ідентифікувати та вилучати з аналізу. У випадку ж,

коли відхилення спричинені об'єктивними чинниками, що проявляються випадково й нерегулярно (помилки другого роду), їх усунення неможливе.

Для виявлення тренда зазвичай застосовують метод автокореляційного аналізу, тоді як для фіксації аномальних значень використовують критерій Ірвіна. Крім того, поширеними інструментами ідентифікації тренда є аналіз різниць середніх рівнів та метод Форстера–Стюарта.

Реалізація методу перевірки різниць середніх рівнів складається з таких кроків.

Крок перший. Вихідний часовий ряд y_1, y_2, y_3,…y_n розділяється на дві приблизно однакові за кількістю рівнів частини: у першій частині n_1 перших рівнів вихідного ряду, у другій – n_2 решта рівнів ($n_1+n_2=n$).

Крок другий. Для кожної з цих частин розраховуються середні значення і дисперсії:

$$\bar{y}_1 = \frac{\sum_{t=1}^{n_1} y_t}{n_1};\ \sigma_1^2 = \frac{\sum_{t=1}^{n_1} (y_t - \bar{y}_1)^2}{n_1 - 1} \tag{7.3}$$

$$\bar{y}_2 = \frac{\sum_{t=1}^{n_2} y_t}{n_2};\ \sigma_2^2 = \frac{\sum_{t=1}^{n_2} (y_t - \bar{y}_2)^2}{n_2 - 1} \tag{7.4}$$

Крок третій. Перевірка однаковості (однорідності) дисперсій обох частин ряду за допомогою *F*-критерію Фішера:

$$F = \begin{cases} \frac{\sigma_1^2}{\sigma_2^2}. \textit{якщо}\ \sigma_1^2 > \sigma_2^2 \\ \frac{\sigma_2^2}{\sigma_1^2}. \textit{якщо}\ \sigma_1^2 < \sigma_2^2 \end{cases} \tag{7.5}$$

Отримане розрахункове значення порівнюється з табличним (критичним) значенням критерію Фішера F_α із заданим рівнем значущості (рівнем помилки) α. Якщо розрахункове значення *F*-критерію менше за табличне F_α, то гіпотеза про рівність дисперсій приймається і слід перейти до четвертого кроку. Якщо *F*

більше або дорівнює F_α, гіпотеза про рівність дисперсій відхиляється і робиться висновок, що цей метод не має відповіді про наявність тренда.

На *четвертому кроці* перевіряється гіпотеза про відсутність тренда за допомогою *t*-критерію Стьюдента. Для цього визначається розрахункове значення критерію Стьюдента за формулою:

$$t = \frac{|\bar{y}_1 - \bar{y}_2|}{\sigma\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}} \tag{7.6}$$

де σ – середньоквадратичне відхилення різниць середніх, розраховане за формулою:

$$\sigma = \sqrt{\frac{(n_1 - 1)\sigma_1^2 + (n_2 - 1)\sigma_2^2}{n_1 + n_2 - 2}} \tag{7.7}$$

Якщо розрахункове значення t менше за табличне t_α, то нульова гіпотеза не відхиляється, тобто тренд відсутній, інакше тренд в аналізованому динамічному ряді є. Зазначимо, що в цьому разі табличне значення t_α береться для числа ступенів свободи, яке дорівнює n_1+n_2-2, при цьому цей метод застосовується тільки для рядів з монотонною тенденцією. Недолік полягає у неможливості правильно визначити існування тренда у випадку, коли часовий ряд містить точку зміни тенденції всередині ряду.

Поділимо вихідний часовий ряд на дві частини по 5 елементів в кожній. Для кожної з цих частин розраховуються середні значення і дисперсії:

Таблиця 7.1

1 частина:

1	2	3	4	5	$\bar{y}_1 = \frac{451,8}{5} = 90,36$	σ_1^2
60,9	68,8	84,0	107,4	130,7		832,83

Таблиця 7.2

2 частина:

1	2	3	4	5	$\bar{y}_2 = \frac{1185,1}{5} = 237,02$	σ_2^2
169,7	193,7	237,9	267,8	316,0		3398,66

Оскільки $\sigma_2^2 > \sigma_1^2$, то $F = \frac{3398,66}{832,83} = 4,08$

Табличне значення *F*-критерію Фішера при рівні значущості α =0,05 дорівнює F_α=5,05.

Оскільки $F < F_\alpha$, то переходимо до наступного етапу.

На наступному етапі перевіряється гіпотеза про відсутність тренда за допомогою t-критерію Стьюдента. Для цього визначимо розрахункове значення критерію Стьюдента:

$$t = \frac{|90,36 - 237,02|}{46\sqrt{\frac{1}{5} + \frac{1}{5}}} = 5,04$$

$$\sigma = \sqrt{\frac{(5-1)\cdot 832,83 + (5-1)\cdot 3398,66}{5+5-2}} = 46$$

Табличне значення t_α=2,3060 береться для числа ступенів свободи, яке дорівнює $n_1+n_2-2=8$. Оскільки $t_\alpha < t$, то гіпотеза про відсутність тренда в аналізованому динамічному ряді відхиляється.

Отже, метод перевірки різниць середніх рівнів підтвердив наявність тренда у вихідному часовому ряді.

7.3. Метод Форстера-Стюарта.

Метод Форстера–Стюарта вирізняється ширшими можливостями та забезпечує більш достовірні результати. На відміну від підходів, що враховують лише тренд середніх значень ряду, він дає змогу виявляти також тенденції у зміні дисперсії часового ряду. У випадку відсутності тренда дисперсії розкид значень залишається стабільним, тоді як зростання дисперсії свідчить про поступове «розхитування» ряду. Практична реалізація цього методу передбачає виконання чотирьох послідовних етапів.

Етап перший. Порівнюється кожний рівень вихідного часового ряду y_1, y_2, y_3,…y_n, починаючи із другого рівня, з усіма попередніми, при цьому визначаються дві числові послідовності:

$$k_t = \begin{cases} 1, \text{ якщо } y_t \text{ більше кожного з усіх попередніх рівнів,} \\ 0, \text{ у протилежному випадку,} \end{cases}$$

Зауваження: якщо ряд є монотонно зростаючим, то ця послідовність складається виключно з 1, і навпаки, при монотонно убуваючому динамічному ряді послідовність складається з 0.

$$l_t = \begin{cases} 1, \text{ якщо } y_t \text{ менше кожного з усіх попередніх рівнів,} \\ 0, \text{ у протилежному випадку,} \end{cases}$$

Зауваження: якщо ряд монотонно убуваючий, то дана послідовність складається виключно з 1, і навпаки, при монотонно зростаючому динамічному ряді послідовність складається з 0.

Етап другий. Розраховуються величини s і d:

$$s = \sum_{t=2}^{n} (k_t + l_t) \tag{7.8}$$

$$d = \sum_{t=2}^{n} (k_t - l_t) \tag{7.9}$$

Неважко помітити, що величина s, яка характеризує зміну часового ряду, набуває значення від 0 (усі рівні ряду однакові або ряд монотонно спадаючий) до n-1 (ряд монотонний зростаючий).

Величина d характеризує зміну дисперсії рівнів часового ряду і змінюється від -(n-1) (ряд поступово спадає) до (n-1) (ряд поступово зростає).

Етап третій. Перевіряється гіпотеза про те, чи можна вважати випадковими:

1) відхилення величини s від μ – математичного сподівання величини s для ряду, в якому рівні розташовані випадково;

2) відхилення величини d від нуля.

Ця перевірка проводиться з використанням розрахункових значень *t*-критерію Стьюдента для середньої і для дисперсії:

$$t_s = \frac{|s-\mu|}{\sigma_1}; \sigma_1 = \sqrt{2\ln n - 3{,}4253} \tag{7.10}$$

$$t_d = \frac{|d-0|}{\sigma_2}; \sigma_2 = \sqrt{2\ln n - 0{,}8456}, \tag{7.11}$$

де μ – математичне сподівання величини s для ряду, в якому рівні розташовані випадково;

σ_1 – середньоквадратичне відхилення для величини s;

σ_2 – середньоквадратичне відхилення для величини d.

Для зручності розраховані табульовані значення величин μ, σ_1, σ_2; фрагмент цих значень подано в таблиці 7.3.

Таблиця 7.3

Значення μ, σ_1, σ_2 для розрахунків

Значення	10	20	30	40
μ	3,858	5,195	5,990	6,557
σ_1	1,288	1,677	1,882	2,019
σ_2	1,964	2,279	2,447	2,561

Етап четвертий. Розрахункові значення t_s і t_d порівнюються з табличними значеннями t-критерію Стьюдента із заданим рівнем значущості t_α. Якщо розрахункове значення t менше за табличне t_α, то гіпотеза про відсутність відповідного тренда приймається, у протилежному випадку – тренд існує. Наприклад, якщо t_s більше від табличного значення t_α, а t_d менше від t_α, то для заданого часового ряду існує тренд у середньому, а тренда дисперсії рівнів ряду немає.

Додатково при аналізі динамічного ряду, як вже зазначалося вище, необхідно виявити наявність аномальних рівнів. Для цього використовуються методи, розраховані для статистичних сукупностей.

Аналіз вихідного динамічного ряду свідчить, що він є монотонно зростаючим, отже послідовності k_t та l_t такі:

Таблиця 7.4

Значення для розрахунків

t	2	3	4	5	6	7	8	9	10
k_t	1	1	1	1	1	1	1	1	1
l_t	0	0	0	0	0	0	0	0	0

Розрахуємо, таким чином показники s=9, d=9. На основі цього здійснимо перевірку відхилення величини s від μ – математичного сподівання величини s для ряду, в якому рівні розташовані випадково, та відхилення величини d від нуля. Ця перевірка проводиться з використанням розрахункових значень t-критерію Стьюдента для середньої та для дисперсії:

$$t_s = \frac{|9 - 3{,}858|}{1{,}288} = 3{,}9922,$$

$$t_d = \frac{|9 - 0|}{1{,}964} = 4{,}5825.$$

Розрахункові значення t_s і t_d порівнюються з табличними значеннями t-критерію Стьюдента із заданим рівнем значущості (α=0,05) t_α=2,2622. Табличне значення береться для числа ступенів свободи, яке дорівнює n-1=9. Оскільки t_α<t, то тренд в аналізованому динамічному ряді є.

7.4. Метод Ірвіна

Метод Ірвіна передбачає використання такої формули:

$$\lambda_t = \frac{|y_t - y_{t-1}|}{\sigma_y} \qquad (7.12)$$

Чисельник обчислюється для кожного рівня динамічного ряду, окрім першого. Середньоквадратичне відхилення, в свою чергу, визначається за формулами:

$$\sigma_y = \sqrt{\frac{\sum_{t=1}^{n}(y_t - \bar{y})^2}{n-1}} \quad (7.13)$$

$$\bar{y} = \frac{\sum_{t=1}^{n} y_t}{n} \quad (7.14)$$

Розрахункові значення λ_t, t=2, 3…n порівнюються із табличними значеннями критерію Ірвіна λ_α, і якщо вони більше від табличних, то відповідне значення y_t рівня ряду вважається аномальним. Значення критерію Ірвіна для рівня значимості α=0,05, тобто із 5%-вою похибкою, наведені в таблиці 7.5.

Таблиця 7.5

Табличні значення критерію Ірвіна λ_α, α=0,05

n	2	3	10	20	30	50	100
λ_α	2,8	2,3	1,5	1,3	1,2	1,1	1,0

Після ідентифікації аномальних значень ряду обов'язковим етапом є з'ясування причин їх появи. Якщо достеменно встановлено, що такі відхилення спричинені помилками першого роду, вони підлягають усуненню. Корекція може здійснюватися двома способами: заміною аномального рівня простим середнім арифметичним двох сусідніх значень ряду або ж підстановкою відповідних величин, визначених на основі апроксимувальної кривої даного часового ряду.

У таблиці 7.6 наведена інформація про кількість здійснених інцидентів інформаційної безпеки в Україні протягом останніх 10 років. Необхідно ідентифікувати тренд у наведеному часовому ряді та визначити наявність у ньому аномальних рівнів.

Таблиця 7.6

Ємність ринку

Рік	1	2	3	4	5	6	7	8	9	10
Кількість	60,9	68,8	84,0	107,4	130,7	169,7	193,7	237,9	267,8	316,0

інцидентів інформаційної безпеки, тис.										

Для ідентифікації тренду використаємо два методи.

Аналіз вихідного динамічного ряду свідчить, що він є монотонно зростаючим, отже послідовності k_t та l_t такі:

Таблиця 7.7

Значення для обчислень

t	2	3	4	5	6	7	8	9	10
k_t	1	1	1	1	1	1	1	1	1
l_t	0	0	0	0	0	0	0	0	0

Обчислимо таким чином показники s=9, d=9. На основі цього здійснимо перевірку відхилення величини s від μ – математичного сподівання величини s для ряду, в якому рівні розташовані випадково, та відхилення величини d від нуля. Ця перевірка проводиться з використанням розрахункових значень t-критерію Стьюдента для середньої та для дисперсії:

$$t_s = \frac{|9 - 3{,}858|}{1{,}288} = 3{,}9922,$$

$$t_d = \frac{|9 - 0|}{1{,}964} = 4{,}5825.$$

Розрахункові значення t_s і t_d порівнюються з табличними значеннями t-критерію Стьюдента із заданим рівнем значущості (α=0,05) t_α=2,2622. Табличне значення береться для числа ступенів свободи, яке дорівнює n-1=9. Оскільки $t_\alpha < t$, то тренд в аналізованому динамічному ряді є.

Розрахуємо середньоквадратичне відхилення:

$$\sigma = \sqrt{\frac{70627{,}24}{9}} = 88{,}59$$

Розрахуємо показник Ірвіна:

Таблиця 7.8

Значення показника Ірвіна

t	2	3	4	5	6	7	8	9	10
λ_t	0,09	0,17	0,26	0,26	0,44	0,27	0,5	0,34	0,54

Оскільки всі розраховані значення λ_t менші від табличного значення показника Ірвіна $\lambda_{0,05}$=1,5, що свідчить про відсутність аномальних рівнів динамічного ряду.

Перелік запитань для самопідготовки:

1. Що таке аналітичне вирівнювання тренда?
2. Для чого застосовується аналітичне вирівнювання у прогнозуванні?
3. Які функціональні форми тренда найчастіше використовуються?
4. Як обирається крива зростання для побудови прогнозу?
5. У чому переваги аналітичного вирівнювання над простими середніми?
6. Що таке різниця середніх рівнів у часовому ряді?
7. Як використовується метод перевірки цих різниць для аналізу тренда?
8. Які статистичні критерії застосовуються для перевірки значущості різниць?
9. У яких випадках доцільно застосовувати цей метод?
10. Як інтерпретуються результати перевірки різниць?
11. У чому полягає суть методу Форстера-Стюарта?
12. Які основні припущення лежать в основі цього методу?
13. Як формується тренд за методом Форстера-Стюарта?
14. Які дані необхідні для застосування цього методу?
15. Яке основне призначення методу Ірвіна в аналізі тренда?
16. Як за допомогою цього методу виявляються відхилення в часовому ряді?
17. Які обчислення необхідні для методу Ірвіна?
18. У чому переваги методу Ірвіна над іншими методами контролю тренда?

Тема 8. Методи екстраполяції тенденції та побудови прогнозу по одному часовому ряду

8.1. Поняття тенденції, способи встановлення наявності тенденції.

8.2. Прості методи екстраполяції тенденції:

8.2.1. Екстраполяція на основі аналітичних показників рядів динаміки.

8.2.2. Екстраполяція на основі плинної середньої.

8.3.3. Екстраполяція на основі індексу сезонності.

8.1. Поняття тенденції, способи встановлення наявності тенденції

У кібербезпеці тенденція трактується як загальний напрям розвитку кіберзагроз чи захисних технологій, що відображає їхню довгострокову закономірність. При застосуванні методів екстраполяції вихідним положенням виступає інерційність процесів, які аналізуються й прогнозуються.

Рівень інерційності залежить від масштабу та характеру процесу: на мікрорівні (наприклад, атака на окремий сервер) навіть один фактор може миттєво змінити ситуацію, тоді як на макрорівні (глобальні тренди атак у кіберпросторі) інерційність зберігається більш стабільно завдяки взаємодії великої кількості різноспрямованих чинників.

У випадку високої інерційності розвитку кіберзагроз можна з достатнім ступенем імовірності припустити, що закономірності, які проявилися в попередніх періодах (наприклад, зростання кількості фішингових атак чи шкідливого ПЗ), зберігатимуть свою дію й у прогнозованому періоді, змінюючись лише незначною мірою.

Основою екстраполяційних методів прогнозування є динамічні ряди (наприклад, статистика інцидентів за роками). Для перевірки наявності тенденції у такому ряді застосовують різні підходи. Один із найпростіших методів – порівняння середніх значень. Для цього динамічний ряд поділяють на дві приблизно рівні частини, кожну з яких розглядають як умовно самостійну

вибірку. Якщо тенденція дійсно існує, середні значення цих частин істотно відрізнятимуться. Якщо ж розходження випадкові й незначні — вважається, що тенденція відсутня.

Для оцінки істотності відмінності між середніми значеннями двох динамічних рядів використовується *t*-критерій Стьюдента.

Розходження буде істотним, якщо розрахункове значення *t*-критерію Стьюдента *(t_p)* буде не меншим за його табличне значення *(t_T)*.

Розрахункове значення *t*-критерію обчислюється таким чином:

$$t_p = \frac{|\bar{x}_1 - \bar{x}_2|}{S_{\bar{x}_1 - \bar{x}_2}} \tag{8.1}$$

де $\bar{x}_1, \bar{x}_2$ – середнє значення рівня відповідно до першої і другої частин ряду, розрахованих для інтервальних динамічних рядів як середнє арифметичне:

$$S_{\bar{x}_1 - \bar{x}_2} = \sqrt{\frac{(n_1 - 1)S_1^2 + (n_2 - 1)S_2^2}{(n_1 - 1) + (n_2 - 1)} \cdot \left(\frac{1}{n_1} + \frac{1}{n_2}\right)} \tag{8.2}$$

де n_1, n_2 – кількість елементів відповідно першої і другої частин ряду.

$S_{\bar{x}_1 - \bar{x}_2}$ – стандартне відхилення;

S_1^2, S_2^2 - дисперсія відповідного першої і другої частин ряду.

$$S_1^2 = \frac{\sum(x - \bar{x}_1)^2}{n_1 - 1}; S_2^2 = \frac{\sum(x - \bar{x}_2)^2}{n_2 - 1} \tag{8.3}$$

8.2. Прості методи екстраполяції тенденції

Прості методи прогнозування, що ґрунтуються на екстраполяції тенденцій, широко застосовуються у сфері кібербезпеки, оскільки мають низку переваг. Зокрема, вони дозволяють швидко оцінити динаміку розвитку кіберзагроз, визначити загальні напрями їхньої еволюції та забезпечити базу

для ухвалення управлінських рішень щодо підвищення рівня захисту інформаційних систем.

До переваг простих методів слід віднести:

- достатньо простий апарат дослідження, що привертає до нього широке коло спеціалістів;

- можливість використання для виконання розрахунків портативних і нескладних обчислювальних засобів;

- швидкість виконання розрахунків в оперативному режимі;

- наявність відносно невеликого масиву інформації.

Нижче наведені прості методи екстраполяції тенденції на основі застосування аналітичних показників динамічних рядів, плинної середньої та індексу сезонності.

8.2.1. Екстраполяція на основі аналітичних показників рядів динаміки

Динамічний ряд (ряд динаміки) являє собою послідовність показників, які відображають зміни певного явища, процесу чи об’єкта у часі. Окремі значення цього ряду називають рівнями.

Залежно від способу відображення часу, динамічні ряди поділяють на моментні та інтервальні. У моментних рядах рівні характеризують величину явища на конкретну дату, наприклад, залишки готової продукції на перше число кожного місяця або вартість основних фондів на початок чи кінець року. В інтервальних рядах рівні показують розмір явища за певний проміжок часу, як-от випуск продукції за місяць, квартал чи рік.

Під час побудови динамічних рядів ключову увагу приділяють забезпеченню порівнянності рівнів. Це означає, що всі показники повинні виражатися в однакових одиницях виміру, обчислюватися за єдиною методологією та охоплювати однакове коло об’єктів.

Позначимо:

y_1 — початкове значення рівня динамічного ряду;

y_n — кінцеве значення рівня динамічного ряду;

y_i — умовно прийнятий (i-й) рівень динамічного ряду;

n — кількість елементів динамічного ряду.

Наведемо основні аналітичні показники динамічного ряду, які використовуються у прогнозуванні:

а) абсолютний приріст:

1) ланцюговий

$$\Delta' y_i = y_i - y_{i-1}; \tag{8.4}$$

2) базисний

$$\Delta y_i = y_i - y_1; \tag{8.5}$$

б) середній абсолютний приріст

$$\overline{\Delta y} = \frac{y_n - y_1}{n-1} = \frac{\sum_{i=1}^{n-1} \Delta' y_i}{n-1}; \tag{8.6}$$

в) коефіцієнт росту:

1) ланцюговий

$$K_{P_i} = \frac{y_i}{y_{i-1}}; \tag{8.7}$$

2) базисний

$$K_{P_i} = \frac{y_i}{y_1}; \tag{8.8}$$

3) за весь період

$$K_{P_n} = \frac{y_n}{y_1}; \tag{8.9}$$

г) коефіцієнт приросту

$$K_{np} = k_p - 1; \tag{8.10}$$

д) середній коефіцієнт росту

$$\overline{k}_p = \sqrt[n-1]{\frac{y_n}{y_1}}; \tag{8.11}$$

е) середній коефіцієнт приросту

$$\bar{k}_{np} = \bar{k}_p - 1; \tag{8.12}$$

ж) абсолютний розмір 1% приросту:

1) ланцюговий

$$\Delta y_{1\%} = \frac{y_i - y_{i-1}}{\frac{y_i - y_{i-1}}{y_{i-1}} 100} = \frac{y_{i-1}}{100}; \tag{8.13}$$

2) за весь період

$$\bar{\Delta} y_{1\%} = \frac{\bar{\Delta} y}{\bar{k}_{np}}; \tag{8.14}$$

з) коефіцієнт випередження (відставання)

$$k = \frac{y_i}{y_{i-1}} : \frac{x_i}{x_{i-1}}; \tag{8.15}$$

Добуток ланцюгових коефіцієнтів росту дорівнює базисному коефіцієнту росту за весь період, тобто

$$k'_{p_1} \cdot k'_{p_2} \cdot k'_{p_3} \ldots\ldots k'_{p_n} = k_{p_n} \tag{8.16}$$

що може бути доведено таким чином.

Нехай маємо динамічний ряд за 5 років та рівень показника за базисний рік, що передує рокам п'ятирічки (y_0). Добуток ланцюгових коефіцієнтів росту складе

$$\frac{y_1}{y_0} \cdot \frac{y_2}{y_1} \cdot \frac{y_3}{y_2} \cdot \frac{y_4}{y_3} \cdot \frac{y_5}{y_4} = \frac{y_5}{y_0};$$

тобто підтверджується залежність 8.16.

На основі наведених аналітичних показників, які широко застосовуються для оцінки динамічних рядів, можна вивести залежності, що можуть бути використані для побудови прогнозів:

$$\hat{y}_{n+1} = y_n + \Delta' y_n; \Delta' y_n = y_n - y_{n-1} \tag{8.17}$$

$$\hat{\hat{y}}_{n+T} = y_n + \bar{\Delta} y \cdot T \tag{8.18}$$

$$\hat{\hat{y}}_{n+1} = y_n \cdot k_{p_n}; k_{p_n} = \frac{y_n}{y_{n-1}} \tag{8.19}$$

$$\hat{\hat{y}}_{n+T} = y_n \cdot \bar{k}_p^T \qquad (8.20)$$

де $\hat{\hat{y}}$ – тут і далі таким чином позначаються прогнозні значення показника.

T – величина горизонту прогнозу *(*T = 1; 2; 3 ...)

Практичне застосування зазначених залежностей покажемо на прикладах.

Розклянемо наступний приклад.

В таблиці 8.1 наведені дані про середню кількість зареєстрованих кіберінцидентів по області за рік. Використовуючи рівняння (18; 20), побудуємо прогноз на наступну п'ятирічку.

Таблиця 8.1

Середня кількість зареєстрованих кіберінцидентів по області

Номер року, t	Постачання певного продукту в рік	Номер року, t.	Постачання певного продукту в рік
1	2	3	4
1	10,7	7	15,9
2	11,5	8	17,2
3	12,2	9	18,1
4	13,4	10	19,8
5	15,0	11	21,2
6	15,0		

У таблиці 8.1 наведені дані по двох п'ятирічках та базисному року.

Остаточно про якість прогнозу можна судити лише після того, як подія відбулася. Щоб оцінити надійність застосованого методу, використовують так званий метод «прогноз екс-пост». Такий підхід застосовується і для інших кількісних методів прогнозування.

Використовуючи дані перших шести років – базисний рік та роки першої п'ятирічки, розрахуємо відповідно:

Середній абсолютний приріст

$$\bar{\Delta}y = \frac{y_k - y_0}{k-1} = \frac{15,0-10,7}{6-1} = \frac{4,3}{5} = 0,9_i$$

Середньорічний коефіцієнт росту

$$\overline{k}_p = \sqrt[k-1]{\frac{y_k}{y_1}} = \sqrt[6-1]{\frac{15,0}{10,7}} = \sqrt[5]{1,40} = 1,07.$$

На основі залежності (3.18) складемо прогноз кіберінцидентів по області на період $(\hat{e}+1) \div n$.

$$\hat{\bar{y}}_{k+1} = 15 + 0,9 \cdot 1 = 15,9 \quad \hat{\bar{y}}_{k+2} = 15 + 0,9 \cdot 2 = 16,8$$

$$\hat{\bar{y}}_{k+3} = 15 + 0,9 \cdot 3 = 17,7 \quad \hat{\bar{y}}_{k+4} = 15 + 0,9 \cdot 4 = 18,6$$

$$\hat{\bar{y}}_{k+5} = 15 + 0,9 \cdot 5 = 19,5$$

Результати розрахунків зведені в таблицю та порівняні з фактичними даними.

Складемо прогноз кіберінцидентів по області на основі формули 8.20.

$$\hat{\bar{y}}_{k+1} = 15 \cdot 1,07^1 = 16 \quad \hat{\bar{y}}_{k+2} = 15 \cdot 1,07^2 = 17,2$$

$$\hat{\bar{y}}_{k+3} = 15 \cdot 1,07^3 = 18,4 \quad \hat{\bar{y}}_{k+4} = 15 \cdot 1,07^4 = 19,7$$

$$\hat{\bar{y}}_{k+5} = 15 \cdot 1,07^5 = 21$$

Результати прогнозу порівняні із фактичними даними та оцінена якість прогнозу (табл. 8.2).

Таблиця 8.2

Оцінка якості прогнозу, складеного на основі середнього абсолютного приросту

№	Фактичне значення	Прогнозоване значення	Відхилення	
			Абсолютне (гр2-гр3)	Відносне (гр4:гр2)100%
1	2	3	4	5
1	15,9	15,9	0	0
2	17,2	16,8	0,4	2,3
3	18,1	17,7	0,4	2,2
4	19,8	18,6	1,2	6,1
5	21,2	19,5	1,7	8
	Середнє значення	—	0,7	3,7

Таблиця 8.3

Оцінка якості прогнозу, складеного на основі середньорічного коефіцієнта росту

№	*Фактичне значенняг*	*Прогнозоване значення*	*Відхилення*	
			Абсолютне (гр2-гр3)	*Відносне (гр4:гр2)100%*
1	*2*	*3*	*4*	*5*
1	15,9	16	-0,1	-0,6
2	17,2	17,2	0	0
3	18,2	18,4	-0,3	-1,7
4	19,8	19,7	0,1	0,5
5	21,2	21	0,2	0,9
	Середнє значення	—	0,1	0,7

Порівняння результатів прогнозування, наведених у таблицях 2 та 3, дає підстави стверджувати, що застосування середньорічного коефіцієнта зростання забезпечує вищу точність прогнозів. Це підтверджується величинами відхилень як за окремими роками, так і за підсумками всього п’ятирічного періоду.

З метою побудови прогнозу на перспективу, тобто за межами наявних статистичних даних, доцільно обчислити середньорічний коефіцієнт зростання на основі показників другої п’ятирічки, використовуючи за відправну точку базисний період.

$$\overline{k}_p = \sqrt[5]{\frac{21,2}{15,0}} = \sqrt{1,413} = 1,071$$

Прогноз на наступне п'ятиріччя становитиме:

$$\hat{y}_{n+1} = 21,2 \cdot 1,071^1 = 22,7 \quad \hat{y}_{n+2} = 21,2 \cdot 1,071^2 = 24,3$$

$$\hat{y}_{n+3} = 21,2 \cdot 1,071^3 = 26,0 \quad \hat{y}_{n+4} = 21,2 \cdot 1,071^4 = 28$$

$$\hat{y}_{n+5} = 21,2 \cdot 1,071^5 = 29,9$$

Прогноз зроблено з урахуванням зберігання тенденцій, які склалися в «передісторії».

Важливим недоліком показників середнього абсолютного приросту та середнього коефіцієнта зростання є те, що їхні значення формуються виключно на основі крайніх рівнів динамічного ряду. При цьому проміжні спостереження, які часто суттєво, а подекуди й визначально впливають на загальну тенденцію зміни показників, фактично не враховуються у розрахунках. Усунути цей недолік значною мірою дозволяє застосування методів аналітичного вирівнювання динамічних рядів.

8.2.2. Екстраполяція на основі плинної середньої

Метод плинної середньої базується на використанні залежності:

$$\Delta x_t + 1 = \lambda_t x_t + \lambda_{t-1}\Delta x_{t-1} + \lambda_{t-2}\Delta x_{t-2} + ... + \lambda_{t-(n-1)}\Delta x_{t-(n-1)}, \quad (8.21)$$

де n — кількість років «передісторії».

Коефіцієнт λ_i розраховується за формулою:

$$\lambda_i = \frac{i \cdot \beta}{n}; \quad (8.22)$$

де i — число, яке означає послідовний натуральний ряд «передісторії», починаючи з останнього;

β — визначається за таблицею, поданою нижче:

Таблиця 8.4

Значення показника β

N	3	4	5	6	7	8
β	0,500	0,400	0,333	0,286	0,250	0,222

Визначимо значення λ для п'ятирічки.

Згідно з даними наведеної вище таблиці, при n=5, β=0,333.

Звідси:

$$\lambda_1 = \frac{1 \cdot 0{,}333}{5} = 0{,}067,\ \lambda_2 = \frac{2 \cdot 0{,}333}{5} = 0{,}133,$$

$$\lambda_3 = \frac{3 \cdot 0{,}333}{5} = 0{,}200,\ \lambda_4 = \frac{4 \cdot 0{,}333}{5} = 0{,}267,$$

$$\lambda_5 = \frac{5 \cdot 0{,}333}{5} = 0{,}333.$$

Якщо підставити розраховані значення λ у формулу 8.21, отримаємо:

$$\Delta x_{t+1} = 0{,}333\Delta x_t + 0{,}267\Delta x_{t-1} + 0{,}200\Delta x_{t-2} + 0{,}133\Delta x_{t-3} + 0{,}067\Delta x_{t-4}$$

Особливістю методу плинної середньої є те, що рівень показників, який знаходиться ближче до прогнозованого періоду, чинить більший вплив на значення прогнозованих показників, порівняно з віддаленими періодами. Досягається це завдяки коефіцієнтуλ.

Прогнозні значення показників розраховуються таким чином:

$$\hat{\hat{x}}_{t+1} = x_t + 0{,}333\Delta x_t + 0{,}267\Delta x_{t-1} + 0{,}200\Delta x_{t-2} + 0{,}133\Delta x_{t-3} + 0{,}067\Delta x_{t-4}$$

$$\hat{\hat{x}}_{t+2} = \hat{\hat{x}}_{t+1} + 0{,}333\Delta x_t + 0{,}267\Delta x_{t-1} + 0{,}200\Delta x_{t-2} + 0{,}133\Delta x_{t-3}$$

$$\hat{\hat{x}}_{t+3} = \hat{\hat{x}}_{t+2} + 0{,}333\Delta x_t + 0{,}267\Delta x_{t-1} + 0{,}200\Delta x_{t-2}$$

$$\hat{\hat{x}}_{t+4} = \hat{\hat{x}}_{t+3} + 0{,}333\Delta x_t + 0{,}267\Delta x_{t-1}$$

$$\hat{\hat{x}}_{t+5} = \hat{\hat{x}}_{t+4} + 0{,}333\Delta x_t$$

На основі даних другої п'ятирічки, включаючи також базисний період (табл. 8.1) складемо прогноз настання кіберінцидентів по області на основі методу плинної середньої. Розрахуємо абсолютні прирости:

$$\Delta x_t = 21{,}2 - 19{,}8 = 1{,}4;\ \Delta x_{t-1} = 19{,}8 - 18{,}1 = 1{,}7;\ \Delta x_{t-2} = 18{,}1 - 17{,}2 = 0{,}9;$$

$$\Delta x_{t-3} = 17{,}2 - 15{,}9 = 1{,}3;\ \Delta x_{t-4} = 15{,}9 - 15{,}0 = 0{,}9;$$

Виходячи з наведених вище залежностей

$$\hat{\hat{x}}_{t+1} = 21{,}2 + 0{,}333 \cdot 1{,}4 + 0{,}267 \cdot 1{,}7 + 0{,}200 \cdot 0{,}9 + 0{,}133 \cdot 1{,}3 + 0{,}067 \cdot 0{,}9 = 22{,}5$$

$$\hat{\hat{x}}_{t+2} = 22{,}5 + 0{,}333 \cdot 1{,}4 + 0{,}267 \cdot 1{,}7 + 0{,}200 \cdot 0{,}9 + 0{,}133 \cdot 1{,}3 = 23{,}8$$

$$\hat{\hat{x}}_{t+3} = 23{,}8 + 0{,}333 \cdot 1{,}4 + 0{,}267 \cdot 1{,}7 + 0{,}200 \cdot 0{,}9 = 24{,}9$$

$$\hat{\hat{x}}_{t+4} = 24{,}9 + 0{,}333 \cdot 1{,}4 + 0{,}267 \cdot 1{,}7 = 25{,}8$$

$$\hat{\hat{x}}_{t+5} = 25{,}8 + 0{,}333 \cdot 1{,}4 = 26{,}3$$

Порівняємо розрахунки прогнозу настання кіберінцидентів по області по області, що розраховані на основі середньорічного коефіцієнта росту і плинної середньої (табл. 8.5).

Таблиця 8.5

Прогноз настання кіберінцидентів по області на наступну п'ятирічку, розрахований двома методами

Роки	Результати прогнозу, що розраховані на основі:	
	середнього коефіцієнту росту	плинної середньої
1	22,7	22,5
2	24,3	23,8
3	26,0	24,9
4	28,0	25,8
5	29,9	26,3

Дані таблиці 8.5 демонструють, що прогноз, побудований на основі середньорічних коефіцієнтів зростання, за результатами дещо випереджає прогноз, отриманий за методом ковзної середньої. Водночас наперед визначити, який із прогнозів виявиться точнішим, досить складно.

Суттєвою перевагою методу ковзної середньої є те, що на значення прогнозованих показників певною мірою впливають усі попередні дані, тоді як середньорічний коефіцієнт зростання враховує лише крайні рівні динамічного ряду.

Наявність кількох альтернативних варіантів прогнозу створює можливість для фахівців, спираючись на власний досвід, знання та професійну інтуїцію, обрати найбільш обґрунтований та прийнятний варіант.

8.2.3. Екстраполяція на основі індексу сезонності

У процесі функціонування інформаційних систем та цифрових сервісів організації стикаються з циклічними коливаннями активності кіберзагроз, що зумовлені сезонними або періодичними факторами (наприклад, збільшенням

кількості атак у святкові періоди, під час масових онлайн-продажів чи політичних подій).

Сезонні коливання у кібербезпеці – це відносно сталі внутрішньорічні чи періодичні коливання в динамічних рядах кіберінцидентів, які обумовлені специфікою діяльності організації або зовнішніми чинниками. Для ефективного планування захисту та організації роботи центрів кібербезпеки особливо важливо дослідити сформовані тенденції сезонних коливань атак і побудувати прогноз на найближчу перспективу, насамперед на наступний рік.

Для аналізу таких коливань застосовують спеціальні показники, аналогічні індексам сезонності, що формують так звану «сезонну хвилю кіберзагроз». Вона дозволяє виявити періоди пікового навантаження на системи захисту та підготуватися до них заздалегідь.

На основі даних, що відображають рівень використання та впровадження рішень у сфері кібербезпеки (табл. 8.6), здійснимо розрахунок індексів сезонності, побудуємо сезонну хвилю та спрогнозуємо попит на такі рішення по місяцях наступного року.

Індекс сезонності визначається за формулою:

$$i_c = \frac{\bar{y}_i}{\bar{\bar{y}}}, \ \bar{y}_i = \frac{\sum_i y_i}{k}, \ \bar{\bar{y}} = \frac{\sum_j \bar{y}_i}{n} = \frac{\sum_j \sum_i y_{ij}}{k \cdot n}, \tag{8.23}$$

Таблиця 8.6

Рівень використання та впровадження рішень в сфері кібербезпеки

Рік / місяць	1-й рік	2-й рік	3-й рік	4-й рік	Разом за 4 роки (2+3 + 4+5)	В середньому за 4 роки гр6/4, $\bar{y}_i$	Індекс сезонності $\left(\frac{\bar{y}_i}{\bar{\bar{y}}}\right)100$	Прогноз даного показника на наступний рік
1	2	3	4	5	6	7	8	9
01	5,3	5,4	5,5	6,4	22,6	5,65	74,6	6,1
02	5,4	5,6	5,7	6,7	23,4	5,85	77,2	6,3
03	6,2	6	5,9	6,9	25,0	6,25	82,5	6,7
04	6,4	6,6	6,7	7,3	27,0	6,75	89,1	7,3
05	7,0	7,2	7,5	7,7	29,4	7,35	97,0	7,9
06	7,5	7,7	8,0	8,2	31,4	7,85	103,6	8,5

07	8,0	8,1	8,5	8,7	33,3	8,33	110,0	9,0
08	8,5	8,6	8,8	9,1	35,0	8,75	115,5	9,4
09	8,9	9,0	9,2	9,5	36,6	9,15	180,8	9,9
10	8,3	8,5	9,0	9,1	34,9	8,72	115,1	9,4
11	8,0	8,3	8,6	8,4	33,3	8,33	110,0	9,0
12	7,5	7,9	8,3	8	31,7	7,93	104,7	8,5
Разом	87,0	88,9	91,7	96,0	363,6	-	1200,0	98,0
В середньому						7,575	100,0	

де $\bar{y}_i$ – середнє значення показника за прийнятий проміжок часу (у нашому прикладі середня величина за кожний місяць, гр.7);

$\bar{y}$ – середнє значення показника за весь період;

$к$ – кількість років *(к* = 1, 2, 3, 4);

n – кількість місяців *(n* = 1, 2, 3, ..., 12)

Розраховані для даних таблиці 8.6 індекси сезонності подані у графі 8, а їх ряд створює сезонну хвилю.

Індекс сезонності для складання прогнозу використовується таким чином.

Припустимо, що на наступний рік планується використання та впровадження рішень в сфері кібербезпеки 98 тис. штук. Для того, щоб сформувати помісячний план цієї реалізації можна використати таку залежність:

$$\hat{\bar{Q}}_i = \frac{\frac{\hat{Q} \cdot i_c}{n}}{100}, \qquad (8.24)$$

де $\hat{\bar{Q}}_i$ – очікуваний місячний об'єм реалізації (i = 1, 2, 12);

$\hat{\bar{Q}}$ – очікуваний річний обсяг реалізації;

i_c – індекс сезонності;

n – кількість періодів *(n* = 12).

Результати розрахунків наведені в графі 9 таблиці 8.6. Застосування індексу не обмежується тільки дослідженням сезонного характеру цього показника. В декількох галузях промисловості коливання виробництва продукції пов'язані з особливостями технології, характером сировини та іншими факторами.

Виходячи з цього, важливою задачею є визначення добових норм переробки продукції по декадах на наступний сезон, щоб середній рівень виконання норми переробки становив не менше 100%. Загальна норма переробки на сезон обчислюється як добуток добової потужності підприємства на встановлений коефіцієнт використання цієї потужності.

У таблиці 8.7 представлені дані щодо відсотка виконання добової норми переробки сировини, необхідної для виготовлення ІТ-продукції (тобто співвідношення фактичної добової переробки сировини до встановленої сезонної норми), а також наведено порядок розрахунку прогнозованого (планового) завантаження підприємства-виробника ІТ-продукції.

Якщо хочеш, можу зробити ще більш стислий варіант цього тексту для конспекту, щоб легко запам’ятати формулу та суть розрахунків.

Позначимо:

N – добова потужність заводу;

P_m – добова продуктивність заводу, яка встановлена на друге півріччя.

$\bar{\hat{y}}$ – розрахункове значення виконання плану добового виробництва сировини в середньому за друге півріччя, яке визначається за формулою:

$$\bar{\hat{y}} = \frac{\sum_i (\bar{\hat{y}}_i \sum Q_{ij})}{\sum_j \sum_i Q_{ij}}, \tag{8.27}$$

де Q_{ij} — обсяг переробленої сировини в *i*-ї декаді *j*-го року.

Таблиця 8.7

Прогноз добової пероробки сировини по декадах на наступний сезон

Декади	Виконання плану добової переробки сировини, %			Добова продуктивність, т	
	Фактично в середньому за п’ятиріччя, $\bar{y}_i$	Вирівняне по рівнянню тренда*, $\hat{y}_i$	Індекс сезонності $\left(\bar{y}_i : \bar{\hat{y}}\right)\cdot 100$	При нормі переробки сировини, що склалася $\left(p_0 \cdot \bar{y}_i\right):100$	При прогнозованому завантаженні (гр4-p_0):100
1	2	3	4	5	6
10,09	76,02	79,76	87,5	836,3	962,5

20,09	96,2	91,95	100,88	1058,2	1109,7
01,1	99,34	96,8	106,2	1092,2	1168,2
10,1	102,96	98,17	107,7	1132,6	1184,8
20,1	96,61	97,55	107,24	1062,7	1177,8
01,11	99,46	95,69	106,8	1094,0	1154,8
10,11	90,33	93,03	102,06	993,0	1122,6
20,11	82,67	89,85	98,57	909,3	1084,3
01,12	89,83	86,35	94,73	988,2	1042,1
10,12	78,38	82,66	90,69	862,2	997,5
20,12	64,42	78,88	85,54	686,6	951,9
01,01	97,43	75,07	82,36	1071,7	905,9
2-півріччя	91,63	91,15	100,0	1007,9	1100,0

В таблиці 8.8 $\bar{\bar{y}}$ подана в підсумковому рядку графи 3. Початкові дані для розрахунку граф 5 і 6 є: добова потужність — N = 1220т.;

Установлений коефіцієнт використання потужності на друге півріччя — K = 0,92.

Отже, планова добова продуктивність на друге півріччя становить $P_0 = KN$ = 0,92 - 1220 = 1100 т.

Дані таблиці 8.8 свідчать про те, що норма переробки сировини підприємством не виконувалась. В графі 6 наведені розрахункові, з використанням індексу сезонності, значення добової переробки цукрового буряка по декадах, які дозволяють забезпечити ритмічну роботу заводу і виконання норми загрузки підприємства в цілому на друге півріччя.

Перелік запитань для самопідготовки:

1. Що таке тенденція у часовому ряді?
2. Якими способами визначають наявність тенденції в динамічних даних?
3. Як розрізняються короткострокові коливання і довгострокова тенденція?
4. Чим відрізняється тренд від сезонності?
5. Які критерії підтверджують наявність стійкої тенденції?

6. Які аналітичні показники використовуються в ряді динаміки (абсолютний приріст, темп зростання тощо)?

7. Як здійснюється екстраполяція на основі цих показників?

8. У чому полягає суть індексного способу прогнозування?

9. Як аналітичні показники дозволяють передбачити майбутні значення ряду?

10. Що таке плинна (ковзна) середня? Як вона використовується для згладжування та прогнозування часових рядів?

11. У чому полягає перевага методу ковзної середньої над простим середнім?

12. Що таке індекс сезонності та як він розраховується?

13. Яку роль відіграє сезонність у прогнозуванні часових рядів?

14. Як використовуються сезонні індекси для побудови прогнозу?

Тема 9. Моделі прогнозування та соціального-економічного демографічного розвитку

9.1. Прогнозування соціального розвитку та рівня життя населення.

9.2. Прогнозування попиту на товари й послуги.

9.3. Демографічні прогнози.

9.4. Модель прогнозування чисельності населення.

9.5. Прогнозування параметрів руху населення.

9.1. Прогнозування соціального розвитку та рівня життя населення

Соціальний розвиток суспільства та рівень життя його населення є важливими показниками загального соціально-економічного стану, оскільки відображають ступінь задоволення фізичних, духовних та соціальних потреб людей. Аналіз цих потреб є невід'ємною частиною соціально-економічного прогнозування та оцінки якості життя.

Вивчення динаміки потреб включає кілька ключових напрямів: визначення структури потреб та факторів, що впливають на їх зміну, встановлення взаємозв'язків між розвитком потреб і цими факторами, проведення ретроспективного аналізу змін у минулому, а також формування прогнозних показників, що дозволяють оцінити соціальний прогрес та очікуваний рівень життя населення.

Традиційні методи планування соціального розвитку мають обмеження, оскільки не враховують усі адміністративні, методологічні та соціальні взаємозв'язки, що впливають на рівень життя. Тому актуальним є створення державної системи прогнозування соціального розвитку, яка забезпечує всебічний аналіз ключових проблем і формування обґрунтованих сценаріїв підвищення якості життя.

Функціонування такої системи передбачає оцінку поточного стану життя населення та формування гіпотез щодо його покращення. Основними

складовими таких гіпотез є приріст національного прибутку, зростання суспільних потреб та наявність ресурсів для майбутнього споживання. Прогнозування темпів збільшення ресурсів здійснюється на основі оцінки ефективності виробничого накопичення та економічного потенціалу країни.

Для забезпечення узгодженості прогнозів використовується комплекс взаємопов'язаних показників, які формують логічну послідовність аналізу та дозволяють всебічно оцінити соціальний розвиток і передбачити тенденції підвищення рівня життя населення. Основними з них є:

соціально-демографічні показники, що характеризують чисельність населення, його склад і соціальну структуру;

показники умов праці, які відображають зміну соціально-економічних і виробничих умов, підвищення професійного рівня працівників та покращення охорони праці;

узагальнюючі показники, що дозволяють оцінити економічні умови життя населення — національний прибуток на душу населення, обсяги споживання ресурсів, динаміку та структуру доходів населення, споживання основних продуктів харчування та непродовольчих товарів, забезпеченість сімей предметами культурно-побутового призначення.

При формуванні прогнозів спочатку визначають цільові показники, наприклад, забезпечення певної насиченості сімей предметами тривалого користування відповідно до національних норм споживання. Потім оцінюються ресурси, необхідні для досягнення цього рівня, та встановлюються вимоги до розвитку відповідних галузей виробництва. Це дозволяє планувати заходи для реалізації прогнозованого рівня життя.

Для оцінки динаміки ресурсів та їх структури у прогнозному періоді застосовуються взаємопов'язані економіко-математичні моделі, зокрема: норми виробничого накопичення з урахуванням нормативів зростання ресурсів для споживання та будівництва, моделі обсягів накопичення та поточного споживання, моделі динаміки фондів та реалізації капітальних вкладень.

Прогнозування передбачає розробку кількох варіантів рішень, що дозволяє обрати оптимальний сценарій розподілу ресурсів між споживанням та інвестиціями у розвиток. Отримані прогнози соціального розвитку та підвищення рівня життя інтегруються у загальні плани економічного та соціального розвитку держави, впливаючи на визначення ключових показників та пріоритетів ресурсного забезпечення.

9.2. Прогнозування попиту на товари й послуги

Попит на товари та послуги є визначальним чинником не лише для розробки загальної стратегії підприємства, але й для планування конкретних тактичних кроків, пов'язаних із організацією виробничих процесів та збутом продукції чи наданням послуг. Вивчення тенденцій змін попиту, а також його науково обґрунтоване прогнозування як на короткостроковий, так і на довгостроковий період, належить до ключових і надзвичайно відповідальних завдань маркетингових служб компаній, оскільки без цього ефективне планування виробництва, оптимізація ресурсів та задоволення потреб споживачів стає практично неможливим.

Склад і рівень попиту на конкретний товар або послугу визначаються багатьма факторами, як економічними, так і природними. До економічних чинників належать рівень виробництва та пропозиції товарів і послуг (позначимо цей фактор загально через П), рівень грошових доходів окремих груп населення (D), а також рівень і співвідношення цін (P). До природних факторів відносяться демографічна структура населення, зокрема розмір і склад сім'ї (S), а також звички, традиції, рівень культури та природно-кліматичні умови.

Економічні чинники відзначаються високою динамічністю, наприклад, змінами у розподілі населення за рівнем доходів. Натомість природні фактори змінюються значно повільніше і протягом коротких періодів (приблизно 3–5 років) майже не впливають істотно на попит. Винятком залишається

демографічна структура населення. Тому при складанні як короткострокових, так і довгострокових прогнозів попиту природні чинники, окрім демографічного, зазвичай розглядають у сукупності.

