

І. С. БЛАГУН

**МОДЕЛЮВАННЯ ТА ПРОГНОЗУВАННЯ
СОЦІАЛЬНО-ЕКОНОМІЧНИХ СИСТЕМ**

Монографія

Івано-Франківськ
Прикарпатський національний університет імені Василя Стефаника
2024

УДК 519.23:519.233.5(075.8)

Б23

*Рекомендовано вченою радою Прикарпатського національного університету
імені Василя Стефаника, протокол № 01 від 30.01.2024 р.*

Рецензенти:

Дмитришин Л. І., доктор економічних наук, професор, Прикарпатський національний університет імені Василя Стефаника

Дзюбановська Н. В., доктор економічних наук, професор, Західноукраїнський національний університет

Благун І. С.

Б23 Моделювання та прогнозування соціально-економічних систем : монографія. Івано-Франківськ : Прикарпатський національний університет імені Василя Стефаника, 2024. 292 с.

ISBN 978-966-640-556-5

В монографії розглянуто теоретичні і методологічні аспекти моделювання та прогнозування соціально-економічних систем. Проаналізовано теоретичні положення їх функціонування та методи емпіричного дослідження. Розглянуто статистичні основи вибірових досліджень, особливості застосування випадкового відбору, способів формування вибірових сукупностей, методи визначення стандартних помилок та оцінки істотності вибірових характеристик. Визначено особливості статистичного моделювання з використанням нейронних мереж. Розглянуто методи прогнозування соціально-економічних систем, серед яких: експертні методи, адаптивного прогнозування та прогнозування часових рядів з використанням нейронних мереж.

УДК 519.23:519.233.5(075.8)

ISBN 978-966-640-556-5

© Благун І. С., 2024

© Прикарпатський національний університет
імені Василя Стефаника, 2024

ЗМІСТ

ПЕРЕДМОВА	4
РОЗДІЛ 1. ЕМПІРИЧНІ МЕТОДИ ДОСЛІДЖЕННЯ СОЦІАЛЬНО-ЕКОНОМІЧНИХ СИСТЕМ	5
1.1. Методи емпіричного дослідження соціально-економічних систем.....	5
1.2. Первинна і вторинна інформація у соціально-економічних системах ...	10
1.3. Збір первинної інформації за допомогою спостереження і опитування .	13
1.4. Експеримент як метод збору первинної інформації	17
РОЗДІЛ 2. СТАТИСТИЧНІ ОСНОВИ ВИБІРКОВОГО ДОСЛІДЖЕННЯ СОЦІАЛЬНО-ЕКОНОМІЧНИХ СИСТЕМ	22
2.1. Особливості застосування вибіркового випадкового відбору.....	22
2.2. Способи формування випадкових вибірових сукупностей.....	27
2.3. Визначення стандартних помилок і необхідного обсягу вибірки.....	30
2.4. Оцінка істотності вибірових характеристик.....	34
РОЗДІЛ 3. СТАТИСТИЧНЕ МОДЕЛЮВАННЯ СОЦІАЛЬНО-ЕКОНОМІЧНИХ СИСТЕМ	38
3.1. Кореляційний та регресійний аналіз	38
3.2. Статистичний аналіз і моделі часових рядів	64
3.3. Методи багатовимірного статистичного аналізу	114
3.4. Статистичний аналіз і моделювання з використанням нейронних мереж	163
РОЗДІЛ 4. ПРОГНОЗУВАННЯ СОЦІАЛЬНО-ЕКОНОМІЧНИХ СИСТЕМ.....	180
4.1. Експертні методи прогнозування	180
4.2. Методи адаптивного прогнозування	208
4.3. Багатофакторне прогнозування на основі класичної регресії та регресійних моделей на змішаних факторних множинах	254
4.4. Прогнозування часових рядів з використанням нейронних мереж	266
СПИСОК ВИКОРИСТАНИХ ДЖЕРЕЛ.....	285

ПЕРЕДМОВА

Сучасний світ є складним лабіринтом взаємопов'язаних між собою соціальних та економічних процесів, які формують соціально-економічну систему людства. Соціально-економічна система є важливим базисом функціонування сучасного суспільства. Від економічних ринків до політичних процесів, від соціокультурних взаємодій до екологічних викликів - усі ці аспекти створюють непередбачувані та динамічні умови, які детермінують наше суспільство. Соціально-економічна система визначає рівень життя людей, ефективність державних політик, стабільність ринків та багато інших аспектів нашого життя. В процесі своєї еволюції, соціально-економічна система може як створювати можливості, так і не передбачувані виклики. В цьому контексті, моделювання та прогнозування соціально-економічних систем виступає ключовим інструментом для розуміння та управління цими складними процесами.

Монографія присвячена дослідженню основних методів моделювання та прогнозування соціально-економічних систем, а саме: різноманітних підходів до побудови моделей, починаючи з емпіричних способів збору інформації та вибіркового спостереження, котрі дозволяють проводити систематизацію та обробку вхідних даних з метою перетворення їх на релевантну інформацію та закінчуючи методами прогнозування, які дозволяють передбачати можливі наслідки прийнятих раніше рішень та подій на основі нейронних алгоритмів.

Завдяки впровадженню інновацій, управлінню невизначеністю та використанню передових статистичних методів можна отримати цінну інформацію, що сприяє сталому розвитку та фінансовій стабільності економік країн.

РОЗДІЛ 1. ЕМПІРИЧНІ МЕТОДИ ДОСЛІДЖЕННЯ СОЦІАЛЬНО-ЕКОНОМІЧНИХ СИСТЕМ

1.1. Методи емпіричного дослідження соціально-економічних систем

Дослідження ринку ґрунтується на таких базових поняттях як «методологія», «метод» і «методика».

Методологія (від гр. *methodos* – спосіб, метод і *logos* – наука, пізнання) – це концептуальний виклад мети і змісту дослідження, сукупність принципів, методів і процедур логічної організації, які використовуються при вивченні об'єктів, явищ і процесів.

Методи дослідження – це прийоми, процедури, операції, які використовуються для розв'язання конкретних задач.

У найширшому трактуванні метод – це шлях, спосіб досягнення мети і завдань дослідження. Вибір методу дослідження визначається інформаційною базою, умовами і цілями дослідження.

Методика (технологія) дослідження – це сукупність підходів і способів проведення дослідження, правил використання методів дослідження для досягнення поставленої цілі. Вона визначає засоби, ресурси забезпечення і послідовність проведення дослідження.

Важливу роль у методології дослідження відіграють поняття “об'єкт дослідження” і “предмет дослідження”.

Об'єктом дослідження прийнято вважати те, на що орієнтована пізнавальна діяльність дослідника, тобто – це процес або явище, які породжують проблемну ситуацію і обрані для дослідження. Об'єктами дослідження ринку є тенденції його розвитку (структури, ємності, динаміки продажів, конкуренції, кон'юнктури, ризиків).

Об'єкти дослідження ринку, як правило, ділять на три групи:

- можливості підприємства (рівень конкурентоспроможності продукції та підприємства загалом, імідж підприємства, результати діяльності підприємства);

- фактори ринку (кон'юнктура ринку, місткість ринку, рівень попиту і пропозиції, конкуренція, динаміка продажів);

- суб'єкти ринку (споживачі, конкуренти, постачальники, посередники, кредитори, товари, ціни, форми і методи продаж).

Предмет дослідження – це значимі з точки зору вивчення окремі властивості і сторони об'єкта дослідження.

Виділяють два різновиди дослідження ринку:

- теоретичний – розроблення нових підходів до вирішення ринкових проблем, формулювання гіпотез і законів розвитку ринку, виведення логічних висновків;

- емпіричний – сукупність процедур, спрямованих на отримання достовірної інформації про реальний ринковий об'єкт (процес) на основі прямої або опосередкованої реєстрації подій, а також на встановлення та узагальнення соціальних фактів, їх систематизацію і класифікацію.

Потрібно зауважити, що емпіричне дослідження не є простим збором даних (який може бути тенденційним та суб'єктивним), а науковими процедурами, у рамках яких з використанням спеціальних методів і технологій здійснюється формування інформаційної бази про стан і тенденції розвитку ринку.

Існують три стадії добування емпіричних знань про ринок товарів і послуг:

- отримання знань у виді конкретних даних досліджу;

- первинна обробка даних досліджу, в результаті якої отримують складніші знання (знання про зв'язки між даними, що забезпечує розчленування даних на групи);

- узагальнення даних досліджуваної групи, у результаті чого отримують знання про закономірності в групах, що вважається вищою формою емпіричного знання.

У випадку емпіричного дослідження має місце взаємодія об'єкта і дослідника. Теоретичне дослідження не передбачає такої взаємодії. Об'єкт вивчається абстрагованим чином, використовуючи результати емпіричного дослідження.

Теоретичне та емпіричне дослідження взаємопов'язані і доповнюють одне одного. Теоретичне дослідження, удосконалюючи та розвиваючи понятійний апарат, відкриває нові можливості пояснення і передбачення фактів, формує напрями емпіричного дослідження. З іншої сторони, нові дані, отримані в результаті емпіричного дослідження, стимулюють розвиток теоретичних досліджень, ставлять нові задачі.

Теоретичні і емпіричні дослідження різняться отриманими результатами. На теоретичному рівні – це концепції, закони і закономірності, відкриття, винаходи, тощо. Результати емпіричного дослідження фіксуються у виді рекомендацій, тверджень, правил.

Емпіричне дослідження спрямоване безпосередньо на об'єкт дослідження і проводиться за допомогою методів:

- *спостереження;*
- *експеримент;*
- *вимірювання;*
- *порівняння.*

Особливості емпіричного і теоретичного досліджень наочніше проявляються при їх порівнянні (табл.1.1).

Таблиця 1.1

Порівняльна характеристика емпіричного і теоретичного дослідження

Рівень дослідження	Компонента дослідження		
	Мета дослідження	Засоби і методи дослідження	Результати дослідження
Емпіричний	Пошук актуальних знань (осмислених даних)	Спостереження. Вимірювання. Експеримент. Порівняння. Методи експертних оцінок	Актуальне значення (емпіричні дані)
	Узагальнення емпіричних даних (виявлення емпіричних законів)	Аналогія. Аналіз і синтез. Індукція і дедукція	Емпіричний закон
	Висунення припущень, щодо властивих об'єкту ознак (описових гіпотез)	Систематизація. Класифікація	Емпірична гіпотеза
Теоретичний	Побудова ідеалізованого об'єкта	Абстрагування. Ідеалізація. Формалізація. Моделювання	Аксиома. Поняття. Принципи. Ідеальні (теоретичні) моделі. Закони
	Побудова наукової теорії	Гіпотетико-індуктивний метод. Конструктивно-генетичний метод	Гіпотеза. Наукове поняття
	Обґрунтування і доведення наукової теорії	Логічні методи. Індукція і дедукція. Аналогія. Системний підхід	Концепт. Теорія

Основними складовими методики емпіричного дослідження ринку товарів і послуг вважаються:

- правила проведення вибіркового дослідження;
- методи збору первинних даних;
- методи обробки, аналізу та узагальнення емпіричних даних.

Методи обробки емпіричних даних можна поділити на такі групи:

- методи розвідувального аналізу даних і описової статистики – виявлення найзагальніших властивостей і тенденцій даних, систематизація і наглядне представлення даних, обчислення основних статистичних показників;

- методи кореляційно-регресійного аналізу – оцінка щільності і встановлення форми зв'язку між ознаками, оцінювання параметрів регресійних рівнянь, тестування адекватності моделей;

- методи кластерного і дискримінантного аналізу – класифікація багатовимірних статистичних сукупностей, перевірка існування відмінностей між групами, визначення ознак, які дають найбільший внесок у міжгрупові відмінності;

- факторний аналіз – опосередковане оцінювання латентних показників, зменшення кількості параметрів для моделювання об'єктів, заміна корельованого набору ознак на некорельований.

Основними характеристиками ефективності методів, які використовуються у процесі емпіричного дослідження, є:

- діагностична сила (роздільна спроможність) – спроможність методу диференціювати досліджувані об'єкти, розділяти їх на групи;

- надійність – спроможність методу забезпечити однакові результати при дослідженні однакових об'єктів в однакових умовах;

- об'єктивність – зведення до мінімально можливого рівня або уникнення суб'єктивного впливу дослідника на результати дослідження;

- репрезентативність – можливість перенесення результатів, отриманих для частини об'єктів (вибіркової сукупності) на всі об'єкти, що входять до даної групи (генеральної сукупності);

- ефективність – отримання необхідної інформації при мінімальних витратах часових і фінансових ресурсів.

Комплексна характеристика методу, яка вказує на його придатність до використання для досягнення поставленої цілі, називається валідністю методу (методики).

1.2. Первинна і вторинна інформація у соціально-економічних системах

Для того, щоб успішно працювати на ринку товарів і послуг, недостатньо виробляти якісну продукцію, необхідно володіти певною інформацією про ринкове середовище. Інформація вважається придатною для використання, якщо вона наділена такими властивостями:

- *актуальності* – необхідним ступенем своєчасності і вагомості відносно проблеми, що вивчається;
- *достовірності* – правдивого (без спотворення даних) відображення стану об'єкта або процесу, який досліджується;
- *повноти* – достатності для вирішення поставлених задач і проблем, прийняття правильних рішень;
- *порівнянності* – можливості зіставлення даних за рахунок єдності предмета дослідження, кола включених показників, методології проведення дослідження і методик вимірювання показників;
- *доступності* – стану інформаційних ресурсів, при якому користувачі, які мають право доступу, можуть використовувати їх безперешкодно;
- *релевантності* – відповідності наявної інформації саме досліджуваному об'єкту або процесу і саме у тій частці, яка складає предмет дослідження;
- *економічності* – використанні таких методів збору і обробки інформації, за яких витрати не перевищують результат впровадження.

Інформацію, яку використовують у процесі дослідження ринку товарів і послуг, поділяють на:

- *первинну* – інформацію, яка збирається вперше, отримується безпосередньо від джерела її виникнення для вирішення конкретних проблем;

- *вторинну* – інформацію, яка зібрана раніше для цілей, відмінних від дослідженої конкретної проблеми, і може бути частково використано для вирішення актуальної конкретної на даний момент часу проблеми.

Методи, спрямовані на отримання первинної інформації, називають методами *польових досліджень*, а вторинної - методами *кабінетних досліджень* (або методами роботи з документами).

Джерело первинної інформації – це безпосередньо сам об’єкт або суб’єкт, який продукує інформацію, що відповідає предмету дослідження. Прикладом джерел первинної інформації, яка використовується для дослідження ринку товарів і послуг, слугують споживачі продукції, постачальники сировини і матеріалів, канали розподілу продукції, торгові агенти, спеціальні аналітичні служби тощо.

До методів збору первинної інформації відносять:

- *спостереження;*
- *опитування;*
- *експеримент.*

Характерними позитивними сторонами первинної інформації є такі:

- збір нової і достовірної інформації відповідно до поставлених цілей дослідження;
- відома методика збору і контролю даних;
- відсутність протиречивих даних, отриманих з різних джерел;
- можливість забезпечення конфіденційності зібраної інформації (прихованні інформації від конкурентів).

До недоліків первинної інформації відносять:

- відносно високу вартість збору інформації;
- відносно тривалий період часу, необхідний для збору інформації;
- для збору та обробки даних часто вимагається залучення кваліфікованого персоналу.

Дослідження на основі вторинної інформації переважно є оглядовими, носять постановочний і описовий характер. Джерела вторинної інформації поділяються на:

- внутрішні – фінансова і статистична звітність підприємства;
- зовнішні – спеціалізована література за тематикою дослідження, статистичні періодичні видання, бази даних тощо.

До позитивних характеристик вторинної інформації відносять такі:

- відносно невелика вартість у порівнянні з первинною інформацією;
- швидкість отримання у порівнянні з первинною інформацією;
- можливість порівняння даних із різних джерел;
- джерела вторинної інформації можуть містити дані, які не можна отримати самостійно;
- формує повніше представлення про досліджувані проблеми.

Водночас вторинна інформація важко піддається оцінці з погляду надійності і достовірності через те, що не завжди вказується методика її отримання, часто буває застарілою і не повністю відповідає цілям дослідження.

Сьогодні основними джерелами вторинної інформації є комп'ютерні бази даних та інформаційні послуги дослідницьких маркетингових компаній.

Процес збору інформації ґрунтується на розробленні відповідної концепції, яка передбачає реалізацію таких процедур:

- визначення проблеми дослідження і факторів (чинників), що впливають на її вирішення;
- формулювання цілей дослідження, спрямованих на вирішення проблеми;
- визначення типів і характеристик даних, які необхідно зібрати;
- обґрунтування і вибір методики збору даних;
- обґрунтування і визначення необхідних матеріальних, фінансових і часових ресурсів.