Наприклад, у сфері кібербезпеки попит на послуги захисту інформації та цифрової інфраструктури також підпорядковується цим закономірностям. Економічні чинники тут представлені бюджетами компаній на IT-безпеку та інвестиціями в модернізацію мереж (П, D, P), тоді як природні та соціальні чинники — це структура організацій, чисельність персоналу та рівень цифрової грамотності співробітників (S), які змінюються значно повільніше. При плануванні стратегій кіберзахисту важливо враховувати ці закономірності, прогнозуючи попит на послуги безпеки на найближчий та віддалений періоди. З цією метою вводять фактор часу *t*. Це дає можливість у загальному вигляді попит подати у вигляді такої функції вказаних факторів $y = f(П,D,P,S,t)$.

Оскільки найбільший вплив на попит має фактор доходу, більшість обчислень попиту та споживання виконують у вигляді функції від грошового доходу на душу населення $y=\varphi(D)$.

Найпростіший метод прогнозування попиту на короткий період базується на використанні структурних моделей попиту. При їх побудові виходять із припущення, що для кожної соціально-економічної групи населення, на основі статистики сімейних бюджетів, можна визначити характерну для неї структуру споживання. При цьому вважається, що протягом прогнозованого періоду змінюється переважно рівень доходів, тоді як ціни, склад домогосподарств та інші фактори залишаються сталими. Зростання доходів розглядається як процес переходу певної кількості сімей із нижчих до вищих доходних груп, тобто частка населення у нижчих інтервалах доходів зменшується, а у вищих – зростає. Домогосподарства, що переходять до нового рівня доходів, набувають таку ж модель споживчої поведінки та попиту, яка була характерна для сімей з аналогічними доходами раніше.

Таким чином, структурні моделі дозволяють розглядати попит як функцію

розподілу споживачів за рівнями доходів. Використання статистично визначених структур споживання у поєднанні з даними про доходи населення дає змогу обчислити загальну структуру попиту.

Наприклад, у сфері кібербезпеки та IT-послуг подібний підхід можна застосувати для прогнозування попиту на різні пакети захисту даних чи програмне забезпечення. Організації з низьким бюджетом на IT безпеку демонструють один тип поведінки у придбанні послуг, а компанії з більшими бюджетами – інший. Збільшення фінансування безпеки аналогічно «переходу» до вищої доходної групи, що дозволяє прогнозувати зростання попиту на більш складні та комплексні рішення для захисту інформаційних систем.

Структурні моделі попиту є одним із ключових видів економіко-математичних моделей, що використовуються для планування та прогнозування рівня споживання і попиту. Одним із поширених прикладів є компаративні (порівняльні) структурні моделі, у яких аналізується структура попиту конкретного об’єкта та порівнюється з аналогічними характеристиками іншого об’єкта. Як правило, у ролі такого аналога виступає регіон або соціальна група населення з оптимальними показниками споживчої поведінки, що дозволяє оцінити очікувані зміни у попиті.

Окрім порівняльних моделей, у прогнозуванні попиту застосовують і конструктивні моделі. Їхня суть полягає у використанні рівнянь сімейного бюджету – математичних залежностей, які відображають рівність між загальними грошовими витратами (що фактично відповідає обсягу споживання) та сумою добутків кількості кожного придбаного товару на його ціну. Такий підхід дозволяє точно визначати структуру попиту і прогнозувати її зміни залежно від доходів, цін та інших факторів.

Наприклад, у сфері кібербезпеки конструктивна модель може допомогти передбачити попит на різні типи послуг або програмного забезпечення, виходячи із бюджету організацій на інформаційну безпеку та співвідношення вартості різних рішень. Це дозволяє ефективно планувати розробку і

пропозицію продуктів відповідно до потреб різних категорій клієнтів.

У сфері прогнозування попиту широко застосовуються також аналітичні моделі споживання та попиту. Ці моделі формулюються у вигляді рівнянь, які описують залежність обсягів споживання товарів і послуг від визначених факторів. Завдяки цьому функціональні залежності набувають конкретизованого вигляду, що дозволяє точніше прогнозувати поведінку споживачів.

Аналітичні моделі можуть бути однофакторними або багатофакторними, залежно від кількості змінних, що враховуються. У практичній діяльності найчастіше використовуються лінійні статистичні моделі на основі кореляційно-регресійного аналізу, які дозволяють кількісно оцінити вплив кожного фактору на рівень попиту.

Наприклад, у сфері кібербезпеки такі аналітичні моделі можна застосовувати для прогнозування попиту на програмне забезпечення для захисту даних або послуги з інформаційної безпеки, виходячи з бюджету компаній, масштабу їхньої цифрової інфраструктури та рівня загроз кібернападів. Це дозволяє виробникам та постачальникам безпекових рішень планувати обсяги виробництва, розробку нових продуктів і пропозицій під різні категорії клієнтів.

9.3. Демографічні прогнози

Демографічні прогнози займають ключове місце в системі інтегрованого прогнозування і широко використовуються для розробки та реалізації комплексних цільових програм. Вони є основою для створення балансів трудових ресурсів, планування соціального забезпечення населення та визначення пріоритетів розвитку соціальної інфраструктури. На їх основі визначають чисельність та структуру робочої сили, прогнозують потреби суспільства у товарах і послугах, а також формують концепції розміщення освітніх, медичних та інших соціально важливих об’єктів.

Об’єктами демографічного прогнозування є такі ключові показники, як народжуваність, смертність, загальна чисельність населення, обсяги та

структура трудових ресурсів, а також характеристики соціальних контактів і взаємодій у суспільстві. Прогнози включають кілька основних видів: обчислення природного руху населення (народжуваність і смертність), прогноз змін у складі працездатного населення, оцінку міграційних процесів, а також прогнозування чисельності населення та його соціально-демографічних параметрів як на загальнонаціональному, так і на регіональному рівнях.

Демографічні прогнози формуються з урахуванням різних характеристик і масштабів дослідження та можуть мати глобальний, регіональний або локальний рівень. Точність таких прогнозів визначається багатьма чинниками, однак ключовим є часовий горизонт оцінки: для довгострокових прогнозів допустиме відхилення становить до 10%, а короткострокові прогнози повинні забезпечувати точність на рівні до 5%. У сфері кібербезпеки аналогічний підхід використовується при прогнозуванні загроз та ризиків: наприклад, оцінка кількості потенційних кібератак на мережі організації на наступні 1–3 місяці повинна бути максимально точною, тоді як довгострокові прогнози загроз на 1–3 роки допускають більшу похибку через змінність технологій і поведінки зловмисників.

Вихідні дані для демографічного прогнозування надходять із статистичної звітності та спеціально організованих спостережень. У кібербезпеці аналогічними джерелами є дані систем моніторингу, журнали подій, інформація про інциденти безпеки та поведінку користувачів. Надійність прогнозів у цьому випадку визначається обґрунтованістю початкових припущень і гіпотез щодо розвитку кібератак, поведінки користувачів та технологічних змін. Наприклад, якщо передбачити підвищення активності фішингових атак під час святкових періодів, це допомагає заздалегідь підготувати системи захисту.

Для проведення прогнозування, як у демографії, так і в кібербезпеці, застосовують методи моделювання, експертні оцінки та екстраполяції. Важливо мати статистичні дані за різними категоріями: у демографії – вік, стать, територія проживання; у кібербезпеці – типи пристроїв, підрозділи компанії,

рівень доступу користувачів, частота спроб доступу до ресурсів. Комбінування цих даних дозволяє отримати точні прогнози щодо чисельності населення або потенційних кібератак, що є критично важливим для планування ресурсів, формування стратегій захисту та запобігання ризикам.

9.4. Модель прогнозування чисельності населення

Прогнозування майбутньої чисельності населення здійснюється на основі аналізу статистичних динамічних рядів, таких як середній абсолютний приріст та середній темп зростання, а також із використанням аналітичних функцій. Найпоширенішими і водночас простими моделями для опису змін чисельності населення є експоненційна та логістична функції. Особливу популярність здобула експоненційна модель, яка застосовується тоді, коли коефіцієнт зростання залишається відносно сталим, тобто рівні динамічного ряду змінюються за законом геометричної прогресії.

Аналогічний підхід використовують і в сфері кібербезпеки для прогнозування динаміки кібератак або кількості інцидентів у мережах організації. Наприклад, якщо число спроб несанкціонованого доступу стабільно зростає протягом декількох місяців, експоненційна модель дозволяє спрогнозувати ймовірну кількість атак у наступних періодах. Це допомагає ефективно планувати заходи захисту, оптимізувати розподіл ресурсів та налаштовувати системи моніторингу.

Варто зазначити, що природний приріст населення зазвичай не залишається сталим, тому експоненційна модель придатна лише для короткострокових прогнозів. Аналогічно, розрахунки, засновані на середньому абсолютному прирості або середньому темпі зростання, також належать до категорії короткотермінових оцінок.

Для більш детальних прогнозів використовують регресійні моделі, які дозволяють оцінити майбутній стан демографічних показників залежно від сукупності чинників, що впливають на динаміку населення. Рівень

демографічних процесів формується під впливом комплексу взаємопов’язаних факторів, до яких належать соціально-економічні умови, історичні традиції, біологічні особливості, зміни у сфері суспільних відносин та інші. Основним завданням є виявлення взаємодії цих факторів, кількісне визначення їх впливу та використання результатів для розробки заходів демографічної політики.

Багатовимірні регресійні моделі дозволяють врахувати вплив різних соціальних і економічних умов на чисельність населення, що особливо важливо при регіональному аналізі. Окрім цього, широко застосовуються моделі з одним або двома факторами взаємозв’язку для спрощеного аналізу.

На практиці регресійні моделі використовують для дослідження соціального розвитку сім’ї, рівня зайнятості населення у сфері домашнього господарства, міграційних процесів, а також чинників, що впливають на репродуктивну поведінку. Наприклад, народжуваність визначається комплексом умов, серед яких культурний і матеріальний рівень населення, забезпеченість дитячими закладами, соціально-економічні фактори тощо. У сфері кібербезпеки подібним чином застосовують регресійні моделі для оцінки впливу різних факторів – таких як кількість користувачів, оновлення систем, рівень навченості персоналу – на частоту кібератак і ефективність заходів захисту.

9.5. Прогнозування параметрів руху населення

У розрахунках показників народжуваності та смертності значне значення мають демографічні моделі. Вони дозволяють аналізувати процеси відтворення населення як перехід від початкової вікової структури до нової структури через певний проміжок часу.

Найпростішою серед таких моделей є модель стаціонарного населення. Вона передбачає, що порядок смертності залишається сталим, а співвідношення кількості новонароджених до тривалості періоду не змінюється, тобто інтенсивність народжуваності є постійною. За таких умов чисельність кожної вікової групи населення залишається сталою, а отже, стабільною є і загальна

чисельність населення. Приріст дорівнює нулю, оскільки кількість народжень збігається з кількістю смертей. При цьому модель виходить із припущення про повну відсутність міграційних процесів.

Аналогічний підхід застосовують у сфері кібербезпеки для моделювання стабільної структури загроз або інцидентів у мережі організації. Наприклад, можна розглядати "стаціонарну модель атак", де кількість спроб несанкціонованого доступу у певній категорії залишається приблизно сталою протягом часу, а склад і тип атак не змінюються. Така модель дозволяє оцінити базовий рівень ризику, визначити необхідні ресурси для моніторингу та планувати заходи захисту без додаткових корекцій на короткострокові коливання.

Перелік запитань для самопідготовки:

1. Які основні індикатори використовуються для оцінки соціального розвитку?
2. Що таке рівень життя населення і як він вимірюється?
3. Які фактори впливають на зміни соціальних показників у прогнозах?
4. Які моделі застосовуються для прогнозування рівня життя?
5. Як пов’язуються соціальні прогнози з економічною політикою?
6. Які методи використовуються для прогнозування попиту?
7. Які фактори враховуються при побудові моделей попиту?
8. У чому різниця між короткостроковим і довгостроковим прогнозом попиту?
9. Як зміни доходів населення впливають на попит?
10. Які моделі застосовуються для прогнозування попиту в окремих секторах економіки?
11. Що таке демографічне прогнозування?
12. Які основні типи демографічних прогнозів існують?
13. Які фактори впливають на точність демографічних прогнозів?

14. У яких галузях використовуються демографічні прогнози?

15. Які моделі використовуються для прогнозування чисельності населення?

16. Як враховуються народжуваність, смертність і міграція у прогнозних розрахунках?

17. Що таке когортний метод прогнозування?

18. Як обирається база для побудови демографічної моделі?

19. Які сценарії можуть будуватись при прогнозуванні чисельності населення?

20. Які параметри руху населення є об’єктом прогнозування?

21. Як прогнозується народжуваність і смертність?

22. Які моделі дозволяють враховувати вікову структуру населення?

23. Як враховується вплив соціальних і економічних чинників на демографічні зміни?

24. У чому складність довгострокового прогнозування параметрів руху населення?

Тема 10. Метод згладжування і сезонне прогнозування

10.1. Наївна модель в прогнозуванні.

10.2. Способи усунення тренда.

10.3. Моделі згладжування для часових рядів, що не мають тренда: модель ковзного середнього, модель експоненційно зваженого ковзного середнього, комбінована модель.

10.4. Визначення початкових значень моделі.

10.5. Моделі згладжування з трендом: модель Холта, модель Брауна.

10.6. Сезонні моделі.

10.1 Наївна модель в прогнозуванні

Розглянемо найпростішу модель, яка будь-коли використовувалася при прогнозуванні – так звану «наївну модель». Значення моделі задаються наступною формулою:

$$\hat{Y}_{t+1} = Y_t . \qquad (10.1)$$

Ідея наївної моделі (10.1) полягає в припущенні, що дане значення у момент часу t приблизно дорівнює попередньому значенню.

Стовпець значень Y_t виходить з початкових даних просто зрушенням на один рядок вниз. На рис. 10.1 представлений графік цієї наївної моделі, який виходить відповідно з початкового графіка шляхом зрушення на одиницю вправо. Прогноз для всіх кварталів, починаючи з 18-го, один і той же і дорівнює останньому значенню ряду даних: 248.



Рисунок 10.1 – Графік наївної моделі

Наївна модель використовується переважно для короткострокового прогнозування у випадках, коли не очікуються суттєві зміни в характері процесу. Її основна цінність полягає у тому, що разом із лінійною моделлю та моделлю постійного зростання наївна модель слугує своєрідним орієнтиром для оцінки ефективності будь-якої іншої прогнозної моделі. На практиці це означає, що перед застосуванням нової моделі її характеристики доцільно порівняти з показниками вже відомих і перевірених моделей.

10.2. Способи усунення тренда

Якщо уважно подивитися на графік наївної моделі (рис. 10.1), то можна зауважити, що крім зрушення на одиницю вправо він в більшій своїй частині (у 5 випадках з 16) розташований нижче за початкову криву. Це пов'язано з тим, що наївна модель не припускає існування лінійного тренда. Тобто наївна модель виявиться неефективною, якщо значення коефіцієнта b_2 в лінійному рівнянні регресії значно відрізнятиметься від нуля. Існує досить багато інших моделей, набагато складніших ніж наївна, які також припускають, що процес не має тренда. Як показник відсутності тренда у ряду Y_i іноді береться коефіцієнт k:

$$k = \frac{b_2}{|\bar{Y}|}. \qquad (10.2)$$

Залишки лінійного рівняння регресії звичайно служать прикладом процесу, що не має лінійного тренда. Як правило, в бізнесі ми маємо справу з

процесами, що мають тренд. Для того, щоб в цьому випадку можна було застосувати модель, призначену для процесів, що не мають тренда, можна вчинити одним з таких двох способів.

Спосіб 1. Розглянемо різниці або частки початкових даних. Таким чином, якщо є ряд даних Y_1, Y_2, ..., Y_n, то розглядається ряд: $\Delta_t, \ldots, \Delta_n$, де $\Delta_t = Y_t - Y_{t-1}$, або ряд $p_t, \ldots, p_n$, де

$$p_t = \frac{Y_t}{Y_{t-1}} \tag{10.3}$$

Насправді прогнозисти звичайно діють трохи інакше. До результатного ряду даних застосовується оператор *ln*. Тобто розглядається ряд: $lnY_1, \ldots, lnY_n$, після чого до нього застосовується оператор приросту Δ. При значеннях p_i, близьких до одиниці, можна нехтувати другим членом в розкладанні Тейлора і вважати, що $ln(p_t) \approx p_t - 1$.

Наприклад: $ln(p_t) = ln(1) + (ln(p_t))'(p_t - 1) + \sout{(ln(p_t))''/2!*(p_t-1)^2} \approx 0 + 1/p_t*(p_t-1) = 0 + 1*(p_t-1) = p_t - 1$

Оператори Δ_i і p_i називають операторами початкового ряду. При цьому абсолютно не гарантовано, що ряди даних, одержані шляхом застосування операторів Δ_i або p_i, не матимуть тренда (в такому разі іноді застосовують повторно оператор Δ).

На рис. 10.2 представлене графічне зображення ряду pt,..., pn.

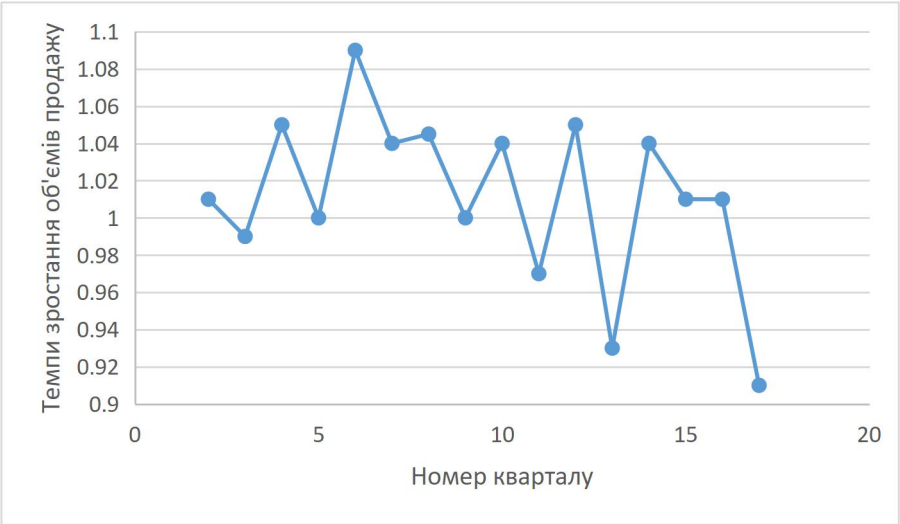


Рисунок 10.2 – Темпи зростання об'ємів продажів

Якщо менеджер вважає, що можна нехтувати зміною в змодельованих квартальних темпах зростання, то він може розглядати p_t, ..., p_n як ряд, що не має тренда, і застосувати до нього наївну або будь-яку іншу модель, призначену для рядів, що не мають тренда.

Спосіб 2 полягає в розгляді залишків лінійної регресії і є одним з найчастіше використовуваних методів усунення тренда. Його називають методом детрендалізації.

Якщо коефіцієнт b_2 лінійного рівняння регресії для ряду e_i має нулі в перших чотирьох десяткових розрядах, то він може вважатися рівним нулю. Тепер можна побудувати модель для ряду без тренда для залишків e_i, потім знайти відповідне змодельоване значення e_k при $k>17$, і додати його до відповідного значення тренда. Оскільки наївна модель припускає, що $e_{t-1} \approx e_t$, то її слід застосовувати у тому випадку, коли коефіцієнт кореляції $r(e_{t-1}, e_t)$ при $t=2,...,n$, позитивний і значно відрізняється від нуля. У такому разі говорять, що в моделі лінійної регресії є позитивна автокореляція першого порядку. Існує спеціальний тест (тест Дарбіна-Уотсона) на наявність автокореляції першого порядку.

Прогноз визначається шляхом додавання змодельованої помилки e_{19} до відповідного значення тренда.

10.3. Моделі згладжування для часових рядів, що не мають тренда: модель ковзного середнього, модель експоненційно зваженого ковзного середнього, комбінована модель

Головне завдання експерта, який складає прогноз, полягає у виборі моделі або набору моделей, які максимально точно відображали б досліджуваний процес. Саме на основі цих моделей надалі формуються прогнози. Моделі згладжування застосовуються тоді, коли необхідно виявити загальний тренд процесу, не зважаючи на коливання, спричинені випадковими чинниками. У цьому вони подібні до моделей підгонки, однак на відміну від останніх, у

моделях згладжування більша вага надається останнім наявним даним, що визначає їх адаптивність.

Ще однією важливою відмінністю моделей згладжування від кривих підгонки є спосіб їх задання: вони визначаються алгоритмічно, а не аналітично як функції від часу. Завдяки рекурсивній природі цих моделей виникає одна важлива властивість – значення моделі збігаються з відповідними expost-прогнозами. Це пояснюється тим, що expost-прогнози також формуються виключно на основі попередніх спостережень.

Значення моделі ковзного середнього (movingaveragemodel) визначаються такою формулою:

$$\widehat{Y}_t = \frac{Y_{t-1}+Y_{t-2}+\ldots+Y_{t-k}}{k} \qquad (10.4)$$

Іншими словами, Y_t дорівнює середньому k попередніх значень. Число k називається порядком ковзного середнього. Середнє значення завжди знаходиться між мінімальним і максимальним значеннями ($Min(Y_1(t-1), Y_1(t-2), \ldots, Y_1(t-k)) \leq Max(Y_1(t-1), Y_1(t-2), \ldots, Y_1(t-k))$ тому модель можна ефективно застосовувати тільки до ряду даних, що не має лінійного тренда. На рис. 10.3 показана модель для ряду p_t, при $k = 5$.

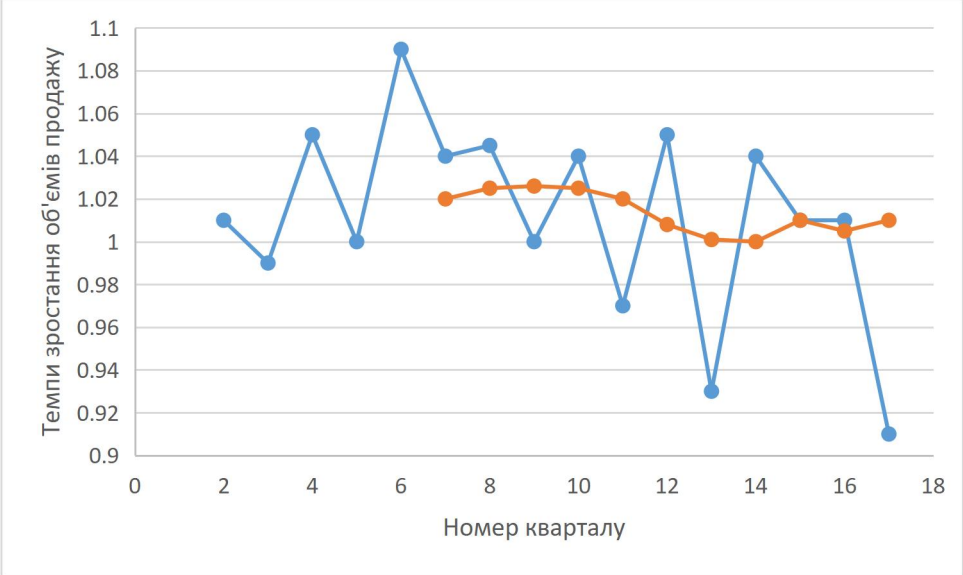


Рисунок 10.3 – Модель ковзного середнього при $k = 5$

Ми бачимо, що графік моделі більш гладкий, ніж початковий. При усереднюванні частково зникли флуктуації початкових даних, викликані випадковими чинниками. На графіку моделі відсутні перші п'ять значень. Це викликано тим, що для моментів часу t = 2, 3, 4, 5, 6 у формулі (10.4) значення p_t не визначені.

При прогнозуванні не враховуються можливі коливання або збурення, що виникають у процесі. Тому прогноз слід сприймати як усереднене значення майбутніх показників.

Щоб отримати прогноз, який враховував би такі збурення, необхідно застосовувати додаткові методи для їх оцінки. Одним із прикладів такого підходу є виділення сезонних компонент.

Для того, щоб знайти прогнози на більш віддалені в майбутнє квартали, відсутні члени у формулі (10.4) замінюють на відповідні одержані раніше прогнози.

Модель експоненційно зваженого ковзного середнього (exponentially weighted moving average (EWMA) model) є природним узагальненням моделі ковзного середнього. Вона будується на таких двох припущеннях.

– кожне нове значення визначається сукупністю всіх попередніх значень;

– вплив попередніх даних слабшає в геометричній прогресії, у міру того як вони відстають далі за часом.

Якщо виходити з вищесказаного, то модель EWMA повинна визначатися формулою:

$$\widehat{Y}_t = Y_{t-1} + \alpha Y_{t-2} + \alpha^2 Y_{t-2} + ... де\, 0 \le \alpha < 1. \qquad (10.5)$$

Проте при визначенні коефіцієнтів при Y_{t-1} потрібно також, щоб їх сума дорівнювала одиниці. В цьому випадку вони називаються вагами. Оскільки сума коефіцієнтів при Y_{t-1} в правій частині формули приблизно дорівнює $1 +$

$\alpha + \alpha^2 + \alpha^{n-1} + ... = \frac{1}{(1-\alpha)}$, *коли* $n \to \infty$ при достатньо великих n, то модель EWMA визначається такою формулою:

$$\hat{Y}_t = (1-\alpha)Y_{t-1} + (1-\alpha)\alpha Y_{t-2} + (1-\alpha)\alpha^2 Y_{t-2} + ... + (1-\alpha)\alpha^{n-1} Y_{t-2} + ... \quad (10.6)$$

Тепер сума коефіцієнтів: $\sum (1-\alpha)\alpha^{n-1} \approx 1$.

З формули (10.6) можна зробити дуже важливий висновок про пам'ять моделі EWMA, тобто про вплив попередніх даних на значення моделі. Чим ближче α до нуля, тим швидше модель «забуває» минулі дані і тим більшу вагу має остання інформація. З другого боку, коли α наближається до одиниці, то ваги при Y_i розподілені більш рівномірно і, отже, графік моделі виглядатиме більш згладженим.

Неважко бачити, що формула (10.6) еквівалентна такій формулі:

$$\hat{Y}_t = (1-\alpha)Y_{t-1} + \alpha\hat{Y}_{t-1}. \quad (10.7)$$

Формула (10.7) більш зручна для обчислень, порівняно з формулою (10.6), тому для визначення значень моделі краще використовувати саме її. Число *(1-α)* називається константою згладжування. Коли a = 0, формула (10.7) збігається з формулою (10.1) для наївної моделі. В правій частині формули (10.7) присутній член aY_{t-1}. Це означає, що обчислення значень моделі носить рекурсивний характер. З (10.7) також витікає, що модель EWMA слід застосовувати тільки для рядів, що не мають тренда.

Розглянемо модель EWMA на прикладі залишків лінійної регресії. Її можна записати в такому вигляді:

$$\hat{e}_t = (1-\alpha)e_{t-1} + \alpha\hat{e}_{t-1}. \quad (10.8)$$

Для того, щоб модель можна було застосувати, потрібно задати значення a і перше значення моделі.

Потрібно відзначити один дуже істотний недолік моделей згладжування. На зміни в ході процесу вони реагують із запізненням. Пояснимо це з допомогою діаграми на рисунку 10.4.

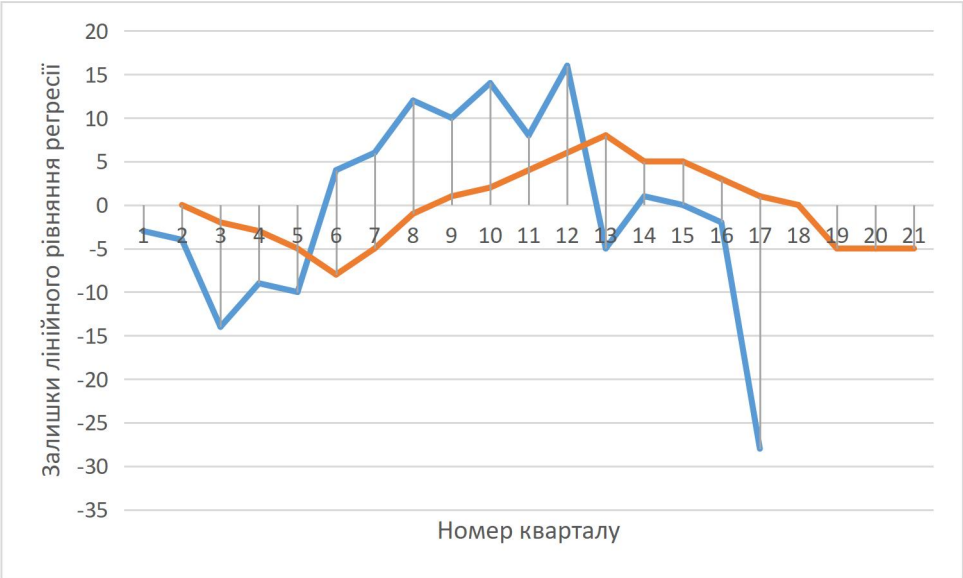


Рисунок 10.4 – Модель EWMA

Спостерігалося зниження показників аж до 3-го кварталу. Починаючи з 4-го кварталу, тенденція почала поступово змінюватися у бік зростання. У 5-му кварталі відбулося невелике тимчасове падіння, після чого графік демонстрував стабільне зростання аж до 8-го кварталу. Розглянемо, як на ці зміни реагувала наша модель. Вона продовжувала демонструвати спад аж до 6-го кварталу. Коли ж її графік нарешті перейшов до зростання, воно тривало до 13-го кварталу, тобто до моменту, коли початкові дані раптово пішли вниз. Пояснити цей факт можна характером формули (10.6), в якій значення моделі визначаються в основному декількома попередніми даними.

Як вже говорилося раніше, вид графіка моделі залежить від величини *a*. При порівнянні двох діаграм, неважко помітити, що графік моделі при більшому значенні (a=0,8) має велику згладжену (рис. 10.4), тоді як графік знизу (при a=0,2) слідує за початковими даними набагато щільніше (рис. 10.5). Також значно відрізняються і прогнозовані значення.

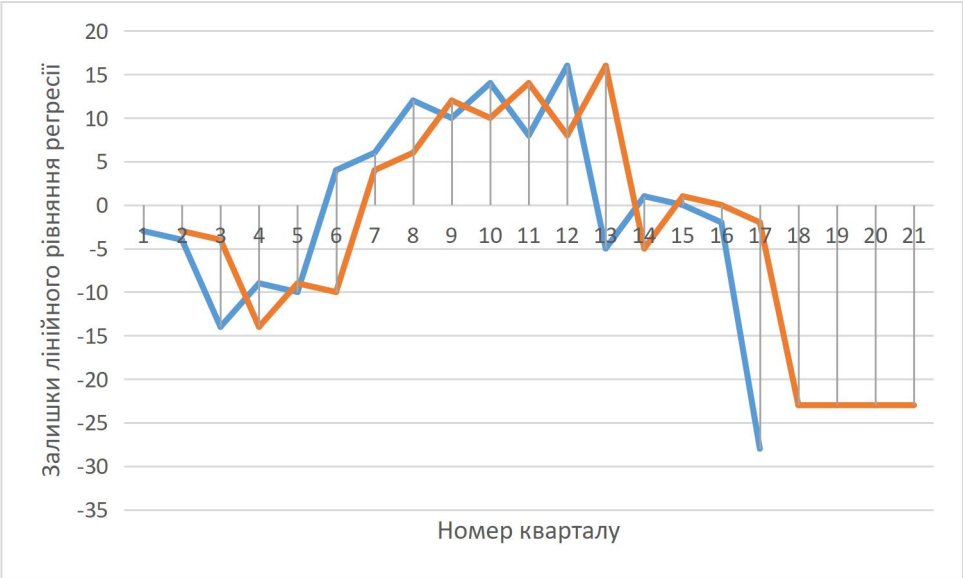


Рисунок 10.5 – Модель EWMA ($\hat{e}_1 = 0$)

Модель експоненціального згладжування (exponential smoothing model) задається формулою:

$$\widehat{Y}_t = (1 - \alpha)Y_t + \alpha\widehat{Y}_{t-1}. \qquad (10.9)$$

Вираз (10.9) відрізняється від формули (10.7) моделі EWMA лише заміною Y_{t-1} на Y_t в правій частині. Насправді модель експоненціального згладжування виходить з моделі EWMA шляхом зсуву початкових даних на одиницю часу. Модель експоненціального згладжування дає прогнози, мало відмінні від тих, що одержані за допомогою моделі EWMA.

Модель подвійного експоненційного згладжування (double exponential smoothing model) виходить за допомогою подвійного застосування моделі експоненційного згладжування. Тобто спочатку модель експоненційного згладжування застосовується до початкових даних, а потім – до змодельованих значень, одержаних на першому етапі. При цьому при повторному застосуванні моделі можна узяти інше значення a.

Для вибору відповідного значення a прогнозист звичайно керується або інтуїцією, або попереднім досвідом. Більш строгим є такий підхід. Оскільки кожна експоненціальна модель визначається своїм значенням a, то можна

розглядати всю сукупність експоненціальних моделей для цього ряду даних. Отже, *a* вважається параметром моделі. Тепер можна знайти оптимальне значення *a*, що мінімізує суму квадратів залишків, тобто вчинемо точно так, як ми робили в першому розділі для знаходження параметрів кривої підгонки.

Тут важливо пам'ятати, що оскільки оптимальне *a* визначається на основі всього ряду даних, то відповідні expost значення *a* не повинні збігатися і їх потрібно знаходити окремо на кожному кроці. Хорошого прогнозу слід чекати при стабільних значеннях *a*, тобто коли expost значення *a* утворюють невеликі коливання навколо деякого середнього значення.

10.4 Визначення початкових значень моделі

Оскільки формула (10.7) є рекурсивною, тобто для обчислення значення моделі у момент *t* ми повинні знати її значення в попередній момент *t-1,* то перед нами постає задача визначення початкових значень моделі (в нашому випадку у момент часу *t=1*). Дійсно, формулу (10.7) можна застосувати тільки тоді, коли *t>1*. Але якщо ми не знаємо Y_1 ,ми не зможемо знайти Y_2 і т.д. Задача визначення початкових значень часто трапляється в прогнозуванні. У частині моделей згладжування вплив початкового значення моделі залежить від величини *a* і від довжини *n* ряду початкових даних. Чим ближче *a* до одиниці, тим більш помітною буде дія початкової умови як на значення моделі, так і на прогноз. В той же час, зважаючи на здатність моделі «забувати» попередні дані вплив початкового значення моделі слабшає із збільшенням довжини ряду даних.

Найпростіший спосіб – узяти за початкове значення $Y_1 = \overline{Y}$ або $Y_1 = Y_1$.

У сучасному прогнозуванні досить поширений метод прогнозування назад (back casting). Наприклад, можна розглянути модель, подібну до моделі (10.7), з тією лише різницею, що рекурсивність здійснюється у зворотному напрямі:

$$\widehat{Y}_{t-1} = (1-\alpha)Y_t + \alpha\widehat{Y}_t, \qquad (10.10)$$

де, *t=n, n-1,…,2*

Значення *t* беруться в зворотному порядку. Як початкове значення для моделі (10.10) береться саме останнє за часом значення Y_n. Після того як значення Y_1, моделі (10.10) визначено, ми знову повертаємося до початкової моделі (10.7). Значення моделі (10.10) полягає в тому, щоб як початкове значення узяти згладжене, визначене загальним ходом процесу, число. Наприклад, якщо на початку процесу відмічалися значні коливання, то тепер для визначення остаточного прогнозу вони не виконуватимуть істотної ролі. Можна було б подумати, що в результаті ми повернемося до того ж значенню Y_n. Але це не так. Рівняння (10.10) можна записати в наступному вигляді:

$$\hat{e}_{t-1} = (1-\alpha)e_t + \alpha\hat{e}_t \qquad (10.11)$$

де, *t=n, n-1,…,2*

Процес прогнозування здійснюється таким чином. Знаходиться значення згідно з формулою (10.10), рухаючись вгору, починаючи з останнього значення. Таким чином, знаходиться згладжене значення e_1, відповідне 1-му кварталу. Потім переходимо до моделі (10.8) з набутим початковим значенням. Приклад прогнозування назад представлений на рис. 10.6.

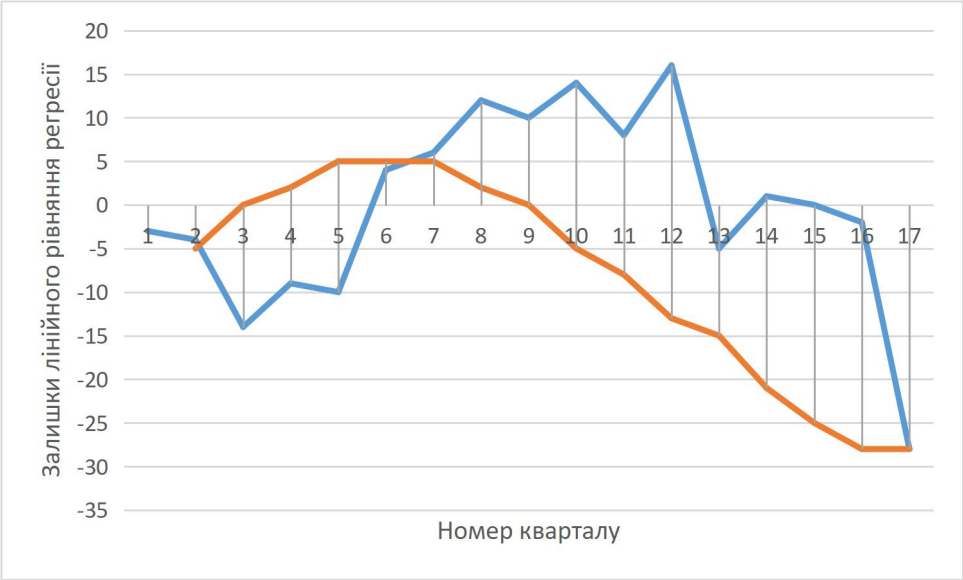


Рисунок 10.6 – Модель прогнозування назад

Початкове значення моделі можна визначити за допомогою кривої підгонки. Оскільки e_i є залишками моделі лінійної регресії, то їх коефіцієнт b_2 буде рівний нулю і лінійне рівняння регресії зведеться до тотожності $e = 0$.

Коли є достатня кількість інформації, її можна розбити на дві групи. Потім значення прогнозу моделі для першої групи можна використовувати як початкове значення моделі для другої групи. Як початкове значення для першої групи можна узяти $Y_1 = Y_1$ або $Y_1 = \overline{Y_1}$. У обох групах беруться оптимальні значення a.

Початкове значення можна взяти відповідно до досвіду і розуміння прогнозистом процесу.

Найкращий спосіб визначення початкових значень, особливо для прогнозистів-початківців, полягає у виборі опції, «автоматично» присутньої у всіх статистичних пакетах.

10.5. Моделі згладжування з трендом: модель Холта, модель Брауна. Модель Холта

В лінійній моделі різниця $Y_t - \widehat{Y}_{t-1}$ дорівнювала b_2, тобто служила характеристикою наявного тренда. Звичайно, було б помилкою припускати, що в рівнянні (9) відповідна різниця буде константою. Проте вираз $Y_t - \widehat{Y}_{t-1}$ може розглядатися як величина тренда для нашого процесу в даний момент часу.

Застосуємо модель експоненціального згладжування до різниць $\widehat{\Delta} = Y_t - \widehat{Y}_{t-1}$ з константою згладжування, рівною $1-\beta$, де $0 \leq \beta < 1$. При цьому значення моделі позначимо як T_t і назвемо їх значеннями тренда. Таким чином, одержуємо перше рівняння Холта:

$$T_t = (1 - \beta)(\widehat{Y}_t - \widehat{Y}_{t-1}) + \beta T_{t-1} \qquad (10.12)$$

Основна ідея Холта полягала в тому, щоб при визначенні Y_t формулі (10.9) додати до попереднього значення моделі Y_{t-1} відповідне значення тренда T_{t-1}. Отже, одержуємо друге рівняння Холта:

$$\widehat{Y}_t = (1-\alpha)Y_t + \alpha(\widehat{Y}_{t-1} + T_{t-1}) \qquad (10.13)$$

Побудова моделі зводиться до трьох основних етапів.

1. Ухвалюють рішення щодо величин параметрів моделі α і β. Якщо немає особливих переваг, то оптимальні значення α і β визначають як завжди, виходячи з мінімізації MSE.

2. Визначають початкові значення Y_t і T_t. Що ж до T_t., то можна узяти

$$T_1 = (\Delta_2 + \Delta_4)/2.$$

3. Поперемінно знаходять Y_t і T_t. При цьому спочатку обчислюють Y_t і лише потім – T_t.

Для прогнозу на m кроків вперед, який позначимо P_{n+m}, використовується третє рівняння Холта:

$$P_{n+m} = \widehat{Y}_n + mT_n \qquad (10.14)$$

Іншими словами, передбачається, що останнє значення тренда залишається незмінним для всіх майбутніх прогнозів.

Діаграма моделі Холта представлена на рис. 10.7.

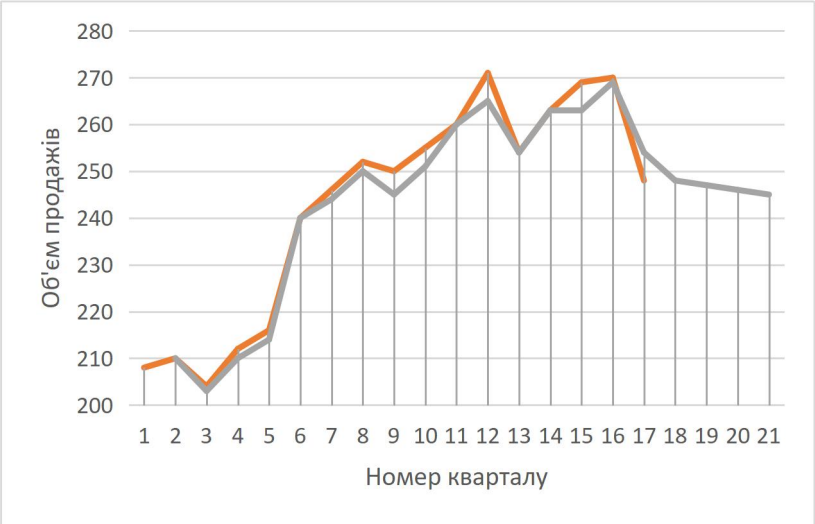


Рисунок 10.7 – Модель Холта, включаючи прогнози на три квартали вперед

Як ми бачимо, ця модель досить добре апроксимує початкові дані.

10.6 Сезонні моделі

Процес, що має кількісні характеристики, визначається такими трьома основними чинниками: тренд (*T*), сезонність (*S*) і випадкова помилка (*e*). Трендовий чинник іноді називають також трендоциклічним. Тоді як трендовий чинник відповідальний за загальну тенденцію процесу, сезонний чинник пов'язаний з впливом певних часових інтервалів на бізнес-процес. Тут важливо пам'ятати, що «сезон» в прогнозуванні є технічним терміном і не обов'язково означає квартал. Ось декілька найхарактерніших прикладів показників, що відповідають різним сезонам:

1) об'єм продажів одягу по кварталах;
2) кількість заявок в туристичній фірмі по місяцях;
3) завантаженість наземного транспорту в місті по днях тижня;
4) кількість покупців в супермаркеті погодинно.

Третій тип випадкових чинників, який також називають помилкою, виникає через непередбачені обставини. У сфері кібербезпеки це, наприклад, може бути раптовий вихід нової вразливості в популярному програмному забезпеченні або непередбачена зміна політики безпеки компанії. Часто такі випадкові чинники проявляються абсолютно несподівано й непередбачувано. Наприклад, це може бути масова кібератака на постачальника хмарних послуг або різкий сплеск фішингових кампаній у соціальних мережах. Навіть детальний подальший аналіз не завжди дає змогу виявити всі фактори, які могли вплинути на безпекову ситуацію або працездатність системи.

Величина:

$$d_t = \frac{Y_t}{S_t} \tag{10.15}$$

називається десезоналізованим значенням. Тут S – сезонний чинник (його називають також сезонним коефіцієнтом), відповідний до моменту часу *t*.

Кажучи менш формальною мовою, десезоналізовані дані – це ті гіпотетичні значення, які могли б прийняти наші дані, якби не було сезонних впливів.

З формули (10.15) випливає:

$$Y_t = d_t S_t. \tag{10.16}$$

Іншими словами, початкові дані є добутком своїх десезоналізованих значень і сезонних чинників. З формули (10.15) також слідує формула:

$$S_i = \frac{Y_t}{d_t} \tag{10.17}$$

Задача полягає в тому, щоб знайти сезонні чинники S_i. Але для цього необхідно спочатку визначити d_t.

З формули (10.17) виходить очевидний, але важливий результат: всі коефіцієнти сезонності не можуть одночасно бути більше або, навпаки, менші за одиницю. Інакше їх сума не буде дорівнювати чотирьом.

Для знаходження десезоналізованих значень ми поступимо так само, як чинити при побудові моделі ковзного середнього. Щоб нейтралізувати сезонні впливи, порядок ковзного середнього k береться рівним 4. Щоб нейтралізувати вплив тренда, ми розміщатимемо кожне ковзне середнє у середині відповідної групи. Числа, що є десезоналізованими значеннями в першому наближенні, ми позначимо d'.

На даний момент ми ще не можемо скористатися формулою (10.10) для знаходження сезонних компонент, оскільки немає точної відповідності між початковими і десезоналізованними даними. Ця проблема легко розв'язується шляхом утворення центрованих ковзних середніх другого порядку. Ці дані ми позначили d'', які виходять шляхом додаткового згладжування десезоналізованих значень i, природно, так само будуть десезоналізованими значеннями для початкових даних. Тепер можемо застосувати формулу, щоб одержати коефіцієнти сезонності.

Для того, щоб отримати однакові значення сезонних чинників за одні і ті ж квартали різних років, сезонні значення сортують по кожному кварталу по роках, а потім знаходять їх середні.

Числа *i* приймають за остаточні значення для сезонних чинників. Тепер можна знайти остаточні десезоналізовані значення.

На рис. 10.8 можна побачити початкові значення до *i* після десезоналізації.

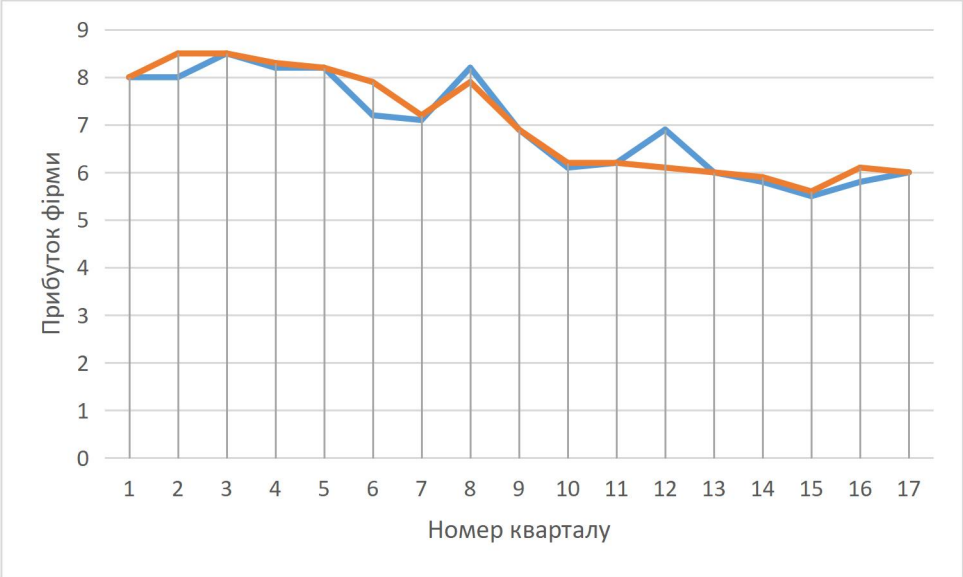


Рисунок 10.8 – Початкові значення до *i* після десезоналізації

Видно, що піки і западини, викликані сезонними впливами, після десезоналізації зробилися менш помітними. Тепер підсумовуємо все вищесказане у вигляді алгоритму.

Алгоритм для знаходження сезонних чинників і десезоналізованих значень:

- знаходимо ковзні середні d' 1-го порядку для початкового ряду Y_t;
- знаходимо (центровані) ковзні середні d'' 2-го порядку для ряду d'_t;
- знаходимо неусереднені коефіцієнти сезонності $S_i^{'} = \frac{Y_t}{d_t^{''}}$;
- усереднюємо значення S', щоб одержати коефіцієнти сезонності S_1, S_2,S_3 і S_4;
- знаходимо остаточні десезоналізовані значення d_t.

Тепер ми можемо побудувати сезонну модель *S*. Робиться це таким чином. Спочатку визначається яка-небудь модель для десезоналізованих значень. Наприклад, як таку модель ми можемо узяти одну з кривих підгонки. Потім значення моделі множаться на відповідні коефіцієнти сезонності. Набуті значення дадуть нам сезонну модель.

Тепер, щоб одержати сезонну модель *S,* помножимо значення моделі на відповідні коефіцієнти сезонності. На рис. 10.9 представлені початкові дані і змодельовані значення S_t.