1.3. Збір первинної інформації за допомогою спостереження і опитування

Метод збору первинної інформації про об'єкт дослідження, який ґрунтується на організованому і цілеспрямованому сприйнятті даних із фіксацією їх на якому-небудь носію, називають *спостереженням*.

За ступенем формалізації розрізняють:

- *структуроване спостереження* – вид спостереження, яке проводиться за чітко визначеною схемою, яка визначає форму фіксації і перелік параметрів, що підлягають реєстрації;

- *неструктуроване спостереження* – вид спостереження, яке передбачає реєстрацію подій і фактів, пов'язаних з об'єктом дослідження, без наперед визначеного плану дій.

Структуроване спостереження проводиться тоді, коли проблема дослідження апріорно визначена і передбачає розробку плану спостереження і форми реєстрації даних.

Неструктуроване спостереження проводиться з використанням лише загального плану, а результати фіксуються у довільній формі.

За способом сприйняття параметрів об'єкта дослідження розрізняють *особисті спостереження* (дані фіксуються самим спостерігачем) і *спостереження з використанням технічних засобів*.

За регулярністю проведення спостереження поділяють на:

- *одноразові* – проводяться лише один раз відповідно до поставленої цілі (у подальшому дана схема дослідження не використовується);

- *епізодичні* – регламент реєстрації параметрів об'єкта дослідження строго не встановлюється;

- *систематичні* – проводяться на регулярній основі, що забезпечує можливість визначення тенденції розвитку об'єкта дослідження.

Залежно від повноти охоплення одиниць досліджуваної сукупності розрізняють:

- *суцільне спостереження* – дослідження всіх без виключення елементів (одиниць) сукупності і, як наслідок, отримання вичерпної статистичної інформації;

- *несуцільне спостереження* – дослідження лише підмножини елементів загальної сукупності, в результаті якого можна отримати узагальнені характеристики всієї сукупності з деяким ступенем точності.

Суцільне спостереження ринку товарів і послуг застосовується у випадках, коли досліджувана сукупність складається із порівняно невеликої кількості одиниць спостереження, або у випадку необхідності отримання точної інформації, зокрема по кожній одиниці сукупності.

При розробленні методики проведення спостереження потрібно розрізнити поняття *одиниці спостереження* і *одиниці сукупності*. Одиниця спостереження – це джерело, з якого отримують інформацію. Окремі елементи сукупності, які є носіями реєстраційних ознак, називають одиницями (елементами) сукупності.

Розрізняють такі види несучільного спостереження:

- *монографічне* - спостереження окремих типових одиниць загальної сукупності з метою їх поглибленого і всестороннього вивчення;

- *основного масиву* – спостереження за одиницями сукупності, які за основною ознакою мають найбільшу питому вагу у загальній сукупності або вносять основний вклад у характеристику об'єкта дослідження;

- *вибіркове* – спостереження, яке ґрунтується на випадковому відборі одиниць із загальної сукупності.

Спостереження проводять на основі плану спостереження, який складається із двох частин – програмно-методологічної та організаційної.

Програмно-методологічна частина передбачає вирішення таких проблем:

- визначення об'єкта спостереження і одиниці спостереження;

- розроблення програми спостереження – переліку ознак, які підлягають реєстрації;

- розроблення формуляра (анкети) спостереження;
- складання інструкцій до заповнення формулярів (анкет);
- визначення часових характеристик проведення спостереження.

У другій частині плану спостереження представлені організаційні питання:

- мета і задачі спостереження;
- підготовчі роботи (підбір і навчання персоналу, складання списків одиниць сукупності тощо);
- порядок проведення спостереження;
- матеріально-технічне і фінансове забезпечення робіт.

Спосіб спостереження, який полягає у тому, що дані про об'єкт дослідження реєструються спеціально підготовленими особами при участі респондентів або безпосередньо самими респондентами, називають *опитуванням*.

У переважній більшості випадів використовують структуроване опитування – розробляється структурована анкета і питання ставляться у заздалегідь визначеному порядку. Питання типової анкети – це питання із заданим варіантом відповіді, коли респонденту пропонують вибрати відповідь із декількох запропонованих варіантів.

Респондент як джерело інформації займає особливе місце у процесі дослідження і може володіти або не володіти інформацією, необхідною для дослідження, розуміти або не розуміти зміст питання, бажати або не бажати давати відповіді на питання. Отримання достовірної і надійної інформації від респондентів забезпечується:

- використанням можливості проведення анонімного опитування;
- чіткою постановкою і правильним порядком формування питань;

- організацією інструктажу інтерв'юерів і ефективним контролем за їх роботою.

Сьогодні опитування вважається найрозповсюдженішим засобом збору первинної інформації про ринок товарів і послуг і реалізується у двох основних формах:

- *інтерв'ю* – збір необхідної інформації шляхом безпосередньої бесіди інтерв'юера з респондентом за схемою “питання - відповідь”;

- *анкетування* – опитування у письмовій формі (як правило, без прямого контакту інтерв'юера з респондентом, при цьому останній має змогу ґрунтовно продумати відповіді на поставлені питання).

Особливим видом опитування є панельне обстеження – такий спосіб збору первинної інформації, при якому протягом відносно тривалого відрізка часу періодично обстежується певна група суб'єктів (типових представників цільової аудиторії) на предмет їх відношення до того чи іншого питання або проблеми. Базовим поняттям при цьому є поняття “панелі” – частина одиниць загальної сукупності, відібраної для багатократного спостереження (опитування) за умови, що предмет дослідження не змінюється.

Елементами панелі слугують торгові і збутові організації, окремі експерти, споживачі тощо. За допомогою панельних обстежень виявляють фактори, які впливають на вирішення проблеми і визначають їх динаміку. Особливість панельного обстеження полягає в тому, що воно завдяки багатократного отримання інформації від одних і тих же суб'єктів, порівняння інформації, яка стосується різних моментів часу, дозволяє оцінити тенденцію розвитку об'єкта дослідження.

1.4. Експеримент як метод збору первинної інформації

Ще одним методом збору первинної інформації про ринок товарів і послуг є *експеримент* – метод збору первинної інформації шляхом активного втручання дослідника у керований процес з метою виявлення та оцінювання впливу незалежних факторів на залежні фактори за умови виключення впливу сторонніх факторів.

Для виконання експериментальних процедур необхідно визначити план експерименту:

- одиниці спостереження і їх поділ на однорідні підгрупи (об'єкти спостереження, реакція яких на незалежні фактори підлягає вивченню);
- незалежні фактори – фактори, значеннями яких маніпулює дослідник (ціна товару, якість упаковки, рівень сервісу, витрати на рекламу тощо);
- залежні фактори – фактори, значення яких формується під впливом незалежних факторів (розмір прибутку, обсяг реалізованої продукції, частка ринку тощо);
- сторонні (незаплановані, побічні) фактори – всі інші, крім незалежних факторів, які впливають на залежні фактори, але у даному експерименті не розглядаються;
- методику виключення впливу сторонніх факторів.

Необхідно розуміти, що одні і ті ж фактори у різних експериментах можуть виступати як незалежні, так і залежні фактори.

Отже, експеримент – це керований процес зміни одного або декількох незалежних факторів з метою вимірювання їх впливу на один або декілька залежних факторів при виключенні впливу сторонніх факторів.

Проведення експериментів ставить дві цілі:

- отримати достовірні оцінки впливу незалежних факторів на залежні фактори для сукупності одиниць, які приймають участь в експерименті;
- зробити достовірні висновки щодо всієї генеральної сукупності.

Якщо досягається перша ціль, то має місце внутрішня достовірність результатів експерименту, тобто зміни залежних факторів дійсно спричинені зміною незалежних факторів. Внутрішня достовірність є мінімальною вимогою до проведення експериментів. Досягнення другої цілі називають зовнішньою достовірністю результатів експерименту. У цьому випадку виявлені у процесі експерименту причинно-наслідкові зв'язки піддаються узагальненню.

Необхідною умовою забезпечення внутрішньої і зовнішньої достовірності результатів експерименту є контроль сторонніх факторів. До основних сторонніх (побічних) факторів відносять:

- історичні фактори – зовнішні по відношенню до експерименту специфічні події, які відбуваються одночасно з ним;
- фактори зрілості – фактори, пов'язані зі змінами, які відбуваються в одиницях спостереження протягом проведення експерименту;
- фактори ефекту тестування – фактори впливу попереднього спостереження на результати наступного спостереження;
- фактори вибуття – фактори, пов'язані з вибуттям частини одиниць спостереження під час експерименту;
- фактори похибки вибірки – фактори невірному вибору складу одиниць спостереження.

Існують такі способи контролю сторонніх факторів:

- випадковий відбір одиниць із генеральної сукупності для участі в експерименті;
- групування одиниць генеральної сукупності на основі деяких ознак, після чого з кожної групи вибираються елементи для проведення експерименту;
- статистичний контроль, який ґрунтується на вимірюванні впливу сторонніх факторів за допомогою статистичних методів;

- методичний контроль, який забезпечується створенням спеціальних методик для проведення експерименту.

Типовими ситуаціями, які потребують експериментальних досліджень, вважаються:

- необхідність виявлення нових (невідомих) властивостей об'єкта, які не витікають із наявних на даний час знань про нього%

- перевірка достовірності тих чи інших теоретичних положень.

Розрізняють два види експериментів:

- *лабораторні* – експерименти, які проводяться у штучно створених умовах;

- *натурні (польові)* – експерименти, які проводяться в реальних умовах.

Прикладом лабораторних експериментів може слугувати імітація поведінки споживачів (вибір товарів у супермаркетах, оцінювання якості товарів тощо) з використанням сучасних комп'ютерних технологій. Типовим прикладом проведення натурного експерименту на ринку товарів і послуг є пробні продажі окремих видів продукції різним цільовим групам споживачів.

Існують певні обмеження щодо проведення натурних експериментів:

- за часом – як правило, натурні експерименти вимагають тривалого періоду часу для встановлення тенденції розвитку об'єкта дослідження і виявлення взаємозв'язків між його параметрами;

- за матеріальними і фінансовими витратами – у переважній більшості натурні експерименти обходяться дорого;

- за можливостями реалізації – багато експериментів складно або неможливо провести у польових умовах.

Експеримент як метод збору первинної інформації має свої переваги і недоліки у порівнянні із спостереженням і опитуванням.

До переваг експерименту відносять:

- високу об'єктивність отриманої інформації, так як результати експерименту - це фактичні дані і події;
- можливість вивчення причинно-наслідкових зв'язків між факторами;
- можливість вивчення явища у “чистому” виді, тобто сторонні фактори, які спотворюють основний процес, підлягають вилученню;
- можливість перевірки ефективності вироблених рішень;
- можливість контролювати оточуюче середовище.

Недоліками експерименту вважаються:

- висока трудомісткість, великі витрати часових і фінансових ресурсів;
- складність організації і забезпечення необхідних умов для проведення експериментів;
- невпевненість у застосуванні результатів експерименту за інших умов оточуючого середовища.

Основні положення та вказані моделі використовувалися при аналізі:

- міжрегіональної диференціації грошових доходів населення за регіонами України, через призму інтегральних показників, внаслідок чого було оцінено загальні тенденції в просторовому розподілі грошових доходів населення. На основі використання методів порівняльного і непорівняльного шкалювання розроблено напрями і механізми зниження міжрегіональної диференціації, які вимагали детального аналізу соціально-економічного стану перш за все тих регіонів України, які мають значний внесок при диференціації за рівнем грошових доходів населення.

Благун І. С., Дмитришин Л. І. Просторово-структурний аналіз доходів населення. Регіональна економіка. 2013. № 1 (67). С. 99–106.

- комплексних підходів щодо визначення рівня інтернаціоналізації суб'єктів економічної діяльності. Розглянуто методики розрахунку окремих груп показників оцінювання рівня інтернаціоналізації підприємств, визначено переваги та недоліки їхнього використання. Здійснено апробацію

застосування окремих груп показників для оцінювання рівня інтернаціоналізації суб'єктів економічної діяльності як для країн з економіками, що розвиваються, так і для країн з перехідними економіками, а також найбільших корпорацій світу. Для визначення рівня інтернаціоналізації суб'єктів економічної діяльності використовувалися комплексні показники, котрі розраховуються на основі рівня концентрації, диверсифікації та інших характеристик видів діяльності у ланцюгу вартості.

Благул І. С., Ільчук П. Г. Методи оцінювання рівня інтернаціоналізації підприємств. Економіка і прогнозування. 2014. № 4. С. 97–109.

- концепції рейтингового управління регіональними економічними системами, що забезпечують розробку взаємозв'язаних управлінських рішень в процесі аналізу об'єктів рейтингування. Встановлено, що складність аналізу полягає в тому, що для формування інформаційної бази використовується велика кількість взаємозв'язаних і, нерідко, різновекторних показників. Розроблено автором алгоритмічну схему та сценарії реалізації рейтингового управління в регіонах України на основі принципів гнучкого управління, що дозволяє врахувати особливості трансформаційних процесів в контексті інформаційної економіки.

Благул І. С., Боднарук І.Л. Теоретичні основи рейтингового управління в регіональних економічних системах. Інноваційна економіка. 2015. № 5 (60). С. 124–130.

РОЗДІЛ 2. СТАТИСТИЧНІ ОСНОВИ ВИБІРКОВОГО ДОСЛІДЖЕННЯ СОЦІАЛЬНО-ЕКОНОМІЧНИХ СИСТЕМ

2.1. Особливості застосування вибіркового випадкового відбору

Загальну сукупність однорідних об'єктів у статистиці називають *генеральною сукупністю*. Відібрана для дослідження частина генеральної сукупності має назву вибірки або *вибіркової сукупності*.

Способи відбору і статистичного оброблення відносно невеликої кількості одиниць із генеральної сукупності, які забезпечують формування висновків про характеристики генеральної сукупності, складають *методику вибіркового дослідження*.

Вибіркові методи дослідження поділяють на дві групи:

- *детерміновані* – вибір одиниці із генеральної сукупності проводиться на підставі індивідуальних висновків дослідника;
- *випадкові (ймовірнісні)* – одиниці генеральної сукупності попадають до вибірки з певною ймовірністю.

До детермінованих методів відбору належить:

- *нетиповий(нерепрезентативний) відбір* – дослідник формує вибірку із зручних, доступних для відбору одиниць генеральної сукупності;
- *експертний відбір* – різновид нетипового відбору, при якому одиниці вибирають із генеральної сукупності на основі деяких суджень дослідника;
- *квотний (груповий) відбір* – двохетапний відбір (на 1-му етапі створюються контрольні групи, на 2-му – використовують типовий або експертний відбір);
- *метод “сніжної лавини”* – спочатку, як правило, випадковим чином підбирають групу респондентів, опитують їх і просять допомогти виявити інших кандидатів для дослідження.

Основна сфера використання нетипового відбору – пошукові дослідження, спрямовані на розробку нових ідей і гіпотез. Нетиповим відбором не рекомендують користуватися при вивченні причинно-наслідкових зв'язків, при дослідженнях, які передбачають знаходження параметрів усієї генеральної сукупності.

Недолік експертного відбору – суб'єктивний характер формування вибірки, а тому його ефективність залежить від компетентності дослідника. Експертний відбір не дозволяє узагальнювати і розповсюджувати отримані результати на генеральну сукупність. Експертний відбір часто слугує засобом визначення складу учасників групових дискусій (фокус-груп).

Необхідною умовою застосування квотного відбору є чітке розуміння характеристик респондентів, поведінка яких підлягає дослідженню. Застосування квотного відбору повинно ґрунтуватися на попередньому вивченні генеральної сукупності з метою виявлення пропорцій одиниць з різними характеристиками і зв'язками.

Метод “сніжної лавини” застосовується у випадках, коли представників досліджуваної сукупності важко вибрати іншими методами. Найчастіше його використовують для фокус-груп і глибинних інтерв'ю. Так як респонденти вибираються за рекомендацією, то забезпечується повнота і достовірність відповідей. Проте значне розкручування “лавини” може спотворювати вибірку (до вибірки можуть входити лише схожі респонденти).

Методи випадкового відбору мають ряд переваг перед суцільним спостереженням, оскільки потребують менших затрат часу і матеріальних ресурсів на дослідження явища, забезпечують менше помилок реєстрації (рис. 2.1).



Рис. 2.1. Особливості і приклади застосування вибіркового випадкового відбору

Однак вказані переваги випадкового відбору мають місце лише при дотриманні наукових принципів його проведення, що дає можливість отримати репрезентативну (представницьку) вибірку, яка відтворює характеристики генеральної сукупності. Вибірка вважається *репрезентативною*, якщо всі основні ознаки генеральної сукупності, з якої вона сформована, представлені приблизно з тією ж частотою (у тій же пропорції), що й у генеральній сукупності. Репрезентативна вибірка забезпечує можливість перенесення результатів дослідження, отриманих на її основі, на генеральну сукупність, з якої вона сформована.

Різниця між значенням деякого показника, отриманого на основі вибірки, і значенням цього показника в генеральній сукупності називається *помилкою вибірки*. Розрізняють *систематичну* і *випадкову помилки*. Перша виникає внаслідок порушення правил випадкового вибору елементів із генеральної сукупності, а друга – це результат випадкового вибору. Систематичні помилки небезпечніші, чим випадкові, але їх можна уникнути. Випадкові помилки носять об'єктивний характер і притаманні вибірковому спостереженню, а тому уникнути їх не вдається. Величина випадкової помилки залежить від обсягу вибірки, способу формування вибірки, коливання ознаки в генеральній сукупності.

У процесі вибіркового дослідження розглядають два типи вибіркових оцінок – точкові та інтервальні. *Точкові оцінки* параметрів розподілу генеральної сукупності визначаються одним числом (вибіркова середня, вибірка частка, вибірка дисперсія) і є не випадковими величинами.

Інтервальна оцінка – це довірчий інтервал (інтервал значень параметра для заданої ймовірності), який визначається двома числами (кінцями інтервалу). Вона дає змогу встановити точність та надійність точкових оцінок.

Нехай Q^* – це знайдена за даними вибірки точкова оцінка невідомого параметра Q . Тоді ймовірність $\gamma = P\{|Q - Q^*| < \Delta\}$, з якою виконується

нерівність $|Q - Q^*| < \Delta$, де Δ – число, що характеризує точність оцінки, будемо називати *довірчою ймовірністю* (надійністю) оцінки параметра Θ за Θ^* . Інтервал $(\Theta^* - \Delta; \Theta^* + \Delta)$ називається довірчим, якщо він покриває невідомий параметр із заданою надійністю γ . Кінці довірчого інтервалу є випадковими величинами.

Якщо генеральна сукупність має нормальний розподіл, то розподіл вибірових характеристик також буде нормальним. Навіть у випадку невідповідності розподілу генеральної сукупності нормальному розподілу для вибірок великого розміру ($n \geq 30$) розподіл вибірових середніх матиме приблизно нормальний розподіл.

У подальшому розгляді матеріалу, який стосується методів вибіркового дослідження ринку товарів і послуг, будемо користуватись позначеннями, представленими у табл.2.1.

Таблиця 2.1

Основні характеристики генеральної і вибіркової сукупності

Характеристики	Позначення	
	Генеральна сукупність	Вибіркова сукупність
Кількість одиниць	N	n
Кількість одиниць яким властива деяка ознака	M	m
Частка одиниць, яким властива деяка ознака	W	\tilde{W}
Середня величина ознаки	\bar{x}	\tilde{x}
Дисперсія ознаки	σ^2	$\hat{\sigma}^2$

Помилки, які виникають через не суцільний характер спостереження, поділяються на *стандартні* (середні) і *граничні*.

Стандартні і граничні помилки пов'язані між собою таким співвідношенням:

$$\Delta = t\mu, \tag{2.1}$$

де μ – стандартна (середня) помилка вибірки; Δ – гранична помилка вибірки; t – коефіцієнт довіри, який визначається на підставі заданого значення ймовірності.

При достатньо великій кількості незалежних спостережень імовірність того, що розбіжність між вибірковою і генеральною середніми або частками не перевищить величини $\Delta = t\mu$, рівна:

$$P = \{|\bar{x} - \tilde{x}| \leq t\mu\} = 2F(t), \quad (2.2)$$

де $F(t)$ – нормована функція Лапласа.

Відповідність окремих значень імовірності коефіцієнту довіри є такою:

$P = 0,683$	$P = 0,866$	$P = 0,954$	$P = 0,950$	$P = 0,997$
$t = 1,0$	$t = 1,5$	$t = 2,0$	$t = 1,96$	$t = 3,0$

Граничну помилку середньої позначимо через Δ_x , а частки - через Δ_w .
Тоді довірчі інтервали рівні:

- для середньої – $\tilde{x} - \Delta_x \leq \bar{x} \leq \tilde{x} + \Delta_x$; (2.3)

- для частки – $\tilde{W} - \Delta_w \leq W \leq \tilde{W} + \Delta_w$. (2.4)

2.2. Способи формування випадкових вибірових сукупностей

За способами відбору розрізняють:

- *повторний відбір* (відбір за схемою поверненої кулі) – одиниця, яка потрапила до вибірки, після реєстрації повертається до генеральної сукупності і зберігає однакову можливість з іншими одиницями знову потрапити до вибірки;

- *безповторний відбір* (відбір за схемою неповерненої кулі) – одиниця, яка потрапила до вибірки, в генеральну сукупність не повертається.

Розрізняють такі способи формування випадкових вибірових сукупностей:

- *простий випадковий відбір;*
- *механічний (систематичний) відбір;*
- *типовий (районований) відбір;*
- *серійний (кластерний) відбір.*

Простий випадковий відбір передбачає формування вибіркової сукупності жеребкуванням без попереднього розчленування генеральної сукупності на окремі групи. Цей спосіб вибіркового дослідження дає хороші результати, якщо між одиницями генеральної сукупності немає різких відмінностей. Метод простого випадково відбору забезпечує повне дотримання принципу випадковості і, як наслідок, – уникнення систематичних помилок.

При механічному відборі досліджують одиниці, які розташовані на однаковій відстані та у певній послідовності серед упорядкованої генеральної сукупності (в алфавітному порядку, за датами народження тощо). Для формування механічної вибірки треба задати початок, крок відліку і обсяг вибірки. Початок відліку може бути встановлений довільно або жеребкуванням, а крок відліку рівний $h=N/n$.

Якщо в генеральній сукупності одиниці розташовані випадковим чином (генеральна сукупність не впорядкована), то механічний відбір можна розглядати як різновид випадкового неповторного відбору.

Репрезентативність механічної вибірки залежить від способу групування об'єктів генеральної сукупності, зведених в єдиний список. Елементи з близькими значеннями кількісних ознак у списку повинні розташовуватися поруч, а елементи, значення яких різняться суттєво, – в різних частинах списку. Механічний відбір застосовувати не можна, якщо у списку елементів спостерігається природна періодичність.

Типовий відбір передбачає формування вибіркової сукупності на основі попередньої структуризації генеральної сукупності і незалежного відбору

елементів із кожної групи (типу, району). Групи можуть утворюватися штучно або використовуватися ті, що сформувалися природно.

Якщо сформовано K типових груп, а $N_i (i = \overline{1, K})$ – обсяг генеральної сукупності i -ї групи, то кількість одиниць, які потрапляють до загальної вибірки з i -ї групи за умови пропорційного відбору обчислюється за формулою:

$$n_i = n \frac{N_i}{N}. \quad (2.5)$$

За непропорційного відбору з кожної групи вибирається однакова кількість одиниць:

$$n_i = \frac{n}{K}. \quad (2.6)$$

Оптимальної вибірки досягають, вибираючи чисельність i -ї групи за такою формулою:

$$n_i = n \frac{N_i \sigma_i}{\sum N_i \sigma_i}, \quad (2.7)$$

де σ_i – середньоквадратичне відхилення i -ї групи.

Типова вибірка забезпечує більшу точність вибіркових статистик, чим проста випадкова вибірка. Крім того, використання типової вибірки дає змогу дослідити характеристики окремих підмножин.

При серійній вибірці випадковим чином формують групи (серії) одиниць із генеральної сукупності, які досліджують повністю. Серії можуть бути сформовані природно (регіони, партії продукції) або штучно. Серійний відбір має деякі переваги організаційного характеру.

Вибір того чи іншого способу формування вибіркової сукупності здійснюється з урахуванням можливостей організації дослідження і його мети. Найкращі результати отримують поєднуючи різні способи відбору.

Схема послідовності проведення вибіркового спостереження наведена на рис.2.2.



Рис. 2.2. Послідовність етапів вибіркового спостереження

2.3. Визначення стандартних помилок і необхідного обсягу вибірки

Значення граничної помилки вибірки залежить від значень коефіцієнта довіри і стандартної помилки вибірки. Для кожного способу формування вибіркової сукупності існують свої формули обчислення стандартної помилки вибірки.

Для вибірок великого розміру ($n > 30$) стандартні помилки середньої (μ_x) та частки (μ_w) при простому випадковому відборі за схемою неповерненої кулі визначають відповідно за такими формулами:

$$\mu_x = \sqrt{\frac{\hat{\sigma}^2}{n} \cdot \frac{N-n}{N}}; \quad (2.8)$$

$$\mu_w = \sqrt{\frac{\tilde{W}(1-\tilde{W})}{n} \cdot \frac{N-n}{N}}. \quad (2.9)$$

Для вибірових сукупностей, які отримані за схемою поверненої кулі, стандартні помилки розраховують так:

$$\mu_x = \sqrt{\frac{\hat{\sigma}^2}{n}}, \quad (2.10.)$$

$$\mu_w = \sqrt{\frac{\tilde{W}(1-\tilde{W})}{n}}. \quad (2.11.)$$

Якщо обробляють невеликі за обсягом вибірки ($n \leq 30$), до формул (2.8) – (2.11) вносять певні поправки. У випадку простого неповторного відбору стандартні помилки вибіркової середньої і вибіркової частки обраховується відповідно за формулами (2.12) і (2.13), а для відбору за схемою поверненої кулі – за формулами (2.14) і (2.15):

$$\mu_x = \sqrt{\frac{\hat{\sigma}^2}{n} \cdot \left(\frac{N-n}{N-1}\right)}; \quad (2.12)$$

$$\mu_w = \sqrt{\frac{\tilde{W}(1-\tilde{W})}{n} \left(\frac{N-n}{N-1}\right)}; \quad (2.13)$$

$$\mu_x = \sqrt{\frac{\hat{\sigma}^2}{n-1}}; \quad (2.14)$$

$$\mu_w = \sqrt{\frac{\tilde{W}(1-\tilde{W})}{n-1}}. \quad (2.15)$$

При цьому коефіцієнт довіри для заданого значення ймовірності та ступеня вільності $V = n - 1$ визначають за розподілом Стьюдента.

Зменшення помилки вибірки пов'язане із збільшенням чисельності вибірки. Враховуючи співвідношення $\Delta_x = t \cdot \mu_x$ і $\Delta_w = t \cdot \mu_w$, можна отримати формули для визначення обсягу вибірки, яка забезпечить

необхідну точність розрахунків показників генеральної сукупності на основі заданих помилок вибірки та їх ймовірностей.

У випадку безповторної вибірки використовують формулу:

$$n = \frac{t^2 \cdot \hat{\sigma}^2 \cdot N}{N \cdot \Delta_x^2 + t^2 \cdot \hat{\sigma}^2}, \quad (2.16)$$

а у випадку повторної вибірки – таку формулу:

$$n = \frac{t^2 \cdot \hat{\sigma}^2}{\Delta_x^2}. \quad (2.17)$$

Необхідну чисельність вибірки при дослідженні частки із заданою точністю для безповторної та повторної вибірки встановлюють відповідно за допомогою таких формул:

$$n = \frac{\tilde{W}(1-\tilde{W}) \cdot t^2 \cdot N}{N \cdot \Delta_w^2 + \tilde{W}(1-\tilde{W}) \cdot t^2}, \quad (2.18)$$

$$n = \frac{\tilde{W}(1-\tilde{W}) \cdot t^2}{\Delta_w^2}. \quad (2.19)$$

При визначенні необхідної чисельності вибірки величина $\hat{\sigma}^2$ є невідомою. Для обчислення її наближеного значення можна скористатися одним із наведених нижче способів:

а) якщо розподіл ознаки в генеральній сукупності нормальний, то можна прийняти

$$\hat{\sigma} \approx \frac{x_{max} - x_{min}}{6}; \quad (2.20)$$

б) для частки взяти максимальне значення дисперсії $\hat{\sigma}^2 = 0,25$;

в) провести пробне спостереження і знайти:

$$\hat{\sigma}^2 = \frac{\sum(x - \bar{x}_{проб})^2}{n_{проб} - 1}; \quad (2.21)$$

г) скористатися значенням дисперсії минулих спостережень.

Для механічного відбору стандартну помилку середньої і частки при впорядкованій генеральній сукупності обчислюють так само, як для безповторної простої вибірки.

Якщо генеральна сукупність невпорядкована, то механічний відбір розглядають як модифікацію повторної простої вибірки.

Стандартну помилку вибіркової середньої типової вибірки у випадку безповторного відбору для кожної групи обчислюють за формулою

$$\mu_x = \sqrt{\frac{\bar{\sigma}^2}{n} \left(1 - \frac{n}{N}\right)}, \quad (2.22)$$

де $\bar{\sigma}^2$ – середня з вибірових дисперсій середніх типових груп.

Для визначення помилки вибіркової частки використовують таку формулу:

$$\mu_w = \sqrt{\frac{\overline{\tilde{W}(1-\tilde{W})}}{n} \left(1 - \frac{n}{N}\right)}, \quad (2.23)$$

де $\overline{\tilde{W}(1-\tilde{W})}$ – середня із вибірових дисперсій часток типових груп.

Стандартна помилка вибіркової середньої типової вибірки менша за стандартну помилку простої випадкової вибірки або дорівнює їй, якщо міжгрупова дисперсія дорівнює нулю.

Загальну чисельність типової вибірки для розрахунку середньої і частки із заданою точністю встановлюють відповідно за формулами:

$$n = \frac{t^2 \cdot \bar{\sigma}^2 \cdot N}{\Delta_x^2 \cdot N + t^2 \cdot \bar{\sigma}^2}, \quad (2.24)$$

$$n = \frac{t^2 \cdot \overline{\tilde{W}(1-\tilde{W})} \cdot N}{\Delta_w^2 \cdot N + t^2 \cdot \overline{\tilde{W}(1-\tilde{W})}}, \quad (2.25)$$

У випадку безповторного відбору серій помилка вибіркової середньої і частки становить

$$\mu_x = \mu_w = \sqrt{\frac{\hat{\sigma}_M^2}{k} \left(1 - \frac{k}{K}\right)}; \quad (2.26)$$

де $\hat{\sigma}_M^2$ - міжсерійна дисперсія середніх або часток;

K - кількість серій генеральної сукупності;

k - кількість серій, які підлягають відбору.

Необхідну чисельність вибірки визначають за формулою:

$$k = \frac{t^2 \cdot \hat{\sigma}_M^2 \cdot K}{K \cdot \Delta^2 + t^2 \cdot \hat{\sigma}_M^2}. \quad (2.27)$$

2.4. Оцінка істотності вибірових характеристик

Дослідження істотності вибірових характеристик (ступеня наближення до параметрів генеральної сукупності) проводиться за допомогою спеціальних статистичних показників, які називаються критеріями узгодженості.

Критерій узгодженості – це випадкова величина з відомим точним або наближеним розподілом, яка використовується для перевірки гіпотези. Значення критерію, обчислене за даними вибірки, називають *критеріальною статистикою*. Під гіпотезою в статистиці розуміють припущення про розподіл випадкової величини.

Перевірка гіпотез полягає у тому, щоб встановити, випадковими чи не випадковими є розбіжності між даними спостереження і деяким припущенням. Гіпотеза, відхилення від якої приписують випадку, називається *нульовою гіпотезою* (її позначають H_0). Кожній нульовій гіпотезі протиставляють *альтернативну гіпотезу* – логічне заперечення нульової гіпотези (позначають H_1).

Після формулювання гіпотези задають рівень істотності α (ймовірність ризику відхилення нульової гіпотези), і за таблицями відповідних функцій розподілу знаходять критичні значення – статистичні характеристики для ймовірності $(1 - \alpha)$. За вибраним статистичним критерієм розраховують вибірове значення критеріальної статистики. Якщо воно перевищує критичне (табличне) значення вибраної функції розподілу, то нульова гіпотеза відхиляється.

Залежно від того, як задається альтернативна гіпотеза, вибирають межі між критичною областю і областю довірчих значень. Якщо перевіряється гіпотеза про відповідність між генеральною і вибірковою сукупністю, то застосовують двосторонній критерій. У разі перевірки за допомогою двостороннього критерію значення статистичної характеристики

обчислюють для рівня істотності (значущості) $\alpha/2$. Якщо ж досліджується істотність відмінностей у певному напрямку, то використовують односторонній критерій.

Статистична перевірка гіпотез пов'язана з ризиком прийняття помилкового рішення:

- помилкового рішення 1-го роду – відхилення правильної гіпотези H_0 (ймовірність такої помилки дорівнює α);
- помилкового рішення 2-го роду – гіпотеза H_0 приймається, хоча насправді правильною є альтернативна гіпотеза H_1 .

У табл.2.2 наведені співвідношення для розрахунку критеріальних статистик вибірових спостережень.