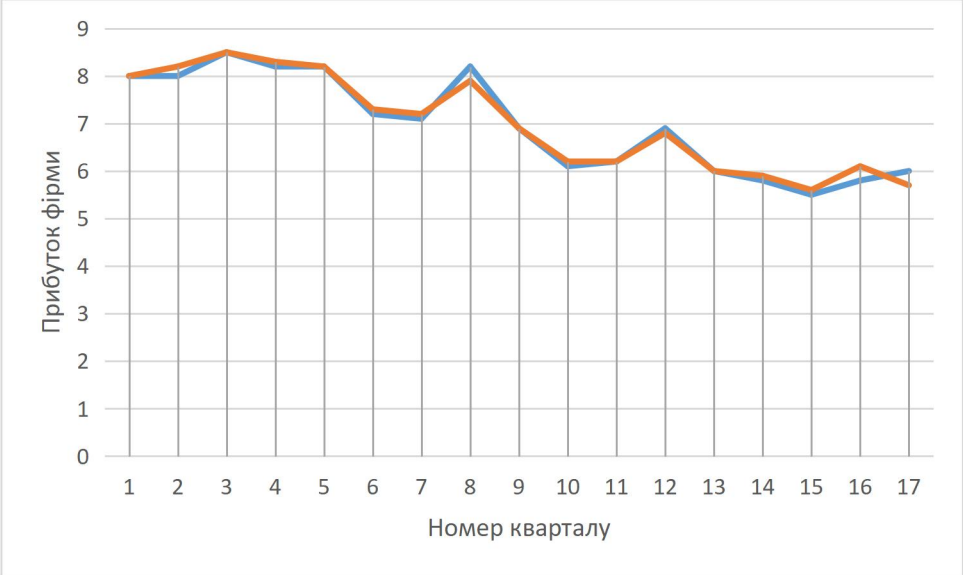


Рисунок 10.9 – Графік сезонної моделі

Прогноз визначається: спочатку ми повинні визначити прогноз для відповідного значення тренда, який помножимо на коефіцієнт сезонності, відповідний певному кварталу.

Перелік запитань для самопідготовки:

1. Що таке наївна модель прогнозування?
2. У яких випадках доцільно використовувати наївну модель?
3. Які є варіанти наївної моделі (наприклад, з урахуванням сезонності)?
4. Які переваги та недоліки наївної моделі?

5. Які методи існують для усунення тренда в часових рядах?

6. Що таке детрендування, і як воно здійснюється?

7. У чому різниця між фільтрацією та диференціюванням?

8. Як впливає усунення тренда на якість прогнозу?

9. У чому суть методу ковзного середнього?

10. Як працює експоненційно зважене ковзне середнє (EWMA)?

11. Що таке комбінована модель згладжування?

12. Як обрати оптимальну довжину вікна або коефіцієнт згладжування?

13. Які переваги експоненційного згладжування порівняно з простим ковзним середнім?

14. Чому важливо правильно задати початкові значення?

15. Які методи використовуються для встановлення початкових значень?

16. Як початкові значення впливають на точність моделі?

17. У чому полягає модель Холта? Як вона враховує тренд?

18. Чим відрізняється модель Брауна від моделі Холта?

19. Як задаються параметри цих моделей?

20. Які ситуації підходять для використання кожної з моделей?

21. Що таке сезонність у часових рядах?

22. Які методи врахування сезонності існують у згладжувальних моделях?

23. Як працює сезонна модель Холта-Вінтерса?

24. Як визначити сезонну компоненту в ряді?

25. Як обирати період сезонності?

Тема 11. Використання ковзного середнього при прогнозуванні

11.1. Основні відомості.

11.2. Основні методи згладжування часових рядів.

11.1. Основні відомості

При застосуванні методу ковзного середнього прогноз на певний період визначається як середнє значення кількох останніх спостережень тимчасового ряду. Наприклад, якщо обрано ковзне середнє за три місяці, прогноз кількості виявлених кіберінцидентів у травні буде обчислюватися як середнє значення показників за лютий, березень і квітень. Якщо ж використовується ковзне середнє за чотири місяці, прогноз травневих інцидентів обчислюється як середнє за січень, лютий, березень і квітень.

Такий метод дозволяє досить точно відобразити зміни основних показників попереднього періоду. Іноді він може виявитися ефективнішим за методи, що базуються на довготривалих спостереженнях. Наприклад, якщо ваша компанія протягом кількох років відстежувала середню кількість щомісячних фішингових атак на корпоративну мережу і цей показник стабільно становив 100 інцидентів, але тепер запроваджено нову систему багатофакторної аутентифікації, логічно очікувати, що кількість атак тимчасово зменшиться.

Якщо для прогнозу травневих інцидентів скористатися середнім значенням за останні 4 місяці, результат може бути дещо завищеним. Натомість ковзне середнє за останні два місяці дозволить точніше відобразити наслідки впровадження нової системи безпеки.

При використанні двомісячного ковзного середнього кожен показник двох останніх місяців має однакову вагу – по половині у формуванні прогнозу. У разі 4-місячного ковзного середнього ці ж останні два місяці впливають лише на четверту частину прогнозного значення.

Отже, чим менша кількість попередніх спостережень, на основі яких обчислюється ковзне середнє, тим точніше воно відображає зміни базового

рівня показника. Проте якщо базою прогнозу є лише одне або два спостереження, такий прогноз може виявитися надто спрощеним і відображати тенденції даних не краще, ніж сама базова лінія.

Метод ковзного середнього має три основні обмеження. По-перше, він ефективний лише там, де щомісячні зміни невеликі; хоча можливі деякі коливання, протягом прогнозного горизонту показники повинні залишатися відносно стабільними. По-друге, прогноз слід будувати на короткий часовий горизонт. По-третє, якщо прогнозована змінна (наприклад, кількість кібератак або інцидентів у мережі) сильно коливається, метод може бути непридатним.

Існують і простіші способи аналізу часових рядів, такі як графічне відображення даних, але вони не дозволяють одержати точний прогноз. Більш точні результати забезпечує метод лінійних трендів із використанням найменших квадратів, який дозволяє адекватніше оцінити майбутні зміни в рівні загроз або інцидентів у сфері кібербезпеки.

11.2. Основні методи згладжування часових рядів.

Незалежно від специфіки кожного часового ряду, можна виділити основні типи завдань, які зазвичай вирішують при аналізі початкових даних.

На першому етапі прагнуть створити просту математичну модель, яка коротко описує поведінку часового ряду. У контексті кібербезпеки це може бути модель динаміки кількості виявлених кібератак або інцидентів у мережі протягом певного періоду.

Далі робиться спроба пояснити спостережувану поведінку ряду за допомогою інших змінних і оцінити ступінь їх взаємозв’язку – як між окремими спостереженнями одного ряду, так і між різними рядами. Наприклад, можна досліджувати зв’язок між кількістю виявлених вразливостей і частотою фішингових атак або між оновленнями систем безпеки та інцидентами в мережі.

Виходячи з цілей аналізу, кожний часовий ряд зазвичай розглядають як комбінацію кількох компонент, що впливають на його поведінку.

– тренд (trend) або довгострокова тенденція в розвитку ряду;

–сезонна компонента або, іншими словами, деякий ефект в динаміці ряду, який повторюється через певний період.

Розподіл динаміки часового ряду на окремі компоненти визначає й вибір груп математичних методів для аналізу кожної з них.

Для виявлення та аналізу тренду зазвичай застосовують регресійний аналіз і методи ковзного середнього. У сфері кібербезпеки це може бути використано для оцінки довгострокової динаміки кількості кібератак, інцидентів або виявлених вразливостей. Для аналізу сезонних ефектів використовують спеціальні моделі сезонного згладжування та сезонної авторегресії, що допомагає, наприклад, передбачати зростання фішингових атак під час святкових періодів або масових кампаній. Існують навіть спеціальні моделі для прогнозування наслідків втручань, наприклад, оцінка впливу нового патчу безпеки або запуску антивірусної кампанії.

Коливання відносно тренда виявляються за допомогою спектрального аналізу, а для їх опису й прогнозування застосовують гармонійні моделі або моделі авторегресії, включно з методом ковзного середнього. У теорії обробки часових рядів існує безліч способів їх згладжування: фільтрація за допомогою перетворення Фур’є, кускова апроксимація многочленами тощо.

Для прогнозування часові ряди ефективні, якщо є значна кількість історичних даних і якщо минулі тенденції чіткі та відносно стабільні. При цьому передбачається, що минуле добре відображає майбутнє. Аналіз часових рядів дозволяє визначити, що відбудеться за відсутності зовнішніх втручань, але він не здатний передбачити раптові зміни тенденції. Тому такий аналіз переважно застосовується для короткострокових прогнозів.

Криві тренда згладжують динамічний ряд даних, виділяючи загальну тенденцію. Вибір кривої тренда, що є непростим завданням, значною мірою визначає результати прогнозу.

У більшості випадків динамічний ряд, крім тренда і випадкових відхилень, містить також сезонні та циклічні компоненти. Циклічні коливання відрізняються більшою тривалістю та нестабільністю амплітуди. Сезонна складова зазвичай вимірюється днями, тижнями або місяцями, а циклічна – роками чи десятиліттями. У кібербезпеці сезонні компоненти можуть відображати періодичне зростання атак, а циклічні – більш довгострокові тенденції, наприклад, появу нових класів загроз.

Основними методами аналізу часових рядів є ковзне середнє, експоненціальне згладжування та проектування тренда. Для ілюстрації роботи цих методів розглядається один і той же часовий ряд, наприклад, щомісячна кількість зафіксованих кіберінцидентів.

Приклад 1

Припустимо, що гількість виявлених інцидентів інформаційної безпеки в локальній мережі фірми, описуються в перебігу тижня часовим рядом:

Таблиця 11.1

Вхідні показники

День тижня	Кількість виявлених в локальній мережі фірми інцидентів інформаційної безпеки
Понеділок	10
Вівторок	6
Середа	5
Четвер	11
П’ятниця	9
Субота	8
Неділя	7

Або дещо по-іншому

Таблиця 11.2

Вхідні показники

x	1	2	3	4	5	6	7
t	10	6	5	11	9	8	7

Метод рухомого (ковзного) середнього

Цей метод розділяють на 2 типи: метод рухомого (ковзного) середнього і метод зваженого (ковзного) середнього.

Метод рухомого (ковзного) середнього.

Цей метод полягає в тому, що розрахунок показника на прогнозований момент часу будується шляхом усереднювання значень цього показника за декілька попередніх днів.

Припустимо, що у нас є дані показника тільки за перші три дні. Обчислимо прогнозоване число випадків на четвертий день тижня (четвер). Для цього визначимо середнє значення числа випадків за попередні три дні – понеділок, вівторок і середу – і знайдемо їх середнє арифметичне:

$$f_4 = \frac{10+6+5}{3} = \frac{21}{3} = 7. \quad (11.1)$$

Прогнозований об'єм інцидентів на п'ятницю обчислюється аналогічним чином за реальними показниками за три попередні дні – вівторок, середу і четвер:

$$f_5 = \frac{6+5+11}{3} = \frac{22}{3} \approx 7{,}33. \quad (11.2)$$

Подібним способом розраховуються прогнози на суботу, неділю і черговий понеділок:

$$f_6 = \frac{5+11+9}{3} = \frac{25}{3} \approx 8{,}33$$

$$f_7 = \frac{11+9+8}{3} = \frac{28}{3} \approx 9{,}33$$

$$f_8 = \frac{9+8+7}{3} = \frac{24}{3} \approx 8$$

В результаті можна отримати таку таблицю:

Таблиця 11.3

Результати

x	1	2	3	4	5	6	7	8
t	10	6	5	11	9	8	7	-
f	-	-	-	7	7,33	8,33	9,33	8

Порівняльні результати наведені на рисунку 11.1: синьою лінією відзначені реальні значення, а помаранчевою – прогнозовані.

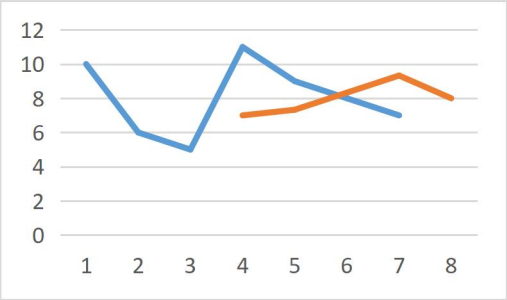


Рисунок 11.1 – Порівняльні результати прогнозу методом ковзного середнього

Для загального випадку розрахункова формула виглядає так:

$$f_k = \frac{x_{k-N}+x_{k-N+1}+\ldots+x_{k-1}}{N} \quad (11.3)$$

або

$$f_k = \frac{1}{N}\sum_{i=1}^{N} x_{k-i}, \quad (11.4)$$

де x_{k-i} – реальне значення показника у момент часу t_{k-i}; N – число попередніх моментів часу; f_k– прогноз на момент часу t_k.

Метод зваженого (ковзного) середнього.

При прогнозуванні методом усереднювання часто доводиться спостерігати, що ступінь впливу використаних при розрахунку реальних показників виявляється неоднаковим, при цьому звичайно більш «свіжі» дані мають більшу вагу. Математично метод зваженого рухомого середнього можна записати як:

$$f_k = \frac{\sum_{i=1}^{N} w_{k-i}x_{k-i}}{\sum_{i=1}^{N} w_{k-i}}, \quad (11.5)$$

де x_{k-i} – реальне значення показника у момент часу;

N – кількість попередніх моментів часу, що використовуються при розрахунку; f_k – прогноз на момент часу t_k; w_i – вага, з якою використовується показник x_{k-i} при розрахунку.

Вага – це завжди додатне число. У разі, коли вся вага однакова, ми одержуємо формулу (11.1).

Для розрахунків звернемося до початкового часового ряду з прикладу 1, вважаючи, що при складанні прогнозу на завтрашній день кількість сьогоднішніх випадків береться з вагою 60, вчорашніх – з вагою 30, а позавчорашніх – з вагою 10.

Маємо:

$$f_4 = \frac{10 * 10 + 30 * 60 + 60 * 5}{100} = \frac{580}{100} \approx 5{,}80$$

$$f_5 = \frac{10 * 6 + 30 * 5 + 60 * 11}{100} = \frac{870}{100} \approx 8{,}70$$

$$f_6 = \frac{10 * 5 + 30 * 11 + 60 * 9}{100} = \frac{920}{100} \approx 9{,}20$$

$$f_7 = \frac{10 * 11 + 30 * 9 + 60 * 8}{100} = \frac{860}{100} \approx 8{,}60$$

$$f_8 = \frac{10 * 9 + 30 * 8 + 60 * 7}{100} = \frac{750}{100} \approx 7{,}50$$

Результати розрахунків наведені в таблиці.

Таблиця 11.4

Вхідні показники

x	1	2	3	4	5	6	7	8
t	10	6	5	11	9	8	7	-
f	-	-	-	5,7	8,7	9,2	8,6	7,5

На рисунку 10.2 синьою лінією відзначені реальні значення, а помаранчевою – прогнозовані.

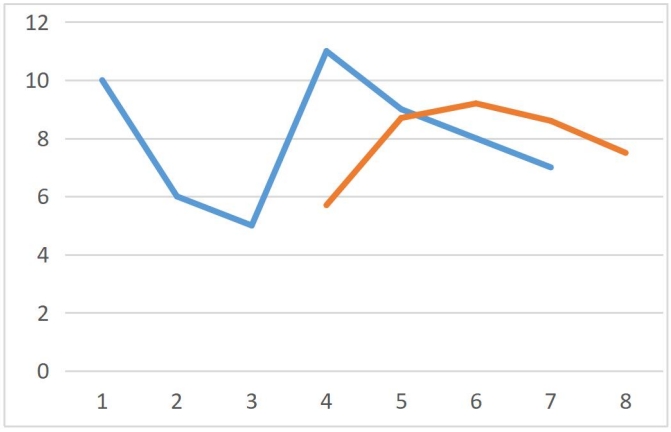


Рис. 11.2 – Порівняльні результати прогнозу методом зваженого ковзного середнього

Метод експоненціального згладжування

Експоненціальне згладжування є дуже популярним методом прогнозування різних часових рядів і історично було незалежно відкрите Броуном і Холтом.

Цей метод передбачає згладжування даних тимчасового ряду з урахуванням «старіння» попередніх спостережень: чим давніше спостереження, тим менший вплив воно має на прогноз. Інакше кажучи, більш свіжі дані мають більшу вагу, а старі – поступово зменшують свій вплив.

Метод експоненціального згладжування вважається ефективним та універсальним. Його основні переваги – можливість врахування ваги вихідної інформації, простота обчислень, гнучкість у відображенні різних динамік процесів. Він дозволяє оцінити параметри тренда, що відображають не середній рівень процесу, а актуальну тенденцію на момент останнього спостереження.

У сфері кібербезпеки цей метод можна застосовувати, наприклад, для прогнозування кількості кібератак або інцидентів у мережі, оцінюючи сучасні тенденції, а не усереднені значення за довгий період. Найбільше експоненціальне згладжування підходить для середньострокових прогнозів. Ключовими аспектами є правильний вибір параметра згладжування та визначення початкових умов для ряду даних.

При розрахунку прогнозу методом експоненціального згладжування (exponential smoothing) враховується відхилення попереднього прогнозу від реального показника, а сам розрахунок проводиться за такою формулою

$$f_k = f_{k-1} + \alpha(x_{k-1} - f_{k-1}). \quad (11.6)$$

де α – стала згладжування ($0 < \alpha < 1$).

Коли ця формула застосовується рекурсивно, то кожне нове згладжене значення (яке є також прогнозом) обчислюється як зважене середнє поточного спостереження і згладженого ряду. Рекомендується як f_0 брати початкове значення, що дає якнайкращий прогноз.

Очевидно, результат згладжування залежить від параметра α. Якщо α дорівнює 1, то попередні спостереження повністю ігноруються. Якщо α дорівнює 0, то ігноруються поточні спостереження. Значення α між 0 і 1 дають проміжні результати.

З формули (10.6) виходить, що α має потрапляти в інтервал між 0 і 1 . На практиці звичайно рекомендується брати α менше 0,3. Крім того, параметр згладжування α часто знаходиться пошуком на сітці. Можливі значення параметра α розбиваються сіткою з певним кроком. Наприклад, розглядається сітка значень α від α = 0,1 до α = 0,9, з кроком 0,1. Потім вибирається той α, для якого сума квадратів (або середніх квадратів) залишків (значення, що спостерігаються, мінус прогнози на крок вперед) є мінімальною.

Для розрахунків знов звернемося до початкового часового ряду, вважаючи, що α = 0,2 і що прогноз на понеділок дорівнює 8.

8,00+0,2(10-8,00)=8,40

8,40+0,2(6-8,40)=7,92

7,92+0,2(5-7,92)=7,34

7,34+0,2(9-7,34)=8,07

8,07+0,2(9-8,07)=8,26

8,26+0,2(8-8,26)=8,21

8,21+0,2(7-8,21)=7,93

Результати зведені в таблицю.

Таблиця 11.5

Вхідні показники

x	1	2	3	4	5	6	7	8
t	10	6	5	11	9	8	7	-
f	-	8,4	7,92	7,34	8,07	8,26	8,21	7,93

На рисунку 11.3 синьою лінією відзначені реальні значення, а помаранчевою – прогнозовані.

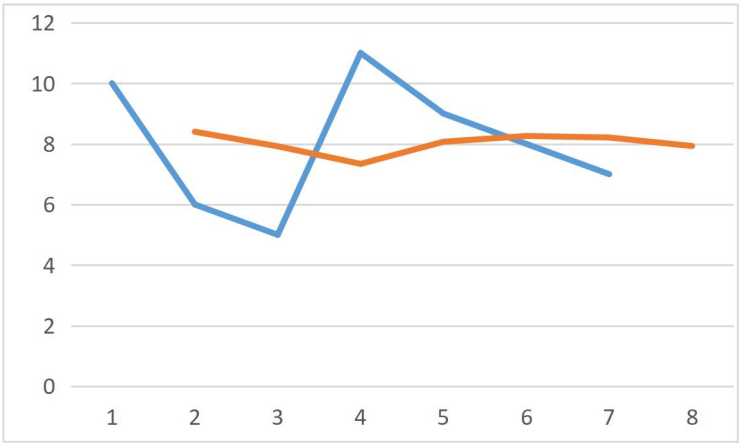


Рисунок 11.3 – Порівняльні результати прогнозу методом експоненціального згладжування

Варто пам’ятати, що при розв’язанні реальної задачі прогнозування часовий ряд формується поетапно, і фактичне значення показника на прогнозований момент часу зазвичай заздалегідь невідоме. Проте перед тим, як робити прогнози за допомогою одного з методів, зазвичай виконують обчислення на основі повного історичного ряду даних. У сфері кібербезпеки це означає, що для прогнозування кількості кібератак або інцидентів спочатку аналізують всю наявну інформацію про попередні інциденти протягом певного періоду часу. Це робиться для того, щоб підібрати відповідне значення f_0 і порівняти результати прогнозу з реальними даними (метод простого ковзного середнього), підібрати відповідні значення f_0 та вагів і порівняти результати прогнозу з реальними даними (метод зваженого ковзного середнього).

Метод проектування тренда

Основною ідеєю методу проектування тренда є побудова прямої, яка «в середньому» якнайменше відхиляється від масиву точок (*t*,*x*) заданого часового ряду (рис. 11.4), що описується рівнянням:

$$x=at+b. \qquad (11.7)$$

де *a*, *b* – сталі коефіцієнти.

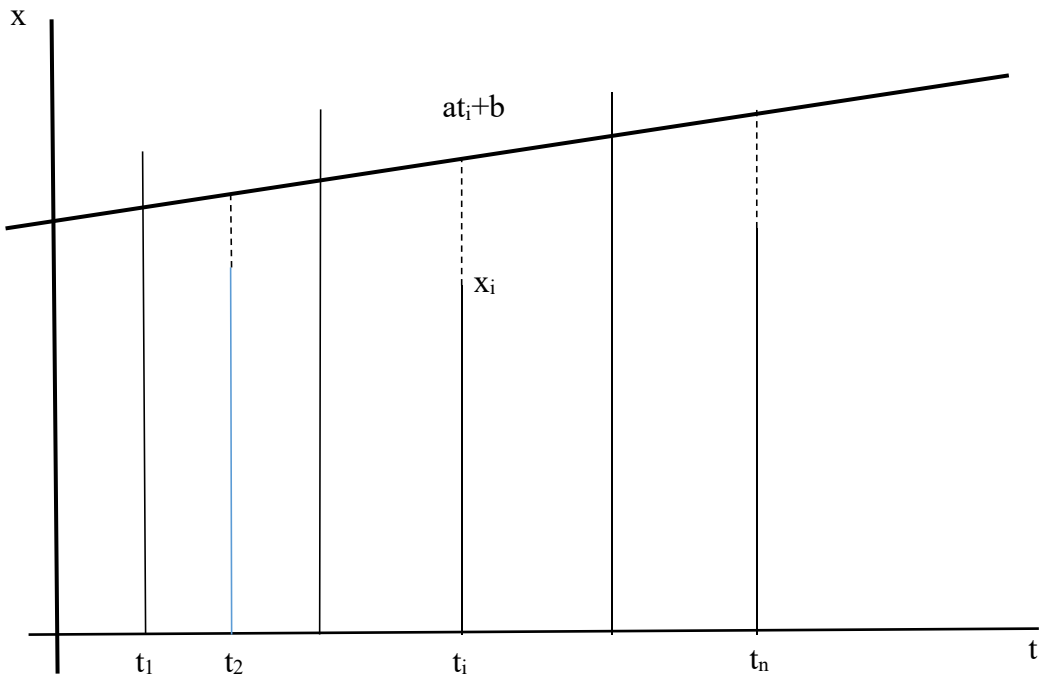


Рисунок 11.4 – Тестовий часовий ряд

Розрахунок коефіцієнтів *a* і *b* ведеться за методом найменших квадратів.

Щоб знайти коефіцієнти *a* і *b*, виконують такі дії: для кожного значення *t* і змінної *x*, користуючись формулою (11.7), обчислюють відповідне значення змінної x_i: $x_i+b, i=1,2,...n$.

Потім знаходять різницю $at_i+b-x_i,\ i=1,2,...n$ яку потім підносять до квадрату (щоб не думати про знак): $(ati+b-xi)^2,\ i=1,2,...n$.

Функція набуває мінімального значення у тому випадку, коли величини *a* і *b* задовольняють таку лінійну систему:

$$\varphi(a,b)=\sum_{i=1}^{n}(at_i+b-x_i)^2\rightarrow min$$

$$\begin{cases} a\sum_{i=1}^{n} t_i + bn = \sum_{i=1}^{n} x_i \\ a\sum_{i=1}^{n} t_i^2 + b\sum_{i=1}^{n} t_i = \sum_{i=1}^{n} t_i x_i \end{cases} \quad (11.8)$$

Ця система завжди має єдине рішення.Розглянемо конкретний приклад, знов використавши часовий ряд з прикладу 1. Складемо допоміжну таблицю:

Таблиця 11.6

Допоміжна таблиця

t_i	x_i	$t_i x_i$	t_i^2
1	10	10	1
2	6	12	4
3	5	15	9
4	11	44	16
5	9	45	25
6	8	48	36
7	7	49	49
$\sum_{i=1}^{n} t_i = 28$	$\sum_{i=1}^{n} x_i = 56$	$\sum_{i=1}^{n} t_i x_i = 223$	$\sum_{i=1}^{n} t_i^2 = 140$

В такому випадку система рівнянь для обчислення коефіцієнтів *a* і *b* набуде такого вигляду:

$$\begin{cases} 28a + 7b = 56 \\ 140a + 28b = 223 \end{cases}$$

Розв’язавши систему, отримуємо: $a = -\frac{1}{28} \approx -0{,}04$, $b = \frac{57}{7} \approx 8{,}14$.

Тоді *x=-0.04t+8,14* – рівняння шуканого тренда. Розрахунок показника для наступного дня за допомогою отриманого рівняння тренда проводиться таким чином: f_8*=-0,04+8,14=7,82*.

На рисунку 11.5 показано реальну та прогнозовану криві, побудовані за допомогою методу проектування тренда.

Точність прогнозу далі можна оцінити за допомогою кореляційного аналізу. Наведені вище методи не вичерпують різноманіття методів аналізу

часових рядів, з яких багато таких, що не тільки основані на простих розрахунках, але й потребують значної аналітичної і комп'ютерної бази.

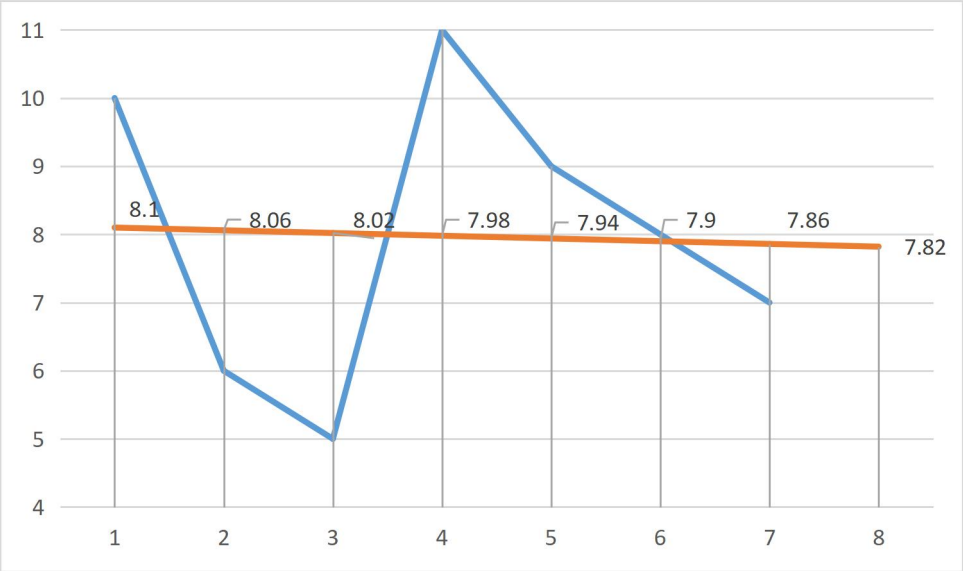


Рисунок 11.5 – Порівняльні результати прогнозу методом проектування тренда

Для складання середньострокових та довгострокових прогнозів застосовуються казуальні та якісні методи, які значно складніші від наведених вище методів аналізу часових рядів.

Перелік запитань для самопідготовки:

1. Що таке ковзне середнє та як воно розраховується?
2. У яких випадках доцільно застосовувати ковзне середнє для аналізу часових рядів?
3. Як ковзне середнє допомагає виявляти тренди та сезонні коливання?
4. Які бувають типи ковзного середнього (просте, зважене, експоненційне)?
5. У чому полягає різниця між ковзним середнім і експоненційним згладжуванням?
6. Які основні методи згладжування існують у статистиці?

7. Як відбувається згладжування за допомогою простого ковзного середнього?

8. У чому полягає метод експоненційного згладжування?

9. Які переваги й недоліки має кожен метод згладжування?

10. Як обрати оптимальний метод згладжування для конкретного типу даних?

Тема 12. Експертні методи прогнозування

12.1. Принципи формування експертних систем прогнозування.

12.2. Індивідуальні і колективні експертні методи.

12.3. Етапи проведення колективної експертної оцінки.

12.1. Принципи формування експертних систем прогнозування

Вплив непередбачуваних подій або факторів, які неможливо врахувати у статистичній моделі, досліджувався, і було показано, що залучення експертних оцінок зазвичай підвищує точність прогнозів.

Для побудови надійної системи прогнозування, яка поєднує судження експертів із статистичними методами, необхідно визначити всі ключові компоненти, важливі для успішного прогнозу, та детально їх охарактеризувати. У контексті кібербезпеки це може включати оцінку ймовірності нових вразливостей, потенційних атак або змін у поведінці користувачів, що не відображаються в історичних даних.

1. Характеристики часових рядів.

1.1 Тренд. Наявність трендової компоненти досить чітко виявляється експертами в порівнянні зі статистичними методами, які можуть бути неточними у разі великих збурень.

1.2. Сезонність. Наявність сезонності пов'язана з більш високою помилкою при прогнозуванні як для статистичних, так і для експертних методів. Дуже важко визначити, який з типів методів є кращим у загальному випадку.

1.3. «Шум». «Шум», або збурення, як правило, заважають цілісному сприйняттю часового ряду людиною. Особливо це помітно у разі великого «шуму», що становить більше 10% від середнього значення часового ряду. В такому випадку точність прогнозу підвищують статистичні методи.

1.4. Нестабільність. Нестабільність може бути пов’язана з відсутністю чіткого закону розвитку процесу або часовими змінами. У таких випадках експертне прогнозування значно підвищує точність прогнозу.

1.5. Кількість наявних спостережень. Цей аспект є надзвичайно важливим для більшості статистичних методів, оскільки для коректної ідентифікації моделі необхідна достатня база даних часового ряду. У той же час наявність великого обсягу даних може знизити точність експертних прогнозів. Причина полягає в тому, що велика кількість якісних і кількісних ознак, які важко або неможливо виміряти, змушує експертів оцінювати значення атрибутів і взаємозв’язки між ними, що ускладнює процес прогнозування. У контексті кібербезпеки це може стосуватися, наприклад, прогнозування кіберзагроз: якщо в системі накопичено надто багато даних про різні види атак, мережеві журнали та поведінку користувачів, експертам стає складніше точно оцінити ймовірність конкретного інциденту.

1.6. Кількість періодів для прогнозування. Практика показує, що короткотермінове прогнозування краще від довгострокового для обох типів методів. При довгостроковому прогнозуванні перевагу мають статистичні методи.

1.7. Представлення даних. Воно може здійснюватися у табличній або графічній формах. У літературі існує гіпотеза про деякий вплив типу представлення на експертні прогнози.

1.8. Можливість аналізу попередніх прогнозів. Суть цього підходу полягає в тому, що дослідник перед побудовою прогнозу, використовуючи статистичні методи або експертні оцінки, аналізує їхні помилки за попередній період. Це дозволяє виявити найбільш надійні джерела інформації. Дослідження показують, що в експериментальних умовах такий аналіз значно підвищує точність прогнозів. Імовірно, у реальних умовах цей ефект також спостерігається, хоча він може бути менш очевидним. Такий підхід дозволяє ранжувати експертів або статистичні методи за точністю прогнозів хоча б на

один період, виявляти найбільш невдалих і, за необхідності, відмовлятися від їхніх оцінок у майбутньому. У сфері кібербезпеки це може означати оцінку надійності аналітиків або алгоритмів прогнозування кібератак, щоб визначити, які джерела інформації давали найбільш точні результати. Основним недоліком є складність отримання експертних оцінок у реальному часі: наприклад, для прогнозу кількості інцидентів щоквартально потрібно збирати тих самих експертів щоразу, що є досить витратним і трудомістким.

2. Характеристики експертів.

2.1. Досвід. Цей фактор включає оцінку накопиченого експертами досвіду при прогнозуванні часових рядів. Розрізняють два його аспекти:

2.1.1.Технічні знання, які стосуються методів аналізу та формальних процедур прогнозування.

2.1.2.Причинні знання, які стосуються розуміння експертом причинних ефектів взаємозв'язків. Ці знання набуваються із загального досвіду прогнозування.

Ці два аспекти не аналізувалися окремо, тому що їх майже неможливо розділити.

2.2. Контекстуальна інформація. Визначається як додаткові дані, що не входять до часового ряду, але допомагають пояснити, інтерпретувати та передбачити його поведінку, а також спираються на досвід експертів. Це визначення включає навіть назви самих часових рядів: експерт знає, який показник аналізується, і може враховувати фактори, які, на його думку, впливають на процес. У сфері кібербезпеки такими факторами можуть бути різкі зміни у законодавстві щодо безпеки, появи нових вразливостей або так звана «м’яка інформація», наприклад, чутки про плани кібератак або нові експлойти. Перевагою контекстуальної інформації є те, що вона дозволяє пояснити минулі зміни та прогнозувати майбутні. Без неї експерти зазвичай не можуть точно оцінити рівень загроз у випадку різкої зміни ситуації.

2.3. Мотивація. Вона визначає ті фактори, які впливають на вибір стратегії прогнозування експертом. Огляд літератури з цієї проблематики дає можливість виділити такі чотири фактори:

2.3.1. Якщо експерт отримує додаткову винагороду при точному прогнозуванні.

2.3.2. Якщо прогноз може бути змінений при отриманні додаткової інформації.

2.3.3. Якщо прогноз є частиною роботи експерта і тому в разі поганого прогнозу може постраждати його репутація.

2.3.4. Якщо експерт адекватно сприймає інформацію про мотивацію, оскільки в разі неточного прогнозу він втрачає обіцяні стимули. Високий рівень мотивації визначає додатковий ефект у роботі експертів, що особливо проявляється у разі наявності контекстуальної інформації.

Отож, експертне прогнозування має такі характеристики.

1. Відомі як інформація про часовий ряд, так і контекстуальна інформація.

2. Немає причин припускати незмінність поведінки часового ряду в майбутньому.

3. Існує можливість:

3.1. Організаційного або політичного впливу на утворення прогнозу.

3.2. Впливання прогнозу на економічне середовище, наприклад, економічний прогноз Міністерства економіки, має вплив на суб'єкти економічної діяльності.

3.3. Отримання прогнозу в результаті взаємодії групи експертів.

4. Експерт може:

4.1. Використовувати свій досвід та знання для вироблення прогнозу.

4.2. Впливати на розвиток процесу, тобто на значення прогнозованої змінної.

4.3. Робити дослідження щодо вивчення взаємовпливу деяких економічних змінних на даний процес.

4.4. Бути зацікавленим у тому, щоб процес розвивався деяким визначеним чином, наприклад, він є оптимістом або песимістом.

4.5. Бути матеріально та/або морально зацікавленим у точності прогнозу.

4.6. Перевіряти та аналізувати результати свого прогнозування.

Багато з цих характеристик втрачають значення під час досліджень у лабораторних умовах. Перевагою таких експериментів є можливість багаторазового контролю та повторення однакових умов. Проте в таких умовах майже неможливо оцінити вплив політичних, економічних або структурних факторів, які в реальному світі можуть суттєво впливати на ситуацію. Крім того, зменшується значення індивідуальних особливостей експертів – їхніх оптимістичних або песимістичних очікувань, а також мотивація для отримання максимально точного прогнозу.

Наступним етапом побудови системи прогнозування є визначення методів інтеграції статистичних моделей та експертних оцінок. У сфері кібербезпеки їх взаємодія може відбуватися, на нашу думку, шістьма основними способами, дозволяючи поєднувати історичні дані про кібератаки з судженнями аналітиків про ймовірність нових загроз.

1. Побудова моделі, коли експерту надається можливість вказати всі необхідні змінні, визначити структуру моделі та деяким чином перевести наявну контекстуальну інформацію у параметри моделі. Експертні оцінки в цьому випадку грають важливу роль у чотирьох напрямах:

1.1. Вибір змінної.

1.2. Специфікація моделі.

1.3. Оцінка параметрів моделі.

1.4. Аналіз даних.

2. Імовірнісне експертне прогнозування, коли експерти вказують імовірність здійснення деякої події, наприклад, зростання величини змінної в наступному періоді. На основі групи таких прогнозів будується загальний імовірнісний прогноз.

3. Комбінування прогнозів передбачає поєднання статистичного прогнозу та експертної оцінки певним чином для отримання більш точного прогнозу або зменшення невизначеності щодо майбутніх подій. Використання комбінованих прогнозів не означає, що окремі прогнози втрачають свою вагу. Вони надають цінну інформацію про переваги конкретних моделей, методів та взаємозв'язок між прогнозами. У практиці кібербезпеки це може означати поєднання автоматизованих алгоритмів прогнозування кількості кібератак із судженнями експертів-аналітиків. При цьому доцільно подавати як індивідуальні, так і комбіновані прогнози, щоб оцінити різницю та варіативність між ними. Це важливо, оскільки помилки прогнозування різними методами часто сильно корельовані.

4. Зміна статистичного прогнозу експертом передбачає, що аналітик коригує значення прогнозу, отриманого на основі статистичної моделі, враховуючи контекстуальну інформацію, яку модель не відображає. Цей підхід, мабуть, є найпоширенішим у практичній роботі. У сфері кібербезпеки це може означати, що аналітик змінює прогнозовану кількість кібератак на підставі додаткових факторів – наприклад, інформації про нові вразливості, плани хакерських груп або зміни в законодавстві. Слід зазначити, що підвищення точності прогнозу відбувається лише за наявності достовірної контекстуальної інформації. Без неї зміни прогнозу стають інтуїтивними і не науковими, що є головною критикою цього методу через його неформальний характер.

5. Зміна моделі прогнозування передбачає, що експерт на основі свого досвіду або контекстуальної інформації обирає саме ту статистичну модель, яка, ймовірно, дасть найменшу похибку для наступного періоду. Таким чином, знаючи поведінку різних методів на основі історичних даних часових рядів, експерт визначає не лише модель, а й її параметри. У сфері кібербезпеки це може означати, що аналітик обирає модель прогнозування кількості інцидентів або атак залежно від специфіки мережі, типу атак та попередніх спостережень, щоб мінімізувати похибку прогнозу для наступного місяця або кварталу. Цей

підхід поєднує в собі риси коригування статистичного прогнозу експертом та використання контекстуальної інформації. Таке прогнозування може виконуватися й виключно статистичними методами, як описано у підрозділі «Моделювання часових рядів при зміні економічної ситуації».

6. Експертна декомпозиція є триетапним процесом, у якому статистичні та експертні методи можуть взаємодіяти. Спершу часовий ряд поділяється (аналітиком або програмно) з урахуванням минулої контекстуальної інформації, наприклад, даних про попередні кібератаки, вразливості або зміни у мережевій інфраструктурі. Потім декомпозований ряд прогнозується окремо для кожної складової. Нарешті, прогноз інтегрується з урахуванням майбутньої контекстуальної інформації, наприклад, очікуваних змін у безпековому законодавстві чи планованих оновлень систем. Головна відмінність між зміною статистичного прогнозу експертом і експертною декомпозицією полягає в підході до контекстуальної інформації. Зміна прогнозу враховує минулі та майбутні фактори вже після формування прогнозу, тоді як експертна декомпозиція спочатку усуває вплив контекстуальних факторів із часового ряду перед прогнозуванням і лише потім інтегрує майбутні контекстуальні дані у кінцевий прогноз. Таким чином, декомпозиція дозволяє оцінювати конкретну ситуацію окремо від змін у зовнішньому середовищі, що особливо корисно для прогнозування динаміки кібератак або оцінки потенційних загроз. Така декомпозиція буде неефективною, якщо:

6.1. Декомпозиція є механічною та/або експерт відноситься скептично до методів, що використовуються.

6.2. Експерт не має досвіду у проведенні декомпозиції.

6.3. Неможливо виділити точну відповідність між контекстуальною інформацією та номером періоду, до якої вона насправді відноситься. Така ситуація можлива, наприклад, при прийнятті закону, дія якого починається лише через декілька періодів спостережень, але опосередкований вплив на

суб'єктів економічної діяльності, а значить, і на величину досліджуваної змінної, відбувся.

Останній з підходів є найбільш складним, але він є найбільш повним, оскільки враховується структура ряду, причому її можна використовувати багаторазово.

Для кожної події можна визначити її «істинний» показник. Експерти можуть його не знати, але завдяки досвіду та наявним статистичним даним вони можуть отримати приблизну оцінку події. Наприклад, у сфері кібербезпеки: якщо система реєструє 100 інцидентів, але через помилки в моніторингу або неповні дані враховано лише 80, експерт при складанні прогнозу буде виходити саме з цієї цифри. Далі аналітик може врахувати, що обрана статистична модель не враховує циклічність атак, і скоригувати прогноз: наприклад, прогноз на наступний період у 90 інцидентів слід інтерпретувати як 85. Таким чином, створення експертного прогнозу залежить від кількох чинників. У літературі зазвичай виділяють п’ять основних факторів, що впливають на точність та зміст таких прогнозів.

1. Передбачуваність навколишнього середовища визначає верхню границю точності виробки прогнозів, а тому показує, наскільки можливо підвищити їх точність завдяки іншим компонентам.

2. Прогностичні навички можуть бути нівельовані, якщо інформаційна система вимірює часовий ряд неточно.

3. Зв'язок між середовищем та експертом є оцінкою потенціальних навичок, яких експерт може досягти при обраній стратегії за умови ідеального вимірювання середовища та незміщених і надійних прогнозів. Очевидним шляхом для підвищення цього зв'язку є навчання експерта взаємозв'язкам у середовищі.

4. Надійність статистичних даних та додаткової інформації. Практика показує, що ненадійність даних є поширеним явищем. Вона може виникати як через прості помилки збору та обробки інформації, так і через готовність

системи (або організації) перейти на новий рівень функціонування. Надійність прогнозів підвищується зі збільшенням обсягу інформації, хоча іноді її кількість перевищує здатність людини обробляти дані. Водночас більший обсяг інформації дозволяє виявляти приховані взаємозв'язки між змінними. У сфері кібербезпеки це означає, що точність прогнозу кількості кібератак залежить від якості журналів подій, інформації про відомі вразливості та контекстуальних даних (наприклад, повідомлень про активність хакерських груп). Надійність таких даних може підвищуватися при інтеграції різних джерел, але водночас експертам доводиться обробляти великий потік інформації.

5. Зміщення. Воно складається з двох компонентів:

5.1. Умовне (регресійне зміщення).

5.2. Безумовний рівень (природний рівень зміщення).

Перший виникає внаслідок занадтої впевненості або невпевненості експертів в прогнозних значеннях, що призводить до заниження чи завищення рівня даних. Другий є звичайним рівнем зміщення прогнозу.

Взагалі слід відмітити, що зміщення майже неможливо подолати. У експериментальних умовах зміщення можна знизити за умови можливості повтору прогнозів або програмного коригування прогнозних значень.

Сучасне прогнозування неможливе без широкого поєднання статистичних методів та експертних оцінок. У останні роки статистичні методи значно розвинулися, проте наразі важливо зосередитися на розвитку та впровадженні методів створення та коригування експертних прогнозів.

Ми запропонували класифікацію як характеристик експертних суджень, так і факторів, необхідних для ефективного прогнозування. Наступним етапом має стати експериментальна робота, де експерти роблять прогнози та аналізують свої помилки. Вони можуть:

– складати власні прогнози на основі наявної контекстуальної інформації,

– коригувати статистичні прогнози,

– комбінувати статистичні моделі та експертні оцінки.

У сфері кібербезпеки це дозволяє точніше прогнозувати кількість та тип кібератак, оцінювати потенційні загрози та планувати заходи безпеки. Загалом розвиток методів комбінованого прогнозування відкриває великі перспективи для підвищення точності оцінок у будь-якій галузі, особливо у високодинамічних та ризикових сферах.

12.2. Індивідуальні і колективні експертні методи

Експертні, або евристичні, методи зазвичай застосовуються при розробці нових напрямків, концепцій розвитку або при складанні політичних, соціальних, економічних та макроекономічних прогнозів. Евристичні методи передбачають використання інтуїції та досвіду висококваліфікованих фахівців і аналітиків, які на основі аналізу та синтезу різних джерел інформації здатні формувати варіанти розвитку того чи іншого процесу.

У сфері кібербезпеки це може означати, що аналітики прогнозують потенційні нові вектори атак, сценарії розвитку кіберзагроз або впровадження нових технологій захисту, використовуючи свій досвід та знання поточної ситуації.

Перевага експертних методів полягає в тому, що фахівець добре розуміє специфіку досліджуваного процесу, його «вузькі місця» та особливості функціонування системи. Завдяки цьому експерт може надати глибокий і змістовний прогноз розвитку подій.

Разом із тим, існують три важливі обмеження, які слід враховувати для успішного прогнозування:

1). Експерт не завжди є професіоналом у прогнозуванні, тому часто застосовує власні методи, які можуть бути грубими або недостатньо наочними.

2). Експерти схильні ігнорувати або недооцінювати роль зовнішніх факторів, які можуть суттєво впливати на процес.

3). Висновки експертів зазвичай точні щодо технічного або фізичного аспекту процесу, але не завжди враховують економічні чи соціальні наслідки.

У основі застосування цих методів лежить припущення, що експерти здатні з високим ступенем ймовірності оцінити: важливість і значення проблеми, перспективність розвитку певного напрямку, час настання подій, доцільність вибору одного з альтернативних шляхів розвитку та інші ключові аспекти прогнозування.

Застосування експертних методів у прогнозуванні засновано на таких передумовах.

1. Кожен експерт створює свою інтуїтивну модель аналізованого явища на основі свого розуміння причинно-наслідкових зв'язків, що визначають еволюцію цього явища, що і дозволяє йому формувати наближені кількісні оцінки.

2. При наданні експерту нової інформації він може творчо переробити цю інформацію й удосконалити свою модель аналізованого явища, тим самим підвищуючи вірогідність оцінок.

Основна задача організаторів експертного опитування полягає в тому, щоб допомогти аналітику або експерту створити максимально правдоподібну модель кіберзагроз або кіберінцидентів і забезпечити умови для її уточнення та удосконалення. Це висуває високі вимоги до досвіду, кваліфікації та професійних знань організаторів опитування. Вони повинні добре розуміти специфіку кібербезпеки на рівні, співставному з рівнем залучених експертів, глибоко вивчити проблему, з’ясувати взаємозв’язки між різними аспектами кіберзагроз і активно впливати на хід опитування.

У прогнозуванні кіберзагроз застосовуються два основні види експертних методів: індивідуальні та колективні.

Індивідуальний прогноз виконується одним експертом. Такий підхід простіший в організації і має чітку логічну схему, оскільки експерт послідовно вибудовує причинно-наслідкові зв’язки в конкретній вузькій галузі, де він є

висококласним фахівцем. Недоліком є сильний суб'єктивний фактор – можливість нав'язування експертом власного бачення проблеми.

Колективний прогноз застосовується для зниження суб'єктивності індивідуальних оцінок. Методи колективних експертних оцінок базуються на принципі колективної думки щодо перспектив розвитку об'єкта прогнозування, у даному випадку – потенційних кібератак, вразливостей або розвитку систем захисту.

Для організації системи колективних оцінок у сфері кібербезпеки необхідно:

– чітко сформулювати цілі та завдання прогнозування,

– встановити порядок взаємодії експертів та організаторів на кожному етапі,

– правильно розподіляти потоки інформації відповідно до плану.

Принципи побудови системи групових експертних оцінок:

1). Обмеження розмаїтості суджень шляхом вирівнювання інформаційної неоднорідності у групі на етапі формування власних моделей причинно-наслідкових зв'язків потенційних кіберзагроз.

2). Використання ітеративного підходу: колективна думка уточнюється періодично на основі нової інформації про зміну кіберсередовища.

3). Забезпечення циркуляції інформації без перекручувань через створення психологічно сприятливого клімату, який стимулює творчі можливості кожного експерта.

4). Кількісна оцінка явищ: кібератаки та їх наслідки характеризуються набором ознак, яким можна присвоїти числові значення для подальшого аналізу та прогнозування.

Розглянемо методи колективних експертних оцінок, що зустрічаються частіше за все.

Просте узгодження думок експертів. Найбільш простими способами генерації колективного решета є процедури голосування і метод експертних груп.

Голосування є одним із найпоширеніших методів прийняття рішень у конкретних ситуаціях. Відсутність універсальної процедури, яка була б оптимальною у всіх випадках і для широкого кола завдань, пояснює наявність майже сотні різних методів голосування, що розроблені та використовуються сьогодні.

Спроба створити аксіоматичну систему, яка б синтезувала «ідеальний» метод голосування, не мала успіху, оскільки навіть базовий набір логічних правил виявився суперечливим.

У практичному застосуванні голосування для вирішення конкретної проблеми, наприклад, визначення пріоритетних напрямів захисту інформаційної системи або вибору кращого сценарію реагування на кіберінциденти, можливі два підходи:

1). Провести порівняльний аналіз існуючих процедур та обрати найбільш придатну для конкретної ситуації.