Таблиця 2.2

Розрахунок критеріальних статистик

Випробування гіпотези	Критерій	Критеріальна статистика
1. На основі вибір-кової середньої а) генеральна дисперсія відома σ_{Γ}^2 ; б) генеральна дисперсія не відома.	нормальний розподіл t -розподіл (кількість ступенів вільності $V = n - 1$)	$Z_{розр} = \frac{\bar{x} - \bar{x}}{\sigma_{\Gamma} / \sqrt{n}} \quad (2.28)$ $t_{розр} = \frac{\bar{x} - \bar{x}}{\hat{\sigma} / \sqrt{n-1}} \quad (2.29)$
2. На основі вибір-кової частки.	нормальний розподіл	$Z_{розр} = \frac{w - \bar{w}}{\sqrt{\frac{\bar{w}(1-\bar{w})}{n}}} \quad (2.30)$
3. На основі порівняння двох дисперсій.	F -розподіл (кількість ступенів вільності $V_1 = n_1 - 1$ і $V_2 = n_2 - 1$)	$F_{розр} = \frac{n_1 \hat{\sigma}_1^2}{n_1 - 1} \cdot \frac{n_2 - 1}{n_2 \hat{\sigma}_2^2} \quad (\hat{\sigma}_1^2 > \hat{\sigma}_2^2) \quad (2.31)$
4. На основі порівняння середніх величин двох вибірок: а) генеральні дисперсії відомі ($\sigma_{\Gamma_1}^2$ і $\sigma_{\Gamma_2}^2$); б) генеральні дисперсії не відомі.	нормальний розподіл t -критерій (кількість ступенів вільності $V = n_1 + n_2 - 2$)	$Z_{розр} = \frac{(\bar{x}_1 - \bar{x}_2)}{\sqrt{\frac{\sigma_{\Gamma_1}^2}{n_1} + \frac{\sigma_{\Gamma_2}^2}{n_2}}} \quad (2.32)$ $t_{розр} = \frac{(\bar{x}_1 - \bar{x}_2)}{\sqrt{\frac{\hat{\sigma}_1^2}{n_1 - 1} + \frac{\hat{\sigma}_2^2}{n_2 - 1}}} \quad (2.33)$
5. На основі порівняння двох вибірових часток.	нормальний розподіл	$Z_{розр} = \frac{(\bar{w}_1 - \bar{w}_2)}{\sqrt{\frac{\bar{w}_1(1-\bar{w}_1)}{n_1} + \frac{\bar{w}_2(1-\bar{w}_2)}{n_2}}} \quad (2.34)$

Основні положення та вказані моделі використовувалися при аналізі:

- факторів ризику в процесі формування та реалізації інвестиційних проєктів в регіоні, що дозволило сформуванню інструментарій для інвестора, який знижує рівень ризику інвестиційних рішень та підвищує рівень їх інформаційної обґрунтованості, зменшує ймовірність погіршення фінансового стану та підвищує премію за ризик в контексті реалізації інвестиційних проєктів в економічному просторі регіонів.

Blahun I. S., Leshuk H. V. Probabilistic and statistical methods of risk assessment of investment projects of a region. *Austrian Journal of Humanities and Social Sciences*. 2017. No. 5-6. P. 53–59.

- методик оцінки економічного розвитку, в основі яких лежать методи синтезу та поступової конкретизації. Розроблено автором алгоритм впровадження системи інтегральної оцінки економіки, як на регіональному так і національному рівнях. Проведено декомпозицію алгоритму на чотири основні етапи. Розроблено процедури виявлення взаємозалежностей у динаміці економічного розвитку на регіональному та національному рівнях.

Blahun I. S., Savchyn I. Z. Detection method of interdependences between dynamics of economic development indicators at regional and national levels. *European Journal of Economics and Management*. 2017. No. 3. P. 47–56.

- потенціалу регіональної конкурентоспроможності інструментарієм теорії гравітації. Визначено, що потенціал регіональної конкурентоспроможності – це комплекс синтезованих чинників економічного характеру у їх взаємозв'язку з іншими детермінантами. Проведено картографічне моделювання потенціалу регіональної конкурентоспроможності. На основі використання статистичного методу дослідження ідентифіковано латентні групи чинників з іманентною їм

структурою регіональної конкурентоспроможності за джерелами їх утворення.

Благун І. С., Кацедан А. В. Моделювання взаємозв'язків чинників регіональної конкурентоспроможності з її потенціалом розвитку. Бізнес Інформ. 2014. № 11. С. 103–115.

- зв'язків між багатовимірними змінними, вивчення яких проводиться за допомогою різних показників (індикаторів). З цією метою було розроблено автором діагностичні змінні. Визначено рівень їх кореляції із процесом економічного зростання регіонів. Побудовано модель головних зв'язків у процесі економічного зростання в регіонах з урахуванням просторової структури економіки.

Благун І. С., Квасній З. В. Моделювання зв'язків сектора МСП та іноземних інвестицій із працевлаштуванням та економічним зростанням у регіонах. Бізнес Інформ. 2014. № 7 (438). С. 71–76.

РОЗДІЛ 3. СТАТИСТИЧНЕ МОДЕЛЮВАННЯ СОЦІАЛЬНО-ЕКОНОМІЧНИХ СИСТЕМ

3.1. Кореляційний та регресійний аналіз

Закономірності функціонування ринку товарів і послуг характеризуються взаємозв'язками між показниками, які перебувають у певних відношеннях як незалежні або залежні ознаки. Ознаки, які впливають на деякий результат, називають *незалежними (факторними, пояснювальними)* ознаками, а ознаки, значення яких формуються під впливом незалежних ознак, називають *залежними (результатними, пояснюваними)* ознаками. Залежність між ознаками може проявлятися у *функціональній* або *стохастичній формі*. За функціональної залежності кожному значенню факторної ознаки або набору значень декількох факторних ознак відповідає єдине (фіксоване) значення результатної ознаки. За стохастичного зв'язку кожному значенню факторної ознаки або набору декількох значень факторних ознак відповідає певна множина значень результатної ознаки. Різновидом стохастичних зв'язків є кореляційні зв'язки.

Кореляційний зв'язок – це зв'язок, за якого дія окремих факторів проявляється лише як тенденція (в середньому) при масовому спостереженні фактичних даних. Він може встановлюватися для пари показників (*парна кореляція*) або для сукупності показників (*множинна кореляція*). Характерною особливістю кореляційного зв'язку, на відміну від функціонального, є його незворотність. Кореляційна залежність вважається неповною, оскільки на значення результату впливають не лише закономірні, але й випадкові чинники, які не відображаються при його обчисленні.

Важливо розуміти, що кореляційна залежність відображає лише взаємозв'язок між ознаками і не розглядає причинно-наслідкові зв'язки.

Метод обробки статистичних даних для вимірювання щільності зв'язку між двома і більше ознаками називають *кореляційним аналізом*.

До основних задач кореляційного аналізу відносять:

- вимірювання щільності (сили, інтенсивності, зв'язності) двох і більше ознак;
- відбір факторів, які істотно впливають на результатну ознаку, на основі оцінювання ступеня зв'язності між ознаками (істотні у даному аспекті фактори використовуються у подальшому в регресійному аналізі).

Залежність математичного сподівання (середнього значення) випадкової величини від однієї або декількох інших випадкових величин називається *регресією*. Для дослідження впливу одного або декілька факторів на результатний показник використовують статистичний метод, який називається *регресійним аналізом*.

Основна мета регресійного аналізу полягає у побудові аналітичної форми зв'язку, за якої зміна результатної ознаки обумовлюється впливом однієї або декількох істотних факторних ознак. При цьому множина всіх інших факторів, які також впливають на результатну ознаку, фіксується на рівні постійних і середніх значень. Регресійний аналіз забезпечує моделювання причинно-наслідкових зв'язків.

До основних задач регресійного аналізу відносять:

- встановлення форми залежності між факторами і результатною ознакою;
- визначення параметрів функції регресії у виді математичного рівняння;
- оцінювання ступеня впливу факторів на результатну ознаку;
- оцінювання значень результатної ознаки на підставі фіксованих значень факторів.

Методи кореляційного і регресійного аналізу використовуються в комплексі і складають *методи кореляційно-регресійного аналізу*. Методи

кореляційно-регресійного аналізу знаходять широке застосування у сфері дослідження ринку товарів і послуг. Наведемо декілька типових прикладів:

- оцінювання впливу витрат на рекламу на обсяг реалізованої продукції;
- оцінювання щільності зв'язку між ціною і обсягом реалізації товару;
- обчислення коефіцієнтів еластичності;
- аналіз залежності обсягів реалізації продукції від витрат на рекламу, ціни і рівня каналів збуту;
- аналіз залежності сприйняття товару покупцями від іміджу товарної марки, реклами і ціни;
- дослідження варіації частки ринку в залежності від чисельності торгового персоналу, витрат на рекламу і бюджету просування товару;
- оцінювання вкладу видатків на рекламу у пояснення варіації обсягів продажів при контрольованому рівні цін;
- прогнозування попиту на продукцію;
- прогнозування цін;
- прогнозування доходів від реалізації продукції.

Нехай залежність результатного показника від факторної ознаки за даними генеральної сукупності описується моделлю вигляду:

$$y = \beta_0 + \beta_1 x + \varepsilon, \quad (3.1)$$

де β_0 і β_1 – невідомі значення параметрів зв'язку, а ε – випадкова величина.

Параметри β_0 і β_1 називають відповідно *вільним членом моделі* та *коефіцієнтом регресії*. Випадкова величина ε акумулює в собі стохастичні збурення, помилки спостереження і вимірювання тощо.

Модель, що побудована на підставі даних вибіркової сукупності та має вигляд

$$\hat{y} = b_0 + b_1 x, \quad (3.2)$$

де b_0 і b_1 – оцінки параметрів β_0 і β_1 , називають *рівнянням парної лінійної регресії*.

У загальному випадку модель, яка описує парний взаємозв'язок між результатним показником (y) і факторною ознакою (x), може бути нелінійною.

Попереднє уявлення про форму залежності між ознаками можна зробити на підставі аналізу графічного представлення сукупності емпіричних даних. Якщо результати кожного спостереження представити графічно парою чисел (x, y), то сукупність таких точок на площині називають *кореляційним полем*. За характером розміщення точок можна зробити певні висновки щодо форми функції зв'язку, а також щільності зв'язку (рис.3.1).

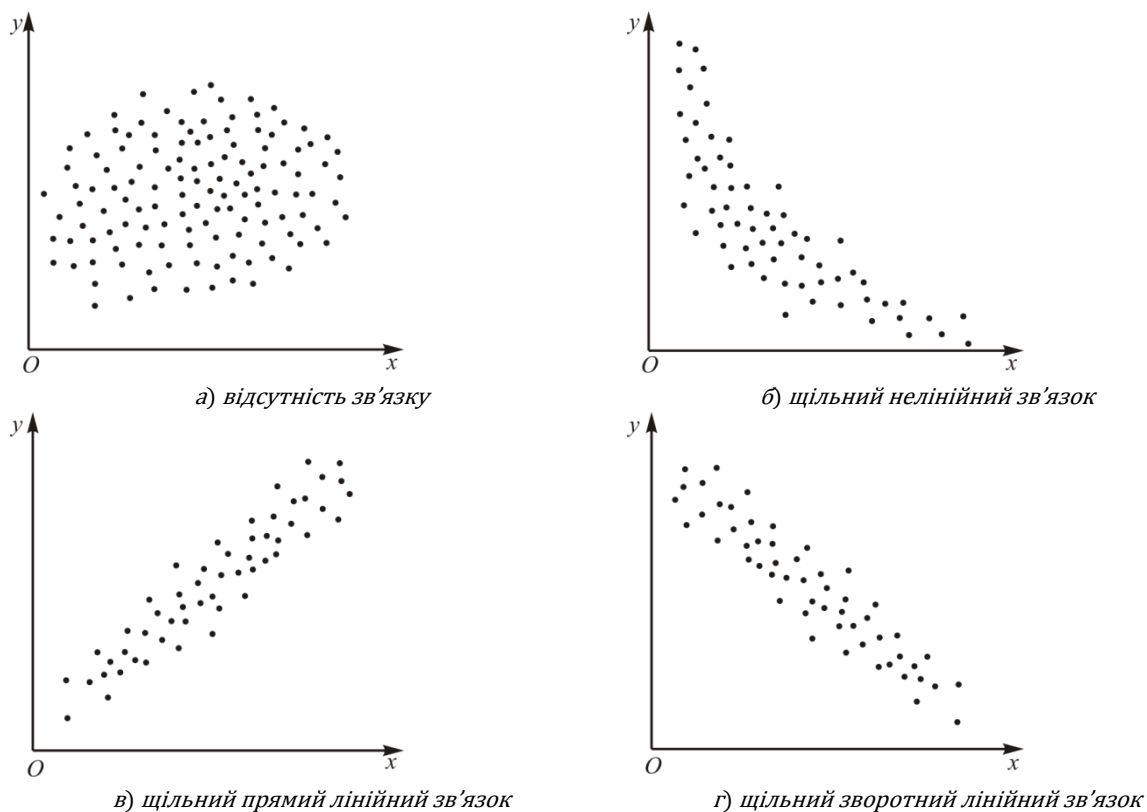


Рис. 3.1. Характеристики форми і щільності зв'язку за виглядом кореляційного поля

Припустимо, що характер розміщення точок кореляційного поля підтверджує наявність лінійного зв'язку. Для побудови парної (однофакторної) регресійної моделі необхідно оцінити числові значення параметрів β_0 і β_1 моделі (3.1).

Завдання оцінювання параметрів вибіркової регресійної моделі полягає у встановленні таких числових значень b_0 і b_1 , які забезпечують мінімальні відхилення емпіричних значень залежної випадкової величини (y_i) від теоретичних значень (\hat{y}_i). Оскільки значення відхилень $e_i = y_i - \hat{y}_i$ можуть бути як додатними, так і від'ємними, то їх сума $\left(\sum_{i=1}^n e_i\right)$ може бути близькою до нуля навіть при значних відхиленнях e_i . Цього недоліку позбавлені відхилення e_i^2 , тому сформульовану умову представимо у такому вигляді:

$$S(b_0, b_1) = \sum_{i=1}^n e_i^2 = \sum_{i=1}^n (y_i - \hat{y}_i)^2 \rightarrow \min. \quad (3.3)$$

Оскільки $\hat{y}_i = b_0 + b_1 \cdot x_i$, то вираз (3.3) перетвориться до вигляду

$$S(b_0, b_1) = \sum_{i=1}^n [y_i - (b_0 + b_1 \cdot x_i)]^2 \rightarrow \min, \quad (3.4)$$

де $S(b_0, b_1)$ – функція двох невідомих величин (b_0, b_1) ; x_i, y_i – емпіричні дані.

Необхідною умовою існування екстремуму функції $S(b_0, b_1)$ є рівність нулю частинних похідних $\frac{\partial S}{\partial b_0}$ і $\frac{\partial S}{\partial b_1}$, тобто:

$$\begin{cases} \frac{\partial S}{\partial b_0} = -2 \left(\sum_{i=1}^n y_i - b_1 \cdot \sum_{i=1}^n x_i - b_0 \cdot n \right) = 0, \\ \frac{\partial S}{\partial b_1} = -2 \left(\sum_{i=1}^n x_i \cdot y_i - b_1 \sum_{i=1}^n x_i^2 - b_0 \cdot \sum_{i=1}^n x_i \right) = 0. \end{cases} \quad (3.5)$$

Із системи нормальних рівнянь Гауса (3.5) знаходимо шукані параметри b_0 та b_1 :

$$b_1 = \frac{n \cdot \sum_{i=1}^n x_i \cdot y_i - \sum_{i=1}^n x_i \sum_{i=1}^n y_i}{n \sum_{i=1}^n x_i^2 - \left(\sum_{i=1}^n x_i \right)^2} = \frac{\overline{xy} - \bar{x} \cdot \bar{y}}{\overline{x^2} - (\bar{x})^2} = \frac{\text{cov}(x; y)}{\text{var}(x)}, \quad (3.6)$$

$$b_0 = \frac{\sum_{i=1}^n y_i - b_1 \cdot \sum_{i=1}^n x_i}{n} = \bar{y} - b_1 \cdot \bar{x}, \quad (3.7)$$

де $\text{cov}(x; y) = \frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})$ і $\text{var}(x) = \sigma_x^2 = \frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^2$.

Ідея знаходження параметрів b_0 і b_1 полягає у побудові такої прямої лінії на кореляційному полі, для якої мінімізується сума квадратів відхилень фактичних значень результатного показника від теоретичних, а метод реалізації цієї ідеї називають *методом найменших квадратів* (МНК) (рис. 3.2).

Якщо між двома ознаками існує лінійний зв'язок, який описується рівнянням регресії вигляду (3.2), то фактичне значення залежної змінної (y) буде дорівнювати сумі теоретичного значення залежної змінної (\hat{y}) та значення залишку (e), зумовленого дією інших факторів, не відображених у моделі, тобто $y = \hat{y} + e$.

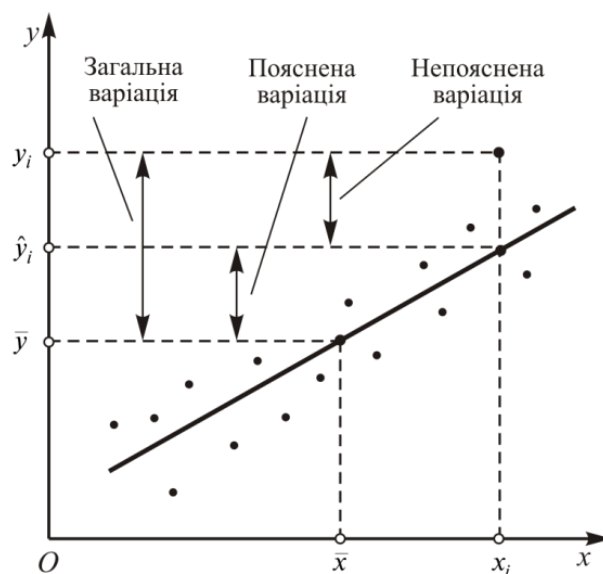


Рис. 3.2. Геометрична інтерпретація варіації залежної змінної

Проведемо n незалежних спостережень за факторною і результатною ознаками. Нехай x_i та y_i — результати i -го спостереження ($i = \overline{1, n}$).

Загальна варіація залежної змінної дорівнює $SST = \sum_{i=1}^n (y_i - \bar{y})^2$ і не залежить від значень факторної ознаки. Натомість, варіація теоретичних (розрахункових) значень (\hat{y}) з урахуванням лінійної форми зв'язку (пояснена варіація) виражається величиною $SSR = \sum_{i=1}^n (y_i - \hat{y}_i)^2$, а варіація, що не пояснюється лінійним зв'язком, — величиною $SSE = \sum_{i=1}^n (y_i - \hat{y}_i)^2$. Непояснена варіація — це результат дії факторів, що не враховані в лінійній моделі.

За теоремою додавання дисперсій сума поясненої і непоясненої варіацій дорівнює загальній варіації $SST = SSR + SSE$.

Розділивши обидві частини цієї рівності на SST , отримаємо

$$\frac{SSR}{SST} + \frac{SSE}{SST} = 1$$

Відношення поясненої варіації, що зумовлена лінійним рівнянням регресії, до загальної дисперсії називають коефіцієнтом детермінації (R^2):

$$R^2 = \frac{SSR}{SST} = \frac{\sum_{i=1}^n (\hat{y}_i - \bar{y})^2}{\sum_{i=1}^n (y_i - \bar{y})^2}, \quad R^2 \in [0; 1]. \quad (3.8)$$

Коефіцієнт детермінації характеризує частку загальної варіації залежної змінної, яка пояснюється лінійною формою зв'язку.

Коефіцієнт детермінації R^2 можна виразити і у такому вигляді:

$$R^2 = \frac{SSR}{n} : \frac{SST}{n} = \frac{\sigma_{\hat{y}x}^2}{\sigma_y^2}, \quad (3.9)$$

де σ_y^2 — загальна дисперсія залежної змінної; $\sigma_{\hat{y}x}^2$ — дисперсія, що пояснюється рівнянням регресії.

За значенням R^2 знаходять коефіцієнт кореляції $R = \sqrt{R^2}$, який характеризує щільність зв'язку між ознаками

Кількісно ступінь наближення кореляційного зв'язку до функціональної лінійної залежності також можна оцінити за допомогою коефіцієнта лінійної кореляції Пірсона:

$$r = \frac{\overline{xy} - \bar{x} \cdot \bar{y}}{\sigma_x \cdot \sigma_y} = \frac{n \sum_{i=1}^n x_i y_i - \sum_{i=1}^n x_i \sum_{i=1}^n y_i}{\sqrt{n \sum_{i=1}^n x_i^2 - \left(\sum_{i=1}^n x_i \right)^2} \cdot \sqrt{n \sum_{i=1}^n y_i^2 - \left(\sum_{i=1}^n y_i \right)^2}}, \quad r \in [-1; +1]. \quad (3.10)$$

За значенням коефіцієнта лінійної кореляції можна зробити такі висновки:

- якщо r набуває значення, яке близьке до -1 , то між факторами існує щільний зворотний зв'язок;
- якщо $r = 0$, то зв'язок відсутній;
- якщо r близьке до $+1$, то між показниками існує щільний прямий зв'язок;
- якщо $|r| = 1$, то між досліджуваними показниками існує функціональний зв'язок.

Зазначимо, що знак коефіцієнта лінійної кореляції Пірсона вказує на напрям зв'язку між ознаками, у той час як $|r|$ характеризує щільність зв'язку (аналогічно як і R).

Оскільки вибірку регресійну модель будують на підставі певної вибірки з генеральної сукупності даних. то з неї можна утворити різні вибірки і відповідно побудувати ряд вибірових регресійних моделей. Постає питання, яка з вибірових моделей є найближчою до моделі генеральної сукупності, тобто як оцінити відповідність оцінок b_0 і b_1 параметрам β_0 і β_1 .

Певні статистичні (імовірнісні) висновки про значення параметрів β_0 і β_1 можна зробити за виконання таких основних припущень:

1. Регресійну модель специфіковано вірно, тобто існуючий зв'язок між ознаками є лінійним.

2. Математичне сподівання випадкової величини ε дорівнює нулю:

$$M(\varepsilon_i) = 0, \quad (i = \overline{1, n})$$

де $M(\varepsilon_i)$ – математичне сподівання випадкової величини ε , що зумовлене значенням x_i факторної ознаки. Зміст припущення 2 полягає у тому, що не враховані у моделі чинники, дія яких проявляється через значення випадкової величини ε , систематично не впливають на математичне сподівання результатного показника.

3. Випадкові величини є незалежними між собою, тобто автокореляція між випадковими величинами ε_i і ε_j відсутня:

$$\text{cov}(\varepsilon_i, \varepsilon_j) = 0, \quad i \neq j.$$

4. Дисперсія випадкових величини ε_i є сталою для усіх значень факторної ознаки x_i (*гомоскедастичність* випадкових величин):

$$D(\varepsilon_i) = \sigma^2 = \text{const}, \quad (i = \overline{1, n}),$$

де $D(\varepsilon_i)$ – дисперсія випадкової величини ε_i , що зумовлена значенням x_i факторної ознаки.

Якщо припущення 4 не виконується, то наявна *гетероскедастичність*.

5. Випадкова величина ε розподілена нормально і має математичне сподівання $M(\varepsilon) = 0$ та дисперсію $D(\varepsilon) = \sigma^2$.

Якщо виконуються припущення (1) – (5), то модель називають *класичною парною лінійною регресійною моделлю*.

Для побудови рівняння регресії достатньо, щоб виконувалися припущення 1–4. Виконання припущення 5 необхідне для оцінювання точності рівняння регресії та адекватності його параметрів.

При тестуванні моделі на адекватність використовують критерій Фішера (F -критерій). Перевірку наявності (відсутності) лінійного зв'язку між змінними здійснюють у такій послідовності:

1. Формулюють нульову гіпотезу H_0 : коефіцієнт регресії для генеральної сукупності дорівнює нулю ($\beta_1 = 0$).

2. Встановлюють розрахункове значення F -критерію ($F_{\text{розрах}}$):

$$F_{\text{розрах}} = \frac{MSR}{MSE} = \frac{SSR}{V_1} : \frac{SSE}{V_2} = \frac{V_2}{V_1} \cdot \frac{SSR}{SSE} = \frac{n-2}{1} \cdot \frac{R^2}{1-R^2}. \quad (3.11)$$

3. З таблиць розподілу Фішера для вибраного рівня значущості α і $V_1=1$ та $V_2 = n-2$ ступенів вільності знаходять $F_{\text{крит}}$.

4. Перевіряють виконання умови $F_{\text{розрах}} > F_{\text{крит}}$:

- якщо умова виконується, то вважаємо, що регресійна модель з імовірністю $(1-\alpha)$ відповідає існуючому зв'язку між ознаками x і y (*лінійна форма зв'язку підтверджується*), гіпотезу H_0 відхиляють;

- якщо умова не виконується, то модель є неадекватною (*лінійна форма зв'язку не підтверджується*), тобто гіпотезу H_0 приймають.

Якщо припущення (1) – (5) виконуються, то оцінки b_0 і b_1 параметрів узагальненої регресійної моделі β_0 і β_1 будуть BLUE-оцінками (найкращими лінійними оцінками), тобто:

- незміщеними, оскільки $M(b_0) = \beta_0$, $M(b_1) = \beta_1$;

- ґрунтовними, оскільки при збільшенні обсягу вибірки оцінки b_0 і b_1 будуть наближатися до значень параметрів β_0 і β_1 ;

- ефективними, оскільки дисперсії оцінок $\sigma_{b_0}^2$ і $\sigma_{b_1}^2$ набувають мінімальних значень порівняно з будь-якими іншими оцінками параметрів β_0 і β_1 .

Відзначимо також, що у разі виконання припущень (1) – (5) математичне сподівання і дисперсія залежної змінної та параметрів вибіркової регресійної моделі набувають таких властивостей:

$$1. M(y_i) = \beta_0 + \beta_1 x_i; \quad \sigma_{y_i}^2 = \sigma_\varepsilon^2 = \text{const};$$

$$2. M(b_0) = \beta_0; \quad \sigma_{b_0}^2 = \sigma_\varepsilon^2 \cdot \frac{\sum_{i=1}^n x_i^2}{n \cdot \sum_{i=1}^n (x_i - \bar{x})^2};$$

$$3. M(b_1) = \beta_1; \quad \sigma_{b_1}^2 = \sigma_\varepsilon^2 \cdot \frac{1}{\sum_{i=1}^n (x_i - \bar{x})^2}.$$

Враховуючи, що ε — неспостережувана випадкова величина, дисперсію якої встановити неможливо, замінюють σ_ε^2 на її оцінку

$$\hat{\sigma}_\varepsilon^2 = MSE = \frac{1}{n-2} \cdot \sum_{i=1}^n e_i^2, \quad (3.12)$$

що дасть змогу обчислити оцінені дисперсії параметрів b_0 і b_1 :

$$\hat{\sigma}_{b_0}^2 = \hat{\sigma}_\varepsilon^2 \cdot \frac{\sum_{i=1}^n x_i^2}{n \cdot \sum_{i=1}^n (x_i - \bar{x})^2}, \quad (3.13)$$

$$\hat{\sigma}_{b_1}^2 = \hat{\sigma}_\varepsilon^2 \cdot \frac{1}{\sum_{i=1}^n (x_i - \bar{x})^2}. \quad (3.14)$$

Параметри рівняння парної регресії розраховують на підставі обмеженої вибірки даних спостережень. Водночас їх значення відіграють важливу роль у процесах планування і прогнозування показників. Дані певної вибірки можуть вказувати на лінійність зв'язку між ознаками, хоча насправді в генеральній сукупності він не існує. Тому необхідно встановити ймовірність того, що лінійний зв'язок у вибірковій сукупності свідчить про той же зв'язок у генеральній сукупності, тобто оцінити значущість коефіцієнтів регресії.

Випадкові параметри b_0 і b_1 розподілені за нормальним законом відповідно з математичними сподіваннями $M(b_0)$ і $M(b_1)$ та дисперсіями $\sigma_{b_0}^2$ і $\sigma_{b_1}^2$. Для знаходження довірчих інтервалів параметрів моделі, тобто

інтервалів, у які із заданою ймовірністю потрапляють значення β_0 і β_1 , першочергово перевіряють значущість параметрів b_0 і b_1 вибіркової моделі за t -критерієм (критерієм Стьюдента).

Послідовність процедури тестування за t -критерієм є такою:

1. Формулюють статистичні гіпотези:

$$\begin{cases} H_0 : \beta_0 = 0 \\ H_1 : \beta_0 \neq 0 \end{cases}, \quad \begin{cases} H_0 : \beta_1 = 0 \\ H_1 : \beta_1 \neq 0 \end{cases}.$$

2. Обчислюють розрахункові значення t -критерію:

$$t_{\text{розр}} = \frac{b_0}{\sigma_{b_0}}, \quad t_{\text{розр}} = \frac{b_1}{\sigma_{b_1}}$$

3. Знаходять критичні значення $t_{\text{крит}}$ з урахуванням рівня значущості α і $\nu = n - 2$ ступенів вільності за таблицею розподілу Стьюдента.

4. Перевіряють умову $t_{\text{розр}} > t_{\text{крит}}$:

- якщо умова виконується, то нульову гіпотезу *відхиляють*, тобто відповідний параметр є *статистично значущим*;

- якщо умова не виконується, то нуль-гіпотеза *приймають*, тобто відповідний параметр є статистично незначущим.

Зауважимо, що за спрощеної процедури тестування використовують умову ($t_{\text{розр}} > 2$), тобто нульову гіпотезу відхиляють, якщо $t_{\text{розр}} > 2$.

З урахуванням формул (3.13) – (3.14) знаходять граничні помилки коефіцієнтів регресії Δb_0 і Δb_1 :

$$\Delta b_0 = t_\alpha \cdot \sigma_{b_0}, \tag{3.15}$$

$$\Delta b_1 = t_\alpha \cdot \sigma_{b_1}, \tag{3.16}$$

де t_α – значення коефіцієнта Стьюдента (коефіцієнта, знайденого за таблицею розподілу Стьюдента для обраного рівня значущості α і $\nu = n - 2$ ступенів вільності).

Отже, межі довірчих інтервалів параметрів рівняння регресії становлять $(b_0 \pm \Delta b_0)$ і $(b_1 \pm \Delta b_1)$, тобто значення параметрів β_0 і β_1 можуть перебувати у таких межах:

$$b_0 - \Delta b_0 \leq \beta_0 \leq b_0 + \Delta b_0, \quad (3.17)$$

$$b_1 - \Delta b_1 \leq \beta_1 \leq b_1 + \Delta b_1. \quad (3.18)$$

Аналогічно до тестування оцінок параметрів регресійної моделі проводять статистичне тестування коефіцієнта лінійної кореляції Пірсона на значущість за t -критерієм.

Для цього обчислюють розрахункове значення t -критерію

$$t_{\text{розра}} = \sqrt{\frac{r^2}{1-r^2}} \cdot (n-2) \quad (3.19)$$

і порівнюють з теоретичним (критичним) значенням ($t_{\text{крит}}$). Теоретичне значення функції розподілу Стюдента ($t_{\text{крит}}$) для заданого рівня значущості α і $\nu = n-2$ ступенів вільності знаходять з відповідної таблиці.

Якщо $t_{\text{розра}} > t_{\text{крит}}$, то гіпотеза про нульове значення коефіцієнта лінійної кореляції у генеральній сукупності *не підтверджується*, тобто в генеральній сукупності *існує лінійний зв'язок*.

Модель $y = \mu + e$, де e – залишок, який відображає вплив неврахованих факторів і невідповідність вибіркової сукупності генеральній сукупності, має ймовірнісний характер. Для забезпечення надійності оцінок взаємозв'язку між статистичними показниками необхідно знайти максимально і мінімально можливі значення залежної змінної для вказаних значень незалежної змінної із заданою ймовірністю, тобто межі довірчого інтервалу.

Якщо припустити, що вибірка є репрезентативною, а відхилення $(y - \mu)$ розподілені нормально, то можна стверджувати, що ймовірність перебування фактичних значень залежної змінної в певних межах дорівнює

$$P\{\mu - t_{\alpha} \sigma_{\varepsilon} \leq y \leq \mu + t_{\alpha} \sigma_{\varepsilon}\} = 1 - \alpha, \quad (3.20)$$

де t_α – значення коефіцієнта_Стюдента.

Оцінка дисперсії σ_ε^2 характеризує непояснену варіацію фактичних значень (y) відносно розрахункових значень (\hat{y}). Непояснена варіація характерна для будь-якої сукупності (вибіркової чи генеральної) у разі існуванні кореляційного зв'язку.

У практиці прогнозування знаходять довірчі інтервали для середнього (y_{n+1}^{cp}) та індивідуального (y_{n+1}^{ind}) значення залежної змінної ознаки при заданому значенні x_{n+1} , які відповідно становлять:

$$y_{n+1}^{cp} = \hat{y} \pm t_\alpha \sigma_\varepsilon \sqrt{1 + \frac{(x_{n+1} - \bar{x})^2}{\sum_{i=1}^n (x_i - \bar{x})^2}}, \quad (3.21)$$

$$y_{n+1}^{ind} = \hat{y} \pm t_\alpha \sigma_\varepsilon \sqrt{1 + \frac{1}{n} + \frac{(x_{n+1} - \bar{x})^2}{\sum_{i=1}^n (x_i - \bar{x})^2}}. \quad (3.22)$$

Як видно із виразів (3.21) і (3.22) на величину довірчих інтервалів для середнього та індивідуального значення залежної змінної ознаки впливає значення незалежної змінної (із віддаленням значення x від \bar{x} межі довірчого інтервалу збільшуються).