2). Розробити нову процедуру голосування, яка враховує специфіку задачі.

У будь-якому випадку необхідно:

– чітко визначити, які властивості процедури голосування є критичними для досягнення надійного рішення;

– встановити критерії оцінки ефективності процедури та способи контролю якості отриманих результатів.

У найбільш загальному вигляді модель процедури голосування містить такі компоненти:

– організатори голосування, що визначають правила, відповідно до яких буде відбуватися процедура голосування;

– пред'явлення, тобто анкета чи виборчий бюлетень;

– виборці, тобто особи, які здійснюють голосування;

– інструкції виборцям, тобто вказівки, що і як вони повинні зробити;

– механізм обробки результатів голосування;

– вирішальний блок, призначений або для вироки колективного рішення, або для продовження процедури голосування, можливо за дещо зміненими правилами.

Метод експертних груп дозволяє залучати знання та досвід фахівців із різних вузьких напрямів кібербезпеки для комплексного вирішення конкретної проблеми, наприклад, оцінки ризиків або вибору стратегії реагування на інциденти. Взаємодія експертів у групі сприяє узгодженню думок, глибшому та більш повному аналізу загроз, вразливостей і потенційних сценаріїв атаки. Основними недоліками методу експертних груп є:

– сильна залежність від особистих взаємин;

– страх підірвати репутацію;

– пристосування до думки більшості.

Саме з метою позбутися цих недоліків і був розроблений підхід, відомий як «мозкова атака».

Метод «мозкової атаки» спрямований на генерування нових ідей і досягнення консенсусу серед експертів за допомогою інтуїтивного мислення. Учасники колективної сесії висловлюють якомога більше пропозицій щодо вирішення проблеми, наприклад, визначення нових стратегій захисту мережі, розробки сценаріїв реагування на кібератаки або виявлення потенційних вразливостей. Особлива увага приділяється нетривіальним ідеям. Оцінка та обговорення запропонованих ідей проводяться після завершення всього процесу генерації. Сутність методу «мозкової атаки» можна представити у вигляді схеми (рис. 12.1).

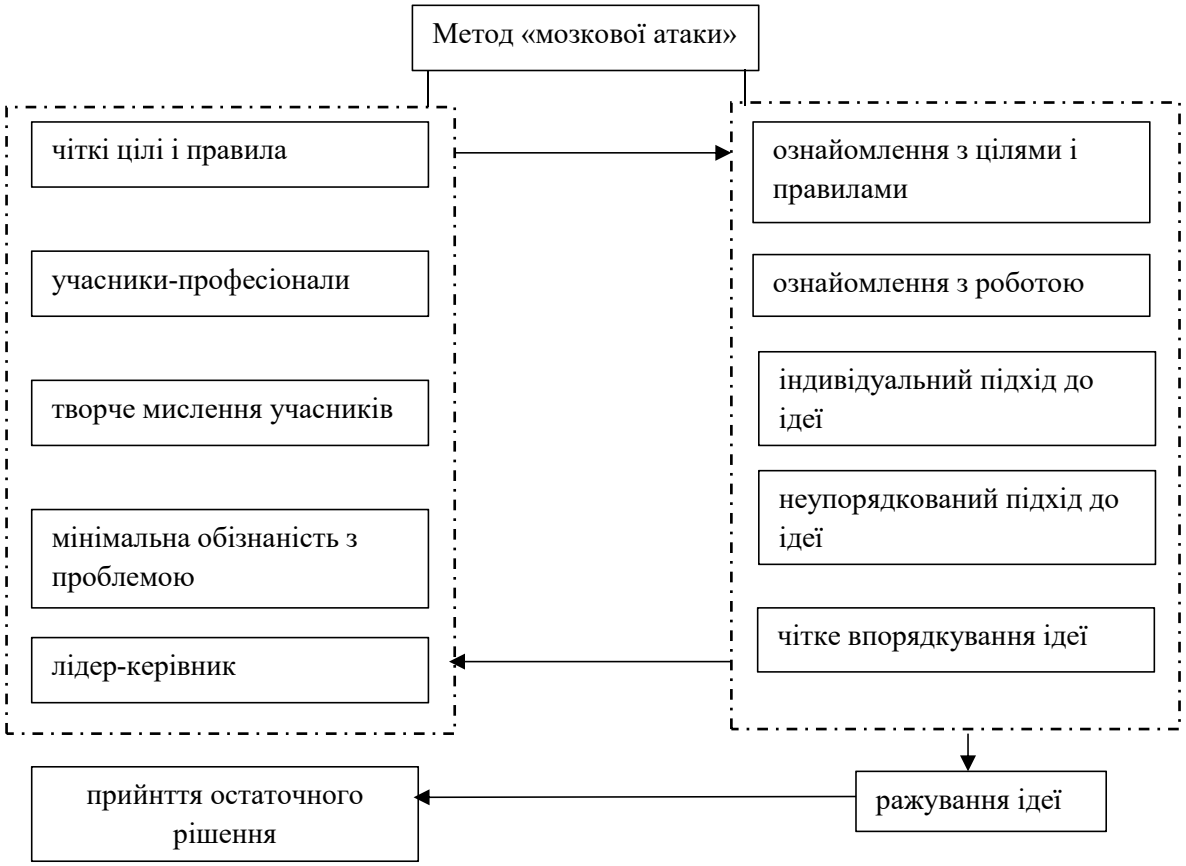


Рисунок 12.1 – Структура методу «мозкової атаки»

Цей метод активізує творчий потенціал експертів під час пошуку рішень складних кіберзагроз або проблем безпеки. Застосування методу передбачає зіткнення різних точок зору та рекомендацій учасників «мозкового штурму» під час аналізу конкретної проблеми, наприклад, вибору оптимальної стратегії захисту мережі або плану реагування на інциденти. Процес включає генерацію ідей та їх подальший критичний аналіз для виявлення найефективніших рішень.

Основа цього методу полягає в тому, що рефері (чи, іншими словами, керівник, координатор) забезпечує виконання таких чотирьох правил учасниками експертної групи:

Правило 1. Зосередження зусиль на дослідженні тільки однієї чітко сформульованої проблеми.

Правило 2. Аналіз будь-якої запропонованої ідеї за принципом: що і як запропонована ідея може дати?

Правило 3. Відсутність критики кожної із запропонованих ідей.

Правило 4. Відсутність витрат часу на дослідження внутрішнього змісту кожної із запропонованих ідей.

У методі «мозкової атаки» прогнозування кіберзагроз або стратегій безпеки здійснюється поетапно. Спершу формується група експертів (10–20 осіб) – висококваліфіковані фахівці з різних напрямів кібербезпеки та загальною обізнаністю у суміжних технологіях. Потім кожен учасник готує проблемну записку, у якій описує конкретну кіберситуацію або потенційну загрозу, а також основні правила проведення «штурму».

Після підготовки починається безпосередній «мозковий штурм», який організовує координатор. Він акцентує увагу на дотриманні правил і фокусує групу на ключовому питанні – наприклад, виборі оптимальної стратегії захисту корпоративної мережі або сценарії реагування на DDoS-атаку. У процесі генерації ідей враховуються всі пропозиції, навіть якщо спочатку вони здаються недоцільними.

Після завершення «атаки» група аналізу систематизує отримані ідеї: створює перелік, формулює їх у стандартній термінології кібербезпеки, об’єднує схожі ідеї, визначає ознаки для класифікації та оцінює загальний діапазон пропозицій.

На наступному етапі відбувається критична оцінка систематизованих ідей щодо їх практичної реалізації. Для цього формується нова експертна група (20–25 осіб) з глибоким досвідом у відповідній галузі кібербезпеки, яка зосереджується на критиці запропонованих рішень. Процес зазвичай триває до 1,5 години і супроводжується записом результатів. У підсумку до реалізації допускаються ті ідеї, які отримали найменшу кількість критичних зауважень та визнані прийнятними для практичного застосування у захисті мережі та інформаційних систем.

Метод Делфі у кібербезпеці передбачає проведення багатотурового опитування групи експертів з питань інформаційної безпеки, кіберзагроз та захисту мереж. Після кожного туру експерти отримують узагальнені результати

попереднього опитування і можуть переглянути свої оцінки та думки щодо ризиків, пріоритетів захисту або стратегій реагування на інциденти.

Опитування продовжується до тих пір, поки зміни в оцінках експертів стають незначними, що свідчить про досягнення стабільного консенсусу. Метод Делфі дозволяє виробити погоджену групову оцінку потенційних загроз, сценаріїв атак або ефективності заходів кіберзахисту.

Цей підхід ефективний тим, що зменшує труднощі, властиві роботі колективної експертної групи, наприклад, вплив домінуючої особистості чи групового тиску, і забезпечує більш об’єктивне та зважене прийняття рішень у сфері кібербезпеки. Сутність методу Дельфі можна представити у вигляді схеми (рис. 12.2).

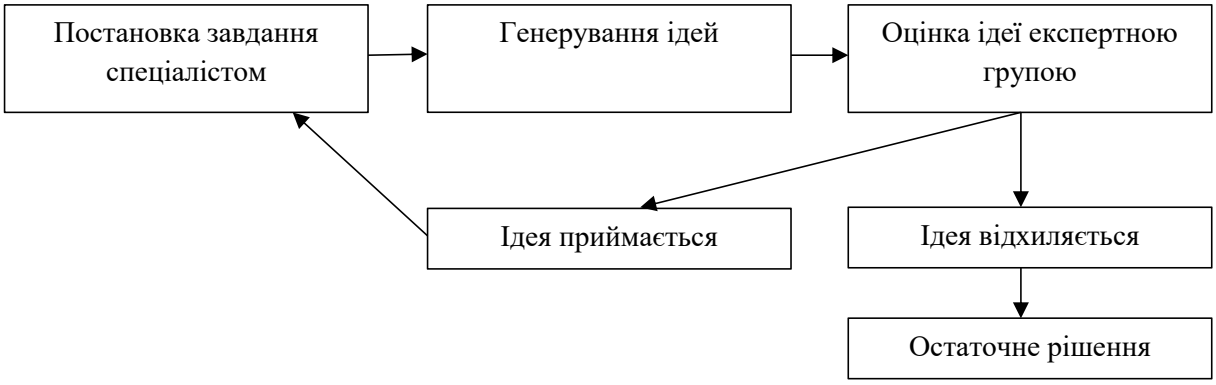


Рисунок 12.2 – Сутність методу Дельфі

Цей метод являє собою розвиток методу експертних груп і включає такі три особливості:

1) анонімність, що досягається шляхом використання запитувальників чи формалізованих методів зв'язку;

2) коректований зворотний зв'язок, що полягає в проведенні декількох турів опитування з послідовним узагальненням результату попередніх турів і представленням цієї інформації експертам при черговому турі опитування з метою зменшення шуму;

3) статистична обробка відповідей.

Стандартна процедура методу Делфі полягає в тому, що спочатку експертів просять назвати всі події, які протягом розглянутого проміжку часу можуть вплинути на досліджувану проблему. Після цього робиться як мінімум чотири опитування.

Мета 1-го опитування – визначити можливі терміни подій. Для кожної події обчислюється медіана (середня дата) і квартилі (дати границь 50-го часового діапазону по 25% у кожен бік від медіани).

Мета інших трьох опитувань – узгодження думок експертів.

12.3. Етапи проведення колективної експертної оцінки

При використанні методу колективної експертної оцінки звичайно виділяють такі етапи проведення експертизи:

1. Формування цілей і питань експертизи.
2. Формування правил проведення чи опитування характеру взаємини експертів.
3. Формування групи експертів.
4. Вибір способу оцінки ступеня компетентності експертів.
5. Формування правил обробки думок експертів.
6. Обчислення експертних узагальнених оцінок і визначення ступеня погодженості експертів.

На першому етапі проведення експертизи особливе значення має чітке визначення її цілей. Наявність ясно сформульованих цілей і зрозумілих потреб є обов’язковою умовою отримання надійних результатів оцінки.

Корисним інструментом формулювання цілей є опис передісторії та поточного стану кіберзагроз або інцидентів у системі, що може слугувати основою для вибору завдань експертизи. Крім того, при формуванні цілей важливо враховувати специфіку та інтереси різних груп фахівців, які братимуть участь у перевірці – наприклад, аналітиків з мережевої безпеки, фахівців із захисту даних та адміністраторів систем. Після збору цих даних можна

переходити до постановки конкретних завдань експертизи та розробки основних правил її проведення.

Вибір цілей і характер процедури експертизи з кібербезпеки значною мірою визначається суттю проблеми (наприклад, оцінкою ризиків у мережевій інфраструктурі, виявленням вразливостей або плануванням заходів реагування на інциденти), очікуваними результатами та способами їх представлення (звіти, рейтинги загроз, рекомендації щодо захисту). Рівень прийняття рішень – державний, галузевий або рівень конкретного підприємства – визначає широту оцінюваного діапазону, кількість альтернатив та ступінь формалізації процедури. Вибір цілей і процедури залежить також від надійності й повноти наявних даних, а також від виду інформації, яку необхідно отримати (наприклад, дані журналів безпеки, результати аудиту або експертні оцінки ймовірності атак).

При формулюванні цілей експертизи насамперед необхідно чітко визначити критерії оцінки та умови використання експертних висновків. Якщо цілей небагато, оцінки за різними шкалами слід узгоджувати, наприклад, встановлюючи більш загальний рівень мети (рейтинг загальної безпеки, а не окремих компонентів).

Структурно-організаційний набір питань у анкеті повинен логічно пов’язуватися з центральним завданням експертизи – наприклад, оцінкою стійкості мережі або ймовірності витоку даних. Хоча форма і зміст питань визначаються специфікою об’єкта аналізу, можна встановити загальні вимоги: питання мають бути сформульовані в зрозумілих термінах, без неоднозначності, логічно відповідати структурі системи, забезпечувати однозначне тлумачення.

За формою питання можуть бути відкритими та закритими, прямими та непрямими. Відкриті питання дозволяють експерту надати вільну відповідь, наприклад, описати можливі сценарії кіберзагроз. Закриті питання містять альтернативні варіанти відповідей, і експерт повинен вибрати один або декілька з них, наприклад, оцінити ймовірність конкретного типу атаки. Непрямі

питання використовуються, коли мету експертизи потрібно замаскувати, наприклад, щоб уникнути упередженості у відповідях або впливу зовнішніх чинників на експерта.

Розглянемо основні групи питань, що використовуються для проведення колективної експертної оцінки.

Питання можуть передбачати відповіді у вигляді кількісної оцінки: наприклад, оцінка часу виявлення та реагування на кіберінцидент, ймовірності успішної атаки або оцінка відносного впливу різних факторів на безпеку інформаційної системи. При визначенні шкали значень кількісних характеристик доцільно використовувати нерівномірну шкалу, оскільки похибка прогнозу може змінюватися залежно від періоду попередження (наприклад, при оцінці загроз на наступний місяць чи рік).

При формуванні групи експертів ключовим є визначення її якісного та кількісного складу. Добір експертів починається з визначення ключових питань, що охоплюють аналіз кіберзагроз і заходів захисту, після чого складається список компетентних спеціалістів.

Для отримання якісного результату до учасників експертизи висуваються такі вимоги: високий рівень загальної ерудиції, глибокі спеціальні знання в галузі кібербезпеки, здатність адекватно оцінювати тенденції розвитку загроз, наявність психологічної установки на передбачення потенційних атак, чисто науковий інтерес до оцінюваного питання без практичної зацікавленості, а також досвід роботи у сфері інформаційної безпеки, аудиті або реагуванні на інциденти.

Існує кілька підходів до оцінки компетентності експертів. По-перше, організатори можуть, зважаючи на попередню участь експерта в кіберекспертизах, оцінити його компетентність.

Іншим методом є тестування експертів. Для цього розробляється спеціальна тест-анкета, яка дозволяє перевірити знання об’єкта дослідження (наприклад, мережевої архітектури, протоколів безпеки, типів атак) та його

аналітичні здібності. Слід зазначити, що розробка такої тест-анкети є складною та трудомісткою процедурою, тому цей метод оцінки компетентності виправданий лише у випадках залучення великої кількості експертів (більше 30 осіб).

Перелік запитань для самопідготовки:

1. Що таке експертна система прогнозування?
2. Які принципи лежать в основі побудови експертних систем?
3. У яких випадках доцільно використовувати експертні методи прогнозування?
4. Які вимоги ставляться до експертів у таких системах?
5. Як забезпечується об’єктивність і надійність експертних оцінок?
6. У чому полягає різниця між індивідуальними та колективними експертними методами?
7. Які є приклади індивідуальних методів експертного прогнозування?
8. Що таке метод Делфі, і до якого типу він належить?
9. Які переваги має колективне експертне оцінювання порівняно з індивідуальним?
10. Які труднощі можуть виникати при застосуванні кожного з підходів?
11. Які основні етапи проведення колективного експертного опитування?
12. Як відбираються експерти для участі в оцінюванні?
13. Які методи використовуються для агрегування експертних думок?
14. Як аналізується надійність і узгодженість оцінок?
15. Які інструменти можуть використовуватись для автоматизації цього процесу?

Тема 13. Критерії визначення якісного прогнозу

13.1. Поняття оптимального прогнозу.

13.2. Оцінювання адекватності прогнозованої моделі.

13.3. Критерії визначення якісного прогнозу.

13.4. Оцінка точності прогнозованої моделі та прогнозів.

13.1. Поняття оптимального прогнозу

Оптимальний прогноз – це зроблене на підставі економічної теорії передбачення, яке використовує всю доступну на момент побудови прогнозу інформацію. Для оптимального прогнозу граничний виграш та граничні витрати збігаються.

Оптимальний прогноз іще називають прогнозом раціональних сподівань. Раціональні сподівання можуть відрізнятися від фактичних значень, але будь-яка різниця має бути випадковою й непередбачуваною. Оскільки раціональні сподівання ґрунтуються на коректній економічній теорії, вони мають властивості незсуненості (за умови квадратичної функції витрат) та ефективності.

Незсуненість означає, що помилка прогнозу має нульове математичне сподівання.

Ефективність передбачає, що в процесі прогнозування буде використана вся доступна інформація, отже, помилка прогнозу не буде корелювати з цією інформацією.

Існують численні критерії перевірки раціональності послідовності прогнозів. Стандартний критерій незсуненості потребує перевірки гіпотези стосовно того, що α= 0 та β = 1 водночас для такої моделі:

$$y_t = \alpha + \beta \hat{y}_t + \varepsilon_t, \qquad (13.1)$$

де y_t – ряд фактичних значень або спостережень;

$\hat{y}_t$ – ряд прогнозованих значень;

ε_t – випадкові залишки.

Перевірка ефективності є складнішою, оскільки неможливо коректно визначити відповідний масив інформації, стосовно якого похибки прогнозу будуть некорельованими.

Узагальнюючи огляд критеріїв визначення якісного прогнозу, можна стверджувати, що варто користуватися системою критеріїв, які мають враховувати:

- кількість зусиль, витрачених на побудову моделі, і наявність готових комп'ютерних програм;
- швидкість, із якою метод уловлює істотні зміни у поведінці ряду, наприклад раптовий зсув математичного сподівання або збільшення кута нахилу лінії тренда;
- існування серійної кореляції у помилках;
- незмінюваність первинних даних;
- повний обсяг роботи в деяких сферах діяльності – тисячі рядів щомісяця потребують оновлення, невеликі витрати й швидкість мають першорядне значення;
- терміновість прогнозування.

13.2. Оцінювання адекватності прогнозованої моделі

Не залежно від виду і способу побудови економіко-математичної моделі питання про можливість її застосування з метою аналізу і прогнозу економічного явища може бути вирішено тільки після встановлення адекватності моделі. Оскільки повної відповідності моделі реальному процесу або об'єкту бути не може, адекватність певною мірою умовне поняття. При моделюванні йдеться про адекватність не загалом, а про відповідність тим властивостям моделі, які вважаються суттєвими для дослідження.

Модель згладжування $\hat{y}_t$ певного часового ряду y_t вважається адекватною, якщо правильно відображає систематичні компоненти часового ряду. Ця вимога еквівалентна вимозі, щоб залишкова компонента $e_t = y_t - \hat{y}_t$ (t = 1, 2, ..., n) відповідала таким властивостям випадкової компоненти часового ряду як: випадковість коливань рівнів залишкової послідовності, відповідність розподілу випадкової компоненти нормальному закону, рівність математичного сподівання випадкової компоненти нулю, незалежність значень рівнів випадкової компоненти. Розглянемо, яким чином здійснюється перевірка цих властивостей залишкової послідовності.

Перевірка випадковості коливань рівнів залишкової послідовності означає перевірку гіпотези про правильність вибору виду тренду. Для дослідження випадковості відхилень від тренду розраховують набір різниць $e_t = y_t - \hat{y}_t$ (t = 1, 2, ..., n). Характер цих відхилень вивчається за допомогою ряду непараметричних критеріїв. Одним з таких критеріїв є «критерій серій», який використовує медіану вибірки, критерій «зростаючих» та «спадних» серій тощо.

Згідно з *критерієм серій*, ряд із величин e_t розташовують у порядку зростання їхніх значень і знаходять медіану e_m одержаного варіаційного ряду, тобто значення, що перебуває в середині для непарного n або середню арифметичну з двох середніх значень для n парного. Повертаючись до вхідної послідовності e_t і порівнюючи значення цієї послідовності з e_m, ставлять знак «плюс», якщо значення e_t, перевищує медіану, і знак «мінус», якщо воно менше за медіану; у випадку однаковості порівнюваних величин відповідне значення e_t пропускають. Отже, одержують послідовність, що складається із плюсів та мінусів, загальна кількість яких не перевищує n. Послідовність розташованих одне за одним плюсів або мінусів називають серією. Щоб послідовність e_t була випадковою вибіркою, довжина найдовшої серії не має бути занадто великою, а загальна кількість серій – занадто малою.

Позначимо довжину найдовшої серії через $K_{\max}$, а загальну кількість серій – через v. Вибірка вважається випадковою, якщо виконуються такі нерівності для 5 %-го рівня значущості:

$$K_{\max} < [3{,}3(\ln n + 1)], \tag{13.2}$$

$$v > [\frac{1}{2}(n + 1 - 1{,}96\sqrt{n-1})] \tag{13.3}$$

де квадратні дужки означають цілу частину числа.

Якщо хоча б принаймні одна з цих нерівностей порушується, то гіпотеза про випадковий характер відхилень рівнів часового ряду від тренду спростовується, а модель тренду визнається неадекватною.

Перевірка відповідності розподілу випадкової компоненти нормальному закону розподілу може бути зроблена лише наближено за допомогою дослідження показників асиметрії (A) і ексцесу (E), оскільки часові ряди, як правило, не дуже довгі. При нормальному розподілі показники асиметрії і ексцесу певної генеральної сукупності дорівнюють нулю. Припустимо, що відхилення від тренду являють собою вибірку з генеральної сукупності, тому можна визначити тільки вибіркові характеристики асиметрії й ексцесу та їхні помилки:

$$\hat{A} = \frac{\frac{1}{n}\sum_{t=1}^{n} e_t^3}{\sqrt{\left(\frac{1}{n}\sum_{t=1}^{n} e_t^2\right)^3}} \quad \sigma_{|\hat{A}|} = \sqrt{\frac{6(n-2)}{(n+1)(n+3)}} \tag{13.3}$$

$$\hat{E} = \frac{\frac{1}{n}\sum_{t=1}^{n} e_t^4}{\left(\frac{1}{n}\sum_{t=1}^{n} e_t^2\right)^2} - 3 \quad \sigma_{\hat{E}} = \sqrt{\frac{24n(n-2)(n-3)}{(n+1)^2(n+3)(n+5)}}, \tag{13.4}$$

де $\hat{A}$ – вибіркова характеристика асиметрії;

$\hat{E}$ – вибіркова характеристика ексцесу;

$\sigma_{\hat{A}}$ і $\sigma_{\hat{E}}$ – відповідні середньоквадратичні помилки.

Якщо одночасно виконуються такі нерівності:

$$\left|\hat{A}\right| < 1{,}5\sigma_{\hat{A}} \quad \left|\hat{E} + \frac{6}{n+1}\right| < 1{,}5\sigma_{\hat{E}} \tag{13.5}$$

то гіпотеза про нормальний характер розподілу випадкової компоненти приймається.

Якщо виконується хоча б одна з цих нерівностей:

$$\left|\hat{A}\right| \geq 2\sigma_{\hat{A}} \quad \left|\hat{E} + \frac{6}{n+1}\right| \geq 2\sigma_{\hat{E}}, \tag{13.6}$$

то гіпотеза про нормальний характер розподілу відхиляється, трендова модель визнається неадекватною. Інші випадки потребують додаткової перевірки за допомогою складніших критеріїв. Для адекватних моделей доцільно ставити запитання щодо оцінювання їхньої точності. Вважається, що моделі з меншою розбіжністю між фактичними й розрахунковими значеннями краще відображають досліджуваний процес у майбутньому. Для характеристики рівня близькості використовують такі описові статистики:

середнє квадратичне відхилення (або дисперсія)

$$\widehat{\sigma}_{\varepsilon} = \sqrt{\frac{1}{n-k-1}} \sum_{t=1}^{n} (y_t - \hat{y}_t)^2, \tag{13.7}$$

середню відносну похибку апроксимації (чим ближче до 0, тим точніша модель):

$$\overline{e}_{відн.} = \frac{1}{n} \sum_{t=1}^{n} \left|\frac{y_t - \hat{y}_t}{y_t}\right| 100\%, \tag{13.8}$$

коефіцієнт сходження:

$$\varphi^2 = \frac{\sum_{t=1}^{n} (y_t - \hat{y}_t)^2}{\sum_{t=1}^{n} (y_t - \bar{y}_t)^2}, \tag{13.9}$$

коефіцієнт детермінації (чим ближче до 1, тим точніша модель):

$$R^2 = 1 - \varphi^2, \tag{13.10}$$

У формулах (13.7-13.10) n – кількість рівнів ряду, k – кількість пояснювальних змінних у моделі, $\tilde{y}_t$ – оцінки рівнів ряду за моделлю, $\bar{y}$ – середнє арифметичне значення вибірки.

На підставі розглянутих показників можна з кількох адекватних моделей обрати найточнішу. Помилка прогнозу, обчисленого для періоду, характеристики котрого вже були використані при оцінюванні параметрів моделі, як правило, буде незначною та мало залежатиме від теоретичної обґрунтованості, застосованої для побудови моделі.

Оскільки формально-статистичний вибір кращої моделі в багатьох випадках не гарантує цілковитої впевненості в його правильності, адже добрий прогноз можна отримати і на підставі поганої моделі, і навпаки, тому про якість застосовуваних методик і моделей у прогнозуванні можна судити лише за сукупністю зіставлень прогнозів і їх реалізації. При цьому незалежно від обраної методики та моделі прогнозування джерелами помилок прогнозу можуть бути:

1) природа змінних (випадковий характер змінних гарантує, що прогноз відхилятиметься від справжніх величин, навіть якщо модель правильно специфікована, її параметри точно відомі);

2) природа моделі (сам процес оцінювання спричиняє похибки оцінок параметрів);

3) помилки, привнесені прогнозом незалежних випадкових величин (пояснювальних змінних);

4) помилки специфікації моделі.

Перевірка рівності математичного сподівання випадкової компоненти нулю, якщо вона розподілена за нормальним законом, виконується на основі *t*-критерію Стьюдента. Розрахункове значення цього критерію задається формулою

$$t = \frac{\bar{e} - 0}{\sigma_e}\sqrt{n}, \qquad (13.11)$$

де $\bar{e}$ – середнє арифметичне значення рівнів залишкової послідовності e_t;

σ_e – стандартне (середньоквадратичне) відхилення для цієї послідовності.

Якщо розрахункове значення t менше від табличного значення t_α статистики Стьюдента із заданим рівнем значущості α і числом ступенів свободи $n - 1$, то гіпотеза про рівність нулю математичного сподівання випадкової послідовності приймається, в протилежному випадку ця гіпотеза відхиляється і модель вважається неадекватною.

Перевірка незалежності значень рівнів випадкової компоненти, тобто перевірка відсутності суттєвої автокореляції в залишковій послідовності, може здійснюватися за рядом критеріїв, найбільш поширеним з яких є *DW*-критерій Дарбіна-Уотсона. Зазначимо, що розрахункове значення критерію Дарбіна-Уотсона в інтервалі від 2 до 4 свідчать про від'ємний зв'язок. У цьому випадку його треба перетворити за формулою $DW' = 4 - DW$ і далі використовувати значення DW'.

Висновок про адекватність трендової моделі робиться, якщо всі розглянуті вище чотири перевірки властивостей залишкової послідовності дають позитивний результат.

13.3. Критерії визначення якісного прогнозу

Якість прогнозу характеризують такі поширені в прогностичній літературі терміни, як точність і надійність. Проте зміст цих термінів часто тлумачать досить неоднозначно. Це можна пояснити тим, що нині поки не знайдено ефективного підходу до оцінювання якості прогнозу, окрім його практичного підтвердження.

Про точність прогнозу прийнято судити за розміром помилки прогнозу – різниці між прогнозовим і фактичним значенням досліджуваного показника.

Але такий підхід можливий лише тоді, якщо дослідник має інформацію стосовно справжніх значень часового ряду, який він оцінював під час розроблення прогнозів. Наприклад, період випередження вже завершився, і дослідник має фактичні значення змінної (це можливо в разі короткотермінового прогнозування) або прогноз перебуває в стадії розроблення, тобто прогнозування здійснюється для певного моменту часу в минулому, для якого існують фактичні дані. Спрощену схему періодів прогнозування показано на рис. 13.1.



Рисунок 13.1 – Спрощена схема періодів прогнозування

В останньому випадку йдеться про використання *expost*-прогнозу. Його сутність полягає у побудові моделі за меншим обсягом даних (n - m) із подальшим порівнянням прогнозових оцінок за останніми m точками (для t від n - m + 1 до n) із відомими фактичними, але спеціально залишеними рівнями ряду. Отримані ретроспективно помилки прогнозу певною мірою характеризують точність застосовуваної методики прогнозування й можуть виявитися корисними в зіставленні кількох прогнозів.

Розмір помилки ретроспективного прогнозу не можна розглядати як остаточний доказ придатності, або навпаки, непридатності застосовуваного методу прогнозування. До неї варто ставитися з відомою обережністю і при її застосуванні в якості міри точності необхідно враховувати, що вона отримана при використанні лише частини наявних даних. Проте ця міра точності має більшу наочність і теоретично більш надійна, ніж похибка прогнозу, обчислена

для періоду, характеристики котрого вже були використані при оцінюванні параметрів моделі.

На основі розглянутих показників можна зробити вибір з декількох адекватних моделей найбільш точної. Помилка прогнозу, обчислена для періоду, характеристики котрого вже були використані при оцінюванні параметрів моделі, як правило, будуть незначними і мало залежатимуть від теоретичної обґрунтованості, застосованої для побудови моделі.

Оскільки формально-статистичний вибір кращої моделі у багатьох випадках не дає повної впевненості у його правильності: гарний прогноз може бути отриманий і по поганій моделі, і навпаки, то, про якість прогнозів застосовуваних методик і моделей можна судити лише за сукупністю зіставлень прогнозів і їхньої реалізації. При цьому незалежно від обраної методики моделі прогнозування джерелами помилок прогнозу можуть бути:

1) природа змінних (випадковий характер змінних гарантує, що прогноз буде відхилятися від справжніх величин, навіть якщо модель правильно специфікована та її параметри точно відомі);

2) природа моделі (сам процес оцінювання спричиняє похибки оцінок параметрів);

3) помилки, що вносяться прогнозом незалежних випадкових величин (пояснюючих змінних);

4) помилки специфікації моделі.

13.4. Оцінка точності прогнозної моделі та прогнозів

Параметричні методи аналізу точності прогнозів. За результатами ex post-прогнозу розраховують такі показники точності прогнозів за *m* кроків:

Середня квадратична похибка:

$$MSE = \frac{\sum_{i=n-m+1}^{n} (y_i - \hat{y}_i)^2}{m}; \qquad (13.12)$$

корінь із середньоквадратичної похибки

$$RMSE = \sqrt{\frac{\sum_{i=n-m+1}^{n}(y_i - \hat{y}_i)^2}{m}}; \quad (13.13)$$

середня абсолютна похибка:

$$MAE = \frac{\sum_{i=n-m+1}^{n}\left|y_i - \hat{y}_i\right|}{m}; \quad (13.14)$$

корінь із середньоквадратичної похибки у відсотках:

$$RMSPE = \sqrt{\frac{100}{m}\sum_{i=n-m+1}^{n}\left(\frac{y_i - \hat{y}_i}{y_i}\right)^2}; \quad (13.15)$$

середня абсолютна похибка у відсотках (MAPE):

$$MAPE = \sum_{i=n-m+1}^{n}\frac{100\left|y_i - \hat{y}_i\right|}{m\left|y_i\right|}. \quad (13.16)$$

Чим менше значення цих величин, тим вища якість ретропрогнозу. На практиці ці характеристики використовують досить часто. Такий підхід дає гарні результати, якщо на періоді ретропрогнозу не виникають принципово нові закономірності. На підставі останніх двох критеріїв можна дійти висновку про загальний рівень адекватності моделі шляхом їх порівняння. Цей рівень наведений у таблиці 13.1.

Таблиця 13.1

Точність пронозу в залежності від MAPE, RMSE

MAPE, RMSE	Точність прогнозу
Менше 10%	Висока
10% - 20%	Добра
20% - 40%	Задовільна
40% - 50%	Погана
Більше 50%	Незадовільна

Вадою обговорених вище характеристик точності прогнозів є їх залежність від обраних одиниць виміру. Було б корисним указати безрозмірний

показник, аналогічний до коефіцієнта кореляції. Одним з таких показників є коефіцієнт невідповідності Тейла, чисельником якого є середньоквадратична похибка прогнозу, а знаменник дорівнює квадратному кореню із середнього квадрата фактичних та оцінних значень:

$$U = \frac{\sqrt{\sum (\hat{y}_i - y_i)^2 / m}}{\sqrt{\frac{1}{m}\sum_{i=1}^{m} y_i^2} + \sqrt{\frac{1}{m}\sum_{i=1}^{m} \hat{y}_i^2}} \qquad (13.17)$$

Перевага коефіцієнта Тейла полягає в тому, що його значення завжди перебувають у межах від нуля до одиниці. Якщо всі прогнози абсолютно точні, то $U = 0$. Якщо всі прогнози дорівнюють нулю, а жодне з фактичних значень не дорівнює нулю або навпаки, U дорівнюватиме одиниці. Таким чином, мале значення U засвідчує, що прогноз є точним, але максимального значення не існує. Значення, яке дорівнює одиниці, відповідає ситуації, коли всі прогнозові значення дорівнюють нулю, що нереально під час прогнозування номінальних величин, але під час розгляду змін такий прогноз відповідає моделі «без змін». Більші за одиницю значення вказують на те, що прогноз гірший, ніж прогноз «без змін».

Коефіцієнт невідповідності Тейла (U) може бути розкладений на три частини:

пропорцію зсунення

$$U^M = \frac{(y_i - \hat{y}_i)^2}{\frac{1}{m}\sum_{i=1}^{m} \frac{(y_i - \hat{y}_i)}{y_i}}; \qquad (13.18)$$

пропорцію дисперсії

$$U^S = \frac{(\sigma_y - \sigma_{\hat{y}})^2}{\frac{1}{m}\sum_{i=1}^{m} \frac{(y_i - \hat{y}_i)}{y_i}}; \qquad (13.19)$$

пропорцію коваріації

$$U^C = \frac{2(1-\rho)(\sigma_y \cdot \sigma_{\hat{y}})}{\frac{1}{m}\sum_{i=1}^{m}(y_i - \hat{y}_i)^2}. \quad (13.20)$$

Зазначимо, що $U^M + U^S + U^C = 1$. Критерій зсуву пропорції (U^M) використовується, щоб перевірити, чи є систематичне відхилення середніх розрахованих та фактичних рядів, тобто чи дає модель систематично завищені або занижені прогнози. Чим менше значення, тим краще. Якщо U^M дорівнює нулю, у розрахованих (прогнозних) значеннях немає зсунень, тобто з моделлю все гаразд. Пропорція дисперсії (U^S) використовується, щоб переконатися, що модель має достатні динамічні властивості для відтворення дисперсії фактичних рядів. Наприклад, модель може відтворювати систематично менші коливання, ніж фактичні. Як і у випадку критерію U^M, менше значення U^S вказує на менше зсунення. Пропорція коваріації вказує, як корелюють фактичні та розраховані ряди. Якщо U^C дорівнює 1, то фактичні та розраховані ряди корелюють ідеально.

Критичні точки важливі як критерії якості, оскільки деякі моделі можуть бути точними, але погано передбачати зміни тенденції (наприклад, поворотні точки в циклах), тобто погано відтворювати критичні точки. Інші моделі можуть бути неточними, але мати гарний динамічний характер. Загалом може бути певний компроміс між точністю та динамічними властивостями моделі. Формального тесту для оцінки цієї властивості не існує. Проте візуальний огляд розрахованих та фактичних рядів звичайно одразу виявляє, здатна модель відтворювати критичні точки чи ні.

Обговорені характеристики точності прогнозів є параметричними в тому сенсі, що вони потребують виконання заданих припущень щодо властивостей математичного сподівання та дисперсії, чинних за умов нормальності відповідних розподілів. Наприклад, використовуючи *MSE*, ми неявно припускаємо, що всі похибки прогнозу мають однакові й постійні математичні сподівання та дисперсії. У реальних економічних ситуаціях найчастіше

порушуються припущення гомоскедастичності та відсутності автокореляції. Можна стверджувати, що кожного разу прогноз будується у новій ситуації, отже, порівняння числової точності прогнозів, зроблених у різні моменти часу, не зовсім коректне. Наведені міркування зумовили використання непараметричних методів аналізу точності прогнозів.

Непараметричні методи аналізу точності прогнозів

Непараметричні методи не залежать від вигляду розподілу, тож не потребують припущення щодо нормальності розподілів. Це особливо корисно, коли йдеться про дані, які унеможливлюють використання числових шкал. Розглянемо два типи непараметричних критеріїв: критерій знаків та рангові критерії.

Критерій знаків для порівняння точності двох послідовностей прогнозів базується на відсотку випадків, коли метод визначення прогнозу *A* кращий, ніж метод *B*. Таке порівняння здійснюють для індивідуальних прогнозів однакових подій (змінних). Якщо обидва методи дають однакову точність, імовірність відповіді «так» на запитання «чи прогноз А кращий за прогноз *B*» становить 0,5 для кожного з *m* випадків прогнозування. Число *K* випадків, коли прогноз *A* кращий, підпорядковано біноміальному розподілу ймовірностей

$$p(K = x) = C_m^x 0{,}5^x 0{,}5^{m-x}. \qquad (13.21)$$

Отже, можна підрахувати імовірність того, що $K \geq x$. Якщо довжина послідовності прогнозів значна, для оцінювання ймовірностей можна використати нормальну апроксимацію біноміального розподілу.

Критерій знаків можна також використовувати для перевірки значущості описової статистики, відомої під назвою «відсоток кращих результатів», яка показує відсоток випадків, у яких один метод прогнозування кращий за інший і розраховується за формулою:

$$\eta = \frac{m}{m+n}, \qquad (13.22)$$

де *m* – кількість прогнозів, підтверджених фактичними даними;

n – кількість прогнозів, не підтверджених фактичними даними.

Коли всі прогнози підтверджуються, $n = 0$ і $\eta = 1$; якщо всі прогнози не підтвердилися, то m, а отже, і η дорівнюють 0.

Рангові критерії. У разі застосування цих критеріїв чисельна характеристика точності (абсолютна похибка, коли маємо один прогноз, або *MSE*, коли розглядають послідовність прогнозів) замінюється рангами, які потім перевіряють на значущість. Наприклад, якщо послідовності прогнозів показників *A* та *B* одержують за допомогою *k* методів, то спочатку обчислюють *MSE*, потім їхні значення ранжують від *1* (найменша *MSE*) до *k* (найбільша *MSE*) (відповідні ранги позначають через R_A, та R_B, для $i = 1, ..., k$). Після знаходження різниць (d_i) між рангами обчислюють коефіцієнт рангової кореляції Спірмена:

$$r_s = 1 - \frac{6\sum_{i=1}^{n} d_i^2}{n(n^2 - 1)} \qquad (13.23)$$

За нульову гіпотезу приймають відсутність залежності між рангами, тобто жоден з методів не є гіршим за решту. Гіпотеза відкидається, якщо значення r_s досить велике.

Хоча непараметричні методи мають свої переваги, важливо усвідомлювати, що вони ігнорують частину доступної інформації. Так, критерії знаків та рангів не враховують числових значень похибок.

Перелік запитань для самопідготовки:

1. Що означає термін «оптимальний прогноз» у контексті прогнозування?
2. Які умови повинні виконуватись, щоб прогноз вважався оптимальним?
3. Чим відрізняється оптимальний прогноз від точного прогнозу?
4. Яку роль відіграє врахування ризику або невизначеності при формуванні оптимального прогнозу?
5. Що таке адекватність моделі?

6. Які показники використовуються для перевірки адекватності (наприклад, залишки, автокореляція)?

7. Які є статистичні тести для оцінки адекватності (наприклад, тест Дарбіна-Уотсона)?

8. Як визначити, чи модель добре описує динаміку реального процесу?

9. Які основні критерії використовуються для оцінки якості прогнозу?

10. Що таке середня абсолютна похибка (MAE), середня квадратична похибка (MSE), середня абсолютна відносна похибка (MAPE)?

11. Як інтерпретувати значення цих показників?

12. У чому переваги та недоліки кожного з критеріїв?

13. Як обрати найбільш підходящий критерій для конкретного завдання?

14. У чому полягає різниця між точністю моделі і точністю прогнозу?

15. Як оцінити точність прогнозу на навчальній і тестовій вибірках?

16. Що таке перехресна перевірка (cross-validation) у контексті прогнозування?

17. Які помилки слід враховувати при аналізі точності?

18. Як покращити точність моделі на основі результатів оцінки?

Лабраторні роботи

Лабораторна робота 1. Приклад обробки реальних даних в системі Statistica.

Під час виконання лабораторної роботи здобувачі повинні виконати такі завдання:

1. Створити електронну таблицю з початковими даними.
2. Перетворити і доповнити початкові дані:
   Вік пароля = Поточний рік – Рік створення пароля; Ймовірність_Лог = Log (Ймовірність, що пароль уже злитий/відомий).
3. Візуалізація даних.
4. Викликати стартову панель модуля **Multiple Regression** і визначити умови аналізу.
5. Вивід результатів та їх аналіз.

Початкові дані наведені в таблиці Л1.1.

Таблиця Л1.1

Ступінь надійності паролю в залежності від року його створення

Рік створення пароля	Ймовірність, що пароль уже злитий/відомий, %
2010	90
2011	80
2012	70
2013	60
2014	50
2015	54
2016	47
2017	43
2018	40
2019	24
2020	18
2021	10
2022	7
2023	5
2024	3
2025	1

В програмному забезпечення *Statistica* потрібно побудувати електронну таблицю, на основі початкових даних, що наведені в таблиці Л1.1.

1 Рік створення пароля	2 Ймовірність, що пароль уже злитий/відомий	3 Вік пароля	4 Ймовірність_лог
2010	90	15	4,49980967
2011	80	14	4,38202663
2012	70	13	4,24849524
2013	60	12	4,09434456
2014	50	11	3,91202301
2015	54	10	3,98898405
2016	47	9	3,8501476
2017	43	8	3,76120012
2018	40	7	3,68887945
2019	24	6	3,17805383
2020	18	5	2,89037176
2021	10	4	2,30258509
2022	7	3	1,94591015
2023	5	2	1,60943791
2024	3	1	1,09861229
2025	1	0	0

Рисунок Л1.1 – доповнені і перетворені початкові дані

Для того, щоб добавити третю колонку, потрібно двічі клікнути на заголовок цієї колонки і у вікні, що з’явиться (рис. Л1.2), потрібно ввести назву колонки «Вік пароля», формулу «2025-Рік створення пароля» у відповідному полі та натиснути на кнопку ОК.

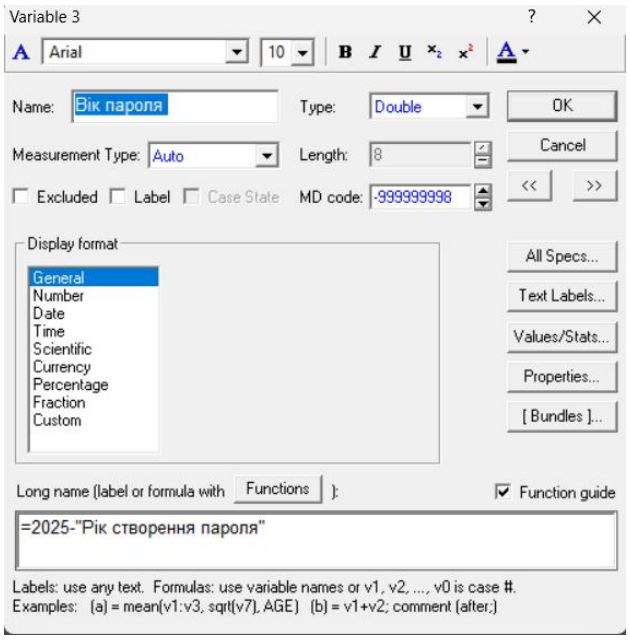


Рисунок Л1.2 – Додавання колонки «Вік пароля»

Аналогічно потрібно добавити колонку Ймовірність_лог (рис. Л1.3).

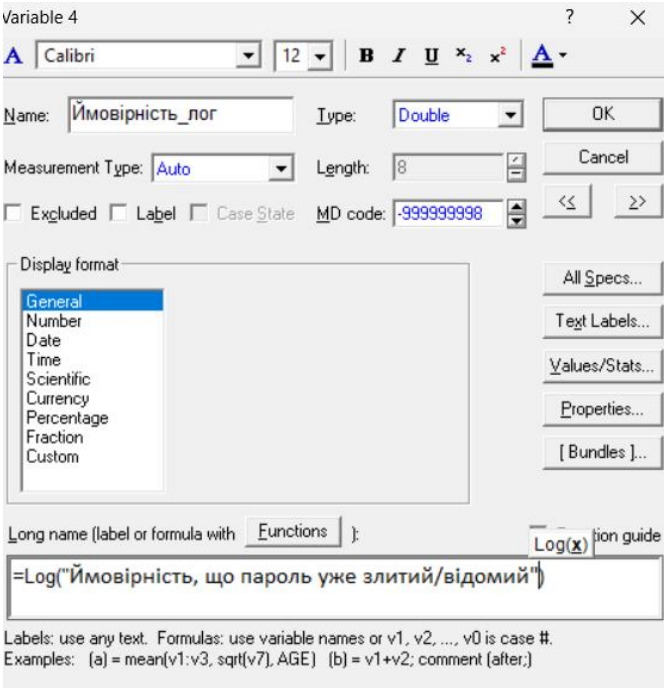


Рисунок Л1.3 – Додавання колонки «Ймовірність_лог»

В цьому вікні, також потрібно виставити число знаків після коми. Наступним кроком є візуалізація даних, тобто побудова графіків.

Щоб побудувати графік необхідно скористатись меню *Statistics/Basic statistics and tables/Descriptive statistics* і у випадаючому вікні (рис. Л1.4) натиснути на кнопку *2M scatterplot* та вибрати змінні для аналізу.

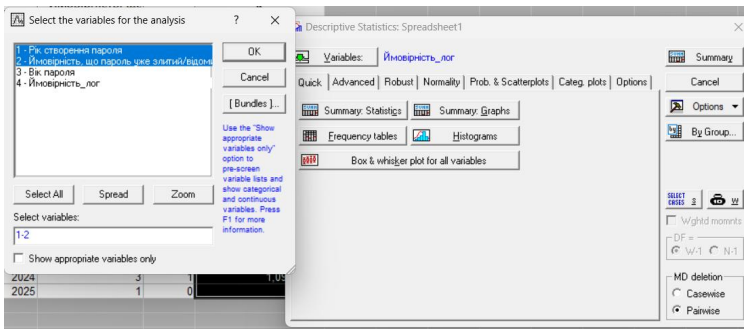


Рисунок Л1.4 – Побудова графіка

Після натиснення на кнопку ОК буде побудовано графік, який відображатиме залежність ціни від року (рис. Л1.5), а другий – ціни (логарифмічної) від віку (рис. Л1.6).

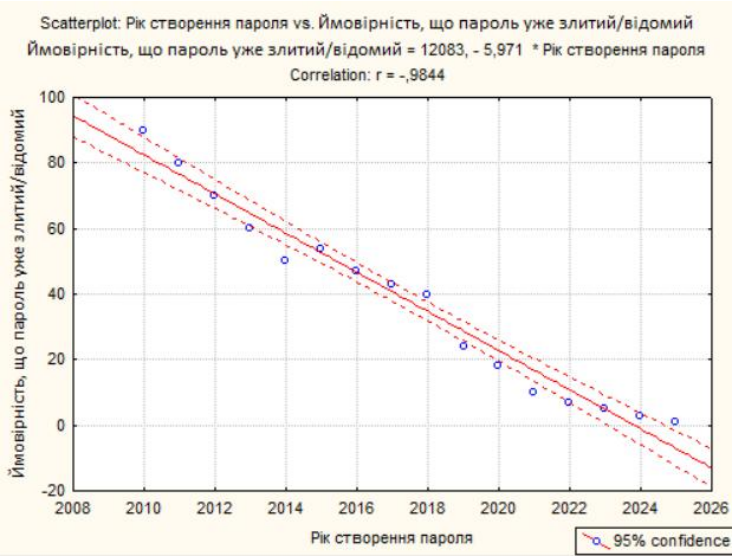


Рисунок Л1.5 – Графік оберненої залежності ймовірності від року створення пароля

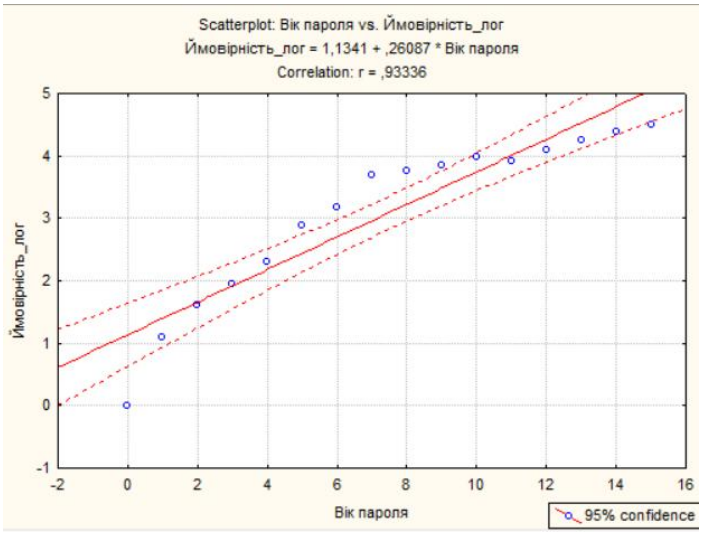


Рисунок Л1.6 – Графік прямої залежності ймовірності (логарифмічної) від віку пароля

З графіка на рис. Л1.5 видно, що чим менший рік створення пароля, тим більша ймовірність, що він вже злитий в інтернет бази та відомий. Залежність між ймовірністю і роком є тісною, оскільки коефіцієнт парної кореляції є високим – R=0,9844. Пояснює коефіцієнт кореляції таблиця Чедока.

Таблиця Л1.2

Таблиця Чедока

Величина коефіцієнта кореляції	0,1 - 0,3	0,3 - 0,5	0,5 - 0,7	0,7 - 0,9	0,9 - 1,0
Ступінь узгоджено- сті думок експертів	слабка	помірна	помітна	висока	дуже висока

Згідно з другим графіком (рис. Л1.6), можна зробити висновок, що чим більший вік пароля, тим більше значення логарифму ймовірності. Зв'язок між цими величинами є досить щільний, що показує коефіцієнт парної кореляції R=0,9333.

Далі необхідно обчислити дескриптивні статистики. Для цього також потрібно використати меню *Statistics/Basic statistics and tables/Descriptive statistics* і у випадаючому вікні (рис. Л1.7) на вкладці *Advanced* вказати, які статистики необхідно відобразити. Також обов'язково необхідно вибрати змінні для аналізу (*Variables*) .

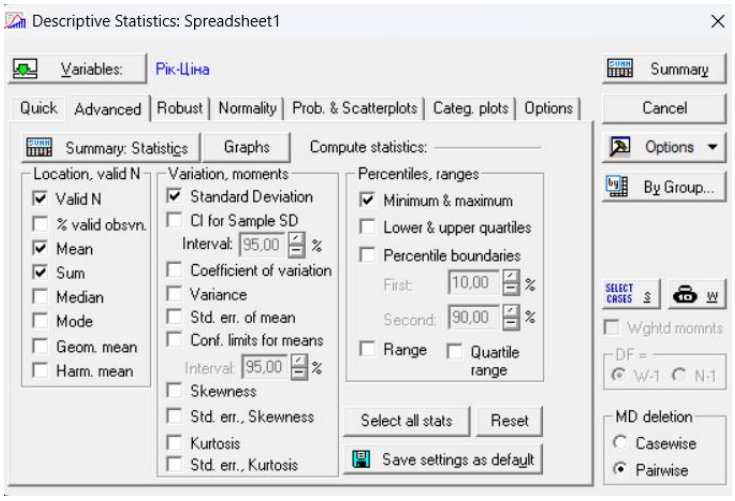


Рисунок Л1.7 – Побудова описових статистик

Після натиснення на кнопку ОК появиться вікно з результатами (рис. Л1.8).

Variable	Descriptive Statistics (Spreadsheet1)					
	Valid N	Mean	Sum	Minimum	Maximum	Std.Dev.
Рік створення пароля	16	2017,500	32280,00	2010,000	2025,000	4,76095
Ймовірність, що пароль уже злитий/відомий	16	37,625	602,00	1,000	90,000	28,87646
Вік пароля	16	7,500	120,00	0,000	15,000	4,76095

Рисунок Л1.8 – Дескриптивні статистики

У цьому вікні показано середнє значення, мода, медіана, мінімум, максимум та дисперсія.

Середнє значення ймовірності становить 37,6%, а середній вік – приблизно 7,5 років. Середні значення року та ймовірності-логарифм враховувати не доцільно.

Наступним завданням є побудова і вибір адекватної моделі залежності ймовірності від року створення пароля чи ймовірності-логарифм від віку пароля. Ці моделі обчислюються за допомогою модуля *Multiple Regression/Множинна регресія* програмного забезпечення *Statistica*. Після запуску цього меню відкриється наступне вікно, в якому необхідно вибрати параметри (рік та ймовірність) і натиснути на кнопку ОК (рис. Л1.9).

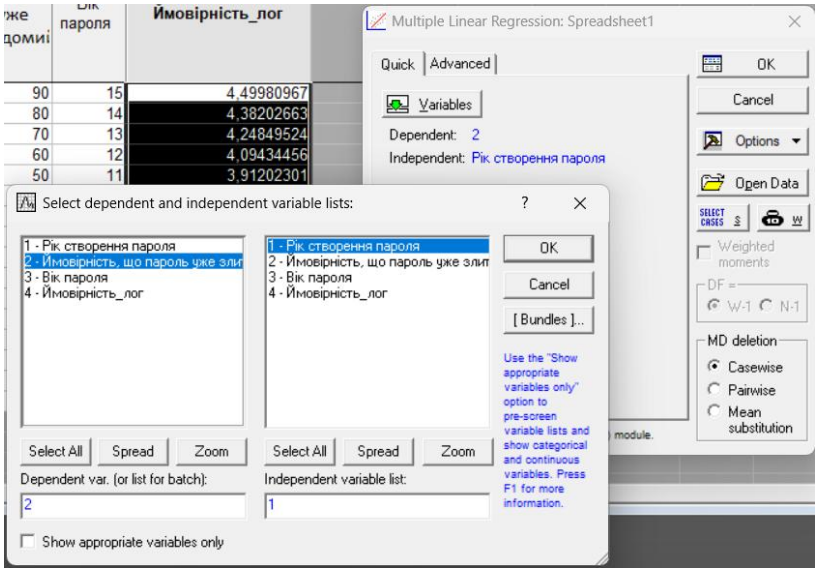


Рисунок Л1.9 – Модуль «Множинна регресія»

Після цього у вікні, яке з’явилось і яке зображене нижче, необхідно натиснути на кнопку ОК для того, щоб отримати необхідні параметри моделі.

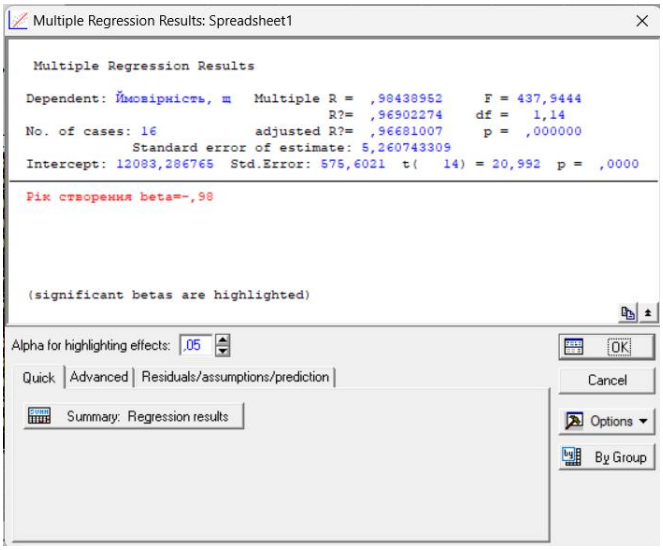


Рисунок Л1.10 – Вікно «Результати множинної регресії»

З’явиться вікно з параметрами моделі.

Regression Summary for Dependent Variable: Ймовірність, що пароль уже злитий
R= ,98438952 R?= ,96902274 Adjusted R?= ,96681007
F(1,14)=437,94 p<,00000 Std.Error of estimate: 5,2607

N=16	Beta	Std.Err. of Beta	B	Std.Err. of B	t(14)	p-level
Intercept			12083,29	575,6020	20,9924	0,000000
Рік створення пароля	-0,984390	0,047039	-5,97	0,2853	-20,9271	0,000000

Рисунок Л1.11 – Параметри моделі

Ця модель матиме вигляд:

Y=12083,29-5,97*x, де

Y – ймовірність,

x – рік створення пароля.

Коефіцієнт детермінації у цій моделі дорівнює 0,963, а кількість значущих параметрів – 2. Можна зробити висновок, що ця модель є адекватною, оскільки чим вищий коефіцієнт детермінації, тим більш адекватніша модель.

Аналогічно будується модель залежності логарифмічної ціни від віку. Після вибору параметрів у модулі *Multiple Regression/Множинна регресія*, отримуємо такі результати (рис. Л1.12).

Regression Summary for Dependent Variable: Ймовірність_лог (Spreadsheet1)
R= ,93336095 R?= ,87116266 Adjusted R?= ,86195999
F(1,14)=94,664 p<,00000 Std.Error of estimate: ,49440

N=16	Beta	Std.Err. of Beta	B	Std.Err. of B	t(14)	p-level
Intercept			1,134120	0,236042	4,804729	0,000280
Вік пароля	0,933361	0,095931	0,260875	0,026813	9,729550	0,000000

Рисунок Л1.12 – Параметри моделі

Друга модель матиме вигляд:

Y = 1,13+0,26*x, де

Y – ймовірність логарифмічна, x – вік пароля.

Коефіцієнт детермінації в цій моделі дорівнює 0,87, а кількість значущих параметрів – 2

Вибір кращої моделі здійснюється за коефіцієнтами детермінації та за кількістю значущих параметрів. Оскільки у двох моделей кількість значущих параметрів однакова, то вибір необхідно здійснювати за коефіцієнтом детермінації. В першій моделі він становить 0,963, в другій – 087. Тому доцільним є вибір першої моделі залежності логарифмічної ймовірності від віку.

Лабораторна робота 2. Прогнозування на основі залежності між двома змінними. Кореляційний аналіз

Одна з найбільш поширених задач статистичного дослідження, яка виконується в процесі прогнозування, полягає у вивченні зв'язку між вибірками. Зазвичай зв'язок між вибірками носить не функціональний, а імовірнісний (або стохастичний) характер. У цьому випадку немає суворої, однозначної залежності між величинами. При вивченні стохастичних залежностей розрізняють **кореляцію** і **регресію**.

Кореляційний аналіз полягає у визначенні ступеня зв'язку між двома випадковими величинами X і Y. В якості міри такого зв'язку використовується коефіцієнт кореляції. Коефіцієнт кореляції оцінюється за вибіркою обсягу *n* пов'язаних пар спостережень (X_i, Y_i) із спільної генеральної сукупності X і Y. Існує кілька типів коефіцієнтів кореляції, застосування яких залежить від вимірювання (способу шкалювання) величин X і Y.

Для оцінки ступеня взаємозв'язку величин X і Y, виміряних в кількісних шкалах, використовується коефіцієнт лінійної кореляції (коефіцієнт Пірсона), який передбачає, що вибірки X і Y розподілені за нормальним законом. При такому розподілі велика частина значень групується біля деякого середнього значення, по обидві сторони від якого частота спостережень поступово знижується.

1. Лінійний коефіцієнт кореляції – параметр, який характеризує ступінь лінійного взаємозв'язку між двома вибірками, розраховується за формулою (Л2.1):

$$r_{xy} = \frac{\sum\left(x_i - \bar{x}\right)\cdot\left(y_i - \bar{y}\right)}{\sqrt{\sum\left(x_i - \bar{x}\right)^2 \cdot \sum\left(y_i - \bar{y}\right)^2}} \qquad \text{(Л2.1)}$$

де X_i – значення, що приймаються у вибірці X,

Y_i – значення, що приймаються у вибірці Y;

$\overline{x}$ – середнє значення по X, $\overline{y}$ – середнє значення по Y.

Коефіцієнт кореляції змінюється від -1 до +1 . Коли при розрахунку виходить величина більша за +1 або менша за -1, – то трапилася помилка в обчисленнях. При значенні 0 лінійної залежності між двома вибірками немає.

Знак коефіцієнта кореляції дуже важливий для інтерпретації отриманого зв'язку. Якщо значення коефіцієнта лінійної кореляції дадатнє, то зв'язок між керуючими ознаками такий, що більше значення однієї ознаки (змінної) відповідає більшому значенню іншої ознаки (іншої змінної). Іншими словами якщо один показник (змінна) збільшується, то відповідно збільшується й інший показник (змінна). Така залежність носить назву прямо пропорційної залежності.

Якщо ж отримане значення коефіцієнта лінійної кореляції від'ємне, то більшому значенню однієї ознаки відповідає менше значення іншої. Інакше кажучи, за наявності знака мінус, збільшенню однієї змінної (ознаки, значення) відповідає зменшення іншої змінної. Така залежність носить назву обернено пропорційної залежності.

Проаналізувати отримані значення коефіцієнта лінійної кореляції можна, скориставшись таблицею Л2.1.

Таблиця Л2.1

Тіснота зв'язку і величина коефіцієнта кореляції

Коефцієнт кореляції r_{xy}	Тіснота зв'язку
$\pm$ 0,91-1,0	Дуже велика
$\pm$ 0,81-0,9	Вельми сильна
$\pm$ 0,65-0,8	Сильна
$\pm$ 0,45-0,64	Помірна
$\pm$ 0,25-0,44	Слабка
До $\pm$ 0,25	Дуже слабка
«+» – пряма залежність «-» – обернена залежність	

Для того, щоб оцінити наявність зв'язку між двома змінними, також можна використовувати t-статистику Стьюдента, яка оцінює відношення величини

лінійного коефіцієнта кореляції до середнього квадратиччного відхилення і розраховується за формулою (Л2.2).

$$t_{розр} = \frac{r_{xy}\sqrt{n-2}}{\sqrt{1-r_{xy}^2}} \qquad (Л2.2)$$

Отриману величину $t_{розр}$ порівнюють з табличним значенням t-критерію Стьюдента з n-2 ступенями вільності. Якщо $t_{розр} > t_{табл,}$ то практично неймовірно, що знайдене значення обумовлене лише випадковими збігами величин X і Y у вибірці з генеральної сукупності, тобто існує залежність між X та Y. І навпаки, якщо $t_{розр} < t_{табл}$, то величини X та Y незалежні.

Приклад. Проведення параметричного кореляційного аналізу

Для 20 компаній було визначено обсяги інвестицій у заходи з кібербезпеки та кількість інцидентів, яких вдалося уникнути після впровадження цих заходів. Необхідно визначити коефіцієнт кореляції між досліджуваними ознаками, використовуючи:

1. Формулу лінійного коефіцієнта кореляції.
2. Статистичну функцію КОРРЕЛ.
3. Пакет аналізу MS Excel.
4. Програмне забезпечення IBM SPSS Statistics.

Розрахунок коефіцієнта кореляції Пірсона з використанням формули (Л2.1).

Відкрийте нову книгу MS Excel і створіть таблицю згідно з рис. Л2.1, збережіть файл під ім'ям Параметричний_Аналіз_Ваше Прізвище.xls.

	A	B	C	D	E	F	G	H	I	J	K	L	M	N
1	№ компанії	Обсяг інвестицій Xi	Кількість інцидентів Yi	Xcep	Xi-Xcep	Ycep	Yi-Ycep	Сума (Xi-Xcep)*(Yi-Ycep)	(Xi-Xcep)^2	(Yi-Ycep)^2	Сума((Xi-Xcep)^2)	Сума((Yi-Ycep)^2)	Корінь(Сума((Xi-Xcep)^2)* Сума((Yi-Ycep)^2)	Rxy
2	1	5	720											
3	2	5	750											
4	3	7	800											
5	4	8	820											
6	5	9	800											
7	6	10	880											
8	7	11	950											
9	8	12	920											
10	9	13	900											
11	10	14	950											
12	11	15	920											
13	12	15	980											
14	13	16	980											
15	14	17	970											
16	15	18	980											
17	16	18	1010											
18	17	19	1100											
19	18	20	1100											
20	19	20	1115											
21	20	21	1100											
22														
23	**Коефіцієнт кореляції**													
24	**Т-статистика Стьюдента**													

Рисунок Л2.1 – Початкові дані

1. В клітинці **D2** обчисліть $\bar{X}$ (середнє значення змінної X_i).

2. В діапазоні клітинок **E2:E21** обчисліть $X_i - \bar{X}$ (відхилення значень змінної X_i від середнього значення $\bar{X}$).

3. В клітинці **F2** обчисліть $\bar{Y}$ (середнє значення змінної Y_i).

4. В діапазоні клітинок **G2:G21** обчисліть $Y_i - \bar{Y}$ (відхилення значень змінної Y_i від середнього значення $\bar{Y}$).

5. В клітинці **H2** обчисліть суму добутку $\sum\left(X_i - \bar{X}\right)\cdot\left(Y_i - \bar{Y}\right)$, скориставшись математичною функцією SUMPRODUCT.

6. В діапазоні клітинок **I2:I21** обчисліть $(X_i - \bar{X})^2$ (квадрати відхилень значень змінної X_i від середнього значення $\bar{X}$).

7. В діапазоні клітинок **J2:J21** обчисліть $(Y_i - \bar{Y})^2$ (квадрати відхилень значень змінної Y_i від середнього значення $\bar{Y}$).

8. В клітинці **K2** обчисліть $\sum\left(X_i - \bar{X}\right)^2$.

9. В клітинці **L2** обчисліть $\sum\left(Y_i - \bar{Y}\right)^2$.

10. В клітинці **M2** обчисліть $\sqrt{\sum\left(x_i-\bar{x}\right)^2\cdot\sum\left(y_i-\bar{y}\right)^2}$.

11. В клітинці **N2** обчисліть коефіцієнт лінійної кореляції R_{xy}.

12. Зробіть висновок про тісноту зв'язку між витратами на систему захисту в організації та кількістю здійснених комп'ютерних атак на цю організацію.

Обчислення коефіцієнта кореляції з використанням функції CORREL.

1. В клітинці D23 обчисліть коефіцієнт кореляції, використовуючи функцію CORREL з категорії Статистичні.

Синтаксис функції CORREL:

CORREL (масив 1; масив 2):

де *масив 1* – посилання на діапазон клітинок першої вибірки (X);

масив 2 – посилання на діапазон клітинок другої вибірки (Y).

23	**Коефіцієнт кореляції**	0,952

Рисунок Л2.2 – Результат виконання функції CORREL

2. Оцініть значимість коефіцієнта кореляції. З цією метою розглядаються дві гіпотези. **Основна H_0**: $\rho_{xy}=0$ і **альтернативна H_1**: $\rho_{xy}\neq 0$. Для перевірки гіпотези H_0 обчисліть **t-статистику Стьюдента** за формулою (Л2.2) в клітинці **D24.** В нашому випадку число ступенів вільності $\nu = n-2=20-2 = 18$ (кількість спостережень мінус два).

3. Порівняйте отримане значення з критичним значенням $t_{\nu,\alpha}$ розподілу Стьюдента. (При $\nu =18$ і ймовірності $\alpha = 0{,}05$, $t_{\nu,\alpha,табл} = 1{,}734$). Зробіть висновки про значимість коефіцієнта кореляції.

24	**T-статистика Стьюдента**	**13,24**

Рисунок Л2.3 – Результат обчислення статистики Стьюдента

Обчислення коефіцієнта кореляції з використанням пакета аналізу MS Excel

1. Для початку переконайтесь, що був активований Пакет аналізу, тобто в меню *Дані* існує команда *Аналіз даних*. Якщо ні, то виконайте команду *Файл/Параметри/Надбудови*. В діалоговому вікні *Надбудови* встановіть галочку Пакет аналізу і натисніть кнопку ОК.

2. Потім виконайте команду *Дані/Аналіз даних*. Виберіть інструмент аналізу Кореляції.

3. В діалоговому вікні *Кореляція* встановіть параметри згідно з рис. Л2.4.

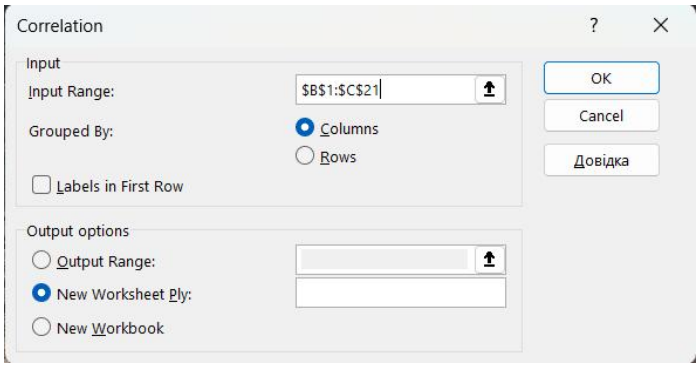


Рисунок Л2.4 – Діалогове вікно «Кореляція»

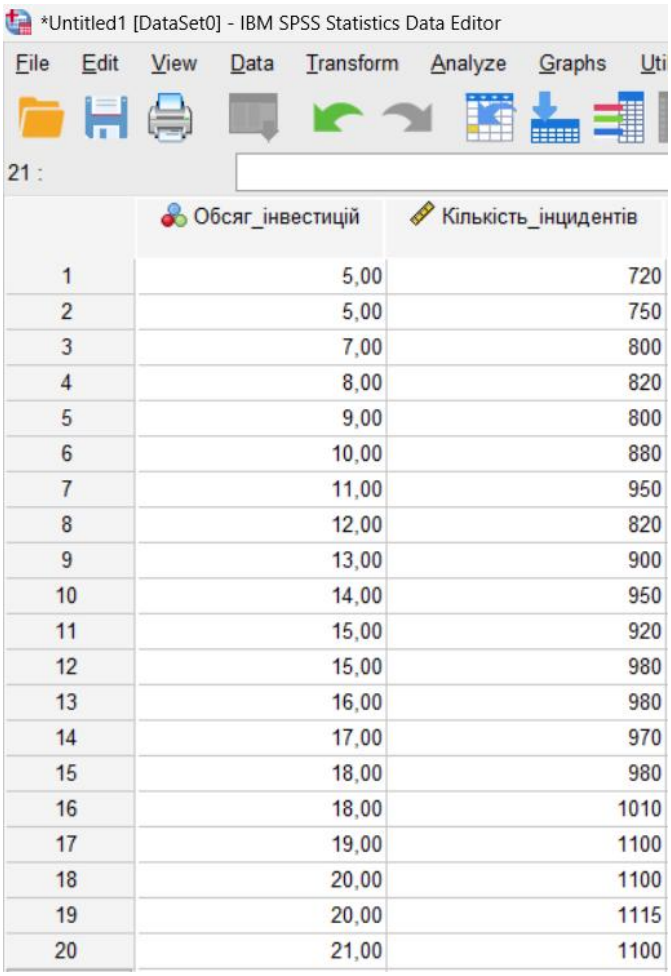
4. На новому рабочому листі отримайте значення коефіцієнта кореляції. Порівняйте це значення, зі значеннями, які були отримані попередніми способами.

5. Збережіть зміни.

Здійснення параметричного кореляційного аналізу в програмному забезпечення IBM SPSS Statistics

1. Отримайте у викладача програмне забезпечення IBM SPSS Statistics, встановіть його на ПК та введіть дані в режимі Data View (Перегляд даних)

згідно з рис. Л2.5. Збережіть під ім'ям Параметричний_Аналіз_Ваше прізвище.Sav (дані можна також скопіювати з файлу MS Excel).



	Обсяг_інвестицій	Кількість_інцидентів
1	5,00	720
2	5,00	750
3	7,00	800
4	8,00	820
5	9,00	800
6	10,00	880
7	11,00	950
8	12,00	820
9	13,00	900
10	14,00	950
11	15,00	920
12	15,00	980
13	16,00	980
14	17,00	970
15	18,00	980
16	18,00	1010
17	19,00	1100
18	20,00	1100
19	20,00	1115
20	21,00	1100

Рисунок Л2.5 – Дані в режимі Data View

Згідно з статистичною теорією, щоб зробити можливим застосування більшості статистичних процедур, дані повинні підкорятися закону нормального розподілу, при якому велика частина значень групується біля деякого середнього значення, по обидві сторони від якого частота спостережень поступово знижується. На діаграмі (рис. Л2.6) показана крива нормального розподілу (Дзвін Гауса).

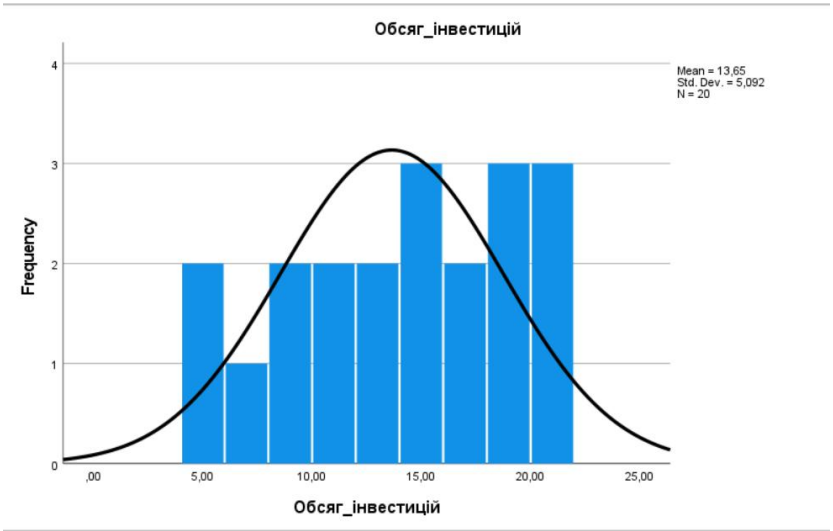


Рисунок Л2.6 – Крива нормального розподілу

Для того, щоб визначити чи підкоряються два емпіричних розподіли одному закону, або визначити чи підпорядковується отриманий розподіл передбачуваної моделі, використовується критерій згоди Колмогорова (також відомий, як критерій згоди Колмогорова-Смірнова).

2. Перш ніж досліджувати наявність кореляції між досліджуваними ознаками протестуємо обидві змінні Витрати і Кількість на закон нормального розподілу за допомогою критерію згоди Колмогорова-Смірнова. Для цього:

✓ Виконайте команду Analyze/Nonparametric Tests/One-Sample (Аналіз/Непараметричні тести/Одновибірковий Колмогорова-Смирнова) .

✓ У діалоговому вікні перетягніть змінну Обсяг_інвестицій в область Test Variable List (Список тестованих змінних).

✓ Виберіть тип розподілу Normal (Нормальний розподіл).

✓ При малих розмірах вибірки застосовуються точні методи. Виберіть використання точного методу при допомозі кнопки [Exact] (Точний). Для практичних цілей слід застосовувати метод Monte-Carlo зі встановленою за замовчуванням кількістю вибірок (10 000). Довірчий рівень 99 % практично завжди є занадто високим, тому змініть його на 95 %, що відповідає довірчому

рівню при розрахунку статистичної помилки вибірки для маркетингових досліджень (рис. Л2.7).

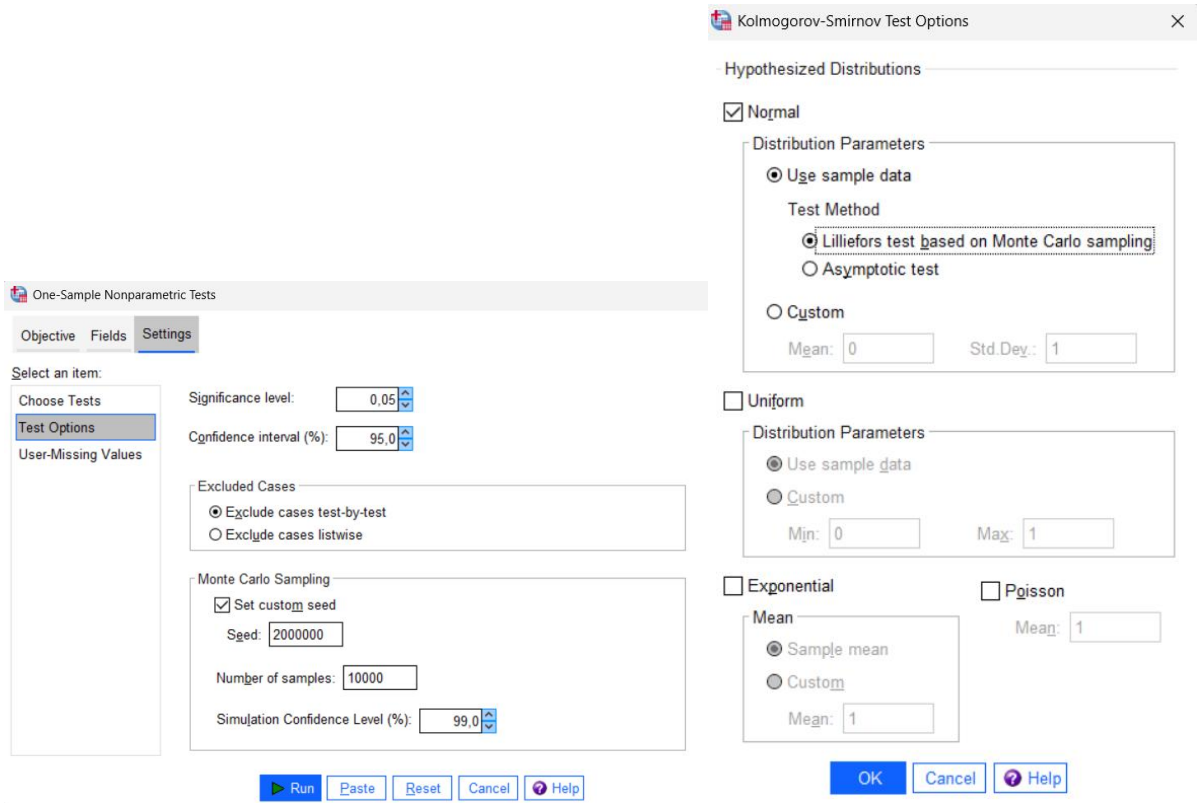


Рисунок Л2.7 – Налаштування параметрів точного методу.

✓ Натисніть на кнопку [Run] (Виконати) и ОК.

3. Проаналізуйте отриманий результат (рис. Л2.8).

Hypothesis Test Summary

	Null Hypothesis	Test	Sig.[a]	Decision
1	The distribution of Кількість_інцидентів is normal with mean 932 and standard deviation 120,334.	One-Sample Kolmogorov-Smirnov Test	,560	Retain the null hypothesis.

a. The significance level is ,050. Lilliefors' method based on 10000 Monte Carlo samples with starting seed 2000000.

One-Sample Kolmogorov-Smirnov Normal Test

Кількість_інцидентів

One-Sample Kolmogorov-Smirnov Normal Test Summary

Total N			20
Most Extreme Differences	Absolute		,125
	Positive		,125
	Negative		-,118
Test Statistic			,125
Asymptotic Sig.(2-sided test)[a]			,200[b]
Monte Carlo Sig.(2-sided test)[c]	Sig.		,560
	99% Confidence Interval	Lower Bound	,547
		Upper Bound	,572

a. Lilliefors Corrected

b. This is a lower bound of the true significance.

c. Lilliefors' method based on 10000 Monte Carlo samples with starting seed 2000000.

Рисунок Л2.8 – Результат тестування на закон нормального розподілу.

Інтерпретація результатів

Вибірка (N): 20 спостережень.

Статистика K-S (D): 0,125.

p-значення (Monte Carlo Sig.): 0,560.

Рівень значущості (α): 0,05.

Оскільки p = 0,560 > 0,05, ми не відхиляємо нульову гіпотезу. Це означає, що дані щодо показника «Кількість_інцидентів» не мають статистично значущих відмінностей від нормального розподілу.

За результатами однофакторного тесту Колмогорова–Смірнова (D = 0,125; p = 0,560), розподіл змінної «Кількість_інцидентів» не відрізняється від нормального на рівні значущості 0,05. Отже, вибірка може розглядатися як така, що має нормальний розподіл.

У ході перевірки нормальності розподілу показника «Кількість_інцидентів» за допомогою тесту Колмогорова–Смірнова було отримано такі результати: середнє значення становить 932, а стандартне відхилення – 120,334, що відображає середній рівень варіативності даних відносно центру розподілу.

Асимптотичне значення ймовірності (Asymp. Sig.) дорівнює 0,200, однак з огляду на малий обсяг вибірки (N = 20) більш коректним є значення, обчислене методом Монте-Карло – p = 0,560. Оскільки p > 0,05, підстав відхиляти нульову гіпотезу немає. Це означає, що розподіл змінної «Кількість_інцидентів» не відрізняється від нормального.

Для візуального перегляду функції нормального розподілу виконайтете команду **Graphs/Histogram** *(Графіки/Гістограма)*. В діалоговому вікні в область **Variable** *(Змінна)* перетягніть змінну **Обсяг інвестицій** і встановіть галочку **Display normal curve** *(Відображати нормальну криву)*.

4. Аналогічним шляхом протестуйте іншу змінну **Кількість інцидентів** на закон нормального розподілу.

5. Після того як змінні протестовані на закон нормального розподілу, можна приступити до визначення лінійного коефіцієнта кореляції Пірсона. Для цього:

✓ Виконайте команду **Analyze/Correlate/Bivariate** *(Аналіз/Кореляція/Парні)*.

✓ В діалоговому вікні в область змінних перетягніть дві змінні, вкажіть коефіцієнт кореляції Пірсона, рівень значимості – двосторонній, натисніть кнопку **[Options]** і встановіть параметри згідно з рис. Л2.9.

✓ Натисніть кнопку **[Continue]** *(Продовжити)* і **ОК**.

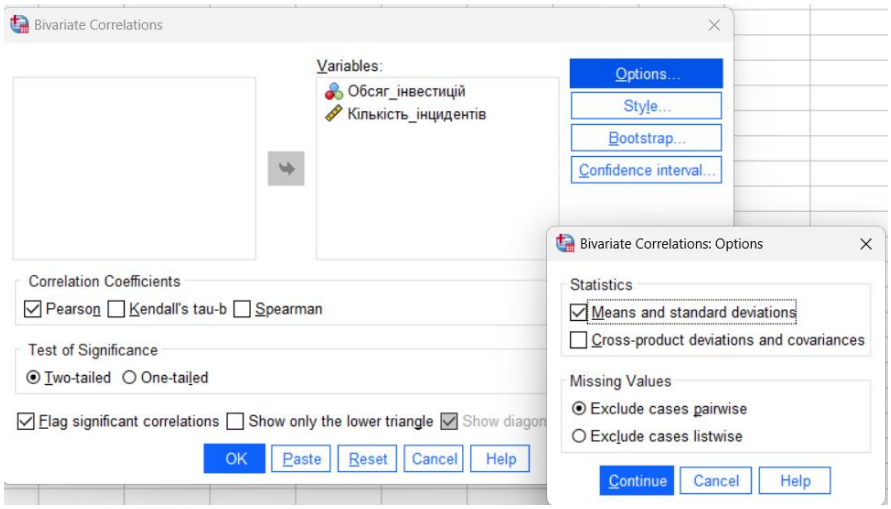


Рисунок Л2.9 – Діалогове вікно «Кореляції»

6. Проаналізуйте отриманий результат (рис. Л2.10).

✓ У верхній області звіту SPSS вивів таблицю **Descriptive Statistics** (*Описова статистика*), оскільки була встановлена галочка **Means and standard deviations** *(Середнє значення та стандартне відхилення).*

✓ З таблиці **Correlations** *(Корреляции)* видно, що коефіцієнт кореляції Пірсона дорівнює 0,952**,** а це вказує на те, що залежність між витратами на рекламу і кількістю залучених туристів прямо пропорційна, а **тіснота зв'язку дуже сильна.**

✓ Статистична значимість (**Sig. (2-tailed)**) є меншою 0,001, і показує ймовірність того, що кореляція не є випадковою, тобто з ймовірністю 95% можна стверджувати, що коефіцієнт кореляції є значущим.

Correlations

Descriptive Statistics

	Mean	Std. Deviation	N
Обсяг_інвестицій	13,6500	5,09153	20
Кількість_інцидентів	932,25	120,334	20

Correlations

		Обсяг_інвестицій	Кількість_інцидентів
Обсяг_інвестицій	Pearson Correlation	1	,952**
	Sig. (2-tailed)		<,001
	N	20	20
Кількість_інцидентів	Pearson Correlation	,952**	1
	Sig. (2-tailed)	<,001	
	N	20	20

**. Correlation is significant at the 0.01 level (2-tailed).

Рисунок Л2.10 – Звіт щодо кореляції Пірсона

7. Для візуального перегляду кореляційного поля виконайте команду **Graphs/Scatter-Dot** (*Графіки/Точкове розсіювання*), потім вкажіть **Simple Scatter** (*Просте розсіювання*), натисніть кнопку **[Define]** (*Визначити*), в область осі X перенесіть змінну **Обсяг інвестицій**, а в область осі Y – змінну

Кількість інцидентів. Натисніть ОК. Результат можна побачити на рисунку Л2.11.

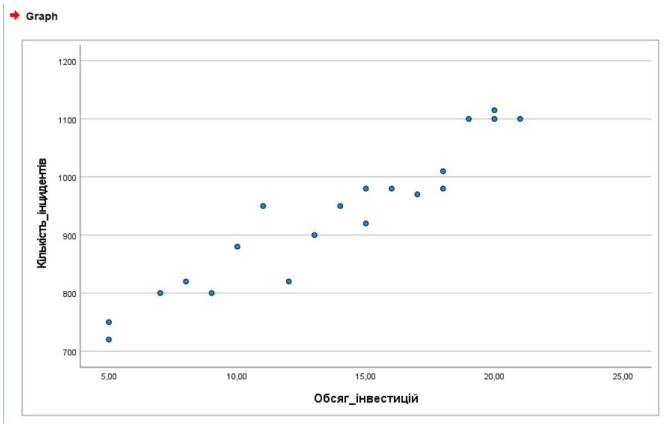


Рисунок Л2.11 – графік кореляції Пірсона

Збережіть отриманий звіт під іменем *Аналіз_Пірсон_Ваше прізвище.spo.*

Завдання

На основі даних таблиці Л2.2 та вищенаведеного алгоритму обчислень необхідно провести параметричний кореляційний аналіз.

Таблиця Л2.2

Вхідні дані

№ організації	Витрати на систему захисту, тис. грн..	Кількість здійснених атак
1	10+N	15+N
2	12+N	14+N
3	13,1+N	14+N
4	14,3+N	13+N
5	13+N	13+N
6	15+N	12+N
7	16+N	11+N
8	16,3+N	14+N
9	16+N	14+N
10	15,9+N	12+N
11	16,7+N	13+N
12	17+N	14+N
13	17,4+N	13+N
14	17,2+N	12+N
15	16,8+N	10+N
16	17,6+N	9+N

17	18+N	10+N
18	19,2+N	8+N
19	18,5+N	11+N
20	18,8+N	13+N

N – порядковий номер здобувача в журналі групи.

Лабораторна робота 3. Побудова прогнозів за допомогою функцій регресії Excel. Регресійний аналіз

Мета регресійного аналізу – визначити кількісні зв'язки між залежними випадковими величинами. Одна з цих величин визначається залежною і називається **відклик**, інші – незалежні, називаються факторами або **регресорами**.

Регресійний аналіз дає змогу:

• Розраховувати регресійні моделі шляхом визначення значень параметрів – постійних коефіцієнтів при незалежних змінних – регресорах.

• Перевіряти гіпотезу щодо адекватності моделей наявними спостереженнями.

• Використовувати модель для ***прогнозування*** значень залежної змінної при нових або неспостережуваних значеннях незалежних змінних.

Серед регресійних моделей зазвичай виділяють:

1. Однофакторні моделі (залежно від однієї змінної).

2. Багатофакторні моделі (залежно від декількох змінних).

3. Лінійні моделі відносно незалежних змінних.

4. Нелінійні моделі за змінними і нелінійні моделі за параметрами.

1. Однофакторні моделі

Найбільш поширені методи регресійного аналізу є параметричними, більша їх частина заснована на припущенні про нормальний розподіл даних спостережень, тому в кожному випадку аналізу необхідна попередня перевірка даних, закону нормальному розподілу.

Лінійна модель.

Якщо коефіцієнт кореляції за абсолютною величиною близький до одиниці, то для побудови залежності використовується ***лінійна*** модель. Для інших випадків використовуються більш складні ***нелінійні*** моделі.

Модель простої лінійної регресії має вигляд:

$$y=b*x+a \qquad (Л3.1)$$

де y – залежна змінна (відклик),

x – незалежна змінна-регресор (чинник),

b – коефіцієнт регресії,

a – зміщення по осі ординат.

Зміщення по осі ординат відповідає точці на осі Y (вертикальної осі), де пряма регресії перетинає цю вісь. Коефіцієнт регресії b через співвідношення $b= tg(a)$ вказує на кут нахилу прямої.

Для отримання параметрів моделі a і b найчастіше використовують метод найменших квадратів, заснований на мінімізації середньоквадратичної помилки моделі. Тобто оптимальним рішенням задачі побудови лінійної регресії є така пряма, для якої сума квадратів вертикальних відстаней до окремих точок даних є мінімальною.

При виконанні регресійного аналізу необхідно отримати оцінки, що дають змогу побудувати прогноз з певною точністю і ймовірністю. При нормальному законі розподілу умови будуть задоволені, якщо оцінити:

1. **Очікувані значення коефіцієнтів *b* и *a*.**

$$b = \frac{\sum (x_i - \bar{x}) \cdot (y_i - \bar{y})}{\sum (x_i - \bar{x})^2} \qquad (Л3.2)$$

$$a = \bar{y} - b \cdot \bar{x} \qquad (Л3.3)$$

де x_i – значення незалежної ознаки;

$\bar{x}$ – середнє значення незалежної ознаки;

y_i – фактичне значення результативного признака, отримані за даними спостережень;

$\bar{y}$ – середнє значення фактичних значень результативної ознаки.

2. **Стандартну (середню квадратичну) похибку моделі.**

$$S_\theta = \sqrt{\frac{\sum (y_i - \hat{y}_i)^2}{n - m}} \qquad \text{(Л3.4)},$$

де y_i – фактичні значення результативного признака, отримані за даними спостережень;

$\hat{y}_i$ – обчислені значення результативного признака;

n – об’єм вибірки;

m – число параметрів в рівнянні регресії.

Приклад

Побудова моделі лінійної регресії

По 20 компаніях було визначено витрати на впровадження сучасних систем моніторингу кіберзагроз та кількість виявлених і заблокованих атак, що зростала зі збільшенням інвестицій. Необхідно побудувати модель лінійної регресії, враховуючи, що змінні підпорядковані нормальному закону розподілу.

Обчислення параметрів рівняння лінійної регресії з використанням функції MS EXCEL LINEST ().

Відкрийте нову книгу MS Excel і створіть таблицю згідно рис. Л3.1, збережіть у своїй папці під ім'ям Регресійний_Аналіз.xls.

	A	B	C	D	E	F	G	H	I
1	№ компанії	Витрати на впровадження Xi	Кількість виявлених атак Yi	Yi^2	(Yi-Yi^)	(Yi-Yi^)2	Сума((Yi-Yi^)2)	So	Yсер
2	1	5	720						
3	2	5	750						
4	3	7	800						
5	4	8	820						
6	5	9	800						
7	6	10	880						
8	7	11	950						
9	8	12	920						
10	9	13	900						
11	10	14	950						
12	11	15	920						
13	12	15	980						
14	13	16	980						
15	14	17	970						
16	15	18	980						
17	16	18	1010						
18	17	19	1100						
19	18	20	1100						
20	19	20	1115						
21	20	21	1100						
22									
23									
24									
25	**Коефіцієнти рівняння**		**b**	**a**					
26									
27	**Рівняння регресії**								

Рисунок Л3.1 – Початкові дані

1. Для отримання коефіцієнтів *a* і *b* лінійного рівняння регресії y = b * x + a, що описує залежність кількості виявлених атак від витрат на впровадження систем моніторингу кібератак скористаємося статистичною функцією MS Excel LINEST (). Для цього виділіть дві клітинки **C26:D26** і виконайте вставку функції LINEST () з аргументами, згідно з рис. Л3.2. Тут Відомі_значення_y – діапазон значень Кількість туристів, Відомі значення_x – діапазон значень Витрати на рекламу. Натисніть комбінацію клавіш SHIFT + CTRL + ENTER.

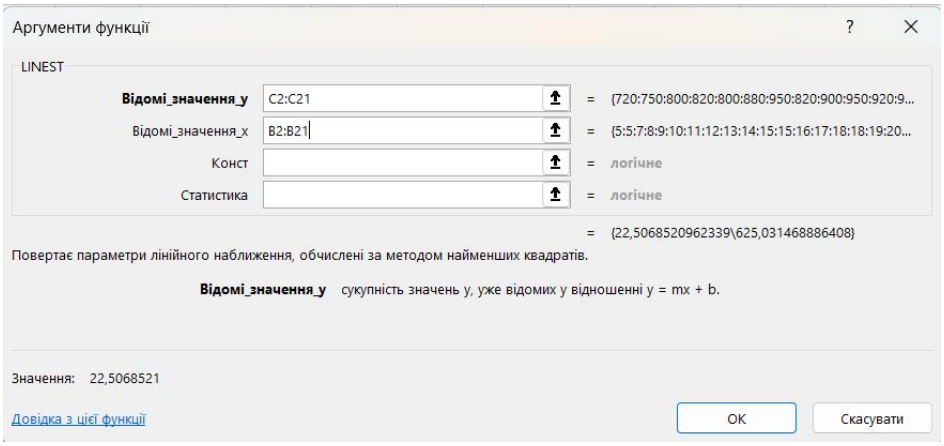


Рисунок Л3.2 – Аргументи функції LINEST.

2. В клітинку **C27** введіть рівняння регресії ***y=b*x+ a***, (замість *b* і *a* підставте отримані коефіцієнти лінійної регресії (рис. Л3.3).

25	**Коефіцієнти рівняння**	**b**	**a**
26		22,507	625,031
27	**Рівняння регресії**	y=22,507*X+625,031	

Рисунок Л3.3 – Результат обчислення рівняння регресії

3. Обчисліть стандартну (середню квадратичну) помилку моделі. Для цього:

• У діапазоні клітинок **D2:D21** обчисліть значення результативної ознаки шляхом підстановки значень незалежної ознаки-регресорів X в рівняння лінійної регресії .

• У діапазоні клітинок **E2:E21** обчисліть відхилення фактичних значень результативної ознаки від розрахованих значень $(y_i - \hat{y}_i)$.

• У діапазоні клітинок **F2:F21** обчисліть квадрати відхилень $(y_i - \hat{y}_i)^2$.

• У комірці **G2** обчисліть суму квадратів відхилень $\sum(y_i - \hat{y}_i)^2$.

• У комірці **H2** обчисліть стандартну помилку моделі S_θ за формулою Л3.4.

• Проаналізуйте величину помилки. Для цього обчисліть середнє значення фактичної результативної ознаки в комірці **I2** і знайдене значення підставте в формулу $S_\theta / \bar{y}_i \cdot 100\%$ (рис. Л3.4).

H	I	J
S_θ	Ycep	Величина похибки
37,726	932,25	4,05%

Рисунок Л3.4 – Аналіз величини помилки

4. Зробіть висновки щодо адекватності побудованої моделі.

Знаходження рівняння лінійної регресії графічним методом

1. Для отримання рівняння регресії побудуємо кореляційне поле змінних X (витрати на рекламу) і Y (кількість туристів).

2. Виділіть діапазон комірок **B2:C21**, запустіть майстра діаграм і виберіть тип діаграми – *Точкова*. Задайте для діаграми ім'я – Кореляційне поле, вісь X – Витрати на впровадження, вісь Y – Кількість туристів. На останньому кроці майстра вкажіть місце розташування – окремий лист.

3. Додайте лінію тренда на точковий графік. Для цього необхідно виділити діаграму і виконати команду меню *Діаграма / Додати лінію тренда*, або виконати дану команду з контекстного меню, клацнувши по будь-якій точці графіка. **Лінія тренда** – це графічне представлення напрямки зміни ряду даних.

4. Виберіть тип тренда *Лінійний*, який використовується для апроксимації даних за методом найменших квадратів відповідно до рівняння y = b * x + a.

5. На вкладці *Параметри* встановіть прапорці *Показати рівняння на діаграмі* і *Помістити на діаграму величину достовірності апроксимації* R^2. Клацніть по кнопці ОК. R^2 – це число від 0 до 1, яке відображає близькість лінії тренда до фактичних даних. Лінія тренда найбільше відповідає дійсності, коли значення R^2 близьке до 1.

6. Порівняйте рівняння регресії, отримане графічним методом (рис. Л3.5), з рівнянням, розрахованим за допомогою функції LINEST ().