Коефіцієнт еластичності застосовують для встановлення міри впливу зміни факторної ознаки на зміну результативної ознаки.

Обчислення коефіцієнта еластичності (E) здійснюють на основі значення параметра b_1 регресійної моделі за такою формулою:

$$E = \frac{\bar{x}}{\bar{y}} \cdot \frac{\partial y}{\partial x} = \frac{\bar{x}}{\bar{y}} \cdot b_1. \quad (3.23)$$

Економічний зміст коефіцієнта еластичності полягає в тому, що його числове значення характеризує відносну зміну результатної ознаки при збільшенні факторної ознаки на 1%.

У загальному випадку модель множинної регресії можна записати у такому вигляді:

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

де y - пояснювана змінна; $x_j (j = \overline{1, m})$ - пояснювані змінні; ε - випадкова складова.

Між пояснюваною і пояснювальними змінними існує стохастичний зв'язок. На початковому етапі дослідження не відомі ні кількість, ні склад пояснювальних змінних, ні аналітична форма зв'язку. Тому необхідно здійснити *специфікацію моделі*:

- *специфікацію факторних ознак* (вибір із загальної сукупності факторних ознак впливових);
- *специфікацію вигляду моделі* (вибір аналітичної форми залежності).

Неправильна специфікація моделі пов'язується із такими можливими причинами:

- у моделі не враховані впливові(суттєві, істотні) факторні ознаки;
- у моделі представлені факторні ознаки, які відграють другорядну роль у формуванні значення результатного показника (суттєво не впливають на результатний показник);
- математична форма зв'язку між змінними у моделі не відповідає такій, що існує реально.

Тому першим кроком, який потрібно зробити у процесі побудови множинної регресійної моделі, є відбір чинників (факторних ознак), які формують результатний показник, тобто визначення множини пояснювальних змінних. Початкову сукупність пояснювальних змінних багатфакторної регресійної моделі визначають на підставі системного теоретичного і практичного аналізу якісного змісту об'єкта дослідження з виокремленням тих змінних, для яких існує можливість прямого чи опосередкованого кількісного вираження за доступною статистичною інформацією. Кінцевий відбір пояснювальних змінних проводиться на етапах оцінювання і тестування побудованої моделі.

Якщо до моделі не увійдуть впливові змінні, то оцінки коефіцієнтів регресії і дисперсія залишків можуть бути зміщеними, а значення дисперсій оцінок коефіцієнтів регресії - заниженими.

У випадку включення до моделі пояснювальної змінної, яка суттєво не впливає на пояснювану змінну, оцінки параметрів рівняння регресії та оцінки їх дисперсій будуть незміщеними, але точність оцінок, отриманих за допомогою методу найменших квадратів (МНК), може знижуватися.

Після того, як сформована множина пояснювальних змінних, з урахуванням виявлених умов функціонування економічної системи, потрібно вказати вигляд співвідношення між результатним показником і факторними ознаками, що його обумовлюють. Вказати точний функціональний вигляд такого співвідношення доволі складно, тому мова йде про апроксимацію зв'язку між змінними відносно простими функціями. При цьому на практиці перевагу віддають лінійним моделям, або таким, що зводяться до лінійних шляхом заміни і перетворення початкових змінних. Використання нелінійних моделей, як правило, пов'язано з ускладненим оцінюванням параметрів моделі та зниженням точності результатів моделювання.

Позначимо:

$$Y = \begin{pmatrix} y_1 \\ y_2 \\ \dots \\ y_n \end{pmatrix}, \quad X = \begin{pmatrix} 1 & x_{11} & x_{12} & \dots & x_{1m} \\ 1 & x_{21} & x_{22} & \dots & x_{2m} \\ \dots & \dots & \dots & \dots & \dots \\ 1 & x_{n1} & x_{n2} & \dots & x_{nm} \end{pmatrix}, \quad \beta = \begin{pmatrix} \beta_0 \\ \beta_1 \\ \dots \\ \beta_n \end{pmatrix}, \quad B = \begin{pmatrix} b_0 \\ b_1 \\ \dots \\ b_m \end{pmatrix}, \quad \varepsilon = \begin{pmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \dots \\ \varepsilon_n \end{pmatrix},$$

$$E = \begin{pmatrix} e_1 \\ e_2 \\ \dots \\ e_n \end{pmatrix},$$

де Y - вектор значень спостережень за пояснюваною змінною;

X - матриця значень спостережень за пояснювальними змінними (до матриці штучно включено стовпець, всі елементи якого рівні 1);

β - вектор параметрів моделі;

ε - вектор не спостережуваних(випадкових) величин;

B - вектор оцінок параметрів моделі;

E - вектор залишків (відхилень) – різниці між фактичними і теоретичними (оціненими) значеннями пояснюваної змінної.

За умови прийнятих позначень загальна і вибіркова моделі лінійної множинної регресії відповідно мають такий вигляд:

$$Y = X\beta + \varepsilon; \quad (3.24)$$

$$Y = XB + E. \quad (3.25)$$

Класична регресійна модель ґрунтується на таких припущеннях:

- модель специфікована правильно (зв'язок між пояснюваною змінною і пояснювальними змінними є лінійним, вибрані факторні ознаки є впливовими);

- математичне сподівання випадкової складової ε для кожного спостереження рівне 0:

$$M(\varepsilon_i) = 0, \quad (i = \overline{1, n});$$

- дисперсія випадкової складової ε є сталою для всіх спостережень:

$$\sigma_{\varepsilon_i}^2 = \text{const} = \sigma^2, \quad (i = \overline{1, n});$$

- значення випадкових складових для i -го і для j -го спостережень не корелюють між собою:

$$\text{cov}(\varepsilon_i, \varepsilon_j) = 0, \quad (i = \overline{1, n}; j = \overline{1, n}; i \neq j);$$

- випадкові складові моделі і пояснювальні змінні розподілені нормально для одних і тих же спостережень, тобто не корелюють між собою (припущення виконується автоматично, якщо пояснювальні змінні є не випадковими величинами):

$$\text{cov}(x_{ij}, \varepsilon_i) = 0, \quad (i = \overline{1, n}; j = \overline{1, m});$$

▪ змінні x_1, x_2, \dots, x_m не залежні між собою (відсутня мультиколінеарність):

$$\text{cov}(x_i, x_j) = 0, \quad (i = \overline{1, n}; j = \overline{1, m}; i \neq j).$$

Для знахоження невідомих параметрів моделі (3.25) використовують 1МНК і отримують:

$$B = (X^T \cdot X)^{-1} X^T \cdot Y. \quad (3.26)$$

Оцінки $b_0, b_1, b_2, \dots, b_m$ невідомих параметрів $\beta_0, \beta_1, \beta_2, \dots, \beta_m$ моделі лінійної множинної регресії (3.24), знайдені з використанням 1МНК за умови дотримання основних припущень, є найкращими лінійними оцінками.

Оцінений коефіцієнт регресії $b_j (j = \overline{1, m})$ характеризує чистий вплив пояснювальної змінної x_j на пояснювану змінну y за умови, що модель містить усі фактори, які впливають на результатний показник.

При дослідженні інтенсивності впливу факторних ознак на результатний показник потрібно враховувати, що коефіцієнти регресії не завжди можна порівнювати між собою, так як їх числові значення залежать від вибраних одиниць вимірювання факторів. З метою встановлення порівняльної сили впливу кожної пояснювальної змінної на пояснювану змінну рівняння регресії представляють у стандартизованому (нормалізованому) вигляді:

$$y^* = \frac{y - \bar{y}}{\sigma_y}, \quad (3.27)$$

$$x_j^* = \frac{x_j - \bar{x}_j}{\sigma_{x_j}} \quad (3.28)$$

де y^* - стандартизовані значення пояснюваної змінної;

$x_j^* (j = \overline{1, m})$ - стандартизовані значення j -ї пояснювальної змінної;

\bar{y} і \bar{x}_j - відповідно середні значення пояснюваної і j -ї пояснювальної змінної;

σ_y і σ_{x_j} - відповідно середньоквадратичні відхилення пояснюваної і пояснювальних змінних.

Вільний член у стандартизованому рівнянні лінійної множинної регресії відсутній, тобто вибірккову модель у стандартизованій формі записують так:

$$\hat{y}^* = b_1^* x_1^* + b_2^* x_2^* + \dots + b_m^* x_m^* \quad (3.29)$$

де $b_j^* (j = \overline{1, m})$ - оцінені стандартизовані коефіцієнти регресії (бета-коефіцієнти).

Чим більшим за абсолютною величиною є значення b_j^* , тим відчутнішим є вплив j -ї пояснювальної змінної на результат.

Іншим показником для порівняння ефектів впливу пояснювальних змінних на пояснювану змінну коефіцієнт еластичності (E_{x_j}):

$$E_{x_j} = b_j \frac{\bar{x}_j}{\bar{y}} \quad , \quad (i = \overline{1, m}) \quad (3.30)$$

який слугує мірою впливу пояснювальної змінної на пояснювану у відсотках.

З метою оцінювання статистичної якості побудованої моделі використовують такі основні статистичні характеристики:

- стандартну помилку моделі;
- множинні коефіцієнти детермінації і кореляції;
- критерії перевірки істотності зв'язку і впливу окремих факторів на результатний показник.

Величину

$$\hat{\sigma}_\varepsilon^2 = \frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{n - m - 1} = \frac{Y^T Y - B^T X^T Y}{n - m - 1}, \quad (3.31)$$

називають оцінкою дисперсії відхилень (оцінкою непоясненої дисперсії).

Для моделі лінійної множинної вибіркової регресії має місце:

$$SST = SSR + SSE \quad (3.32)$$

де SST - загальна сума квадратів відхилень пояснюваної змінної від середнього значення, зумовлена впливом усіх можливих факторів;

SSR - сума квадратів відхилень, зумовлена впливом факторних ознак, включених до моделі;

SSE - сума квадратів залишків, яка характеризує ступінь впливу інтенсивності неврахованих факторів.

Частка загальної варіації, яка пояснюється мінливістю включених до моделі факторів, називається *множинним коефіцієнтом детермінації* (R^2):

$$R^2 = \frac{SSR}{SST} = 1 - \frac{SSE}{SST} = \frac{B'X'Y - n\bar{y}^2}{Y'Y - n\bar{y}^2} \quad (3.33)$$

Арифметичне значення кореня квадратного з коефіцієнта детермінації $R = +\sqrt{R^2}$ називають *коефіцієнтом множинної кореляції*. Він характеризує щільність зв'язку між результатним показником і факторами, включеними до рівняння множинної лінійної регресії.

На основі значення коефіцієнта детермінації проводять тестування адекватності моделі лінійної множинної регресії. З цією метою формулюють такі статистичні гіпотези:

- H_0 : для генеральної сукупності $\beta_1 = \beta_2 = \dots = \beta_m = 0$, тобто між пояснюваною змінною і включеними до моделі пояснювальними змінними відсутній лінійний зв'язок;

- H_1 : для генеральної сукупності не всі коефіцієнти регресії одночасно рівні 0, тобто лінійний зв'язок між пояснюваною і включеними до моделі пояснювальними змінними статистично значущий.

Для тестування вказаних вище гіпотез використовують критерій Фішера. Розрахункове значення F-критерію обчислюють за формулою

$$F_{розр} = \frac{V_2}{V_1} \cdot \frac{R^2}{1-R^2} \quad (3.34)$$

і порівнюють з критичним значенням $F_{кр}(\alpha; V_1; V_2)$ при заданому рівні значущості α і $V_1 = m$ та $V_2 = n - m - 1$ ступенях вільності (m - кількість пояснювальних змінних моделі, n - обсяг вибірки).

Якщо $F_{розр} > F_{кр}$, то гіпотезу H_0 відхиляють і приймають альтернативну.

Для перевірки статистичної значущості параметрів рівняння регресії формуються такі статистичні гіпотези:

- H_0 : для генеральної сукупності $\beta_j = 0$;
- H_1 : для генеральної сукупності $\beta_j \neq 0$.

На рівні значущості α і $V = n - m - 1$ ступенів вільності нульова гіпотеза H_0 відхиляється, якщо має місце:

$$t_{розр}^{e_j} = \frac{|b_j|}{\hat{\sigma}_{b_j}} > t_{кр}(\alpha, V) \quad (3.35)$$

де $\hat{\sigma}_{b_j}$ - стандартна вибіркова помилка j -го коефіцієнта рівняння регресії.

Оцінка дисперсії j -го коефіцієнта рівняння регресії рівна:

$$\hat{\sigma}_{b_j}^2 = c_{jj} \hat{\sigma}_{\varepsilon}^2, \quad (j = \overline{0, m}), \quad (3.36)$$

де c_{jj} - діагональні елементи матриці $C = (X^T X)^{-1}$.

Статистична незначущість коефіцієнта регресії свідчить про те, що на основі вибірки, використаної для побудови рівняння регресії, не можна гарантувати надійність оцінки впливу відповідної пояснювальної змінної на пояснювану змінну.

Знаючи стандартну вибіркoву помилку ($\hat{\sigma}_{b_j}$) коефіцієнта множинної лінійної регресії (b_j), можна розрахувати довірчий інтервал для значення параметра (β_j) у генеральній сукупності:

$$b_j - t_{кр}(\alpha, V) \hat{\sigma}_{b_j} \leq \beta_j \leq b_j + t_{кр}(\alpha, V) \hat{\sigma}_{b_j} \quad (3.37)$$

де $t_{кр}$ - критичне значення розподілу Ст'юдента для рівня значущості α і $V = n - m - 1$ ступенів вільності.

Величина

$$\Delta b_j = t_{кр}(\alpha, V) \hat{\sigma}_{b_j}$$

називається *граничною помилкою параметра β_j* .

Якщо кількість факторних ознак рівна m , то можна побудувати 2^m різних регресійних моделей. Очевидно, що перед дослідником стоїть завдання вибрати оптимальну ("найкращу") модель, яка відображає основні закономірності об'єкта прогнозування з достатнім ступенем статистичної надійності. З поміж існуючих методів вибору кінцевої моделі найчастіше користуються двома методами – методом покрокового включення змінних і методом покрокового виключення змінних.

За методом покрокового включення до моделі на кожному кроці долучається пояснювальна змінна, яка дає найбільший внесок у регресію. Метод покрокового виключення діє у зворотному напрямі (з моделі покроково вилучаються статистично незначущі пояснювальні змінні).

Прогнозування на основі моделі множинної лінійної регресії полягає в оцінюванні очікуваних значень пояснюваної змінної при заданих значеннях пояснювальних змінних.

Точковий прогноз на основі множинної лінійної регресії отримують шляхом підстановки прогнозних (вказаних дослідником) значень пояснювальних змінних в оцінене рівняння регресії. Якщо пояснювальні

змінні у прогнозованому періоді будуть рівними $x_1 = x_1^{прогн}$, $x_2 = x_2^{прогн}$, ..., $x_m = x_m^{прогн}$, то точковий прогноз $y_{точк}^{прогн}$ пояснюваної змінної буде рівним:

$$y_{точк}^{прогн} = b_0 + b_1 x_1^{прогн} + b_2 x_2^{прогн} + \dots + b_m x_m^{прогн} \quad (3.38)$$

Інтервальний прогноз (проміжок, в який при заданій ймовірності і при заданих значеннях пояснювальних змінних попадає прогнозне значення пояснюваної змінної) рівний:

$$y_{точк}^{прогн} - t_{кр} \hat{\sigma}_y \leq y^{прогн} \leq y_{точк}^{прогн} + t_{кр} \hat{\sigma}_y \quad (3.39)$$

де $\hat{\sigma}_y$ - оцінка середньої помилки прогнозного значення пояснюваної змінної.

Величина $\hat{\sigma}_y$ обчислюється так:

$$\hat{\sigma}_y = \hat{\sigma}_\varepsilon \sqrt{1 + X_0^T (X^T \cdot X)^{-1} X_0}, \quad (3.40)$$

де $X_0^T = (1; x_1^{прогн}; x_2^{прогн}; \dots; x_m^{прогн})$ - розширений транспонований вектор; $\hat{\sigma}_\varepsilon$ - стандартна вибіркова похибка рівняння регресії.

Слід відзначити, що прогнозування на основі рівняння регресії є надійним лише тоді, коли значення пояснювальних змінних суттєво не виходять за межі їх вибірових значень. Причому, прогноз є тим точнішим, чим ближчими є прогнозні значення пояснювальних змінних до середніх вибірових значень. Тому використання лінії регресії поза межами досліджуваного діапазону значень пояснювальних змінних може призвести до значних помилок.