Рисунок Л3.5 – Модель лінійної регресії

Побудова моделі лінійної регресії за допомогою інструменту «Регресія»

1. Спочатку переконайтеся, що був активований Пакет аналізу, тобто в меню *Дані* є команда *Аналіз даних*. Якщо ні, то виконайте команду *Сервіс / Надбудови*. У діалоговому вікні Надбудови встановіть прапорець Пакет аналізу і клацніть по кнопці ОК.

2. Далі виконайте команду *Дані / Аналіз даних*. Виберіть інструмент аналізу *Регресія* зі списку Інструменти аналізу. Клацніть по кнопці ОК.

3. На екрані з'явиться діалогове вікно Регресія (рис. Л3.6).

• в текстовому полі Вхідний інтервал Y введіть діапазон зі значеннями залежної змінної **$C$2:$C$21**.

• в текстовому полі Вхідний інтервал X введіть діапазон зі значеннями незалежних змінних **$B$2:$B$21**.

• Переконайтеся, що в полі Рівень надійності введено 95% і перемикач *Параметри виводу* встановлений в положенні *Новий робочий лист*.

• Клацніть на кнопку ОК.

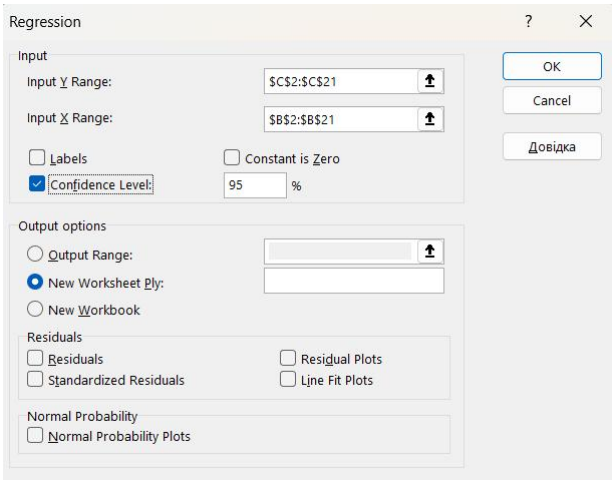


Рисунок Л3.6 – Діалогове вікно інструмента аналізу Регресія.

4. У результаті на новому аркуші будуть відображені результати використання інструменту Регресія (рис. Л3.7).

	A	B	C	D	E	F	G	H	I
1	SUMMARY OUTPUT								
2									
3	*Regression Statistics*								
4	Multiple R	0,95230459							
5	R Square	0,906884032							
6	Adjusted R Square	0,901710922							
7	Standard Error	37,72592269							
8	Observations	20							
9									
10	ANOVA								
11		*df*	*SS*	*MS*	*F*	*Significance F*			
12	Regression	1	249505,3356	249505,3356	175,3073385	1,01979E-10			
13	Residual	18	25618,41437	1423,245243					
14	Total	19	275123,75						
15									
16		*Coefficients*	*Standard Error*	*t Stat*	*P-value*	*Lower 95%*	*Upper 95%*	*Lower 95,0%*	*Upper 95,0%*
17	Intercept	625,0314689	24,68905537	25,31613541	1,59118E-15	573,1616883	676,9012495	573,1616883	676,9012495
18	X Variable 1	22,5068521	1,699866084	13,24036776	1,01979E-10	18,93556598	26,07813822	18,93556598	26,07813822

Рисунок Л3.7 – Вивід результатів інструменту Регресія.

5. В області **Регресійна статистика** отримали:

• Множинний R = 0,952 (коефіцієнт лінійної кореляції Пірсона).

• R – квадрат = 0,9068. Ця величина називається коефіцієнтом детермінації – це частка від суми квадратів, обумовлених регресією і залишковою сумою квадратів. Ця величина характеризує якість регресійної прямої, тобто ступінь відповідності між регресійної моделлю і вихідними даними. Міра визначеності завжди лежить в діапазоні від 0 до 1. У нашому випадку R2 = 0,9068 – це

означає, що регресійною моделлю описано 90,7 % випадків. У простому лінійному регресійному аналізі квадратний корінь з коефіцієнта детермінації, рівний коефіцієнту кореляції Пірсона. Порівняйте отриману величину з величиною достовірності апроксимації, отриманою при побудові моделі лінійної регресії графічним способом.

• Стандартна помилка = 37,73 – оцінка адекватності лінійної моделі. Порівняйте цей результат з величиною S_θ, отриманою вище (за формулою (16.4)).

6. В області **Дисперсійний аналіз** міститься така інформація:

6.1 Стовпець **df –** число ступенів вільності (використовується при перевірці адекватності моделі за статистичними таблицями):

- в рядку **Регресія з**находиться k_1 – кількість коефіцієнтів рівняння, не враховуючи вільного члена b;
- в рядку **Залишок** знаходится k_2=n- k_1 -1, де n – кількість початкових даних.

6.2. Стовпець **SS (сума квадратів):**

- в рядку **Регресія:** $SS^{regr} = \sum_{i=1}^{n} (\hat{Y}_i - \overline{Y})^2$ - наведена частка дисперсії, яка описується регресіє (сума квадратів, обумовлена регресією).

де $\hat{Y}_i$ **–** модельні (розрахункові) значення Y, отримані шляхом підстановки значень X в побудовану модель;

$\overline{Y}$ **–** середнє значення Y;

- в рядку **Залишок:** $SS_{resid} = \sum_{i=1}^{n} (\hat{Y}_i - Y_i)^2$ - наведена частка дисперсії, яка не враховується при записі рівняння (остаточна сума квадратів).

6.3. Стовпець **MS – допоміжні величини:**

- в рядку Регресія: $S_r^2 = SS_{reg} / k_1$;
- в рядку Залишок: $S_e^2 = SS_{resid} / k_2$.

Використовуються для розрахунку критерію Фішера.

6.4. Стовпець **F – критерій Фішера**. $F = \frac{S_r^2}{S_e^2}$.

6.5. Стовпець **Значимість F (критерію Фішера) – оцінка адекватності побудованої моделі**. Знаходиться за значеннями F, k_1 и k_2 за допомогою функції **FРАСП.** Якщо значимість **F** менше 0,05, то модель може вважатись адекватною з ймовірністю 0,95.

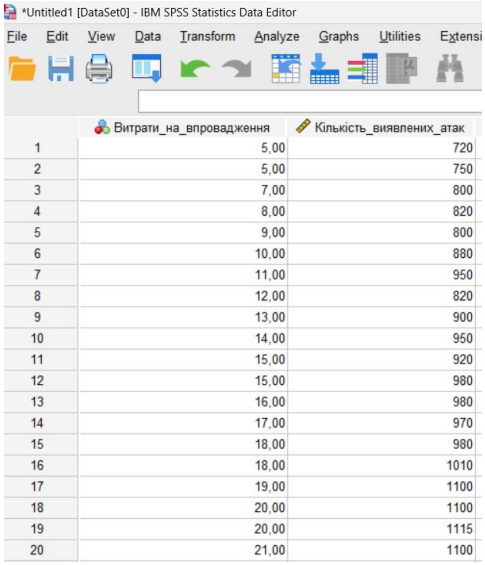
7. Серед отриманих результатів є стовпець «Коефіцієнти», що містить значення a = 625,031 в рядку «Y-перетин», b = 22,507 – в рядку «Мінлива X1». Порівняйте отримані результати з раніше розрахованими коефіцієнтами *a* і *b*.

8. Стандартна помилка, t-статистика – це допоміжні величини, які використовуються для перевірки значущості коефіцієнтів моделі.

9. P-значення – оцінка значущості коефіцієнтів моделі. Якщо P – величина менше 0,05, то з імовірністю 0,95 можна вважати, що відповідний коефіцієнт моделі значущий (тобто його не можна вважати рівним нулю і Y суттєво залежить від відповідного X).

10. Нижні і верхні 95% – довірчі інтервали для коефіцієнтів моделі.

Побудова моделі лінійної регресії за допомогою програмного забезпечення IBM SPSS Statistics



	Витрати_на_впровадження	Кількість_виявлених_атак
1	5,00	720
2	5,00	750
3	7,00	800
4	8,00	820
5	9,00	800
6	10,00	880
7	11,00	950
8	12,00	820
9	13,00	900
10	14,00	950
11	15,00	920
12	15,00	980
13	16,00	980
14	17,00	970
15	18,00	980
16	18,00	1010
17	19,00	1100
18	20,00	1100
19	20,00	1115
20	21,00	1100

Рисунок Л3.8 – Початкові дані

Для побудови моделі лінійної регресії в IBM SPSS Statistics дотримуйтесь тах інструкцій:

1. Виконайте команду Analyze/Regression/Linear (*Аналіз/Регресія/Лінійна*).

2. У діалоговому вікні змінну Кількість виявлених атак перенесіть в область *Dependent* (Залежна), а змінну Витрати на впровадження – в область *Independent* (Незалежна). Натисніть кнопку ОК (рис. 16.9).

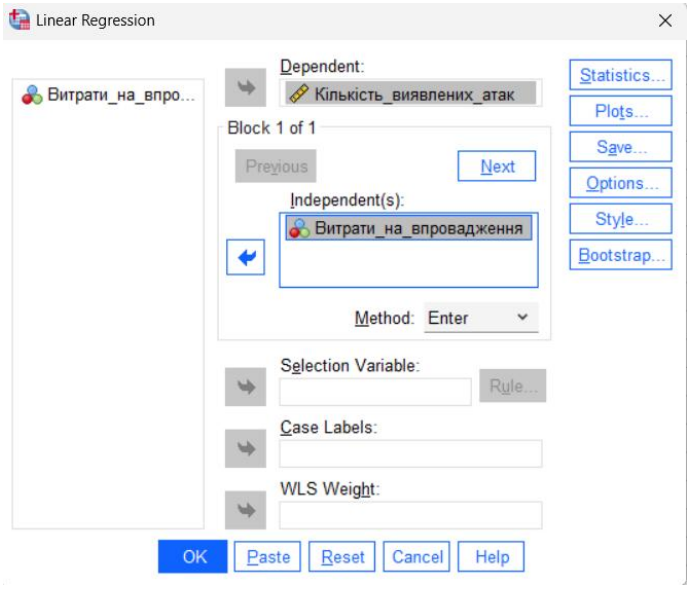


Рисунок Л3.9 – Фрагмент вікна Linear Regression.

3. Проаналізуємо отримані результати.

Перше, на що необхідно звернути увагу – це таблиця **ANOVA**, на якій представлені результати дисперсійного аналізу (рис. Л3.10). Результати цієї таблиці відповідають результатам, що отримані з використанням інструменту аналізу **Регресія** в MS Excel (таблиця *Дисперсійний аналіз*).

– В рядку **Regression** *(Регресія)* предствлена частка дисперсії, яка описується рівнянням регресії (сума квадратів, обумовлена регресією).

– В рядку **Residual** (*Залишок*) – частка дисперсії, яка не враховується при записі рівняння (залишкова сума квадратів).

– Найбільший інтерес стовить **статистична значимість Sig або значимість F** (ймовірність випадковості результату), яка повинна бути меншою або рівною 0,05. Якщо значимість менша за 0,05, то модель може вважатись

адекватною з ймовірністю 0,95. В нашому випадку ймовірність того, що отриманий випадковий результат, дуже мала ($p \leq 0{,}001$), і, відповідно, адекватність побудованої моделі наближається до 100%.

ANOVA[a]

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	249505,336	1	249505,336	175,307	<,001[b]
	Residual	25618,414	18	1423,245		
	Total	275123,750	19			

a. Dependent Variable: Кількість_виявлених_атак

b. Predictors: (Constant), Витрати_на_впровадження

Рисунок Л3.10 – Таблиця ANOVA (Дисперсійний аналіз).

Потім необхідно розглянути таблицю **Model Summary** *(Зведена таблиця по моделі)*, яка містить важливі відомості побудованої моделі (рис. Л3.11). Результати цієї таблиці відповідають результатам, отриманим за допомогою інструменту аналізу **Регресія** (таблиця *Регресійна статистика*) в MS Excel.

– Коефіцієнт кореляції **R**=0,952, що вказує на наявність дуже сильного зв'язку між досліджуваними змінними.

– Величина **R Square** (R^2) **- коефіцієнт детермінації.** В нашому випадку $R^2 = 0{,}907$ а це означає, що регресійною моделлю описано 90,7% випадків.

– Більш точне значення дає **Adjusted R Square** (*Виправлений квадрат R*).

– Третім, практично значимим показником, який визначає якість регресійної моделі, є величина стандартної похибки розрахунків **(Std. Error of the Estimate).** Порівняйте цей *показник з отриманим раніше в MS Excel.*

Model Summary

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	,952[a]	,907	,902	37,726

a. Predictors: (Constant), Витрати_на_впровадження

Рисунок Л3.11 – Таблиця **Model Summary**.

На підставі таблиць ANOVA і Model Summary можна судити про практичну придатність побудованої регресійної моделі. Враховуючи, що ANOVA показує досить високу значимість (меншу за 0,001), коефіцієнт

кореляції дорівнює 0,952, можна зробити висновок про те, що з урахуванням обмеження модель описує приблизно 90% сукупної дисперсії, тобто побудована регресійна модель є статистично значущою і практично прийнятною.

Основні практичні результати регресії містяться в таблиці Coefficients (Коефіцієнти), представлені на рис. 16.12. Результати цієї таблиці відповідають результатам, отриманим за допомогою інструменту аналізу Регресія в MS Excel.

– Стовпець *B* містить коефіцієнт регресії. Вони служать для формування регресійного рівняння, за яким можна розрахувати величину залежної змінної при різних значеннях незалежних. У нашому прикладі **B** = 22,507.

– Рядок *Constant* (Постійна) містить значення залежної змінної при нульовому значенні незалежної змінної (зміщення по осі ординат) і в нашому випадку **A** = 625,031.

– Отже, в результаті побудови регресійної моделі можна сформувати таке регресійне рівняння:

– Y = 22,507 * X +625,031

– Стовпець *Std. Error* (Стандартна помилка) , розрахована для коефіцієнта регресії B. При 95 %-вому довірчому інтервалі коефіцієнт може відхилятися від величини **B** на ± 2 * Std.Error. Це означає, що коефіцієнт **B**, рівний 22,507 в 95% випадків може відхилятися від даного значення на ± 2*1,7 або на ±3,4. Мінімальне значення коефіцієнта дорівнюватиме 22,507 - 3,4 = 19,107; а максимальне – 22,507+3,4=25,907. Таким чином, в 95 % випадків коефіцієнт регресії варіюється в межах від 19,107 до 25,907 (при середньому значенні 22,507).

Coefficients[a]

Model		Unstandardized Coefficients		Standardized Coefficients		
		B	Std. Error	Beta	t	Sig.
1	(Constant)	625,031	24,689		25,316	<,001
	Витрати_на_впровадження	22,507	1,700	,952	13,240	<,001

a. Dependent Variable: Кількість_виявлених_атак

Рисунок Л3.12 – Таблиця Coefficients

На цьому інтерпретація результатів регресійного аналізу може вважатися завершеною.

Побудова лінії регресії

1. Для графічного відображення лінії регресії побудуйте спочатку кореляційне поле способом, описаним вище.

2. Далі у вікні перегляду результатів клацніть двічі на графіку, щоб перенести його в редактор діаграм.

3. У вікні редактора діаграм виконайте команду Elements / Fit line at total (*Елементи / Лінія тренда для всього графіка*).

4. У вікні Properties (*Властивості*) на вкладці Fit Line підтвердіть попередню установку Linear Regression і клацніть по кнопці [Apply] (*Застосувати*).

Результат представлено на рисунку Л3.13.

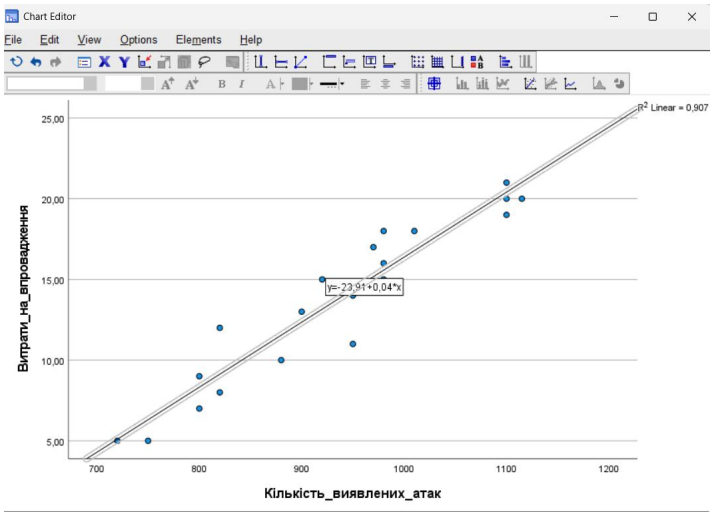


Рисунок Л3.13 – Графік регресії

Завдання

На основі даних таблиці Л3.1 та вищенаведеного алгоритму обчислень необхідно побудувати моделі лінійної регресії.

Таблиця Л3.1

Вхідні дані

№ організації	Витрати на систему захисту, тис. грн..	Кількість здійснених атак
1	10+N	15+N
2	12+N	14+N
3	13,1+N	14+N
4	14,3+N	13+N
5	13+N	13+N
6	15+N	12+N
7	16+N	11+N
8	16,3+N	14+N
9	16+N	14+N
10	15,9+N	12+N
11	16,7+N	13+N
12	17+N	14+N
13	17,4+N	13+N
14	17,2+N	12+N
15	16,8+N	10+N
16	17,6+N	9+N
17	18+N	10+N
18	19,2+N	8+N
19	18,5+N	11+N
20	18,8+N	13+N

N – порядковий номер здобувача в журналі групи.

Лабораторна робота 4. Побудова прогнозів із застосуванням методів екстраполяції тенденції.

Задача 1

Побудувати прогноз на 7-11 роки за допомогою середнього абсолютного приросту та середнього коефіцієнта росту використовуючи дані таблиці Л4.1 та матеріал теми 8. Порівняти результати.

Таблиця Л4.1

Кількість фішингових листів, на які "клікнув" користувач за рік

Номер року, t	Кількість, разів.	Номер року, t.	Кількість, разів.
1	11,7+N	7	16,9+N
2	12,5+N	8	18,2+N
3	13,2+N	9	19,1+N
4	14,4+N	10	20,8+N
5	16,0+N	11	22,2+N
6	16,0+N		

N – порядковий номер здобувача в журналі групи.

Задача 2

На основі таблиці Л4.2 та матеріалів теми 8, необхідно перевірити гіпотезу про існування тенденції у динамічному ряду за допомогою методу порівняння середніх рівнів ряду.

Таблиця Л4.2

Вхідні показники

№	x
1	20.4+N
2	20.5+N
3	20.6+N
4	20.4+N
5	20.8+N
6	21.5+N
7	21.5+N
8	21.6+N
9	22.1+N
10	22.5+N
11	22.4+N

12	22.4+N
13	22.5+N
14	22.6+N
15	22.8+N
16	22.9+N
17	22.8+N
18	22.9+N
19	30.1+N
20	31.2+N

N – порядковий номер здобувача в журналі групи.

Задача 3

На основі таблиці Л4.3 та матеріалів теми 8, необхідно побудувати прогноз на наступні 5 років за допомогою методу екстраполяції на основі плинної середньої.

Таблиця Л4.3

Кількість оновлень безпеки, проігнорованих користувачем за рік

Номер року, t	Кількість, разів.
1	11,7+N
2	12,5+N
3	13,2+N
4	14,4+N
5	16,0+N

N – порядковий номер здобувача в журналі групи.

2. Використовуючи таблицю Л4.4 та матеріали теми 8, побудувати прогноз обсягу випуску оновлень антивірусних баз на наступний рік за допомогою індексу сезонності, враховуючи, що на наступний рік ІТ-фірма передбачає реалізувати 108+(N*48) оновлень.

Таблиця Л4.4

Обсяг випуску оновлень антивірусних баз

Рік / місяць	1-й рік	2-й рік	3-й рік	4-й рік	Разом за 4 роки	В середньому за 4 роки	Індекс сезонності	Прогноз обсягу випуску на наступний рік
01	6,3+N	6,4+N	6,5+N	7,4+N				
02	7,4+N	6,6+N	6,7+N	7,7+N				

03	8,2+N	7+N	6,9+N	7,9+N				
04	9,4+N	7,6+N	7,7+N	8,3+N				
05	8,0+N	8,2+N	8,5+N	8,7+N				
06	8,5+N	8,7+N	9,0+N	9,2+N				
07	9,0+N	9,1+N	10,5+N	10,7+N				
08	9,5+N	9,6+N	11,8+N	11,1+N				
09	9,9+N	10,0+N	10,2+N	10,5+N				
10	9,3+N	9,5+N	10,0+N	10,1+N				
11	9,0+N	9,3+N	9,6+N	9,4+N				
12	8,5+N	0,9+N	9,3+N	9+N				
Разом								
В середньому								

N – порядковий номер здобувача в журналі групи.

Лабораторна робота 5. Побудова регресійних моделей для прогнозування багатовимірних процесів.

На розвиток кібербезпеки певного регіону витрачається 2 види ресурсів: трудові (тис. ІТ-фахівців) і основні виробничі фонди (млн грн). Варіанти вхідних даних про коефіцієнт розвитку кібербезпеки в регіоні та витрати ресурсів за 8 останніх років наведені по варіантах у таблиці Л5.1.

Таблиця Л5.1

Коефіцієнт розвитку кібербезпеки (Y) та затрати ресурсів: праці (L) і основних виробничих фондів (K)

Варіант 1								
Змінна	Номер спостереження, рік							
	1	2	3	4	5	6	7	8
Y	51	52	57	65	70	65	60	62
K	29	32	35	37	41	40	40	41
L	43	46	50	50	63	64	64	63
Варіант 2								
Змінна	Номер спостереження, рік							
	1	2	3	4	5	6	7	8
Y	42	43	45	47	51	55	60	62
K	30	32	35	37	42	45	48	51
L	43	46	50	51	63	65	68	70
Варіант 3								
Змінна	Номер спостереження, рік							
	1	2	3	4	5	6	7	8
Y	42	43	45	47	51	55	60	62
K	30	32	35	37	42	45	48	51
L	43	46	50	50	63	64	64	63
Варіант 4								
Змінна	Номер спостереження, рік							
	1	2	3	4	5	6	7	8
Y	51	52	57	65	70	65	60	62
K	30	33	36	40	42	48	50	52
L	40	42	51	52	60	62	62	64
Варіант 5								
Змінна	Номер спостереження, рік							
	1	2	3	4	5	6	7	8

Y	42	45	47	50	52	53	60	63
K	19	28	30	33	34	38	42	45
L	40	43	52	52	56	60	62	69
Варіант 6								
Змінна	Номер спостереження, рік							
	1	2	3	4	5	6	7	8
Y	51	52	57	65	70	65	66	70
K	29	32	35	37	41	40	40	41
L	43	46	50	50	63	64	64	63
Варіант 7								
Змінна	Номер спостереження, рік							
	1	2	3	4	5	6	7	8
Y	51	52	55	63	68	65	66	68
K	29	33	36	40	42	48	50	52
L	42	46	52	55	60	62	64	67
Варіант 8								
Змінна	Номер спостереження, рік							
	1	2	3	4	5	6	7	8
Y	51	52	55	63	68	65	66	68
K	20	32	35	37	41	40	40	41
L	40	42	51	52	60	62	62	64
Варіант 9								
Змінна	Номер спостереження, рік							
	1	2	3	4	5	6	7	8
Y	51	52	57	65	70	65	68	71
K	25	34	35	40	41	42	45	47
L	43	46	50	50	63	64	64	63
Варіант 10								
Змінна	Номер спостереження, рік							
	1	2	3	4	5	6	7	8
Y	50	52	55	60	66	68	68	70
K	26	32	35	37	41	40	40	41
L	38	42	51	52	60	62	62	64
Варіант 11								
Змінна	Номер спостереження, рік							
	1	2	3	4	5	6	7	8
Y	51	52	57	65	65	65	66	67
K	29	32	33	37	39	40	42	45
L	38	42	51	52	60	62	62	64
Варіант 12								
Змінна	Номер спостереження, рік							

	1	2	3	4	5	6	7	8
Y	38	42	45	49	52	55	60	62
K	31	32	36	40	41	42	45	48
L	43	46	50	50	63	64	64	63
Варіант 13								
Змінна	Номер спостереження, рік							
	1	2	3	4	5	6	7	8
Y	51	52	57	65	70	65	60	62
K	29	32	35	37	41	42	43	45
L	43	46	50	50	63	64	65	68
Варіант 14								
Змінна	Номер спостереження, рік							
	1	2	3	4	5	6	7	8
Y	50	52	60	66	67	68	69	71
K	20	25	29	25	39	40	42	43
L	40	42	43	48	52	56	60	63
Варіант 15								
Змінна	Номер спостереження, рік							
	1	2	3	4	5	6	7	8
Y	49	50	52	55	59	63	65	66
K	29	32	35	37	37	39	40	41
L	41	45	48	50	55	57	64	65
Варіант 16								
Змінна	Номер спостереження, рік							
	1	2	3	4	5	6	7	8
Y	48	52	56	63	68	70	72	75
K	28	30	32	36	41	42	43	44
L	42	45	49	50	52	55	58	60
Варіант 17								
Змінна	Номер спостереження, рік							
	1	2	3	4	5	6	7	8
Y	48	52	57	65	70	71	70	72
K	28	33	35	37	41	42	45	46
L	43	46	50	55	60	62	63	65
Варіант 18								
Змінна	Номер спостереження, рік							
	1	2	3	4	5	6	7	8
Y	50	52	57	65	70	65	60	62
K	28	33	35	37	41	42	45	46
L	42	44	50	52	58	64	66	67
Варіант 19								

Змінна	Номер спостереження, рік							
	1	2	3	4	5	6	7	8
Y	51	53	60	65	68	69	70	72
K	29	32	35	37	41	42	43	45
L	43	46	50	50	63	64	64	63

Здобувача під час виконання лабораторної роботи необхідно:

1) побудувати три запропоновані моделі;

2) розрахувати параметри кожної моделі;

3) вибрати адекватну виробничу функцію (формулу математичної залежності між об'ємом випуску продукції і затратами ресурсів);

4) висвітлити змістовну інтерпретацію одержаних результатів, щодо вибору кращої моделі.

Хід розв'язання завдання

Необхідно внести дані для побудови адекватної моделі та розрахунку її параметрів у програмне забезпечення *Statistica*:

Таблиця Л5.2

Початкові дані

	Y	K	L
1	51	29	43
2	52	32	46
3	57	35	50
4	65	37	50
5	70	41	63
6	65	40	64
7	60	40	64
8	62	41	63

Потрібно побудувати три моделі: лінійну модель Кобба-Дугласа, нелінійну модель Кобба-Дугласа та однофакторну модель.

Спочатку необхідно побудувати лінійну модель Кобба-Дугласа ($Y=A+\alpha K+\beta L$).

Обчислення здійснюється в модулі Multiple regression. Будуть одержані такі дані:

Таблиця Л5.3

Параметри класичної моделі Кобба-Дугласа

№	Beta	Std.Err.	B	Std.Err.	t(5)	p-level
Intercept			5,517639	12,58066	0,43858	0,679268
K	1,587116	0,626207	2,328375	0,91868	2,53449	0,052243
L	-0,761176	0,626207	-0,562104	0,46243	-1,21553	0,278419

Згідно з таблицею Л5.3, виробнича функція матиме такий вигляд: Y=5,52+2,33*K-0,56*L. При цьому коефіцієнт детермінації дорівнює 0,81 (це означає, що зміна K та L на 81% залежить від них самих, а на 19% – від зміни зовнішніх факторів). Коефіцієнт парної кореляції дорівнює 0,89 і це означає, що зв'язок між двома факторами є досить тісний. В моделі нема жодного значущого параметра. Середнє відхилення – 0.

Наступним кроком є побудова нелінійної (класичної) моделі Кобба-Дугласа ($\mathbf{Y=A*K^{\alpha}*L^{\beta}}$). Розрахунок здійснюється в модулі Nonlinear estimation. Будуть одержані такі дані:

Таблиця Л5.4

Параметри нелінійної моделі Кобба-Дугласа

№	Estimate	Standard	t-value	p-level	Lo. Conf	Up. Conf
A	2,805630	2,005835	1,39873	0,220760	-2,35053	7,961793
B	1,416034	0,559965	2,52879	0,052608	-0,02340	2,855471
C	-0,508798	0,414850	-1,22646	0,274627	-1,57520	0,557608

Функція матиме вигляд: Y = 2,8*K^1,4*L^-0,5

При цьому коефіцієнт детермінації дорівнює 0,80 (це означає, що зміна K та L на 80% залежить від них самих, а на 20% – від зміни зовнішніх факторів). Коефіцієнт парної кореляції рівний 0,89 і це означає, що зв'язок між двома факторами є досить тісний. В моделі нема жодного значущого параметра. Середнє відхилення – 0,005

Наступним кроком є побудова однофакторної моделі (**$y=A*k^{\alpha}$**). Для цього необхідно розрахувати значення y та k. y = Y/L, k = K/L. Будуть отримані наступні дані:

Таблиця Л5.5

Значення y та k

№	y	k
1	1,18604651	0,674418605
2	1,13043478	0,695652174
3	1,14	0,7
4	1,3	0,74
5	1,11111111	0,650793651
6	1,015625	0,625
7	0,9375	0,625
8	0,984126984	0,650793651

Обчислення параметрів моделі:

Таблиця 18.6

Параметри однофакторної моделі

№	Estimate	Standard	t-value	p-level	Lo. Conf	Up. Conf
A	2,071281	0,266119	7,783293	0,000237	1,420112	2,722451
B	1,583697	0,324602	4,878886	0,002769	0,789424	2,377969

Виробнича функція матиме вигляд:

y=2,07*k^1,58

При цьому коефіцієнт детермінації дорівнює 0,79 (це означає, що зміна K та L на 79% залежить від них самих, а на 21% від зміни зовнішніх факторів). Коефіцієнт парної кореляції дорівнює 0,89 і це означає, що зв'язок між двома факторами є досить тісний. В моделі два значущих параметри. Середнє відхилення – 0.

При виборі адекватної виробничої функції, враховують коефіцієнт детермінації. Однак у всіх моделях ці коефіцієнти приблизно однакові. Тому наступним фактором вибору адекватної виробничої функції буде кількість

значущих параметрів. В однофакторній моделі два значущих параметри з двох, а в інших моделях – жодного. Ще слід зважати на середні відхилення. Проте у нашому прикладі всі вони приблизно однакові у всіх моделях.

Отже, адекватною виробничою функцією буде однофакторна функція виду: y=2,07*k^1,58.

Лабораторна робота 6. Аналіз часових рядів та побудова прогнозу за допомогою непараметричних методів.

У деяких випадках можна зустрітися з такими явищами, які не піддаються вираженню числом одиниць. У цьому випадку вдаються до непараметричних методів, які дають змогу виміряти інтенсивність зв'язку як між кількісними ознаками, форма розподілу яких відрізняється від нормальної, так і між якісними ознаками.

В основу непараметричних методів покладено принцип нумерації значень статистичного ряду. Кожній одиниці сукупності присвоюється порядковий номер у ряді, який буде впорядкований за рівнем ознаки. Таким чином, ряд значень ранжується, а номер кожної окремої одиниці буде її рангом.

Можна отримати попереднє уявлення про наявність зв'язку між ознаками, якщо зіставити послідовність взаємного розташування рангів факторної та результативної ознаки. Для цього ранги індивідуальних значень факторної ознаки розташовують у порядку зростання, і якщо ранги результативної ознаки виявляють тенденцію до збільшення, можна припускати наявність прямого зв'язку. Якщо ж із збільшенням рангів факторної ознаки, ранги результативної ознаки зменшуються, то це говорить про можливу наявність між досліджуваними ознаками зворотного зв'язку.

Для проведення непараметричного аналізу використовують:

1. Коефіцієнт рангової кореляції Спірмена.

2. Коефіцієнт рангової кореляції Кендела.

3. Коефіцієнт конкордації ω.

1. Коефіцієнт рангової кореляції Спірмена є непараметричним аналогом коефіцієнта кореляції Пірсона і заснований на розгляді різниці рангів значень факторної та результативної ознаки. Формула коефіцієнта рангової кореляції Спірмена має вигляд:

$$\rho = 1 - \frac{6 \cdot \sum d_i^2}{n(n^2 - 1)}, \qquad (Л6.1)$$

де $d_i = x_i - y_i$, $i = \overline{1,n}$ - різниці між рангами змінних *X* та *Y*;

n – кількість признаків.

2. Коефіцієнт рангової кореляції Кендела також є мірою зв'язку між змінними *X* та *Y*. Формула коефіцієнта рангової кореляції Спірмена має вигляд:

$$\tau = \frac{2S}{n(n-1)}, \qquad (Л6.2)$$

де

$$S = P + Q. \qquad (Л6.3)$$

Для обчислення потрібно впорядкувати ряд рангів змінної *X*, привівши його до ряду натуральних чисел. Потім розглядають послідовність рангів змінної *Y*.

P – це міра відповідності послідовності рангів змінної *Y* послідовності рангів змінної *X*. При визначенні доданка *P* треба встановити , скільки числі , що розташовані праворуч від кожного з елементів послідовності рангів змінної *Y*, має величину рангу, що перевищує ранг елемента, який розглядається.

Q – це міра невідповідності послідовності рангів змінної *Y* послідовності рангів змінної *X*. Для визначення *Q* підраховують, скільки чисел, що розташовані праворуч від кожного з членів послідовності рангів змінної Y мають ранг менший за одиницю. Такі величини беруться зі знаком мінус.

При досить великій кількості спостережень між коефіцієнтами кореляції Спірмена і Кендела існує співвідношення (19.4):

$$\rho = \frac{3}{2}\tau \qquad (19.4).$$

Завдання. Використання непараметричних методів аналізу

Умова задачі: Експерти аналітичного центру оцінювали рівень стійкості інформаційних систем компаній на етапі впровадження заходів кіберзахисту таким чином (таблиця Л6.1):

Таблиця Л6.1

№ депутата	1	2	3	4	5	6	7	8	9	10
Ранг компаній за результатами оцінки експертів рівня стійкості інформаційних систем	7	4	1	3	10	5	9	2	8	6
Ранг компаній за фактичними показниками кіберзахищеності після впровадження заходів	5	6	2	7	8	3	10	1	9	4

Задача 1

Обчислити коефіцієнт рангової кореляції Спірмена за формулою (Л6.1) в MS Excel.

Приклад

1. Відкрийте нову книгу MS Excel та створіть таблицю згідно з рис. Л6.1.
2. **Лист 1** перейменуйте в **Спірмен** і збережіть під іменем *Непараметр_аналіз_Ваше прізвище.xls*.

	A	B	C	D	E
1	**№ компанії**	**Ранг компаній за результатами оцінки експертів рівня стійкості інформаційних систем**	**Ранг компаній за фактичними показниками кіберзахищеності після впровадження заходів**	**Різниця рангів |di|**	**|di|^2**
2	1				
3	2				
4	3				
5	4				
6	5				
7	6				
8	7				
9	8				
10	9				
11	10				
12	**Всього**				
13	**Коефіцієнт рангової кореляції Спірмера**				

Рисунок Л6.1 – Дані для задачі

3. В діапазоні клітинок **D2:D11** обчисліть абсолютну різницю рангів $|d_i|$ з використанням математичної функції **ABS.**

4. В діапазоні клітинок **E2:E11** обчисліть квадрат різниці рангів d_i^2.

5. В клітинці **E12** обчисліть $\sum d_i^2$.

6. В клітинці **E13** обчисліть коефіцієнт рангової кореляції Спірмена за формулою (19.1).

13	**Коефіцієнт рангової кореляції Спірмена**	**0,758**

Рисунок Л6.2 – Результат обчислення коефіцієнта

7. Зробіть висновок про напрям і тісноту зв'язку між досліджуваними ознаками.

Задача 2

Обчислити коефіцієнт рангової кореляції Кендалла за формулою (Л6.2) в MS Excel.

Приклад

1. Перейдіть на новий аркуш книги *Непараметр_аналіз_Ваше прізвище.xls* і назвіть його **Кендалл**.

2. Скопіюйте на цей лист дані з діапазону комірок **A1:C11** листа **Спірмен**.

3. Виділіть цей діапазон комірок і проведіть його сортування за стовпцем **B** (ранжування за змінною *X*). Доповніть лист даними згідно з рис. Л6.3.

	A	B	C
1	№ компанії	Ранг компаній за результатами оцінки експертів рівня стійкості інформаційних систем (X)	Ранг компаній за фактичними показниками кіберзахищеності після впровадження заходів (Y)
2	3	1	2
3	8	2	1
4	4	3	7
5	2	4	6
6	6	5	3
7	10	6	4
8	1	7	5
9	9	8	9
10	7	9	10
11	5	10	8
12			
13	P		
14	Q		
15	S		
16	Коефіцієнт рангової кореляції Кендала		

Рисунок Л6.3 – Дані для задачі

4. Обчисліть значення P – ступінь відповідності послідовності рангів змінної Y послідовності рангів змінної X. Наприклад, першому значенню в послідовності рангів змінної Y, тобто числу 2 відповідає вісім чисел (7, 6, 3, 4, 5, 9, 10, 8), які перевищують ранг 2; другому значенню 1 відповідає також вісім чисел (7, 6 , 3, 4, 5, 9, 10, 8), що перевищують ранг 1; третьому значенню 7 відповідають три числа (9, 10, 8) і т.д. Підсумовуючи таким чином отримані числа (8+8+3+.....) отримаєте значення P в клітинці C13 (рис. Л6.4).

5. Обчисліть значення Q – ступінь невідповідності послідовності рангів змінної Y послідовності рангів змінної X. Наприклад, першому значенню 2 послідовності рангів змінної Y, відповідає тільки одне число 1, яке менше рангу 2; другому значенню 1 відповідає 0 чисел, ранг яких менше 1; третьому значенню 7 відповідають чотири числа (6, 3, 4, 5), ранг яких менше 7 і т.д. Підсумовуючи таким чином отримані числа зі знаком мінус (-1 - 0 - 4 - ...), отримуємо значення Q в клітинці C14 (рис. Л6.4).

6 . В клітинці C15 обчисліть значення S (рис. Л6.3).

7. В клітинці C16 обчисліть значення коефіцієнта рангової кореляції Кендалла (рис. Л6.4).

13	P		35
14	Q		-10
15	S		25
16	Коефіцієнт рангової кореляції Кендала		0,556

Рисунок Л6.4 – Обчислення рангового коефіцієнта кореляції Кендела

8. Порівняйте коефіцієнти Спірмена та Кендалла, використовуючи формулу (Л6.4), та наведіть ***змістовні висновки***.

Задача 3

Провести непараметричний аналіз за допомогою програмного забезпечення IBM SPSS Statistics.

Приклад

1. В програмному забезпеченні IBM SPSS Statistics SPSS введіть дані в режимі Data View (*Перегляд даних*) згідно з рис. Л6.5 Збережіть у своїй папці під ім'ям Непараметричний_Аналіз_Ваше прізвище.sav (дані можна також скопіювати з файлу MS Excel).

2. Перейдіть в режим Variable View (Перегляд змінних) і для змінних R1 і R2 задайте Label (Мітка) Ранг за оцінкою експерта і Ранг за отриманими голосами відповідно.

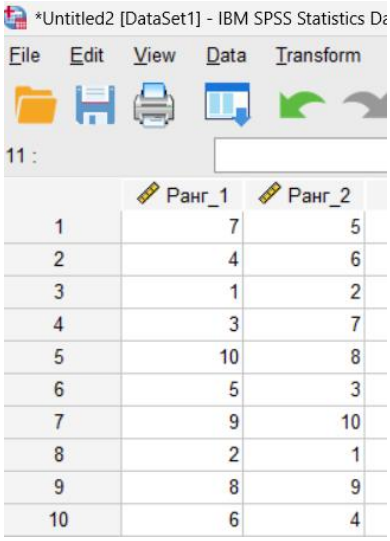


Рисунок Л6.5 – дані в режимі Data View.

3. Виконайте команду Analyze / Correlate / Bivariate (*Аналіз / Кореляція / Парні*).

4. У діалоговому вікні в область змінних перетягніть обидві змінні, вкажіть коефіцієнт кореляції Спірмена і Кендалла, рівень значимості – двосторонній (рис. Л6.6).

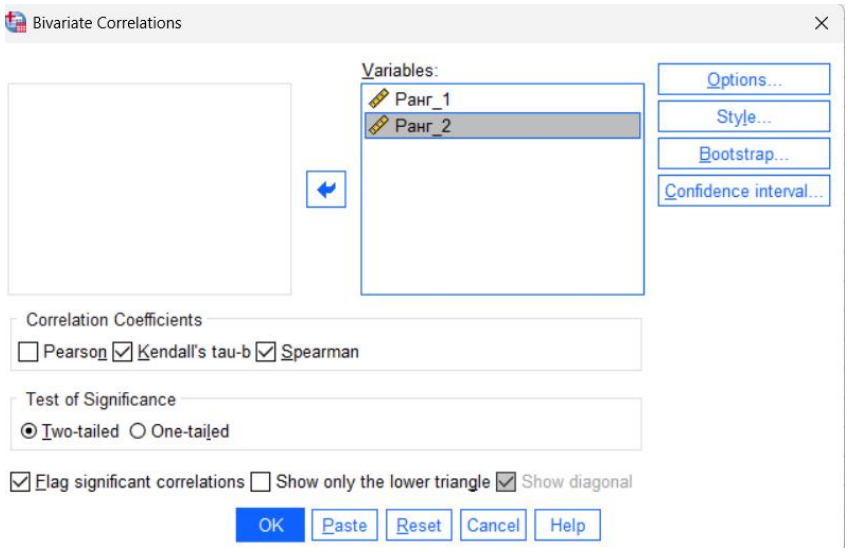


Рисунок Л6.6 – Діалогове вікно кореляції

5. Проаналізуйте отриманий результат (рис. Л6.7) аналогічно попередньому завданню. Порівняйте отримані результати з коефіцієнтами кореляції Спірмена і Кендалла, розрахованими за формулами в MS Excel. Надайте ***змістовні висновки***.

6. Збережіть звіт у своїй папці під ім'ям Непараметр_Аналіз_ Ваше прізвище.spo.

Correlations

			Ранг_1	Ранг_2
Kendall's tau_b	Ранг_1	Correlation Coefficient	1,000	,556*
		Sig. (2-tailed)	.	,025
		N	10	10
	Ранг_2	Correlation Coefficient	,556*	1,000
		Sig. (2-tailed)	,025	.
		N	10	10
Spearman's rho	Ранг_1	Correlation Coefficient	1,000	,758*
		Sig. (2-tailed)	.	,011
		N	10	10
	Ранг_2	Correlation Coefficient	,758*	1,000
		Sig. (2-tailed)	,011	.
		N	10	10

*. Correlation is significant at the 0.05 level (2-tailed).

Рисунок Л6.7 – Звіт по кореляції Спірмена і Кендалла.

У більшості випадків рекомендується застосовувати коефіцієнт кореляції Спірмена. Використання коефіцієнта Кендалла виправдане тільки в тому випадку, коли в структурі даних наявні викиди.

Варіанти завдань наведено в таблиці Л6.2.

Таблиця Л6.2

Варіанти початкових даних

Варіант 1										
№ компанії	1	2	3	4	5	6	7	8	9	10
Ранг компаній за результатами оцінки експертів рівня стійкості інформаційних систем	7	4	1	3	10	5	9	2	8	6
Ранг компаній за фактичними показниками кіберзахищеності після впровадження заходів	5	6	2	7	8	3	10	1	9	4
Варіант 2										
№ компанії	1	2	3	4	5	6	7	8	9	10
Ранг компаній за результатами оцінки експертів рівня стійкості інформаційних систем	1	3	10	5	7	4	6	8	9	2

Ранг компаній за фактичними показниками кіберзахищеності після впровадження заходів	2	7	8	3	5	6	4	9	10	1
Варіант 3										
№ компанії	1	2	3	4	5	6	7	8	9	10
Ранг компаній за результатами оцінки експертів рівня стійкості інформаційних систем	3	1	6	4	2	7	10	5	9	8
Ранг компаній за фактичними показниками кіберзахищеності після впровадження заходів	7	2	4	6	1	5	8	3	10	9
Варіант 4										
№ компанії	1	2	3	4	5	6	7	8	9	10
Ранг компаній за результатами оцінки експертів рівня стійкості інформаційних систем	10	6	2	4	1	9	3	7	5	8
Ранг компаній за фактичними показниками кіберзахищеності після впровадження заходів	8	4	1	6	2	10	7	5	3	9
Варіант 5										
№ компанії	1	2	3	4	5	6	7	8	9	10
Ранг компаній за результатами оцінки експертів рівня стійкості інформаційних систем	6	10	4	1	9	3	2	8	5	7
Ранг компаній за фактичними показниками кіберзахищеності після впровадження заходів	4	8	6	1	10	7	2	9	3	5
Варіант 6										
№ компанії	1	2	3	4	5	6	7	8	9	10
Ранг компаній за результатами оцінки експертів рівня стійкості інформаційних систем	7	4	1	3	6	10	2	5	9	8
Ранг компаній за фактичними показниками кіберзахищеності після впровадження заходів	5	7	1	6	4	8	2	3	10	9

Варіант 7										
№ компанії	1	2	3	4	5	6	7	8	9	10
Ранг компаній за результатами оцінки експертів рівня стійкості інформаційних систем	1	6	2	3	8	7	10	4	9	5
Ранг компаній за фактичними показниками кіберзахищеності після впровадження заходів	1	4	2	6	9	5	8	7	10	3
Варіант 8										
№ компанії	1	2	3	4	5	6	7	8	9	10
Ранг компаній за результатами оцінки експертів рівня стійкості інформаційних систем	2	7	6	1	4	8	9	3	10	5
Ранг компаній за фактичними показниками кіберзахищеності після впровадження заходів	2	5	4	1	7	9	10	6	8	3
Варіант 9										
№ компанії	1	2	3	4	5	6	7	8	9	10
Ранг компаній за результатами оцінки експертів рівня стійкості інформаційних систем	1	10	8	3	5	2	7	4	9	6
Ранг компаній за фактичними показниками кіберзахищеності після впровадження заходів	1	8	9	6	4	2	5	7	10	3
Варіант10										
№ компанії	1	2	3	4	5	6	7	8	9	10
Ранг компаній за результатами оцінки експертів рівня стійкості інформаційних систем	5	10	1	3	6	2	7	8	9	4
Ранг компаній за фактичними показниками кіберзахищеності після впровадження заходів	4	9	1	6	3	2	9	5	8	7

Лабораторна робота 7. Експертні методи прогнозування

За даними таблиці Л7.1, в якій відображають висновки п’яти експертів про ранги десяти факторних ознак, здобувачам необхідно:

1. Проаналізувати за допомогою рангових коефіцієнтів кореляції Спірмена і Кендалла висновки експертів: а) перших двох; б) третього й четвертого; в) останніх двох.

2. Оцінити ступінь узгодженості думок усіх 5 експертів про вплив різних факторних ознак на величину результативної ознаки.

3. Впорядкувати фактори за спаданням сумарних рангів. Порівняти отримані рангові коефіцієнти і навести змістовні висновки.

Варіанти завдань 1-4.

У таблиці Л7.1 наведені експертні оцінки (ранги) якісних факторів, що найбільш суттєво впливають на ефективність системи кіберзахисту компанії: своєчасне виявлення кіберзагроз, наявність актуальних оновлень програмного забезпечення, регулярність резервного копіювання даних, оперативність реагування на інциденти, кваліфікація фахівців із безпеки, культура дотримання політик інформаційної безпеки, розташування дата-центру, достатність кадрового забезпечення, проведення навчань і тренінгів, надання користувачам інструментів для безпечної роботи.

Таблиця Л7.1

Варіанти завдань 1-4

Варіант 1										
Експерти	Фактори									
	1	2	3	4	5	6	7	8	9	10
1	10	9	2	8	7	4	1	3	6	5
2	10	9	8	5	4	3	1	2	7	6
3	10	9	7	5	4	6	2	1	3	8
4	10	9	4	7	8	3	1	2	5	6
5	9	8	2	7	5	4	3	6	10	1
Варіант 2										
Експерти	Фактори									
	1	2	3	4	5	6	7	8	9	10
1	10	4	8	5	3	9	2	6	1	7

2	5	1	8	3	6	2	10	9	4	7
3	1	5	2	7	10	8	3	6	9	4
4	7	2	5	10	9	1	4	8	3	6
5	10	5	8	1	4	6	2	9	7	3
Варіант 3										
Експерти	Фактори									
	1	2	3	4	5	6	7	8	9	10
1	1	6	2	7	3	8	4	9	5	10
2	2	6	4	10	9	5	3	8	7	1
3	3	10	4	1	7	9	5	2	8	6
4	10	4	2	9	8	7	3	5	6	1
5	1	7	2	8	4	9	5	3	10	6
Варіант 4										
Експерти	Фактори									
	1	2	3	4	5	6	7	8	9	10
1	10	8	5	3	2	9	6	4	7	1
2	10	9	5	4	1	2	3	6	7	8
3	4	1	8	5	9	2	6	10	7	3
4	4	10	5	1	8	2	6	9	7	3
5	9	8	3	7	1	2	6	10	4	5

Варіанти завдань 5-7.

Отримані результати ранжування наступних факторів (табл. Л7.2), які знижують ефективність системи кіберзахисту компанії: нерегулярне оновлення засобів безпеки; недостатній моніторинг актуальних кіберзагроз; робота команди безпеки у скороченому складі протягом місяця; слабка взаємодія з зовнішніми експертними організаціями; недостатня інформаційна політика щодо підвищення обізнаності користувачів; нестача ресурсів для організації віддаленого моніторингу; проведення аудиту лише в окремих підрозділах; недоліки в роботі відділу, що відповідає за безпеку онлайн-сервісів; невиконання пріоритетів захисту найбільш критичних активів; неправильний розподіл спеціалізованих засобів кіберзахисту.

Таблиця Л7.2

Варіанти завдань 5-7

Варіант 5										
Експерти	Фактори									
	1	2	3	4	5	6	7	8	9	10
1	10	9	2	8	7	4	1	3	6	5

2	10	9	8	5	4	3	1	2	7	6
3	10	9	7	5	4	6	2	1	3	8
4	10	9	4	7	8	3	1	2	5	6
5	9	8	2	7	5	4	3	6	10	1
Варіант 6										
Експерти	Фактори									
	1	2	3	4	5	6	7	8	9	10
1	1	6	2	7	8	3	10	9	4	5
2	4	9	5	1	10	2	8	6	7	3
3	4	8	2	5	10	3	7	9	6	1
4	4	1	8	10	5	2	9	6	7	3
5	7	4	5	1	10	6	8	2	9	3
7										
Експерти	Фактори									
	1	2	3	4	5	6	7	8	9	10
1	10	9	2	8	7	4	1	3	6	5
2	1	4	7	5	10	9	2	6	8	3
3	1	2	5	10	3	6	8	4	9	7
4	1	9	4	7	8	3	10	2	5	6
5	1	8	4	7	5	2	3	6	10	9

Варіанти завдань 8-10.

Відомі ранги експертів (табл. Л7.3) щодо визначення впливу таких факторів на ефективність функціонування системи кіберзахисту в організації: режим роботи служби безпеки (цілодобовий моніторинг); різноманітність застосованих засобів захисту; якість налаштування й оновлення безпекових систем; наявність спеціалізованих засобів для окремих категорій даних (наприклад, конфіденційних чи персональних); якість реагування на інциденти; стан інфраструктури та безпечності середовища (мережевої й фізичної); швидкість обробки та блокування загроз; наявність резервних засобів захисту; організація захисту віддаленого доступу; можливість попереднього налаштування та автоматизації політик безпеки.

Таблиця Л7.3

Варіанти завдань 8-10

Варіант 8										
Експерти	Фактори									
	1	2	3	4	5	6	7	8	9	10

1	10	7	4	6	9	1	5	2	8	3
2	1	4	7	9	10	2	5	8	3	6
3	10	3	9	6	1	7	4	8	5	2
4	10	4	3	9	5	2	6	7	8	1
5	7	3	10	4	8	1	6	2	9	5
Варіант 9										
Експерти	Фактори									
	1	2	3	4	5	6	7	8	9	10
1	1	5	10	6	2	9	7	3	8	4
2	1	8	6	10	2	4	9	7	3	5
3	2	8	4	7	3	10	9	5	1	6
4	1	3	7	6	10	9	4	2	8	5
5	6	10	4	5	2	8	7	3	9	1
Варіант 10										
Експерти	Фактори									
	1	2	3	4	5	6	7	8	9	10
1	6	4	3	9	7	2	10	8	5	1
2	4	10	3	5	8	1	9	6	2	7
3	5	4	1	10	6	2	9	7	3	8
4	5	4	6	8	1	10	7	2	9	3
5	6	5	1	8	2	7	9	3	10	4

Поставлені завдання необхідно вирішувати в програмному забезпеченні *Statistica*. Для цього введемо початкові дані (рис. Л7.1).

	1 Експерт 1	2 Експерт 2	3 Експерт 3	4 Експерт 4	5 Експерт 5
1	1	1	1	2	1
2	3	2	5	4	7
3	10	7	8	10	8
4	4	5	2	3	6
5	2	4	3	4	2
6	7	9	7	7	10
7	5	3	4	1	5
8	9	6	6	9	9
9	8	8	9	5	3
10	6	10	10	8	4

Рисунок Л7.1 – Початкові дані

Щоб проаналізувати єдність думок експертів за допомогою коефіцієнтів кореляції Спірмена необхідно використати модуль системи *Statistics-Nonparametric statistics/Аналіз-Непараметрична статистика* (рис. Л7.2).

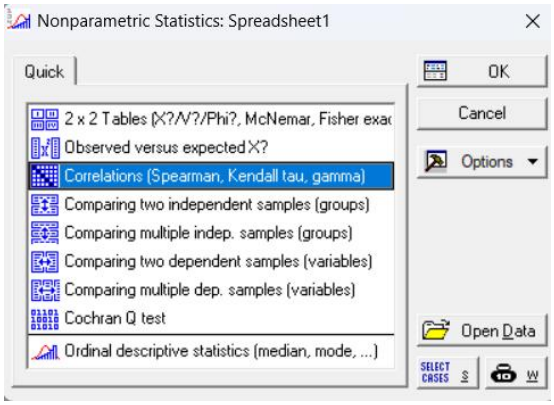


Рисунок Л7.2 – Модуль «Nonparametric statistics»

У даному вікні необідно вибрати меню «*Correlations Spearman, Kendall tau, gamma*» та натиснути кнопку ОК.

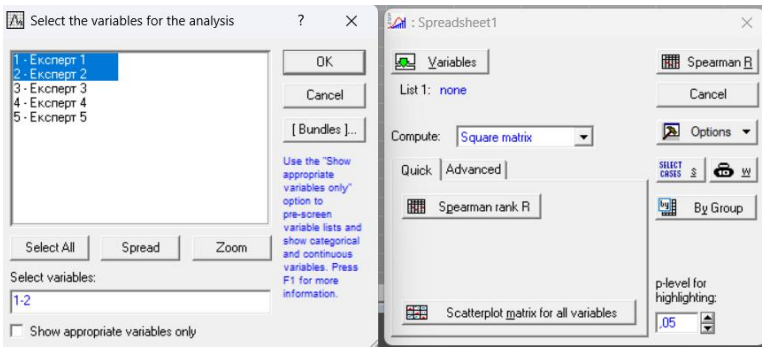
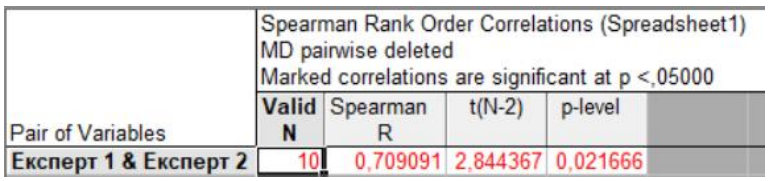


Рисунок Л7.3 – Вибір змінних та аналіз

З даного рисунка видно, що для того, щоб проаналізувати висновки перших двох експертів за допомогою рангового коефіцієнта Спірмена, потрібно вибрати перших два експерти та натиснути кнопку *Spearman* R, щоб проаналізувати висновки перших двох експертів за допомогою рангового коефіцієнта Кендалла – натиснути на кнопки *Kendall tau, gamma*.

Щоб пояснити отримані коефіцієнти, потрібно скористатись таблицею Чедока (табл. Л7.4).

На рис. Л7.4 показано тісноту зв’язку перших двох експертів розраховану за допомогою рангового коефіцієнта Спірмена.



Spearman Rank Order Correlations (Spreadsheet1)
MD pairwise deleted
Marked correlations are significant at p <,05000

Pair of Variables	Valid N	Spearman R	t(N-2)	p-level
Експерт 1 & Експерт 2	10	0,709091	2,844367	0,021666

Рисунок Л7.4 – Тіснота зв’язку між першим і другим експертами

Для перших двох експертів коефіцієнт Спірмена дорівнює 0,71. Згідно таблицею Чедока це означає, що ступінь узгодженості між цими двома експертами є високим. Рівень похибки цього параметра дорівнює 0,0216.

Для оцінки ступень узгодженості думок п’яти експертів потрібно використати коефіцієнт конкордації. Для змістовної інтерпретації цього коефіцієнта можна також використати таблицю Чедока (табл. Л7.4).

Таблиця Л7.4

Таблиця Чедока

Величина коефіцієнта	0,1-0,3	0,3-0,5	0,5-0,7	0,7-0,9	0,9-1,0
Ступінь узгодженості думок експертів	слабка	помірна	помітна	висока	дуже висока

Для обрахунку коефіцієнта конкордації потрібно скористатись меню *Statistics-Nonparametric statistics/Аналіз-Непараметрична статистика* і вибрати в даному меню «Comparing multiple dep. samples» (рис. Л7.5).

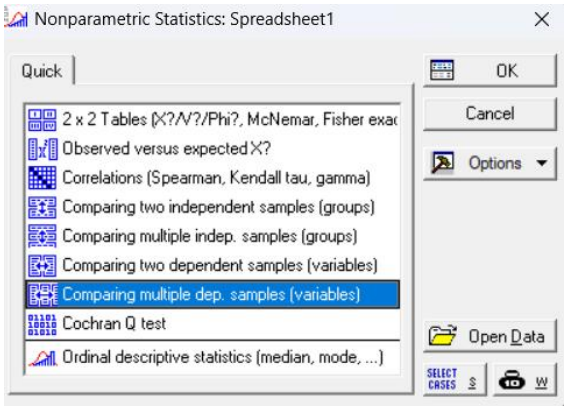


Рисунок Л7.5 – Модуль «Nonparametric statistics»

В наступному вікні необхідно вибрати усіх п’ять експертів і натиснути на кнопку «*Summary: Friedman ANOVA & Kendalls concordance*» (рис. Л7.6).

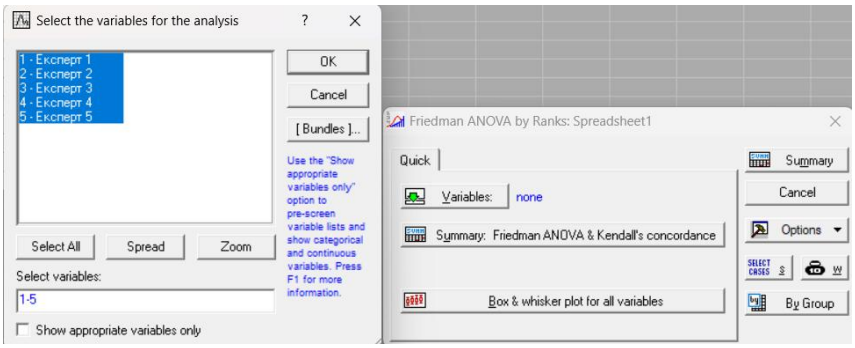


Рисунок Л7.6 – Обчислення узгодженості думок експертної групи

Результати розрахунку узгодженості думок експертної групи показано на рис. Л7.7.

Variable	Friedman ANOVA and Kendall Coeff. of Concordance (Spreadsheet1) ANOVA Chi Sqr. (N = 10, df = 4) = ,3908046 p = ,98322 Coeff. of Concordance = ,00977 Aver. rank r = -,1003						
	Average Rank	Sum of Ranks	Mean	Std.Dev.			
Експерт 1	2,950000	29,50000	5,500000	3,027650			
Експерт 2	2,850000	28,50000	5,500000	3,027650			
Експерт 3	2,900000	29,00000	5,500000	3,027650			
Експерт 4	3,100000	31,00000	5,300000	3,056868			
Експерт 5	3,200000	32,00000	5,500000	3,027650			

Рисунок Л7.7 – Таблиця результатів обрахунку коефіцієнта Конкордації

В нашому випадку коефіцієнт конкордації дорівнює 0,009. Згідно таблиці Чедока він є дуже слабким і тому необхідно шукати експерта, який погано впливає не загальний результат.

Щоб впорядкувати фактори за спаданням сумарних рангів необхідно в пакеті MS Excel просумувати значення кожного фактора і впорядкувати за спаданням. Фактор, який буде мати найбільшу суму буде тим фактором, який на думку експертів найбільше впливає на певне явище.

Лабораторна робота 8. Розрахунок прогнозів із застосуванням методу ковзного середнього

При використанні методу ковзного середнього прогноз будь-якого періоду є не що інше, як отримання середнього показника декількох результатів наглядів тимчасового ряду. Наприклад, якщо ви вибрали ковзке середнє за три місяці, прогнозом на травень буде середнє значення показників за лютий, березень і квітень. Якщо ви вибрали метод прогнозування ковзного середнього за чотири місяці, ви зможете оцінити травневий показник як середнє значення показників за січень, лютий, березень і квітень.

Обчислення за допомогою цього методу достатньо точно відбивають зміни основних показників попереднього періоду. Іноді при складанні прогнозу ці методи більш ефективні, ніж методи, засновані на довготривалих наглядах. Наприклад, ви давно формували прогноз кількості зареєстрованих кіберінцидентів у вашій компанії, і середній показник за останні кілька років становив близько 1000 випадків. Якщо організація планує суттєве скорочення штату фахівців із кібербезпеки, логічно припустити, що кількість вчасно виявлених та нейтралізованих інцидентів зменшуватиметься (принаймні протягом кількох місяців).

Якщо для прогнозування рівня захищеності на наступний місяць скористатися середнім значенням цього показника за останні 4 місяці, то результат, імовірно, буде завищеним, порівняно з фактичним. Натомість прогноз, побудований на основі даних лише за два останні місяці, набагато точніше відобразить наслідки скорочення штату спеціалістів із безпеки.

При вживанні ковзного середнього за два останні місяці кожний з показників (за цей часовий період) відповідає за половину значення прогнозу. При 4-місячному ковзному середньому показники цих же останніх місяців відповідають тільки за четверту частину значення прогнозу.

Таким чином, 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 ковзкового середнього є всього лише одне або два нагляди, то такий прогноз може стати дуже спрощеним. Зокрема, він відображатиме тенденції в даних, на яких він будується, нітрохи не краще, ніж сама базова лінія.

Існує три основні обмеження на застосування методу ковзного середнього. По-перше, цей метод більш підходить там, де в кожному наступному періоді, відбуваються лише невеликі зміни (хоча можуть бути випадки істотної варіації в періоді, протягом часового горизонту має бути відносна стабільність). По-друге, часовий горизонт прогнозу має бути відносно коротким. По-третє, якщо прогнозована змінна, що лежить в основі закономірності, коливається, то метод може бути непридатним. Існують і простіші методи для аналізу рядів даних, наприклад, графік часових рядів, але вони не дозволяють одержати точний прогноз. Однак саме такий прогноз дає метод ліній трендів найменших квадратів.

Приклад

1. Алгоритм складання прогнозів за допомогою надбудов ковзного середнього базується на використанні надбудови MS Excel *Пакет аналізу*.

Встановити цю надбудову можна таким чином.

1.1. Виберіть команду *Файл / Параметри* (рис. Л8.1).

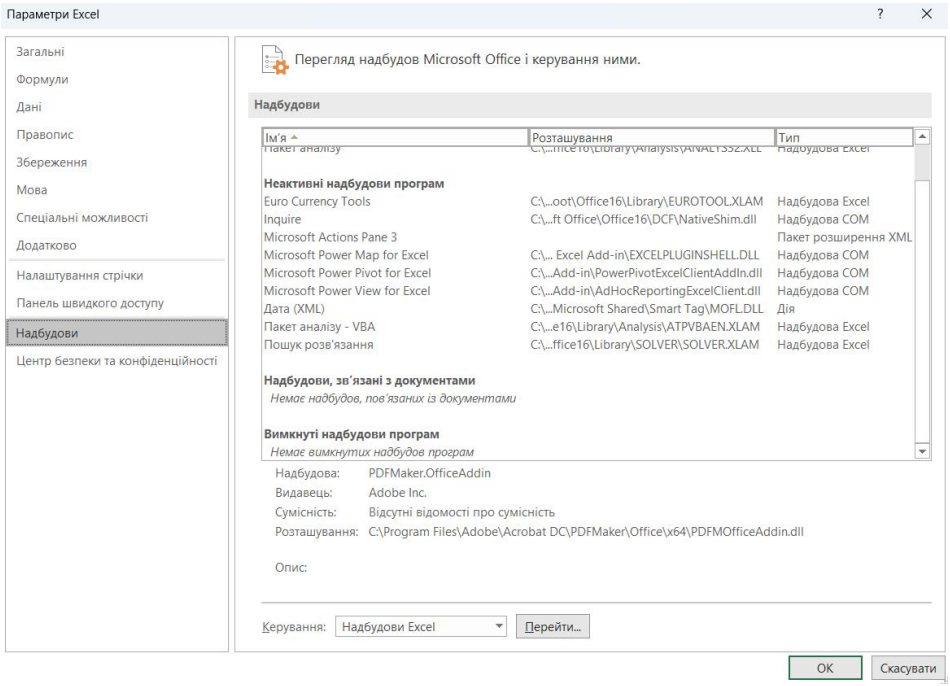


Рисунок Л8.1 – Вибір команди «Надбудови»

1.2. З'явиться діалогове вікно *Надбудови*. Встановіть прапорець *Пакет аналізу* і клацніть на кнопці ОК.

1.3. За необхідності активізуйте робочий лист, що містити дані про вашу базову лінію.

1.4. В меню *Дані* ви знайдете нову команду *Аналіз даних*. Виберіть її.

1.5. З'явиться діалогове вікно *Аналіз даних*, в якому містяться всі доступні функції аналізу даних. Із списка виберіть інструмент аналізу *Ковзне середнє / Moving average* і клацніть на кнопці ОК (рис. Л8.2).

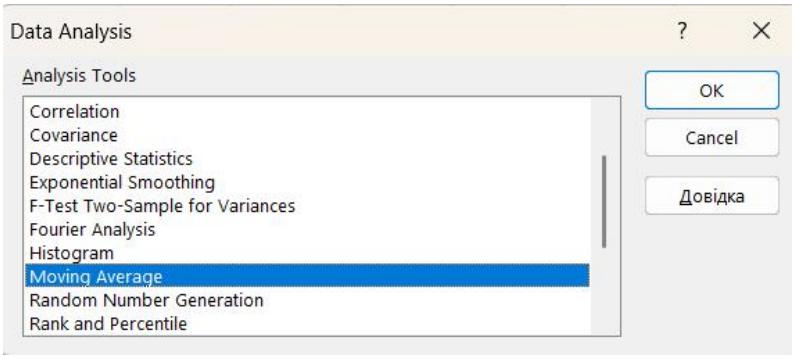


Рисунок Л8.2 – Робота в діалоговому вікні «Аналіз даних»

1.6. З'явиться діалогове вікно *Ковзне середнє*.

1.7. В полі Вхідний інтервал введіть дані про вашу базову лінію або вкажіть діапазон в робочому листі, посилання на нього з'явиться в цьому полі.

1.8. В полі Інтервал введіть кількість місяців, днів, які хочете включити в підрахунок ковзного середнього.

1.9. В полі введення Вихідний інтервал введіть адресу клітинки, у яку поміщають прогнозні дані.

1.10. Клацніть на кнопці ОК.

MS Excel виконує замість вас роботу по внесенню значень у формулу для обчислення ковзного середнього. Значення ковзного починається із значень #Н/Д, які дорівнюють значенню вказаного вами інтервалу мінус один. Це робиться через недостатню кількість даних для обчислення середнього значення декількох перших результатів наглядів, з інтервалом у 3 періоди (рис. Л8.3).

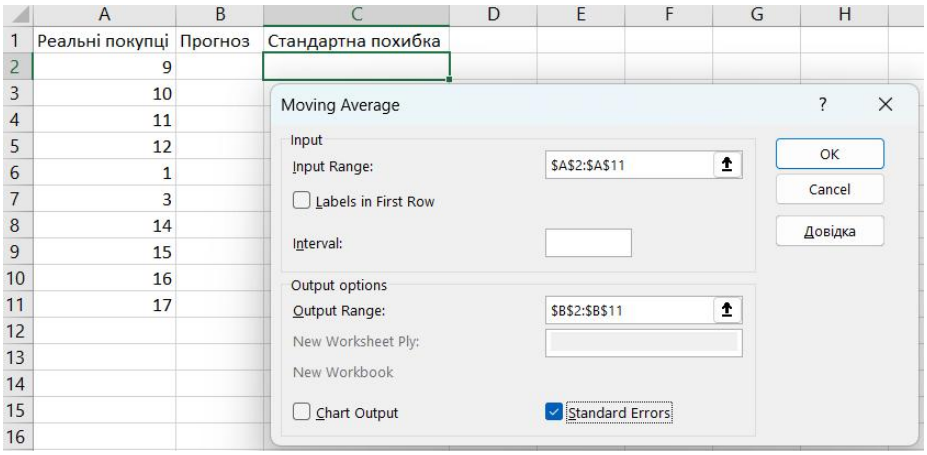


Рисунок Л8.3 – Обчислення ковзкого середнього

2. Алгоритм складання прогнозів ковзкового середнього за допомогою діаграм

Ви можете змінити форму представлення інформації, створивши графік, в якому дані базової лінії використовуються, для того, щоб продемонструвати лінію тренду ковзкого середнього. Цей метод часто зручніший, ніж використання надбудови Ковзке середнє, оскільки Excel завантажує Диспетчер надбудов всього за декілька секунд і ви можете пропустити цей етап, якщо

створюєте лінію тренда за допомогою діаграми. При цьому може виникнути проблема: на графіку не показані фактичні числові значення ковзкового середнього. Крім того, прогноз створюється на один часовий період раніше. Причому в цьому випадку ви не зможете виправити ситуацію, оскільки не існує можливості змінити розташування лінії тренда та графіка.

За допомогою майстра діаграм побудуйте графік-діаграму (рис. Л8.4).

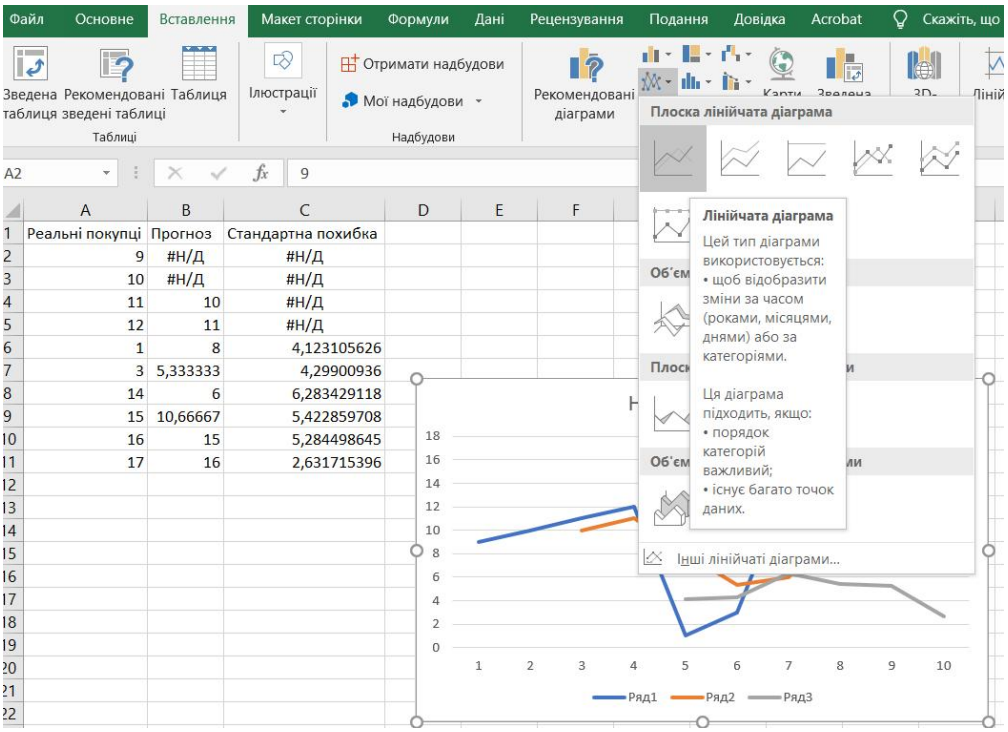


Рисунок Л8.4 – Побудова діаграми

Клацніть правою кнопкою миші на ряду даних діаграми і з контекстного меню, що з'явилося, виберіть команду *Додати лінію тренда* (рис. Л8.5).

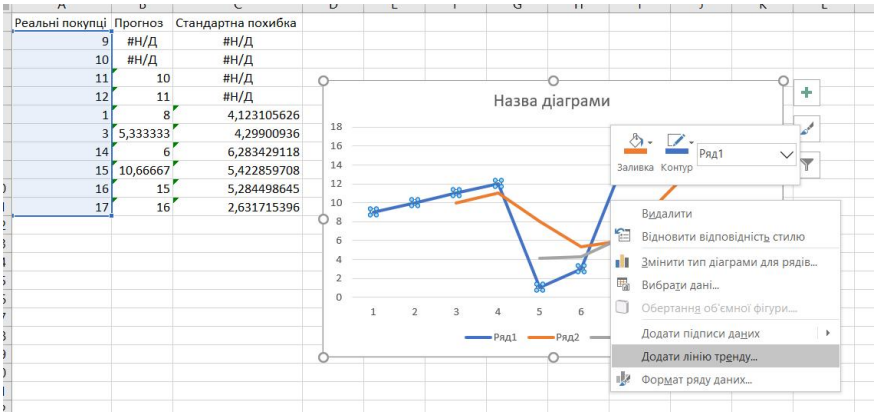


Рисунок Л8.5 – Побудова лінії тренда

В діалоговому вікні *Формат лінії тренда*, що з'явилося, клацніть на вкладці Тип, виберіть лінію тренда. Потім вкажіть, що необхідно показати рівняння, коефіцієнт достовірності, а також зробити прогноз на декілька періодів уперед.

Закінчивши виконання вищеперелічених дій, ви побачите на графіку лінію тренда ковзного середнього (разом з фактичними даними). Перші декілька показників ковзкового середнього відсутні з тієї ж причини, що надбудова *Ковзне середнє* 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.

Завдання

Задача 1

В таблиці Л8.1 наведені дані про витрати користувачів на покупку ліцензованого програмного забезпечення, включаючи операційну систему Windows у Львівській області.

Використовуючи описану вище методику прогнозування, складіть прогноз витрат на купівлю даних програмних продуктів на перший квартал 2025 р. Порівняйте фактичні дані і з вашим прогнозом. Спробуйте знайти пояснення можливим розбіжностям між прогнозними і фактичними даними.

Таблиця Л8.1

Споживацькі витрати на купівлю орераційної системи Windows, 2024 р.

Рік	Квартал 1	Квартал 2	Квартал 3	Квартал 4
2024	3043+N	3274+N	3452+N	4699+N

Задача 2

В таблиці Л8.2 наводяться дані про щомісячний обсяг наданих послуг певною IT-компанією в доларах.

Таблиця Л8.2

Обсяг наданих послуг

Число місяця	Обсяг послуг, тис.дол.	Число місяця	Обсяг послуг, тис.дол.
1	16+N	15	14+N
2	23+N	16	29+N
3	27+N	17	30+N
4	29+N	18	36+N
5	43+N	19	42+N
6	54+N	20	49+N
7	65+N	21	76+N
8	17+N	22	16+N
9	24+N	23	25+N
10	29+N	24	27+N
11	32+N	25	28+N
12	39+N	26	35+N
13	50+N	27	49+N
14	69+N	28	72+N

N – порядковий номер здобувача в журналі групи.

Найдіть ковзне середнє обсягу продаж на наступний місяць. Знайдіть абсолютне та відносне відхилення отриманих даних для кожного з семи днів першого тижня наступного місяця. Побудуйте лінію тренда.

Задача 3

За даними таблиці Л8.3 розрахуйте ковзке середнє, побудуйте лінію тренда та зробіть прогноз обсягу реалізації ІТ-послуг на 3 місяці наступного року. Зробіть оцінку якості прогнозу.

Таблиця Л8.3

Обсяг реалізації ІТ-послуг, тис.дол.

Місяць	Обсяг реалізації
1	5,4+N
2	5,6+N
3	6,0+N
4	6,6+N
5	7,2+N
6	7,7+N
7	8,5+N
8	8,6+N
9	9,0+N

10	8,5+N
11	8,3+N
12	7,9+N

N – порядковий номер здобувача в журналі групи.

Лабораторна робота 9. Розроблення бази економетричних моделей для оцінювання та прогнозування рівня кібербезпеки в Україні

Згідно з умовою, потрібно побудувати прогноз на наступні два роки роки, на основі вказаних даних, за допомогою трьох моделей.

Початкові дані за останні 18 років мають такий вигляд:

Таблиця Л9.1

Початкові дані

	Y	K	L
2007	94,1	1668,10	10580
2008	87,2	1249,30	29500
2009	83,9	849,20	38800
2010	72,4	726,90	38430
2011	79,9	624,40	38070
2012	83,1	538,10	47970
2013	95,1	570,90	47490
2014	95,7	501,70	47010
2015	103,1	482,30	46710
2016	108,40	615,50	56910
2017	138,20	693,70	66440
2018	113,50	879,60	76070
2019	115,10	1202,60	75870
2020	109,20	1489,95	72400
2021	95,20	1602,77	72300
2022	107,50	1774,00	74400
2023	112,90	2185,47	73500
2024	100,80	1962,44	73200

де,

Y – показник інформаційної безпеки підприємства: чим він вищий, тим краще захищена IT-інфраструктура;

K – інвестиції в IT-інфраструктуру та кібербезпеку, млн грн (сума витрат на IT-захист у державному та приватному секторах);

L – кадровий ресурс у сфері кібербезпеки, кількість осіб.

Потрібно обчислити три моделі: лінійну, нелінійну і однофакторну і визначити з них найкращу. Всі обчислення потрібно здійснювати в

програмному забезпечення *Statistica* та MS Excel.

1) Лінійна модель Y=A+αK+βL – обчислюється за допомогою меню *Statistics* у модулі *Multiple Regression/Множинна регресія* програмного забезпечення *Statistica* (рис. Л9.1). Після запуску цього меню відкриється наступне вікно, в якому необхідно вибрати параметри та натиснути на кнопку ОК.

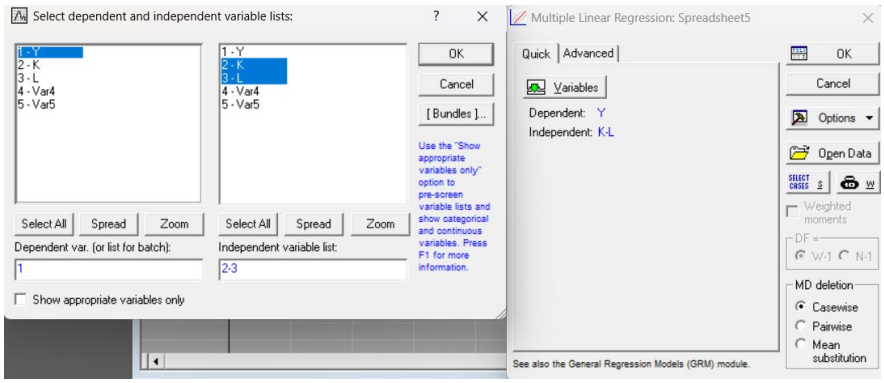


Рисунок Л9.1 – Модуль «Множинна регресія»

Після цього у вікні, яке з'явилось і яке зображене нижче (рис. Л9.2), необхідно натиснути на кнопку ОК для того, щоб отримати необхідні параметри моделі.

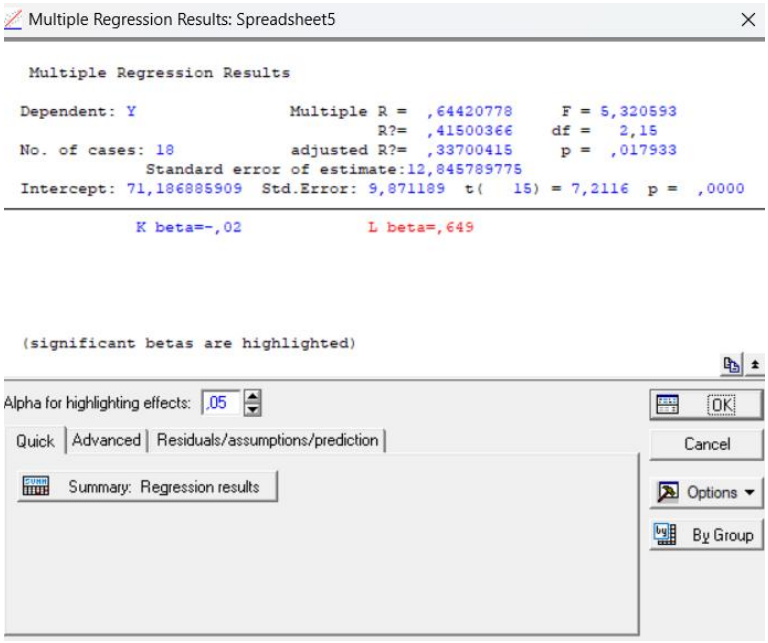


Рисунок Л9.2. – Вікно «Результати множинної регресії»

З'явиться вікно з параметрами моделі (рис. Л9.3).

N=18	Regression Summary for Dependent Variable: Y (Spreadsheet5) R= ,64420778 R?= ,41500366 Adjusted R?= ,33700415 F(2,15)=5,3206 p<,01793 Std.Error of estimate: 12,846 Beta	Std.Err. of Beta	B	Std.Err. of B	t(15)	p-level
Intercept			71,18689	9,871189	7,211582	0,000003
K	-0,015374	0,210279	-0,00043	0,005915	-0,073112	0,942683
L	0,649327	0,210279	0,00053	0,000172	3,087932	0,007500

Рисунок Л9.3 – Параметри моделі

З даного вікна, для подальших розрахунків необхідно використати такі параметри:

A = 71,18

α = -0,00043

β = 0,00053

Функція має вигляд: Y=71,18-0,00043*K+0,00053*L

При цьому коефіцієнт детермінації R^2 дорівнює 0,415, кількість значущих параметрів (параметрів, які відображені червоним кольором) – 2, середнє відхилення – 0.

Для того, щоб знайти середнє відхилення, потрібно у вікні, яке зображене нижче, натиснути на кнопку «*Perform residual analysis / Аналіз залишків*» в закладці «*Residuals/assumptions/predictions / Залишки/передбачені/спостежені значення*» (рис. Л9.4).

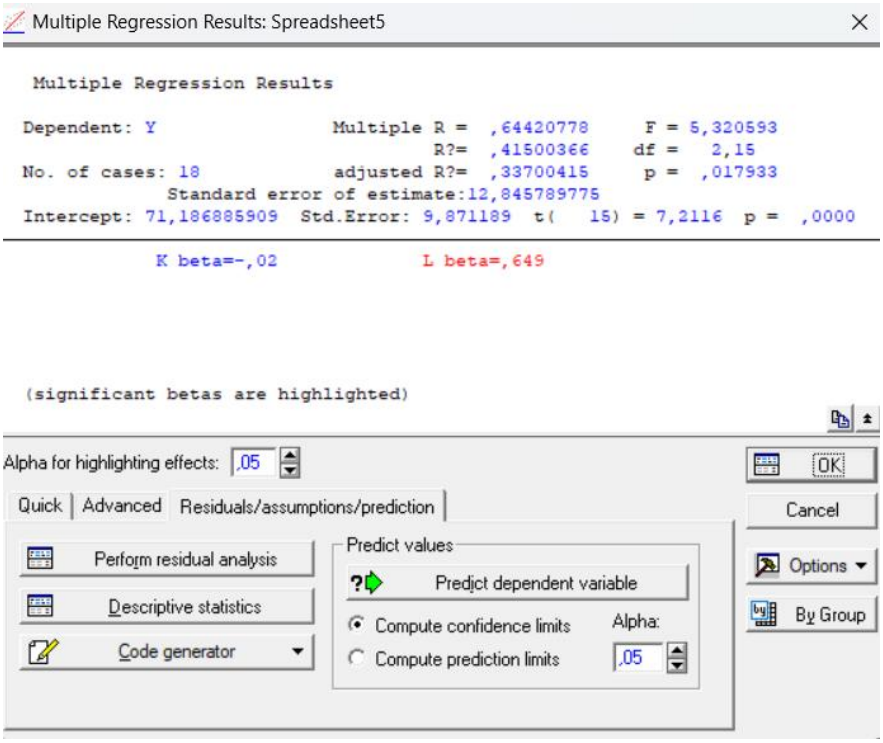


Рисунок Л9.4 – Знаходження середнього відхилення

В наступному вікні потрібно натиснути на кнопку «*Summary: Residuals & predicted / Залишки та передбачені*» (рис. Л9.5).

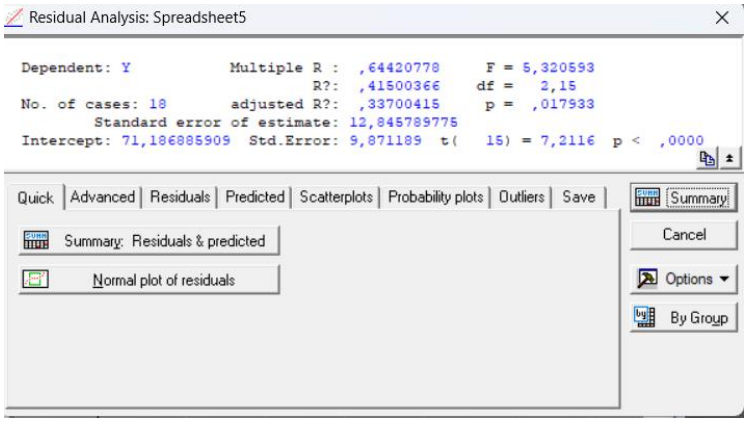


Рисунок Л9.5 – Знаходження середнього відхилення

Внаслідок цього з’явиться наступне вікно (рис. Л9.6).

Predicted & Residual Values (Spreadsheet5)
Dependent variable: Y

Case No.	Observed Value	Predicted Value	Residual	Standard Pred. v.	Standard Residual	Std.Err. Pred.Val	Mahalanobis Distance	Deleted Residual
1	94,1000	76,0731	18,0269	-2,32857	1,40333	9,807652	8,965183	43,2215
2	87,2000	86,2823	0,9176	-1,32404	0,07144	5,627001	2,317535	1,1355
3	83,9000	91,3846	-7,4846	-0,82201	-0,58265	4,002239	0,705743	-8,2893
4	72,4000	91,2414	-18,8414	-0,83610	-1,46674	4,183024	0,858192	-21,0763
5	79,9000	91,0949	-11,1949	-0,85051	-0,87149	4,419791	1,068032	-12,6982
6	83,1000	96,3795	-13,2795	-0,33054	-1,03376	4,308612	0,968058	-14,9628
7	95,1000	96,1109	-1,0109	-0,35697	-0,07870	4,185173	0,860044	-1,1310
8	95,7000	95,8864	-0,1864	-0,37906	-0,01451	4,456558	1,101653	-0,2119
9	103,1000	95,7358	7,3642	-0,39388	0,57328	4,536540	1,175755	8,4135
10	108,4000	101,0845	7,3155	0,13240	0,56949	4,229473	0,898447	8,2050
11	138,2000	106,1018	32,0982	0,62607	2,49873	4,680441	1,312396	37,0117
12	113,5000	111,1255	2,3745	1,12038	0,18484	5,217314	1,859835	2,8435
13	115,1000	110,8798	4,2202	1,09620	0,32852	4,591615	1,227548	4,8383
14	109,2000	108,9164	0,2836	0,90301	0,02208	4,360614	1,014502	0,3206
15	95,2000	108,8146	-13,6146	0,89299	-1,05985	4,600353	1,235822	-15,6176
16	107,5000	109,8536	-2,3536	0,99523	-0,18322	5,247757	1,892656	-2,8251
17	112,9000	109,1986	3,7014	0,93078	0,28814	6,870012	3,917861	5,1841
18	100,8000	109,1361	-8,3361	0,92463	-0,64893	5,882707	2,620737	-10,5482
Minimum	72,4000	76,0731	-18,8414	-2,32857	-1,46674	4,002239	0,705743	-21,0763
Maximum	138,2000	111,1255	32,0982	1,12038	2,49873	9,807652	8,965183	43,2215
Mean	99,7389	99,7389	-0,0000	-0,00000	-0,00000	5,067049	1,888889	1,3230
Median	98,2500	98,7320	0,0486	-0,09907	0,00378	4,564077	1,201651	0,0543

Рисунок Л9.6 – Вікно «Передбачені значення та залишки»

На перетині стовпчика «*Residual / Залишки*» і рядка «*Mean / Середнє*» і знаходиться шукане середнє відхилення.

Всі знайдені параметри першої моделі необхідно вписати у заздалегідь підготовлену в пакеті MS Excel таблицю (табл. Л9.2).

Таблиця Л9.2

Зведена таблиця з результатами обчислень

№з/п	Параметри			Критерії оцінки		
	A	α	β	R^2	К-сть значущих параметрів	Середнє відхилення
1	71,18	-0,00043	0,00053	0,415	2	0
2						
3						

2) Друга модель, параметри якої необхідно знайти, має вираз $Y=A*K^{\alpha}*L^{\beta}$. Обчислення проводяться в модулі *Statistics-Advanced linear-Nonlinear Models - Nonlinear Estimation / Аналіз - поглиблені методи аналізу-нелінійне оцінювання*.

Відкриється наступне вікно (рис. Л9.7):

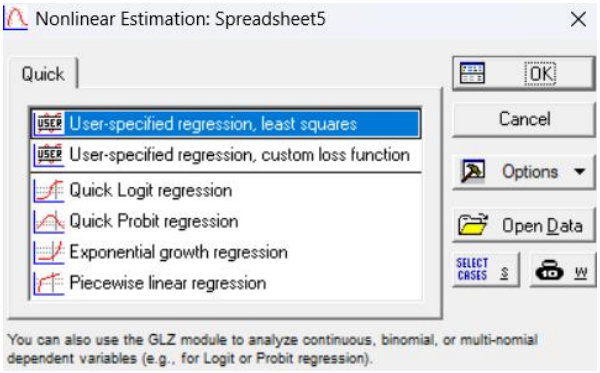


Рисунок Л9.7 – Вікно «*Нелінійне оцінювання*»

У цьому вікні необхідно вибрати перший пункт і натиснути на кнопку ОК. Відкриється наступне вікно, в якому необхідно натиснути кнопку «*Function to be estimated / Оцінювана функція*» і ввести формулу, як показано на рисунку Л9.8.

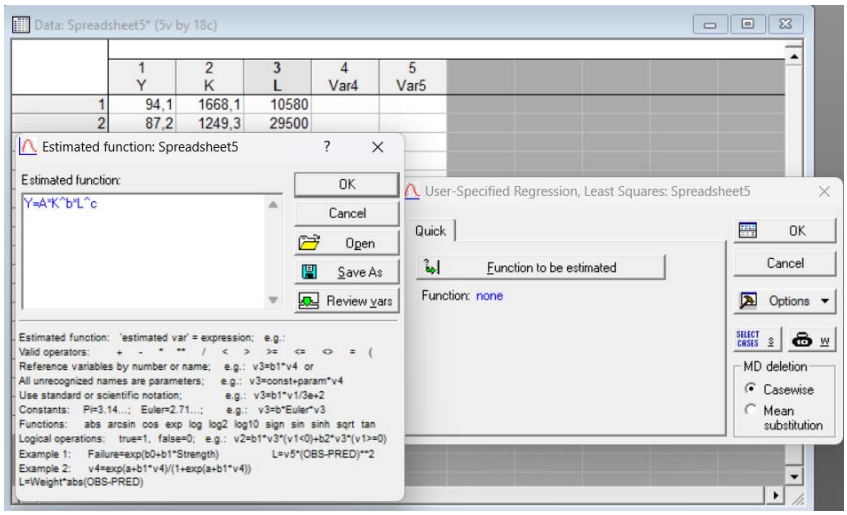


Рисунок Л9.8 – Введення функції

З’явиться наступне вікно (рис. Л9.9).

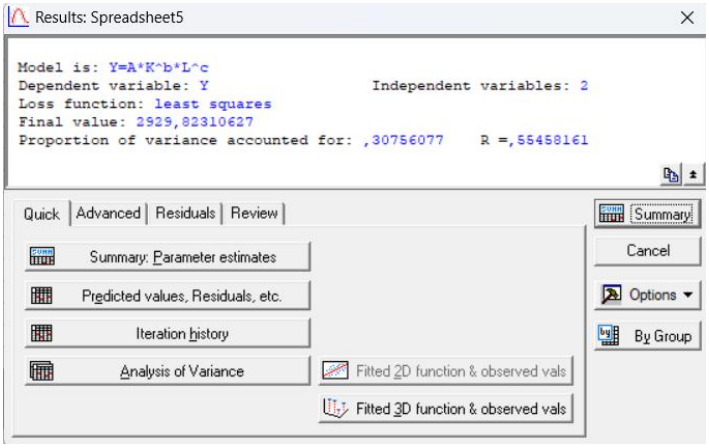


Рисунок 22.9 – Оцінка параметрів моделі

З цього вікна видно, що R2=0,307. Після цього необхідно натиснути на кнопку «*Summary: Parameter estimates / Оцінювання параметрів*» і з'явиться вікно з параметрами моделі (рис. Л9.10).

Model is: Y=A*K^b*L^c (Spreadsheet5)
Dep. Var. : Y
Level of confidence: 95.0% (alpha=0.050)

	Estimate	Standard error	t-value df = 15	p-level	Lo. Conf Limit	Up. Conf Limit
A	10,18064	9,235488	1,102339	0,287687	-9,50434	29,86562
b	0,02498	0,067487	0,370216	0,716397	-0,11886	0,16883
c	0,19457	0,083772	2,322580	0,034675	0,01601	0,37312

Рисунок Л9.10 – Вікно параметрів моделі

З цього вікна беремо необхідні параметри:

A = 10,18

α = 0,0249

β = 0,1945

Функція матиме вигляд: Y=10,18*K^0,0249*L^0,1945

Коефіцієнт детермінації R^2 дорівнює 0,307, кількість значущих параметрів – 1, середнє відхилення – 0,0341.

Якщо у вікні, яке наведене на рис. Л9.9 натиснути у закладці «*Residuals / Залишки*» на кнопку «*Observed, predicted, residuals / Спостережені, передбачені та залишки*», то можна обчислити середнє відхилення.

Всі знайдені параметри першої моделі необхідно вписати у заздалегідь підготовлену в пакеті MS Excel таблицю (табл. Л9.3).

Таблиця Л9.3

Зведена таблиця з результатами обчислень

№з/п	Параметри			Критерії оцінки		
	A	α	β	R^2	К-сть значущих парамерів	Середнє відхилення
1	71,18	-0,00043	0,00053	0,415	2	0
2	10,18	0,0249	0,1945	0,307	1	0,0341

3) Однофакторна модель (y=A* k^α) також обчислюються в модулі *Statistics-Advanced linear-Nonlinear Models - Nonlinear Estimation / Аналіз - поглиблені методи аналізу-нелінійне оцінювання* аналогічно попередній моделі. Єдиною відмінністю при обчисленні є те, що для цього необхідно визначити y та k, та добавити у двох нових колонках в якості вхідних даних. Їх умовно необхідно позначити Y_L та K_L:

$$y = У/L;$$

$$k = K/L$$

Одержані розрахунки представимо у наступній таблиці (табл. Л9.4):

Таблиця Л9.4

Результати обчислень

Y_L	K_L
0,0089	0,1577
0,0030	0,0423
0,0022	0,0219
0,0019	0,0189
0,0021	0,0164
0,0017	0,0112
0,0020	0,0120
0,0020	0,0107
0,0022	0,0103
0,0019	0,0108
0,0021	0,0104
0,0015	0,0116
0,0015	0,0159
0,0015	0,0206
0,0013	0,0222
0,0014	0,0238
0,0015	0,0297
0,0014	0,0268

Результати обчислень показані на рис. Л9.11.

Model is: Y_L=A*K_L^b (Spreadsheet5)
Dep. Var. : Y_L
Level of confidence: 95.0% (alpha=0.050)

	Estimate	Standard error	t-value df = 16	p-level	Lo. Conf Limit	Up. Conf Limit
A	0,032399	0,005509	5,88152	0,000023	0,020721	0,044077
b	0,723358	0,058392	12,38788	0,000000	0,599571	0,847144

Рисунок 22.11 – Параметри однофакторної моделі

A=0,0323

α=0,7233

Коефіцієнт детермінації R^2 дорівнює 0,85, кількість значущих параметрів – 2, середнє відхилення – 0,0001.

Всі знайдені параметри першої моделі необхідно вписати у заздалегідь підготовлену в пакеті MS Excel таблицю (табл. Л9.5).

Таблиця 22.5

Зведена таблиця з результатами обчислень

№з/п	Параметри			Критерії оцінки		
	A	α	β	R^2	К-сть значущих парамен рів	Середнє відхилення
1	71,18	-0,00043	0,00053	0,415	2	0
2	10,18	0,0249	0,1945	0,307	1	0,0341
3	0,0323	0,7233		0,85	2	0,0001

Для побудови прогнозу необхідно вибрати кращу модель. Краща модель та, в якій більший коефіцієнт детермінації, більша кількість значущих параметрів та менше середнє відхилення. В нашому випадку підходить третя модель: оскільки в ній найбільший коефіцієнт детермінації та кількість значущих параметрів – 2, друга модель відпадає через меншу кількість значущих параметрів і менший коефіцієнт детермінації.

Для обчислення прогнозного значення Y на наступні роки, необхідно знайти прогнозні значення k і y. Для цього нам потрібно взнати прогнозні значення K та L.

Прогноз у програмному забезпечення *Statistica* обчислюється у модулі *Statistics-Advanced linear-Nonlinear Models - Time Series&Forecasting / Аналіз - поглиблені методи аналізу-часові ряди та прогнозування* (рис. Л9.12).