Часто передбачення значень обсягів продажів призводить до необхідності включення до прогностичної моделі як кількісних, так і номінальних (якісних) ознак. Уведення номінальних за своєю сутністю факторних ознак у регресійні моделі безпосередньо неможливо здійснити, оскільки їх рівні набувають якісних, а не кількісних значень. У той же час номінальні факторні ознаки можуть істотно впливати на щільність зв'язку

між пояснюваною і пояснювальними змінними. Наприклад, на обсяг реалізації окремих видів продукції впливає номінальна ознака “сезонність”. Викорстання номінальних ознак у регресійному аналізі передбачає їх оцифрування, тобто приписування кожній градації якісної шкали відповідного числа з дотриманням таких умов: шкала градацій повинна бути повною; градації не повинні перетинатися.

Оцифровані і включені до регресійної моделі номінальні ознаки називають *фіктивними* або *структурними змінними*. Фіктивна змінна – це умовний код, який вказує на належність або неналежність одиниці сукупності до певної градації, тобто відображає два протилежних стани номінальної факторної ознаки. Найчастіше використовують *бінарні фіктивні змінні*, які приймають два значення (0 і 1) в залежності від деякої умови. Якщо номінальна ознака має k градацій, то для її представлення в моделі використовують $(k - 1)$ фіктивну змінну.

Фіктивні змінні включають до регресійної моделі з метою:

- оцінювання впливу структурних (групових) відмінностей одиниць сукупності на пояснювану змінну;
- оцінювання впливу номінальної факторної ознаки на значення пояснюваної змінної.

Регресійна модель може містити одну або декілька фіктивних змінних у поєднанні із кількісними пояснювальними змінними або без них, а форма їх поєднання залежить від припущень щодо характеру впливу фіктивних змінних на пояснювану змінну.

Якщо пояснювана змінна залежить лише від однієї номінальної пояснювальної змінної, яка представлена k градаціями, регресійна модель має вигляд:

$$y = \alpha_0 + \alpha_1 d_1 + \alpha_2 d_2 + \dots + \alpha_{k-1} d_{k-1} + \varepsilon, \quad (3.41)$$

де y - пояснювана змінна; $d_1, d_2, \dots, d_{k-1} \in \{0; 1\}$ - фіктивні пояснювальні змінні; $\alpha_0, \alpha_1, \dots, \alpha_{k-1}$ - параметри моделі; ε - випадкова величина.

Нехай значення результатного показника y формується під впливом m факторних ознак x_1, x_2, \dots, x_m , виражених у метричній шкалі вимірювання, і p факторних ознак u_1, u_2, \dots, u_p , виражених у номінальній шкалі вимірювання. Якщо позначити через $q_k (k = \overline{1, p})$ кількість градацій k -ї номінальної ознаки, то до регресійної моделі можна включити $q_1 + q_2 + \dots + q_p = p$ змінних, а сама модель запишеться так:

$$y = \gamma_0 + \sum_{j=1}^m \beta_j x_j + \sum_{i=1}^p \sum_{l=1}^{q_p-1} \alpha_{il} d_{il} + \varepsilon, \quad (3.42)$$

де d_{il} - i -та фіктивна змінна, обумовлена l -ю ознакою; γ_0 - вільний член; β_j - коефіцієнти регресії, які вимірюють чистий ефект впливу фактора x_j ; α_{il} - коефіцієнти регресії, які вимірюють чистий вплив l -ї градації i -го номінального фактора.

Регресійні моделі з фіктивними змінними слугують ефективним засобом прогнозування сезонних явищ і процесів.

Якщо має місце адитивна сезонність (рівні часового ряду подають як суму трендової і сезонної компонент) і відсутня тенденція, то модель для квартальних даних можна записати у такому вигляді:

$$y_t = \alpha_0 + \alpha_1 d_1 + \alpha_2 d_2 + \alpha_3 d_3 + \varepsilon_t, \quad (3.43)$$

де $y_t (t = \overline{1, n})$ - рівень часового ряду для періоду t ; $\alpha_0, \alpha_1, \alpha_2$ і α_3 - параметри моделі; d_1, d_2 і d_3 - відповідно фіктивні змінні для I, II і III кварталів.

При цьому четвертий квартал розглядають як базу порівняння, а фіктивні змінні набувають таких значень:

- $d_1 = 1$, якщо спостереження відноситься до I кварталу, і $d_1 = 0$ в усіх інших випадках;

- $d_2 = 1$, якщо спостереження відноситься до II кварталу, і $d_2 = 0$ в усіх інших випадках;

- $d_3 = 1$, якщо спостереження відноситься до III кварталу, і $d_3 = 0$ в усіх інших випадках.

Вільний член α_0 у моделі (3.43) інтерпретується як середній рівень для IV кварталу, а коефіцієнти регресії α_j - як різниця між середнім рівнем j -го кварталу і середнім рівнем IV кварталу (для IV кварталу $d_1 = d_2 = d_3 = 0$). Використовуючи параметри моделі (3.43), можна оцінити середні значення для кожного кварталу ($\bar{y}_j = \alpha_0 + \alpha_j$) і абсолютну величину сезонних коливань ($S_j = \bar{y}_j - \bar{y}$).

Якщо параметри моделі (3.43) статистично значущі, то в результаті підстановки значень фіктивних змінних можна отримати прогнозні значення.

У випадку наявності тенденції адитивна модель сезонності з фіктивними змінними записується так:

$$y_t = \tilde{y}_t + \alpha_1 d_1 + \alpha_2 d_2 + \alpha_3 d_3 + \varepsilon_t, \quad (3.44)$$

де \tilde{y}_t - деякий тренд або багатofакторна регресійна модель.

Для лінійного характеру тренду модель (3.44) має такий вигляд:

$$y_t = \alpha_0 + \beta_1 t + \alpha_1 d_1 + \alpha_2 d_2 + \alpha_3 d_3 + \varepsilon_t, \quad (3.45)$$

де t - фактор часу, який набуває значень порядкових номерів кварталів.

Параметри моделі (3.45) мають таку інтерпретацію:

- α_0 - рівень IV кварталу для року $t = 0$ (попереднього до періоду, який досліджується);

- β_1 - вплив тенденції при елімінаванні сезонності;

- α_1 , α_2 і α_3 - зміна рівня для відповідного кварталу в порівнянні з IV кварталом.

У випадку мультиплікативної сезонності модель (3.45) трансформується у таку модель:

$$y_t = ab^t c_1^{d_1} c_2^{d_2} c_3^{d_3} \cdot \varepsilon_t. \quad (3.46)$$

Злогарифмувавши праву і ліву частину рівняння (3.46), отримуємо лінійну модель

$$\ln y_t = \ln a + t \ln b + d_1 \ln c_1 + d_2 \ln c_2 + d_3 \ln c_3 + \ln \varepsilon_t,$$

параметри якої можна оцінити за 1МНК. Після цього шляхом потенціювання переходять до моделі (3.46).

У моделі (3.46) параметр b - це середній коефіцієнт зростання рівня часового ряду незалежно від впливу сезонності, а c_j характеризує вплив сезонності j -го кварталу по відношенню до базового кварталу незалежно від тенденції часового ряду.

3.2. Статистичний аналіз і моделі часових рядів

Однією із важливих задач статистичного аналізу і моделювання є дослідження динаміки соціально-економічних показників. Послідовно розташовані у хронологічному порядку статистичні дані, які відображають розвиток у часі процесу (явища), називають *динамічним рядом*. Оскільки ознакою, за якою здійснюють упорядкування значень економічних показників, є час, то динамічні ряди називають також *часовими*.

До складу часових рядів входять елементи двох типів:

- числові значення статистичних показників (рівні ряду);
- моменти або періоди часу, яким відповідають рівні ряду.

За кількістю показників, для яких визначаються рівні, часові ряди поділяють на одновимірні і багатовимірні (векторні).

Якщо значення рівнів ряду визначаються неперервно у часі, то ряд є неперервним, у протилежному випадку – дискретним. У подальшому будемо розглядати лише дискретні часові ряди.

Часовий ряд можна інтерпретувати як спостереження над неперервним випадковим процесом у певні моменти часу. Під *випадковим процесом* розуміють функцію $Y(t)$ з не випадковим аргументом t , яка за довільного значення t є випадковою величиною. Послідовність спостережень $y_{t_1}, y_{t_2}, \dots, y_{t_n}$ за випадковим процесом $Y(t)$ у статистиці називають *реалізацією випадкового процесу* або *часовим рядом*.

Часовий ряд називають *строго стаціонарним* (або *стаціонарним у вузькому розумінні*), якщо спільний розподіл імовірностей n спостережень $y_{t_1}, y_{t_2}, \dots, y_{t_n}$ відповідає розподілу ймовірностей $y_{t_1+\tau}, y_{t_2+\tau}, \dots, y_{t_n+\tau}$.

На практиці частіше користуються поняттям стаціонарності у широкому розумінні («слабкої» стаціонарності).

Часовий ряд називають *стаціонарним у широкому розумінні*, якщо він характеризується сталими математичним сподіванням, дисперсією та автокореляцією будь-якого порядку в усіх проміжках часу. Якщо ці умови не виконуються, то ряд відносять до класу *нестационарних*.

Стаціонарні часові ряди представляють процеси, які характеризуються тим, що не зазнають якісних змін, тобто не змінюють свою структуру, щільність і напрям взаємозв'язків між елементами.

Окремим випадком стаціонарних часових рядів є «білий шум», під яким розуміють послідовність незалежних між собою однаково розподілених величин із нульовим математичним сподіванням і постійною дисперсією.

Нехай задано часовий ряд y_1, y_2, \dots, y_t . Основні показники часових рядів і формули для їх розрахунку представлені в табл.3.1.

Показники динаміки часового ряду

Показники	Абсолютний приріст	Коефіцієнт зростання	Темп зростання, %	Темп приросту, %
Базисні	$\Delta y_t^{\bar{b}} = y_t - y_{\bar{b}}$	$k_t^{\bar{b}} = \frac{y_t}{y_{\bar{b}}}$	$T_{зр\ t}^{\bar{b}} = k_t^{\bar{b}} \cdot 100$	$T_{пр\ t}^{\bar{b}} = T_{зр\ t}^{\bar{b}} - 100$
Ланцюгові	$\Delta y_t^{\bar{l}} = y_t - y_{t-1}$	$k_t^{\bar{l}} = \frac{y_t}{y_{t-1}}$	$T_{зр\ t}^{\bar{l}} = k_t^{\bar{l}} \cdot 100$	$T_{пр\ t}^{\bar{l}} = T_{зр\ t}^{\bar{l}} - 100$

Важливу роль у процесі дослідження тенденції часових рядів відіграють оцінки середнього арифметичного значення, дисперсії та коваріації. Для інтервального ряду з рівновіддаленими у часі рівнями їх обчислюють за такими формулами:

– середнє арифметичне значення

$$\bar{y} = \frac{1}{n} \sum_{t=1}^n y_t ; \quad (3.47)$$

– дисперсія

$$\sigma^2 = \frac{1}{n-1} \sum_{t=1}^n (y_t - \bar{y})^2 ; \quad (3.48)$$

– коваріація k -го порядку

$$\text{cov}(y_t; y_{t+k}) = \frac{1}{n-1} \sum_{t=k+1}^n (y_t - \bar{y})(y_{t+k} - \bar{y}). \quad (3.49)$$

Середній абсолютний приріст ($\Delta \bar{y}$), середній коефіцієнт зростання (\bar{k}), середній темп зростання ($\bar{T}_{зр}$) і середній темп приросту ($\bar{T}_{пр}$) обчислюють відповідно за такими формулами:

$$\Delta \bar{y} = \frac{y_n - y_1}{n-1} = \frac{1}{n-1} \sum_{t=1}^{n-1} \Delta y_t^{\bar{l}}; \quad (3.50)$$

$$\bar{k} = \sqrt[n-1]{\frac{y_n}{y_1}} = \sqrt[n-1]{k_2^{\bar{l}} \cdot k_3^{\bar{l}} \cdots k_n^{\bar{l}}}, \bar{T}_{зр} = \bar{k} \cdot 100\%; \quad (3.51)$$

$$\bar{T}_{пр} = \bar{T}_{зр} - 100\%. \quad (3.52)$$

Середній абсолютний приріст і середній коефіцієнт зростання (або середній темп зростання) як узагальнюючі характеристики відповідно

швидкості та інтенсивності зміни рівнів часового ряду можна використовувати для побудови найпростіших прогнозів.

У випадку, коли тенденція розвитку процесу є близькою до лінійної, тобто значення ланцюгових абсолютних приростів є приблизно однаковими, то за точковий прогноз можна прийняти:

$$y_{n+\tau}^{\text{прогн}} = y_n + \tau \cdot \Delta \bar{y}, \quad (3.53)$$

де τ – період випередження.

Якщо рівні часового ряду наближено відповідають характеру зміни показникової або експоненційної функції, то прогнозне значення можна отримати за формулою

$$y_{n+\tau}^{\text{прогн}} = \bar{k} y_n. \quad (3.54)$$

Для отримання якісної статистичної моделі необхідна повна і достовірна інформація про рівні часового ряду. Часовий ряд об'єктивно відображає тенденцію розвитку соціально-економічного процесу лише за умови можливості зіставлення (порівняння) його рівнів. Основними вимогами щодо зіставлення часових рядів вважають такі:

- рівність періодів (моментів часу), яким відповідають рівні ряду;
- однакова повнота охоплення досліджуваних частин процесу (явища);
- збіг просторових меж процесу (явища);
- спільномірність рівнів ряду (однаковий масштаб виміру);
- єдине тлумачення одиниці об'єкта спостереження.

Рівні часового ряду можуть складатися зі значень, які сформувалися під впливом характерних для досліджуваного процесу чинників, а також отриманих під впливом інших чинників, які не є типовими для цього процесу. Такі значення різко виділяються і, як наслідок, використання методології статистичного аналізу, яка не враховує аномальні значення, може призвести до істотних помилок моделювання. Через це рівні, які суттєво виділяються на фоні інших, потребують додаткового вивчення.

Для виявлення аномальних рівнів найчастіше використовують *метод Ірвіна*, що ґрунтується на такій статистиці:

$$\lambda_t = \frac{|y_t - y_{t-1}|}{\sigma_y}, t = \overline{2, n}, \quad (3.55)$$

де σ_y – середнє квадратичне відхилення рівнів часового ряду.

Розраховані значення статистики λ_t порівнюють з критичними $\lambda_{кр}$ (табл. 3.2). Якщо $\lambda_t > \lambda_{кр}$, то відповідний рівень часового ряду y_t вважають аномальним.

Таблиця 3.2

Критичні значення критерію Ірвіна ($\alpha = 0,05$)

<i>n</i>	10	20	30	50	100
$\lambda_{крит}$	1,5	1,3	1,2	1,1	1,0

Після виявлення аномальних рівнів часового ряду аналізують причини їхнього виникнення. Якщо аномальні значення зумовлені причинами випадкового характеру, то їх переважно вилучають або замінюють апроксимованими значеннями чи середніми арифметичними значеннями двох сусідніх рівнів. Аномальні значення, пов'язані з дією чинників, що мають об'єктивний характер, заміні не підлягають.

У загальному випадку часовий ряд можна розкласти на такі складові:

тренд ($f(t)$) – характеризує основну тенденцію розвитку процесу (явища) і є результатом комплексної дії чинників упродовж тривалого періоду часу;

сезонна компонента (s_t) – характеризує вплив періодичних або близьких до них коливань сезонного характеру (вплив пори року на виробництво та реалізацію певних видів товарів і надання послуг – продаж зимового спортивного інвентарю, споживання безалкогольних напоїв, морозива, транспортні перевезення тощо);

циклічна компонента (c_t) – характеризує вплив циклів економічного розвитку (наприклад, пов'язаних з демографічними процесами або процесами технічного переозброєння виробництва);

випадкова (нерегулярна) компонента (ε_t) – характеризує дію випадкових чинників.

Трендову, сезонну і циклічну компоненти називають *систематичними (регулярними)* складовими рівнів часового ряду. Вони характеризують вплив постійно діючих (основних) чинників на рівні часового ряду.

Випадкова компонента формується під впливом чинників двох видів: раптової (неочікуваної) та (або) поточної дії. Чинники першого виду (різкі зміни курсу валют, стихійні лиха тощо) зумовлюють значні стрибки рівнів часового ряду. Чинники другого виду – випадкові коливання, які є результатом дії певної кількості побічних причин.

Необхідно зазначити, що усі складові (компоненти) часового ряду не піддаються спостереженню і є теоретичними величинами.

Сукупність статистичних методів, які використовуються для виявлення складових часового ряду, називають *аналізом часових рядів (time series analysis)*. Основні задачі аналізу часових рядів зводяться до виявлення складових, які формують рівні ряду, і побудови відповідних математичних моделей.

Основна мета статистичного аналізу часових рядів полягає у встановленні співвідношення між закономірністю та випадковістю в процесі формування значень соціально-економічних показників.

Часовий ряд представляють різними типами моделей.