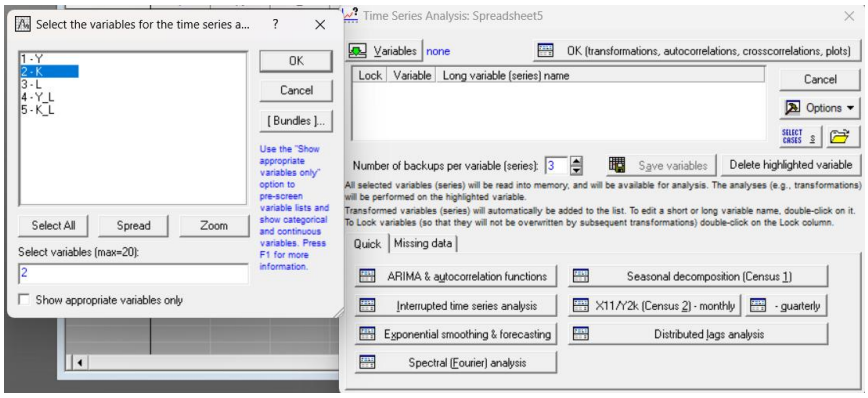


Рисунок Л9.12 – Вікно модуля «Часові ряди та прогнозування»

В даному вікні необхідно вибрати змінну для побудови прогнозу (в нашому випадку спочатку необхідно вибрати змінну K а потім L) та запустити модуль *Exponential smoothing & Forecasting / Експоненційне згладжування та прогнозування* (рис. Л9.13).

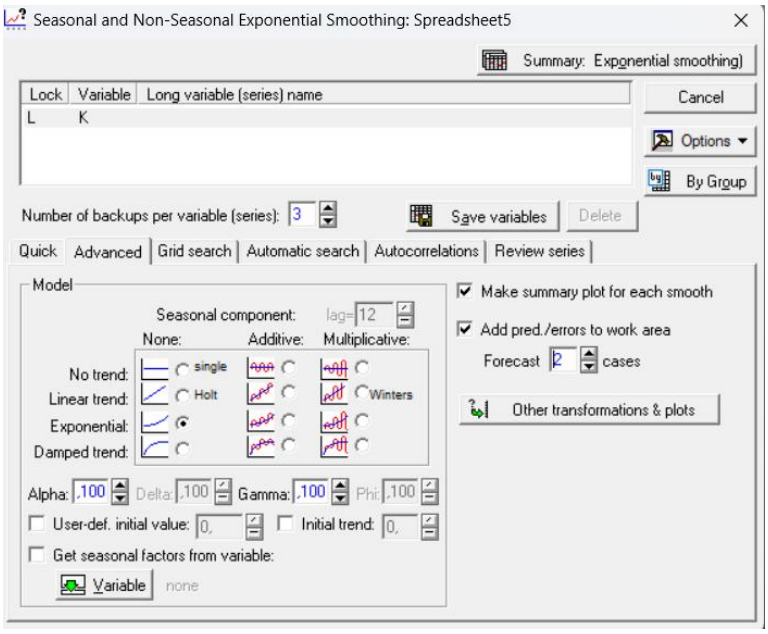


Рисунок Л9.13 – Вікно модуля «Експоненційне згладжування та прогнозування»

В цьому вікні потрібно вибрати модель, за допомогою якої необхідно побудувати прогноз, вказати на скільки років потрібно зробити прогнозування і натиснути на кнопку ОК. Результати прогнозування показані на рис. Л9.14.

Exp. smoothing: S0=1928, T0=,7489 (Spreadsheet5)
Expon.trend,no season; Alpha= ,100 Gamma=,100
K

Case	K	Smoothed Series	Resids
1	1668,100	1443,592	224,508
2	1249,300	1099,680	149,620
3	849,200	837,231	11,969
4	726,900	629,851	97,049
5	624,400	481,193	143,207
6	538,100	373,927	164,173
7	570,900	295,857	275,043
8	501,700	247,366	254,334
9	482,300	210,833	271,467
10	615,500	186,291	429,209
11	693,700	183,561	510,139
12	879,600	193,077	686,523
13	1202,600	223,088	979,512
14	1489,950	285,655	1204,295
15	1602,770	376,561	1226,209
16	1774,000	477,963	1296,037
17	2185,470	597,515	1587,955
18	1962,440	763,565	1198,875
19		905,930	
20		928,981	

Рисунок Л9.14 – Результати прогнозування

Числа у 19 і 20-му рядках, це прогнозні значення показника К на наступні 2 роки.

Аналогічно будується прогноз для показника L.

Результати прогнозування необхідно занести у заздалегідь підготовлену таблицю у пакеті MS Excel (табл. Л9.6).

Таблиця 22.6

Прогнозовані значення K та L на два роки

Прогнозовані значення			
	Y	K	L
1-й рік		905,93	62014
2-й рік		928,98	62014

За допомогою засобів Exlel потрібно підставити ці значення у економетричну функцію, яка була вибрана раніше і отримати прогноз показника інформаційної безпеки підприємства на наступні роки.

Після проведених обчислень показник інформаційної безпеки такий:

Y(1-й рік)=94,22

Y(2-1 рік)=95,94

Отримані результати необхідно занести у таблицю Л9.7.

Таблиця Л9.7

Кінцеві результати

Початкове значення	1-й рік.	2-й рік.
звіт	очікуване	проєкт
100,8	94,22	95,94

Отже, можна стверджувати, що згідно з обраною моделлю, можна побачити, що значення показника безпеки захищеності ІТ-інфраструктури буде зростати впродовж наступних двох років. Також, знаючи заздалегідь рівень інвестицій в ІТ-інфраструктуру та кібербезпеку а також планові можливості кадрового потенціалу кібербезпеки, можна прогнозувати значення показника безпеки.

Завдання

На основі даних таблиці Л9.8 побудувати дві моделі Кобба-Дугласа та одну однофакторну модель. Вибрати серед побудованих моделей кращу, та на її основі побудувати прогноз на наступні два роки.

Таблиця Л9.8

Початкові дані

Y	K	L
93,3+N	1668,10+N	1058,6
88,1+N	1249,30+N	950,4
82,9+N	849,20+N	880
71,4+N	726,90+N	843,4
79,9+N	624,4+N	807+N
78,1+N	538,1+N	797,8+N
94,1+N	570,9+N	749,8+N
92,7+N	501,7+N	701,7+N
102+N	482,3+N	671,6+N

112,4+N	615,5+N	691,0+N
143,2+N	693,7+N	644,5+N
113,5+N	879,6+N	607,7+N
114,1+N	1202,6+N	587,1+N
109,2+N	1489,95+N	724,9+N
93,2+N	1602,77+N	723,2+N
108,5+N	1774,00+N	744,6+N
112,9+N	2185,47+N	735+N
100,8+N	1962,44+N	732,1+N

N – порядковий номер здобувача в журналі групи.

Тестові завдання

1. Що вивчає економічна кібернетика?

a) Лише статистичні закономірності у фінансових системах

b) Розробку програмного забезпечення для банківських структур

c) Загальні принципи управління складними економічними системами

d) Тільки методи збору економічних даних

2. Який із методів є основним у економічній кібернетиці?

a) Математичне моделювання

b) Літературний аналіз

c) Історико-порівняльний метод

d) Емпіричне спостереження

3. Що є предметом економічного прогнозування?

a) Минулі економічні процеси

b) Майбутні тенденції та закономірності розвитку економіки

c) Особисті фінанси громадян

d) Виключно фінансова звітність підприємств

4. Що є об’єктом економічного прогнозування?

a) Природні явища

b) Суспільні та економічні процеси

c) Лише національний бюджет

d) Технологічні процеси виробництва

5. Яка головна роль прогнозування в управлінні?

a) Виявлення помилок у бухгалтерському обліку

b) Обґрунтування управлінських рішень

c) Встановлення зарплатного фонду

d) Контроль виконання планів

6. Моніторинг у прогнозуванні – це:

a) Аналіз поточного стану економічних процесів

b) Складання річного бюджету

c) Контроль за роботою підприємства

d) Побудова математичної моделі майбутнього

7. Що може виступати альтернативою прогнозуванню?

a) Випадкове прийняття рішень

b) Реактивне управління, що ґрунтується на фактах минулого

c) Планування без урахування ресурсів

d) Інтуїтивні передбачення без даних

8. Який принцип прогнозування полягає в урахуванні можливих варіантів розвитку?

a) Принцип адекватності

b) Принцип багатоваріантності

c) Принцип безперервності

d) Принцип цілісності

9. Принцип наукової обґрунтованості прогнозів означає:

a) Використання статистичних і математичних методів

b) Виключно інтуїтивні передбачення

c) Прогнозування тільки на основі досвіду керівника

d) Пошук аналогій у художній літературі

10. Основна суть прогнозування полягає в:

a) Аналізі лише минулих економічних показників

b) Передбаченні можливих змін і тенденцій у майбутньому

c) Підрахунку витрат і доходів підприємства

d) Створенні статистичних звітів

11. За ознакою "період упередження" виділяють прогнози:

a) Детерміновані, змішані, стохастичні

b) Дискретні, аперіодичні, циклічні

c) Коротко-, середньо-, довгострокові

12. Загальний прогноз передбачає:

a) Відповідь на питання щодо можливості настання деякої ситуації

b) Визначення числових оцінок майбутньої ситуації

c) опис майбутньої ситуації

13. Що є суттю прогнозування?

a) Побудова бухгалтерської звітності

b) Наукове передбачення можливих станів об’єкта в майбутньому

c) Випадкове прийняття управлінських рішень

d) Аналіз виключно минулих подій

14. Що є предметом прогнозування?

a) Статистичні дані

b) Майбутні тенденції та закономірності розвитку

c) Планові показники підприємств

d) Економічні помилки минулого

15. Об’єктом прогнозування виступає:

a) Тільки виробництво

b) Економічні та соціальні процеси

c) Фінансові документи

d) Стандарти бухгалтерського обліку

16. Який із прогнозів відноситься до типології за часом?

a) Короткостроковий

b) Галузевий

c) Науково-технічний

d) Просторовий

17. Галузевий прогноз належить до:

a) Територіальної типології

b) Об’єктної типології

c) Часової типології

d) За масштабом

18. За масштабом прогнози поділяються на:

a) Короткострокові, середньострокові, довгострокові

b) Локальні, регіональні, глобальні

c) Соціальні, економічні, політичні

d) Емпіричні та теоретичні

19. Який метод прогнозування ґрунтується на використанні думки експертів?

a) Метод експертних оцінок

b) Метод екстраполяції

c) Балансовий метод

d) Кореляційний аналіз

20. Методи прогнозування поділяють на:

a) Статистичні та динамічні

b) Кількісні та якісні

c) Теоретичні та практичні

d) Наукові та прикладні

21. Що таке динамічний (часовий) ряд?

a) Набір випадкових спостережень

b) Показники, розташовані в логічному порядку

c) Сукупність значень показника, впорядкованих у часі

d) Дані, зібрані методом експертних оцінок

22. Тренд у часовому ряді – це:

a) Випадкові коливання показників

b) Систематична складова, що відображає загальну тенденцію розвитку

c) Помилка вимірювання у статистичних даних

d) Разові збурення в економіці

23. Прогнозування це:

a) Наука про систему нашого мислення про майбутнє, про способи та методи дослідження майбутнього, про знаходження багатоваріантних альтернатив зміни майбутнього, яке носить імовірнісний характер

b) Наукова дисципліна про закономірності розробки прогнозів

c) Система принципів і способів організації та побудови теоретичної і практичної діяльності, а також вчення про цю систему

24. Що розуміється під визначенням "варіант прогнозу"?

a) Один з прогнозів, що становлять групу можливих прогнозів об'єкта прогнозування

b) Спосіб дослідження об'єкта прогнозування, спрямований на розробку прогнозу

c) Сукупність методів і правил розробки прогнозів конкретних об'єктів

d) Система методів прогнозування і засобів їх реалізації, що функціонує відповідно до основних принципів прогнозування

25. Що розуміється під визначенням "Метод прогнозування"?

a) Один з прогнозів, що становлять групу можливих прогнозів об'єкта прогнозування

b) Спосіб дослідження об'єкта прогнозування, спрямований на розробку прогнозу

c) Сукупність методів і правил розробки прогнозів конкретних об'єктів

d) Система методів прогнозування і засобів їх реалізації, що функціонує відповідно до основних принципів прогнозування

26. Модель – це:

a) Штучний, створений людиною об'єкт довільної природи, який замінює або відтворює досліджуваний об'єкт-оригінал

b) Штучний, створений людиною об'єкт довільної природи, який досліджує досліджуваний об'єкт-оригінал

c) Немає правильної відповіді

27. Що вивчає кореляційно-регресійний аналіз?

a) Залежність між якісними ознаками

b) Кількісні залежності між економічними явищами

c) Виключно статистичний розподіл вибірки

d) Взаємозв’язки лише у фінансовій сфері

28. Який тип залежності вивчає регресійний аналіз?

a) Причинно-наслідкову

b) Логічну

c) Хронологічну

d) Просторову

29. Що означає «регресійна залежність»?

a) Взаємозв’язок між двома подіями у минулому

b) Функціональний зв’язок між результативною та факторними змінними

c) Співвідношення якісних ознак

d) Залежність без математичної форми

30. Який метод використовується для побудови лінійної регресії?

a) Метод експертних оцінок

b) Метод найменших квадратів

c) Метод балансових розрахунків

d) Метод Монте-Карло

31. Яка основна ідея методу найменших квадратів?

a) Мінімізація різниці між спостереженнями та прогнозом

b) Максимізація кількості факторів у моделі

c) Усунення мультиколінеарності

d) Використання випадкових прогнозів

32. Що характеризує множинна регресія?

a) Залежність між двома змінними

b) Залежність результативної змінної від кількох факторних

c) Тільки часову динаміку

d) Тільки кореляційні зв’язки

33. Який перший етап побудови моделі множинної регресії?

a) Перевірка залишків на автокореляцію

b) Вибір факторних змінних

c) Розрахунок коефіцієнта детермінації

d) Побудова трендової моделі

34. Сутність кореляції полягає у:

a) Встановленні причинно-наслідкових відносин

b) Вимірюванні тісноти зв’язку між змінними

c) Побудові прогнозів на основі часових рядів

d) Вивченні просторових закономірностей

35. Лінійна кореляція відображає:

a) Прямолінійний взаємозв’язок між змінними

b) Лише нелінійний зв’язок

c) Розподіл випадкових величин

d) Взаємозв’язок якісних характеристик

36. Коефіцієнт лінійної кореляції (r) показує:

a) Точність прогнозування

b) Тісноту і напрямок зв’язку між змінними

c) Величину похибки регресії

d) Виключно силу негативного зв’язку

37. Які завдання стоять перед прогнозуванням соціальних процесів?

a) З'ясування перспективи найближчого або віддаленого майбутнього

b) Сприяння розробленню оптимальних програм і планів економічного та соціального розвитку об’єкта

c) Обидві відповіді правильні

38. Що таке евристичні методи прогнозування?

a) Методи, що базуються лише на статистиці

b) Методи, що використовують виключно математичні моделі

c) Методи, які ґрунтуються на досвіді, інтуїції та експертних оцінках

d) Методи, що застосовуються тільки в технічних науках

39. Що характерно для індивідуального прийняття рішень?

a) Врахування різних думок

b) Відповідальність і вибір однією особою

c) Використання методу «Дельфі»

d) Голосування у групі

40. Групове прийняття рішень відрізняється тим, що:

a) Виключає вплив суб’єктивних факторів

b) Враховує думки кількох експертів

c) Завжди швидше за індивідуальне

d) Використовується лише в кризових ситуаціях

41. Який метод групової роботи ґрунтується на анонімному опитуванні експертів у кілька турів?

a) Метод «мозкового штурму»

b) Метод Дельфі

c) Метод сценаріїв

d) Метод морфологічного аналізу

42. Яка головна ідея «методу мозкового штурму»?

a) Вибір кращого варіанта шляхом голосування

b) Генерація максимальної кількості ідей у групі

c) Використання статистичних даних

d) Визначення середнього прогнозу

43. Метод морфологічного аналізу застосовують для:

a) Пошуку інноваційних рішень шляхом комбінування можливих варіантів

b) Перевірки точності прогнозу

c) Розрахунку коефіцієнта кореляції

d) Визначення оптимальної вибірки

44. Який метод групової роботи передбачає створення альтернативних варіантів розвитку подій?

a) Метод сценаріїв

b) Метод експертних оцінок

c) Метод головних компонент

d) Метод екстраполяції

45. Основна перевага групового прийняття рішень полягає в:

a) Економії часу

b) Поєднанні знань і досвіду різних фахівців

c) Усуненні ризику

d) Виключенні суб’єктивності

46. Який недолік може мати групове прийняття рішень?

a) Зменшення якості прогнозів

b) Вплив авторитетної думки на інших учасників

c) Повна відсутність суб’єктивності

d) Неможливість використання статистики

47. До евристичних методів прогнозування НЕ належить:

a) Метод експертних оцінок

b) Метод екстраполяції

c) Метод Дельфі

d) Метод «мозкового штурму»

48. Який коефіцієнт показує тісноту узгодженості думок двох експертів?

a) Коефіцієнт кореляції

b) Коефіцієнт детермінації

c) Коефіцієнт Спірмена

49. Який коефіцієнт показує тісноту зв'язку експертної групи?

a) Коефіцієнт кореляції

b) Коефіцієнт конкондації

c) Коефіцієнт Спірмена

50. Яка мета регресійного аналізу?

a) Визначити кількісні зв'язки між залежними випадковими величинами

b) Визначити якісні зв'язки між залежними випадковими величинами

c) Обидві відповіді правильні

51. Що є суттю непараметричних методів прогнозування?

a) Використання статистичних параметрів вибірки

b) Побудова прогнозу без попереднього припущення про закон розподілу

c) Використання лінійних регресійних моделей

d) Застосування винятково експертних оцінок

52. Який принцип лежить в основі непараметричних методів?

a) Використання лише середніх значень

b) Гнучкість і мінімум попередніх припущень

c) Побудова функціональних моделей

d) Використання обов’язкової нормальності даних

53. Для яких випадків доцільно застосовувати непараметричні методи прогнозування?

a) Коли відомий закон розподілу даних

b) При малих вибірках або невідомому розподілі

c) Лише для довгострокових прогнозів

d) Виключно для часових рядів

54. Який із наведених методів належить до непараметричних?

a) Метод серій

b) Метод найменших квадратів

c) Лінійна регресія

d) ARIMA-модель

55. Суть параметричних методів прогнозування полягає у:

a) Використанні виключно експертних оцінок

b) Оцінюванні параметрів моделей, що описують залежність між змінними

c) Побудові прогнозів без даних

d) Аналізі трендів без математичних моделей

56. Який принцип лежить в основі параметричних методів прогнозування?

a) Побудова моделей із використанням розрахованих параметрів

b) Виключення математичних залежностей

c) Використання інтуїтивних оцінок

d) Аналіз даних без їх формалізації

57. Що є характерним для параметричних методів?

a) Не потребують жодних припущень

b) Базуються на регресійних та статистичних моделях

c) Використовують тільки рангові шкали

d) Виключають математичний апарат

58. Прикладом параметричного методу прогнозування є:

a) Метод кореляційно-регресійного аналізу

b) Метод знаків

c) Метод серій

d) Метод експертних оцінок

59. Який недолік мають параметричні методи?

a) Потребують значного обсягу даних та виконання певних умов

b) Виключають точність прогнозів

c) Неможливо використовувати за наявності трендів

d) Використовуються лише для якісних даних

60. Головна відмінність між параметричними та непараметричними методами прогнозування полягає у тому, що:

a) Параметричні методи завжди менш точні

b) Непараметричні ґрунтуються на гіпотезах про параметри розподілу

c) Параметричні використовують оцінку параметрів моделей, а непараметричні – ні

d) Обидва види методів повністю ідентичні

61. Що таке модель у контексті економічної динаміки?

a) Схематичне відображення процесу або системи

b) Виключно математична формула

c) Графік динамічного ряду

d) Будь-яка економічна теорія

62. Яка основна суть методу моделювання?

a) Точне копіювання реальної системи

b) Створення спрощеного відображення реального процесу для його аналізу

c) Використання лише експертних оцінок

d) Прогнозування на основі інтуїції

63. До якого етапу моделювання належить формування завдання?

a) Збір вихідних даних

b) Постановка проблеми

c) Перевірка моделі

d) Аналіз результатів

64. Що передбачає етап збору даних у моделюванні соціально-економічних систем?

a) Визначення основних змінних та параметрів

b) Створення прогнозів

c) Формування висновків

d) Вибір типу життєвого циклу

65. Який етап моделювання пов'язаний з перевіркою адекватності моделі?

a) Постановка проблеми

b) Верифікація

c) Збір даних

d) Інтерпретація результатів

66. Що відображає модель типу «життєвий цикл»?

a) Стадії розвитку економічного показника у часі

b) Виключно короткострокові зміни у виробництві

c) Тільки фінансові показники підприємства

d) Залежність між витратами й прибутками

67. Які основні фази життєвого циклу об’єкта?

a) Зародження, розвиток, занепад

b) Аналіз, синтез, перевірка

c) Початок, середина, кінець

d) Планування, виконання, контроль

68. У чому полягає практична цінність моделі життєвого циклу?

a) У можливості передбачити прибуток компанії

b) У відображенні закономірностей розвитку та прогнозуванні змін

c) У визначенні вартості активів

d) У заміні економічного аналізу

69. Який метод найчастіше використовується для побудови динамічних економічних моделей?

a) Метод статистичних порівнянь

b) Математичне моделювання

c) Метод аналогій

d) Логічний аналіз

70. Основна мета моделювання соціально-економічних систем полягає у:

a) Заміні реальної системи

b) Аналізі й прогнозуванні розвитку процесів

c) Зменшенні кількості економічних змінних

d) Виявленні минулих тенденцій

71. Що таке аналітичне вирівнювання тренда?

a) Побудова прогнозу без використання даних

b) Метод, що полягає у згладжуванні динамічного ряду за допомогою аналітичної функції

c) Виключно експертна оцінка майбутніх тенденцій

d) Використання середніх значень без аналізу

72. Мета аналітичного вирівнювання тренда полягає у:

a) Виявленні випадкових відхилень

b) Визначенні основної тенденції розвитку процесу

c) Заміщенні статистичного аналізу

d) Формуванні якісних оцінок

73. Яка функція найчастіше використовується для аналітичного вирівнювання тренда?

a) Логарифмічна

b) Поліноміальна

c) Кусково-лінійна

d) Степенева

74. Метод перевірки різниць середніх рівнів застосовують для:

a) Перевірки адекватності моделі тренда

b) Побудови експертних оцінок

c) Порівняння прогнозів із реальними даними

d) Аналізу автокореляції

75. Який показник розраховується при методі перевірки різниць середніх рівнів?

a) Середня похибка прогнозу

b) Середнє квадратичне відхилення

c) Коефіцієнт варіації

d) Коефіцієнт детермінації

76. У чому суть методу Форстера-Стюарта?

a) У порівнянні прогнозів із фактичними результатами

b) У визначенні випадкових коливань динамічного ряду

c) У перевірці мультиколінеарності

d) У побудові регресійної моделі

77. Що аналізується за допомогою методу Форстера-Стюарта?

a) Тільки середнє значення

b) Ознаки випадковості залишків

c) Лише детермінована тенденція

d) Структурні зрушення у тренді

78. У чому полягає сутність методу Ірвіна?

a) У перевірці правильності вибору трендової функції ✓

b) У пошуку середнього рівня

c) У визначенні коефіцієнта кореляції

d) У моделюванні життєвого циклу

79. Який критерій використовується в методі Ірвіна?

a) Коефіцієнт автокореляції

b) Критерій Ir (відношення різниці рівнів до середнього квадратичного відхилення)

c) Критерій Стьюдента

d) Критерій χ^2

80. До яких методів належить аналітичне вирівнювання тренда?

a) Якісних методів прогнозування

b) Кількісних методів прогнозування

c) Експертних методів прогнозування

d) Імітаційного моделювання

81. Що таке тенденція у часовому ряді?

a) Короткострокові випадкові коливання

b) Довгострокова спрямованість розвитку явища

c) Сезонні коливання показника

d) Разові зрушення у даних

82. Який спосіб використовується для встановлення наявності тенденції?

a) Побудова графіка часових рядів

b) Використання тільки середнього значення

c) Визначення коефіцієнта варіації

d) Застосування лише експертної оцінки

83. Яка особливість методів екстраполяції тенденції?

a) Використовують лише експертні оцінки

b) Базуються на припущенні збереження тенденції в майбутньому

c) Не потребують історичних даних

d) Використовуються лише для коротких рядів

84. На чому ґрунтується екстраполяція на основі аналітичних показників динаміки?

a) Використанні математичних характеристик ряду (середній приріст, темпи зростання)

b) Виявленні випадкових відхилень

c) Побудові експертних оцінок

d) Виключно графічному аналізі

85. Яка основна перевага екстраполяції на основі аналітичних показників?

a) Простота і можливість швидко отримати прогноз

b) Висока точність у будь-яких умовах

c) Використання без попередньої обробки даних

d) Повна незалежність від історичних рядів

86. У чому полягає метод плинної середньої?

a) У використанні фіксованого значення середнього

b) У згладжуванні випадкових коливань за допомогою середніх значень

c) У визначенні тільки тренда за останній період

d) У використанні винятково експертних оцінок

87. Який недолік методу плинної середньої?

a) Втрата інформації про короткострокові коливання

b) Неможливість застосування до часових рядів

c) Використання занадто складних формул

d) Непридатність для сезонних даних

88. Для чого використовується екстраполяція на основі індексу сезонності?

a) Для виявлення разових зрушень

b) Для врахування періодичних коливань у часовому ряді

c) Для перевірки кореляційної залежності

d) Для заміни трендової складової

89. Як розраховується індекс сезонності?

a) Як відношення фактичного рівня до трендового значення

b) Як середнє арифметичне ряду

c) Як різниця між максимальним і мінімальним значенням

d) Як похідна трендової функції

90. До яких методів належать екстраполяція на основі плинної середньої та індексу сезонності?

a) Якісних методів прогнозування

b) Кількісних методів прогнозування

c) Експертних методів прогнозування

d) Імітаційних методів прогнозування

91. Що є основною метою прогнозування соціального розвитку?

a) Формування державного бюджету

b) Визначення потреб у виробництві товарів

c) Забезпечення стабільності рівня життя населення

d) Контроль за міграційними процесами

92. Який метод найчастіше застосовують при прогнозуванні рівня життя населення?

a) Метод сценаріїв

b) Метод проб і помилок

c) Метод експертних оцінок у біології

d) Метод контрольних зрізів

93. Прогнозування попиту на товари та послуги є важливою частиною:

a) Банківської діяльності

b) Маркетингових досліджень

c) Аграрної статистики

d) Митного регулювання

94. Що враховується під час побудови демографічного прогнозу?

a) Виключно економічні показники

b) Лише рівень народжуваності

c) Народжуваність, смертність та міграція

d) Тільки чисельність працездатного населення

95. Яка модель використовується для прогнозування чисельності населення?

a) Лінійна модель пропорцій

b) Екстраполяційна модель

c) Балансова модель капіталу

d) Модель оптимізації витрат

96. Прогнозування параметрів руху населення охоплює:

a) Динаміку освітнього рівня

b) Показники народжуваності, смертності та міграції

c) Економічні показники промисловості

d) Фінансові витрати на охорону здоров’я

97. Який вид прогнозу найчастіше використовується для оцінки змін чисельності населення у найближчі роки?

a) Короткостроковий

b) Середньостроковий

c) Довгостроковий

d) Індивідуальний

98. Що є ключовим у прогнозуванні попиту на послуги?

a) Вартість земельних ресурсів

b) Поведінка споживачів

c) Рівень інфляції у світі

d) Обсяг державного боргу

99. Соціально-економічні прогнози будуються для:

a) Підготовки статистичних збірників

b) Обґрунтування стратегій розвитку суспільства

c) Митних перевірок

d) Оцінки ефективності одного підприємства

100. Який фактор найбільше впливає на довгострокові демографічні прогнози?

a) Природний приріст

b) Валютний курс

c) Обсяги експорту

d) Сезонність попиту

101. Яка з моделей належить до наївних у прогнозуванні часових рядів?

a) Прогноз дорівнює середньому значенню за період

b) Прогноз дорівнює попередньому фактичному значенню

c) Прогноз ґрунтується на середньому зваженому показнику

d) Прогноз визначається за трендовою лінією

102. Який із способів застосовується для усунення тренда з часового ряду?

a) Дисконтування

b) Центрування

c) Кластеризація

d) Індексація

103. Модель ковзного середнього використовується для:

a) Врахування сезонності

b) Усунення короткострокових коливань

c) Побудови лінії регресії

d) Аналізу дисперсії

104. Яка особливість експоненційно зваженого ковзного середнього?

a) Усі спостереження мають однакову вагу

b) Нові спостереження отримують більшу вагу

c) Використовуються тільки перші дані ряду

d) Прогноз завжди збігається із середнім арифметичним

105. Яка мета комбінованих моделей згладжування?

a) Вибір максимально простого методу

b) Підвищення точності прогнозів

c) Усунення потреби у сезонних показниках

d) Зменшення кількості даних для аналізу

106. Як визначають початкові значення для моделей згладжування?

a) Використовують середнє перших спостережень

b) Задають випадковим чином

c) Підбирають на основі останнього значення

d) Формують із даних майбутніх періодів

107. Яка модель належить до моделей згладжування з трендом?

a) Модель ковзного середнього

b) Модель Брауна

c) Наївна модель

d) Комбінована модель

108. У чому полягає суть моделі Холта?

a) Вона враховує лише сезонність

b) Вона враховує рівень і тренд часового ряду

c) Вона усуває всі коливання з ряду

d) Вона базується на простому середньому

109. Яке основне завдання сезонних моделей прогнозування?

a) Усунути довгострокові тренди

b) Врахувати циклічні коливання, що повторюються

c) Побудувати регресійне рівняння

d) Зменшити обсяг статистичних даних

110. Яка ключова різниця між моделлю Холта та моделлю Брауна?

a) Холт враховує сезонність, а Браун – ні

b) Холт використовує два параметри згладжування, а Браун – один

c) Браун враховує тренд, а Холт – ні

d) У моделі Брауна не потрібні початкові значення

111. Що таке ковзне середнє у прогнозуванні?

a) Метод визначення стандартного відхилення

b) Метод усереднення ряду шляхом використання певної кількості останніх спостережень

c) Метод побудови регресійного рівняння

d) Метод розрахунку індексів сезонності

112. Яка основна мета використання ковзного середнього?

a) Усунути вплив тренда

b) Усунути випадкові коливання

c) Визначити циклічність економічних процесів

d) Зменшити вплив міграційних процесів

113. Чим відрізняється просте ковзне середнє від зваженого?

a) У простому кожне спостереження має однакову вагу

b) У простому враховується тренд, а у зваженому – ні

c) У простому використовуються всі дані, у зваженому – тільки частина

d) У простому середньому використовують прогнозні значення

114. Яка головна перевага методу ковзного середнього?

a) Простота розрахунків

b) Висока точність у будь-яких випадках

c) Можливість врахування сезонності

d) Універсальність для всіх типів часових рядів

115. До якого класу методів належить ковзне середнє?

a) Економетричні моделі

b) Методи згладжування часових рядів

c) Методи дисперсійного аналізу

d) Методи кластеризації

116. Що відбувається зі згладженим рядом при збільшенні довжини вікна ковзного середнього?

a) Ряд стає більш деталізованим

b) Ряд стає більш згладженим

c) Ряд втрачає інформацію про сезонність

d) Ряд зберігає всі випадкові коливання

117. Який недолік має метод ковзного середнього?

a) Складність математичного апарату

b) Втрата крайніх значень ряду

c) Неможливість використання для економічних показників

d) Надмірна чутливість до шумів

118. Зважене ковзне середнє на відміну від простого:

a) Призначає більшій давності більшу вагу

b) Призначає новішим даним більшу вагу

c) Використовує лише прогнозні значення

d) Ігнорує початкові значення ряду

119. Які часові ряди найбільш доцільно згладжувати ковзним середнім?

a) Ті, що мають виражений сезонний компонент

b) Ті, що містять випадкові коливання

c) Ті, що описуються рівнянням регресії

d) Ті, що мають постійний тренд

120. Який метод згладжування є подальшим розвитком ковзного середнього?

a) Модель регресії

b) Експоненційне згладжування

c) Метод найменших квадратів

d) Кластерний аналіз

121. Що є основою експертних методів прогнозування?

a) Виключно статистичні дані

b) Математичне моделювання

c) Судження фахівців у певній галузі

d) Історичні документи

122. Який принцип лежить в основі формування експертних систем прогнозування?

a) Автоматизація всіх процесів

b) Використання інтуїтивних рішень

c) Систематизація та узагальнення знань експертів

d) Виключення людського фактора

123. До індивідуальних експертних методів належить:

a) Метод «Дельфі»

b) Анкетування одного спеціаліста

c) Колективна дискусія

d) Метод мозкового штурму

124. Який метод належить до колективних експертних?

a) Експертне інтерв’ю

b) Анкетування одного експерта

c) Метод «Дельфі»

d) Особистий прогноз

125. Перший етап проведення колективної експертної оцінки – це:

a) Опрацювання результатів

b) Вибір експертів

c) Формування сценаріїв

d) Аналіз узагальнених даних

126. Що відрізняє метод «Дельфі» від звичайного анкетування?

a) Анонімність та багатоетапність опитування

b) Використання тільки письмових відповідей

c) Проведення у формі відкритої дискусії

d) Одноразовий збір інформації

127. Колективні експертні методи використовуються для:

a) Прогнозування лише в економіці

b) Узгодження різних точок зору та отримання об’єктивніших прогнозів

c) Зменшення кількості експертів

d) Уникнення статистичної обробки

128. Який недолік індивідуальних експертних методів?

a) Висока вартість

b) Суб’єктивність результатів

c) Неможливість застосування в соціології

d) Тривалість збору даних

129. На якому етапі колективної експертної оцінки відбувається узагальнення думок спеціалістів?

a) Вибір експертів

b) Аналіз і обробка результатів

c) Формулювання проблеми

d) Проведення першого опитування

130. Який метод найчастіше застосовується як колективний експертний для стратегічних прогнозів?

a) Метод аналогій

b) Метод «Дельфі»

c) Експоненційне згладжування

d) Регресійний аналіз

131. Що означає поняття «оптимальний прогноз»?

a) Прогноз, який максимально враховує сезонність

b) Прогноз із мінімальною похибкою за заданих умов

c) Прогноз, що базується лише на експертних оцінках

d) Прогноз із найбільш складною математичною моделлю

132. Який показник найчастіше використовують для оцінювання адекватності прогнозованої моделі?

a) Середньоквадратичне відхилення

b) Медіана ряду

c) Мода

d) Коефіцієнт варіації сезонності

133. Що входить до критеріїв якісного прогнозу?

a) Простота у використанні

b) Точність, надійність і обґрунтованість

c) Виключно високий рівень деталізації

d) Використання експертних оцінок

134. Який критерій прогнозу характеризує його здатність відтворювати реальні процеси?

a) Суб’єктивність

b) Адекватність

c) Математична складність

d) Стабільність алгоритму

135. Який метод використовується для оцінки точності прогнозів?

a) Метод середніх абсолютних відхилень

b) Метод експертних оцінок

c) Метод кластеризації

d) Метод побудови трендових рівнянь

136. Якщо прогнозована модель не відповідає фактичним даним, то вона є:

a) Оптимальною

b) Неадекватною

c) Стабільною

d) Узагальненою

137. Що означає висока надійність прогнозу?

a) Його зручність у використанні

b) Стійкість результатів до зміни вихідних даних

c) Найнижчу вартість розрахунків

d) Використання лише лінійних моделей

138. Який показник точності прогнозу визначається як середнє відхилення прогнозних значень від фактичних?

a) Середня похибка прогнозу

b) Мода

c) Варіаційний ряд

d) Дисперсія тренду

139. Чим відрізняється якісний прогноз від неякісного?

a) Використанням складних моделей

b) Мінімальною похибкою та адекватністю моделі

c) Максимальною кількістю даних

d) Виключно експертним методом

140. Який критерій прогнозу оцінює ступінь його відповідності реальним умовам майбутнього розвитку?

a) Обґрунтованість

b) Формальність

c) Складність

d) Суб'єктивність

141. В якому модулі системи STATISTICA можна знайти параметри лінійної функції?

a) Multiply regretion

b) Factor analysis

c) Basic statistic and tables

d) Nonlinear estimation

142. Параметр функції вважається значущим, якщо:

a) p-level>0.05

b) p-level=0.05

c) p-level<0.05

143. Скільки параметрів має класична модель Кобба-Дугласа?

a) 1

b) 2

c) 3

d) 4

144. Скільки параметрів має однофакторна модель?

a) 1

b) 2

c) 3

d) 4

145. Що показує значення коефіцієнта кореляції «-0,96»?

a) Зв’язок між двома факторами є обернено-пропорційний і значущий

b) Зв’язок між двома факторами є прямо-пропорційний і значущий

c) Зв’язок між двома факторами є прямо-пропорційний і незначущий

146. Щоб обчислити прогнозні значення на 2 роки наперед потрібно у системі STATISTICA зайти у меню:

a) Statistics-claster analysis

b) Statistics-descriptive analisys- factor analysis

c) Statistics-Basic statistic and tables-descriptive statistics

d) Statistics-advansed linear/nonlinear-time series and forecasting-ecsponensial smoothing and forecasting

147. Метод двох крайніх точок – це:

a) Метод прогнозування для значень показників якого-небудь ряду усередині періоду спостережень між двома крайніми точками

b) Метод прогнозування по прямій, проведеній через дві крайні точки

c) Метод прогнозування на основі інтерполяційного полінома

148. Методи простої екстраполяції засновані на припущенні про:

a) Змінний характер процесу, що протікає

b) Практично незмінний характер процесу

c) Постійні зміни у стані зовнішнього та внутрішнього середовища об’єкта прогнозування

149. Що відбувається на першому етапі математичного моделювання?

a) Обирається «еквівалент» об’єкта, що відображає в математичній формі найважливіші його властивості

b) Вибір алгоритму для реалізації моделі на комп’ютері

c) Створюються програми, що «переносять» модель і алгоритм на доступну комп’ютерну мову

150. В чому полягає метод ковзного середнього?

a) Прогноз будь-якого періоду являє собою середній показник кількох попередніх елементів часового ряду

b) Прогноз будь-якого періоду являє собою диференційований показник кількох попередніх елементів часового ряду

c) прогноз будь-якого періоду являє собою детермінований показник кількох попередніх елементів часового ряду

151. Якщо коефіцієнт кореляції r=0,9, то зв'язок між змінними:

a) Слабкий

b) Середній

c) Сильний

d) Відсутній

152. При кореляційному аналізі отримано r=−0,75. Це означає:

a) Сильний прямий зв’язок

b) Сильний обернений зв’язок

c) Слабкий прямий зв’язок

d) Зв’язку немає

153. Якщо коефіцієнт детермінації R^2=0,64, то це означає, що модель пояснює:

a) 36% варіації залежної змінної

b) 64% варіації залежної змінної

c) 16% варіації залежної змінної

d) 80% варіації залежної змінної

154. Коефіцієнт кореляції дорівнює r=0,05. Це означає:

a) Майже відсутній зв’язок

b) Сильний зв’язок

c) Середній зв’язок

d) Обернений зв’язок

155. Якщо рівняння регресії має вигляд y=2+3x, то при x=4 прогнозоване значення y:

a) 10

b) 12

c) 14

d) 20

156. У рівнянні регресії y=5−0,5x коефіцієнт −0,5 означає:

a) Залежна змінна зростає зі зростанням x

b) Залежна змінна зменшується зі зростанням x

c) Зв’язку немає

d) Це константа

157. При коефіцієнті кореляції r=−1 можна сказати:

a) Ідеальний прямий зв’язок

b) Ідеальний обернений зв’язок

c) Відсутній зв’язок

d) Слабкий зв’язок

158. Якщо r^2=0,81, то коефіцієнт кореляції може дорівнювати:

a) 0,9 або -0,9

b) 0,81

c) 0,64

d) 0,1

159. Якщо коефіцієнт регресії дорівнює 4, то це означає:

a) При зміні x на 1, y змінюється на 4

b) При зміні x на 4, y змінюється на 1

c) Зв’язку немає

d) Це коефіцієнт кореляції

160. Для двох експертів ранги об’єктів збігаються повністю. Яке значення коефіцієнта кореляції Спірмена?

a) $r_s=0$

b) $r_s=-1$

c) $r_s=1$

d) $r_s=0$

161. Два експерти дали абсолютно протилежні ранги (повністю обернені). Яке значення коефіцієнта Спірмена?

a) $r_s=1$

b) $r_s=-1$

c) $r_s=0$

d) $r_s=0,25$

162. Якщо коефіцієнт кореляції Кендалла $\tau=0,6$, то це означає:

a) Зв’язок слабкий

b) Зв’язок середньої сили

c) Зв’язок сильний

d) Зв’язку немає

163. При оцінюванні 5 об’єктів трьома експертами отримано коефіцієнт конкордації W=0,9. Це означає:

a) Дуже слабка узгодженість думок

b) Середня узгодженість думок

c) Висока узгодженість думок

d) Взагалі немає узгодженості

164. Якщо при обчисленні коефіцієнта конкордації отримано W=0, то це означає:

a) Абсолютна узгодженість

b) Відсутність узгодженості

c) Часткова узгодженість

d) Ідеальний обернений зв’язок

Рекомендована література:

1. Артеменко В.Б. Моделювання и прогнозування економічних рядів динаміки. - Львів: Вид-во Львівської комерційної академии, 2003. - 228 с.
2. Варфоломєєв В.І. Алгоритмічне моделювання елементів економічних систем: Практикум: Навч. посібник. М .: Фінанси і статистика, 2000.
3. Василенко В.А. Ткаченко Т.І. Стратегічне управління. К .: ЦУЛ, 2003.
4. Вітлинський В. В. Моделювання економіки : навч. посіб. Київ : КНЕУ, 2003. – 408 с.
5. Вітлінський В. В. Моделювання економіки : навч. посіб. Київ : КНЕУ, 2003. 408 с.
6. Вітлінський В. В., Терещенко Т. О., Савіна С. С. Економіко-математичні методи та моделі: оптимізація : навч. посіб. Київ : КНЕУ, 2016. 303 с.
7. Вовк В.М., Дрогомірецька З.Б. Основі системного аналізу - Львів: Видавничий центр ЛНУ ім. Івана Франка, 2002.
8. Галущак М. П. Прогнозування соціально-економічних процесів : навч. посіб. Тернопіль : ТДТУ, 2009. 101 с.
9. Галущак М. П., Галущак О. Я., Кужда Т. І. Прогнозування соціально-економічних процесів: навчальний посібник для економічних спеціальностей. – Тернопіль: ФОП Паляниця, 2021. 160 с.
10. Галущак М. П., Кужда Т. І. Оцінювання соціально-економічних факторів інноваційного розвитку підприємства на основі прогнозування з інтервалом довіри. Галицький економічний вісник. – Тернопіль: П.П. Созанський А.М., 2010. № 2. С. 102-108.
11. Домарадзька Г.С., Гладун Т.М., Фещур Р.В.. Прогнозування і макроекономічне планування : Навч. посібник. Львів : «Магнолія 2006», 2007. 211 с.

12. Дослідження операцій в економіці : підруч. / О. І. Черняк та ін. ; за ред. О. І. Черняк. Миколаїв : МНАУ, 2020. 398 с.

13. Економіко-математичне моделювання : навч. посіб. / В. В. Вітлінський та ін. ; за заг. ред. В. В. Вітлінського. Київ : КНЕУ, 2008. 536 с.

14. Економіко-математичне моделювання : навч. посіб. / за ред. О. Т. Іващука. Тернопіль : ТНЕУ «Економічна думка», 2008. 704 с.

15. Економіко-математичне моделювання : навч. посіб. / Т. С. Клебанова та ін. Харків : ВД «Інжек», 2012. 352 с.

16. Єріна А. М. Статистичне моделювання та прогнозування : навч. посіб. Київ : КНЕУ, 2001. 170 с.

17. Здрок В.В., Лагоцький Т.Я., Паславська І.М. Моделювання економічної динаміки : практикум. Львів : «Магнолія 2006», 2013. 256 с.

18. Кулявець В.О. Прогнозування соціально-економічних процесів : навч. посіб. Київ : Кондор, 2007. 148 с.

19. Кулявець В.О. Прогнозування соціально-економічних процесів: Навч. посіб. / В.О. Кулявець. – К.: Кондор, 2009. – 194 с.

20. Лиса О.І. Аналіз методів статистичного прогнозування соціальноекономічних явищ. Соц.-екон. пробл. сучас. періоду України. Пробл. розв. пром. вир-ва регіону: зб. наук. пр. 2009. Вип. 4. С. 250-259.

21. Лук'яненко І. Г., Краснікова Л. І. Економетрика : підруч. Київ : Знання, 1998. 494 с.

22. Макроекономічне прогнозування та його принципи. Національна економіка : навч. посіб. / [В. І. Мельникова, О. П. Мельникова, Т. В. Сідлярук та ін.]. – 2-ге вид. перероб. та доп. К. : Центр учбової літератури, 2012. 248 с.

23. Моделі і методи соціально-економічного прогнозування : підруч. / В. М Геець та ін. 2-е вид., випр. Харьків : ВД «Інжек», 2008. 396 с.

24. Моделі і методи соціально-економічного прогнозування : підручник / В. М. Геєць, Т. С. Клебанова, О. І. Черняк та ін. – 2 вид., виправ. Харків : ВД "ІНЖЕК", 2008. 396 с.

25. Моделі і методи соціально-економічного прогнозування : підручник / В. М. Геєць, Т. С. Клебанова, О. І. Чкерняк та ін. 2-е вид., виправ. Харків : ВД «ІНЖЕК», 2008. 396 с.

26. Наконечний С. І., Терещенко Т. О., Романюк Т. П. Економетрія : підруч. вид 2-е, допов. та перероб. Київ : КНЕУ, 2000. 286 с.

27. Пілько А. Д. Прогнозування соціально-економічних процесів : конспект лекцій. Івано-Франківськ : Супрун В. П., 2011. 83 с.

28. Планування та прогнозування в умовах ринку. – Навчальний посібник / під ред. д.ф.н., проф. В. Г. Воронкової. К. : ВД «Професіонал», 2006. 608 с.

29. Полотай О.І. Побудова прогнозу сукупного інноваційного показника інформаційної безпеки соціально-економічної сфери. *Вісник ЛДУБЖД* : зб. наук. праць. Львів : ЛДУБЖД, 2023. № 27. С. 60–66.

30. Полотай О.І. Побудова та оцінка прогнозу показників інноваційного сегменту економічної безпеки України. *Інфраструктура ринку. Електронний науково практичний журнал*. 2018. Вип. 15. С. 209–215.

31. Полотай О.І., Кухарська Н.П. Прогнозування індексу захищеності комп'ютерних мереж за допомогою економетричних моделей. *Електронне фахове наукове видання «Кібербезпека: освіта, наука, техніка»*. 2025. Том 4 № 28: С. 365-374.

32. Присенко Г.В., Равікович Є.І. Прогнозування соціально-економічних процесів : навч. посіб. Київ : КНЕУ, 2005. 378 с.

33. Прогнозування соціально-економічних процесів : конспект лекцій / уклад.: О. В. Шебаніна та ін. Миколаїв : МНАУ, 2022. 95 с.

34. Прогнозування соціально-економічних процесів. Методичні рекомендації з вивчення дисципліни та виконання контрольних робіт для студентів заочної форми навчання / О. В. Шебаніна та ін. Миколаїв : МНАУ, 2015. 70 с.

35. Прогнозування та аналіз часових рядів. Методичні вказівки до практичних занять та самостійної роботи студентів спеціальності 051

«Економіка» освітня програма «Економічна кібернетика», «Економічна аналітика» / укл.: Юрченко М. Є. Чернігів: ЧНТУ, 2018. 88 с.

36. Сафронова В.М. Прогнозування та моделювання в соціальній роботі: Навч. посібник для студ. вищ. навч. закладів. М .: Видавничий центр «Академія», 2002.

37. Семяновський В.М. Прогнозування соціально-економічних процесів: Опорний конспект лекцій. К.: КНТЕУ, 2007. 85 с.

38. Статистика : навч. посіб. / за ред. О.В. Раєвнєвої. Харків : Вид. ХНЕУ, 2010. 520 с.

39. Ткачук Р.Л., Полотай О.І., Балацька В.С., Брич Т.Б., Кухарська Н.П. Моделювання захисту операційних систем від реалізації кібератак з використанням критерію Пірсона. *Вісник ЛДУЖД*. 2025. № 31. С. 117-125.

НАВЧАЛЬНЕ ВИДАННЯ

Полотай Орест Іванович
Ткачук Ростислав Львович
Ящук Валентина Ігорівна
Брич Тарас Богданович

МОДЕЛЮВАННЯ ТА ПРОГНОЗУВАННЯ В СОЦІАЛЬНІ СФЕРІ

Навчальний посібник

Літературний редактор **Галина Падик**

Технічний редактор
та ком'ютерна верстка **Маріанна Климус**
Ростислав Ткачук

Друк **Назарій Петролюк**

Підписано до друку 19.12.2025 р.
Формат 60x84/16. Гарнітура Times New Roman.
Друк цифровий. Папір офсетний.
Ум. друк. арк. 16,9

Друк ЛДУБЖД
79007, Україна, м. Львів, вул. Клепарівська, 35
Тел. /факс: (032)233-24-79
E-mail: vnrd@ldubgd.edu.ua
Свідоцтво суб'єкта видавничої справи ДК №7249 від 09.02.2021 р.