Модель часового ряду, яку представлено у вигляді суми тренда, сезонної, циклічної і випадкової компонент, називають *адитивною* моделлю:

$$y_t = f(t) + s_t + c_t + \varepsilon_t. \quad (3.56)$$

Модель часового ряду, яку представлено у вигляді добутку тренда, сезонної, циклічної і випадкової компонент – *це мультиплікативна* модель:

$$y_t = f(t) \cdot s_t \cdot c_t \cdot \varepsilon_t. \quad (3.57)$$

Модель, яка має вигляд

$$y_t = f(t) \cdot s_t \cdot c_t + \varepsilon_t, \quad (3.58)$$

називають *комбінованою* моделлю часового ряду.

Використані у наведених моделях компоненти не обов'язково притаманні будь-якому часовому ряду. Існують ряди, в яких відсутні як тренд, так і сезонні та циклічні коливання. У цьому випадку рівні ряду є функцією випадкової компоненти і коливаються навколо середнього рівня, що характерно для стаціонарних часових рядів.

На основі моделей (3.56)-(3.58) шляхом елімінування (виключення впливу) окремих компонент досліджують інші компоненти, зокрема елімінуючи сезонні, циклічні і випадкові коливання, можна досліджувати *основну тенденцію ряду* – тренд.

У процесі аналізу часових рядів досліджують тенденції трьох видів:

- *тенденцію середнього рівня (основну тенденцію)* – зміну середніх значень часового ряду;
- *тенденцію дисперсії* – зміну відхилень емпіричних значень рівнів часового ряду від обчислених за рівнянням тренду;
- *тенденцію автокореляції* – зміну зв'язків між окремими рівнями часового ряду.

Тренд, як загальний напрям розвитку процесу (явища), переважно подають у вигляді гладкої кривої (траєкторії), що є функцією від часу. У цьому разі передбачають, що вплив усіх основних чинників на процес дослідження можна виразити через функцію часу. Тому під трендом часто розуміють рівняння регресії $\hat{y} = f(t)$. Зазначимо, що представлення тренда лише одним параметром t не дає змоги повністю описати характер розвитку процесу, і його не можна розглядати як єдине джерело формування закономірності процесу.

Перші припущення щодо характеру зміни тренда можна зробити в результаті візуального аналізу графіка часового ряду.

Якщо графічний аналіз не забезпечує чіткого уявлення про характер часового ряду, то проводять його тестування за статистичними критеріями. У випадку, коли в практичних дослідженнях результати одного тесту не дають чіткої підстави для прийняття або відхилення висунутої гіпотези, виникає необхідність використання інших тестів. Сукупність тестів для перевірки основної тенденції часового ряду поділять на дві основні групи:

непараметричні – не передбачають наявності будь-яких припущень щодо закону розподілу рівнів часового ряду;

параметричні – застосовують лише за умови нормального розподілу рівнів часового ряду).

Для перевірки гіпотези про наявність тренда в часовому ряді можна використати *метод перевірки різниць середніх* – параметричний тест на основі критерію Стьюдента. Вигляд математичного виразу для обчислення критеріальної статистики залежить від прийнятого припущення щодо дисперсії часового ряду. Розглянемо послідовність тестування за одним із варіантів методу перевірки різниць середніх.

На першому етапі часовий ряд ділять на дві послідовні приблизно однакові за кількістю рівнів частини, кожен з яких розглядають як окрему нормально розподілену вибірку. Якщо розбіжність між середніми двох сукупностей є неістотною (випадковою), то можна стверджувати про відсутність основної тенденції у часовому ряді, в протилежному випадку – про наявність тренда. Тому перевірка наявності основної тенденції зводиться до перевірки гіпотези про рівність середніх двох нормально розподілених сукупностей.

Прийmemo такі позначення: n_1 , n_2 – обсяги першої і другої частини сукупності відповідно; \bar{y}_1 , \bar{y}_2 – середні значення для першої і другої частини сукупності відповідно.

На другому етапі оцінюють дисперсії першої і другої частини часового ряду (\hat{S}_1^2 і \hat{S}_2^2):

$$\hat{S}_1^2 = \frac{1}{n_1-1} \sum_{t=1}^{n_1} (y_t - \bar{y}_1)^2; \quad \hat{S}_2^2 = \frac{1}{n_2-1} \sum_{t=n_1+1}^{n_2} (y_t - \bar{y}_2)^2. \quad (3.59)$$

Метод різниць середніх ґрунтується на гіпотезі про рівність дисперсій першої і другої частин ($S_1^2 = S_2^2$), тому на третьому етапі перевіряють гіпотезу про рівність дисперсій за F -критерієм. Розрахункове значення критеріальної F -статистики обчислюють за формулою

$$F_{\text{розн}} = \frac{\max(\hat{S}_1^2; \hat{S}_2^2)}{\min(\hat{S}_1^2; \hat{S}_2^2)}. \quad (3.60)$$

Якщо для вибраного рівня значущості α та V_1 і V_2 ступенів вільності виконується умова

$$F_{\text{кр}} \left(1 - \frac{\alpha}{2}; V_1; V_2 \right) < F_{\text{розн}} < F_{\text{кр}} \left(\frac{\alpha}{2}; V_1; V_2 \right), \quad (3.61)$$

то приймають гіпотезу про рівність дисперсій ($H_0: S_1^2 = S_2^2$), у протилежному випадку – приймають альтернативну ($H_1: S_1^2 \neq S_2^2$) гіпотезу. Якщо $\hat{S}_1^2 > \hat{S}_2^2$, то $V_1 = n_1 - 1$ і $V_2 = n_2 - 1$, інакше – $V_1 = n_2 - 1$ і $V_2 = n_1 - 1$.

У випадку прийняття гіпотези про рівність дисперсій приступають до четвертого етапу – перевірки гіпотез про рівність середніх рівнів двох сукупностей ($H_0: \bar{y}_1 = \bar{y}_2; H_1: \bar{y}_1 \neq \bar{y}_2$) за t -критерієм. Для цього обчислюють значення t -статистики:

$$t_{\text{розн}} = \frac{|\bar{y}_1 - \bar{y}_2|}{\hat{S} \cdot \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}}. \quad (3.62)$$

У формулі (3.62) \hat{S}^2 – це загальна оцінка дисперсії, обчислена за оцінками дисперсій \hat{S}_1^2 першої і \hat{S}_2^2 другої частин:

$$\hat{S}^2 = \frac{1}{n_1 + n_2 - 2} [(n_1 - 1)\hat{S}_1^2 + (n_2 - 1)\hat{S}_2^2]. \quad (3.63)$$

З таблиці розподілу Стюдента для вибраного рівня значущості α та $V = n_1 + n_2 - 2$ ступенів вільності знаходять критичне значення $t_{\text{кр}}$ і порівнюють з розрахованою t -статистикою ($t_{\text{розн}}$). Якщо $t_{\text{розн}} < t_{\text{кр}}$, то

нульову гіпотезу H_0 про рівність середніх не відхиляють, тобто розбіжність між середніми рівнями вважають неістотною, що свідчить про відсутність тренда в часовому ряді.

Якщо гіпотезу про рівність дисперсій відхиляють, то t -критерій (3.62) застосовувати не можна.

Розглянутий вище метод дає надійні результати тоді, коли часовий ряд змінюється монотонно. Якщо ж для ряду властива тенденція зміни загального напрямку розвитку, то перевірка гіпотези може не виявити наявності основної тенденції.

Надійнішим методом перевірки наявності тренда в часовому ряді, порівняно із методом перевірки різниць середніх, є застосування непараметричного *методу Фостера–Стюарта*. Крім перевірки наявності тренда, цей метод дає змогу встановити наявність тенденції дисперсії часового ряду. Відсутність тенденції дисперсії означає, що варіація рівнів часового ряду є постійною.

Реалізацію методу Фостера–Стюарта розпочинають з порівняння кожного рівня часового ряду (починаючи з другого, $t = \overline{2, n}$) з усіма попередніми рівнями. У результаті порівняння формують дві числові послідовності:

$$U_t = \begin{cases} 1, \text{ якщо значення } u_t \text{ більше за всі попередні рівні,} \\ 0 - \text{ у протилежному випадку;} \end{cases} \quad (3.64)$$

$$W_t = \begin{cases} 1, \text{ якщо значення } u_t \text{ менше за всі попередні рівні,} \\ 0 - \text{ у протилежному випадку.} \end{cases} \quad (3.65)$$

Після цього обчислюють такі значення:

$$\begin{aligned} m_t &= U_t + W_t; & k_t &= U_t - W_t; \\ M &= \sum_{t=2}^n m_t; & K &= \sum_{t=2}^n k_t. \end{aligned}$$

Статистика M набуває значень від 0 (ряд монотонно спадає) до $(n - 1)$ (ряд монотонно зростає), і використовується для перевірки наявності тенденції дисперсії.

Статистика K набуває значень від мінус $(n - 1)$ (ряд монотонно спадає) до плюс $(n - 1)$ (ряд монотонно зростає або має місце чергування зростання і зменшення рівнів). Її використовують для перевірки наявності тренда.

Для перевірки наявності основної тенденції та тенденції дисперсії використовують критерій Стьюдента:

$$t_{\text{розр}}^1 = \frac{M - \mu}{\sigma_1}; \quad (3.66)$$

$$t_{\text{розр}}^2 = \frac{K - o}{\sigma_2}, \quad (3.67)$$

де μ – математичне сподівання M ; σ_1 – середнє квадратичне відхилення M ; σ_2 – середнє квадратичне відхилення K .

Значення μ , σ_1 та σ_2 табульовані для різних обсягів спостережень (табл. 3.3). Величини $t_{\text{розр}}^1$ і $t_{\text{розр}}^2$ порівнюють з критичним значенням $t_{\text{кр}}$ для вибраного рівня значущості α і значення ступеня вільності $V = n - 1$. Якщо розрахункове значення перевищує критичне, то відповідну гіпотезу про відсутність тенденції відхиляють.

Таблиця 3.3

Табульовані значення μ , σ_1 та σ_2

n	μ	σ_1	σ_2
10	3,858	1,288	1,964
15	4,636	1,521	2,158
20	5,195	1,677	2,279
25	5,632	1,791	2,373
30	5,990	1,882	2,447
35	6,294	1,956	2,509
40	6,557	2,019	2,561
45	6,790	2,072	2,606
50	6,998	2,121	2,645

Із сукупності інших методів і критеріїв тестування тенденції часового ряду виділяють методи Манна–Уїтні та Кокса–Стюарта, метод серій, критерії Кохрена, Бартлетта, Дікі–Фуллера та ін.

За наявності у часовому ряді тренда та сезонних і циклічних коливань значення кожного наступного рівня ряду залежить від попередніх.

Кореляційну залежність між послідовними рівнями часового ряду називають *автокореляцією*.

Кількісною мірою щільності (залежності) лінійного зв'язку між послідовними рівнями часового ряду слугує *коефіцієнт автокореляції* $r(\tau)$ з часовим лагом τ (*коефіцієнт автокореляції порядку* τ), який розраховують за такою формулою:

$$r(\tau) = \frac{(n-\tau) \sum_{t=1}^{n-\tau} y_t y_{t+\tau} - \sum_{t=1}^{n-\tau} y_t \sum_{t=1}^{n-\tau} y_{t+\tau}}{\sqrt{(n-\tau) \sum_{t=1}^{n-\tau} y_t^2 - (\sum_{t=1}^{n-\tau} y_t)^2} \cdot \sqrt{(n-\tau) \sum_{t=1}^{n-\tau} y_{t+\tau}^2 - (\sum_{t=1}^{n-\tau} y_{t+\tau})^2}}. \quad (3.68)$$

Коефіцієнт автокореляції $r(\tau)$ характеризує щільність зв'язку між послідовними рівнями часового ряду y_1, y_2, \dots, y_n та $y_{1+\tau}, y_{2+\tau}, \dots, y_{n+\tau}$ (зміщеними один відносно іншого на τ одиниць) і змінюється в межах $[-1; +1]$.

Близьке до одиниці за абсолютною величиною значення коефіцієнта автокореляції вказує на щільну лінійну залежність між початковими і зсуненими рівнями часового ряду, а близьке до нуля – на їхню незалежність.

Формулу (3.68) використовують лише для оцінювання щільності лінійного зв'язку між рівнями часового ряду, а тому її застосування у випадку чітко вираженої нелінійної тенденції є неправомірним.

На підставі значення коефіцієнта автокореляції можна зробити висновки лише про наявність чи відсутність лінійної (або близької до неї) тенденції, але за знаком коефіцієнта автокореляції не можна зробити висновок про зростаючу чи спадну тенденції в рівнях ряду.

Лаг τ (зсув у часі) визначає порядок автокореляції. Якщо $\tau = 1$, то наявна автокореляція 1-го порядку, якщо $\tau = 2$, то – 2-го порядку і т.д.

Послідовність значень коефіцієнтів автокореляції першого, другого та інших порядків утворює *автокореляційну функцію* – функцію оцінювання автокореляції залежно від часового лага. Графік автокореляційної функції називають *корелограмою*. Аналіз автокореляційної функції і корелограми дає змогу визначити лаг, за якого зв'язок між поточним та попередніми рівнями ряду є найщільнішим.

Обчисливши декілька коефіцієнтів автокореляції, можна визначити лаг, за якого автокореляція є найбільшою, і встановити таким чином структуру часового ряду:

якщо коефіцієнт автокореляції 1-го порядку є статистично значущим і характеризується максимальним значенням серед інших, то ряд містить лише лінійний тренд;

якщо максимального значення набуває статистично значущий коефіцієнт автокореляції p -го порядку, то часовий ряд, крім тенденції, містить коливання з періодичністю в p моментів часу, причому ці коливання можуть бути як сезонними, так і циклічними;

якщо жоден із коефіцієнтів автокореляції не є статистично значущим, то ряд характеризується випадковою структурою або сильною нелінійною тенденцією, виявлення якої вимагає додаткових досліджень.

Зі збільшенням лага (τ) кількість пар $(n - \tau)$ спостережень $(y_t, y_{t+\tau})$ зменшується, тому лаг τ необхідно вибирати таким, щоб величина $(n - \tau)$ була достатньою для знаходження $r(\tau)$. На практиці користуються правилом, за яким максимальне значення τ не повинно перевищувати величину $\frac{n}{4}$, де n – кількість рівнів часового ряду.

З практичного погляду ($\alpha = 0,05$) можна стверджувати, що коефіцієнт автокореляції $r(\tau)$ буде статистично значущим, якщо його абсолютне значення перевищує $\frac{2}{\sqrt{n-\tau}}$.

Корелограма наочно показує, як зміни попередніх рівнів впливають на наступні рівні часового ряду, що слугує основою для виявлення структури

часового ряду. На рис. 3.3 наведено типову корелограму стаціонарного ряду, для якого індивідуальні значення коливаються навколо середнього значення, не виявляючи при цьому ні суттєвої тенденції зміни середнього, ні сезонних особливостей. У зміні значень коефіцієнтів автокореляції відсутня певна закономірність і коефіцієнти автокореляції статистично незначущі (значення коефіцієнтів автокореляції розташовані між паралельними прямими, які визначають межі значущості).

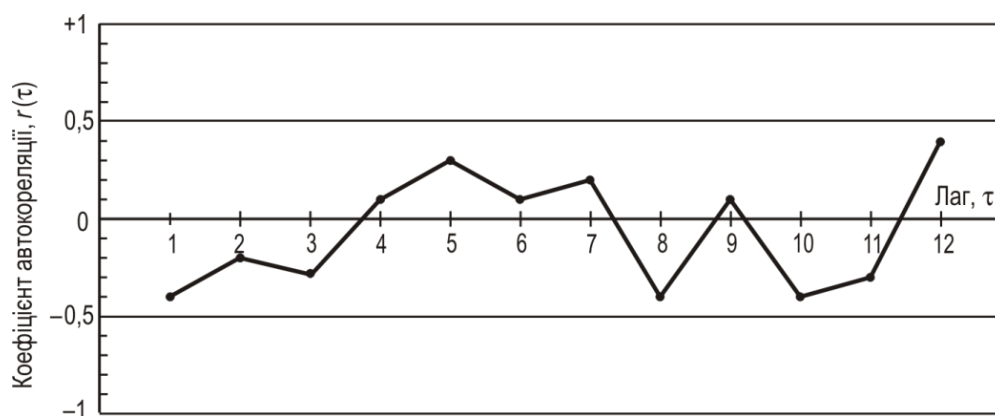


Рис. 3.3. Корелограма стаціонарного часового ряду

Корелограму часового ряду з лінійно-адитивним трендом (зростання середнього значення відбувається приблизно на однакову величину для кожного моменту часу) наведено на рис. 3.4. Візуальний аналіз цієї корелограми вказує на залежність значень коефіцієнта автокореляції від величини лага. Зі зростанням лага абсолютні значення коефіцієнтів автокореляції зменшуються, а максимальне значення (за абсолютною величиною) відповідає лагу $\tau = 1$, причому значущим є лише коефіцієнт автокореляції $r(1)$.

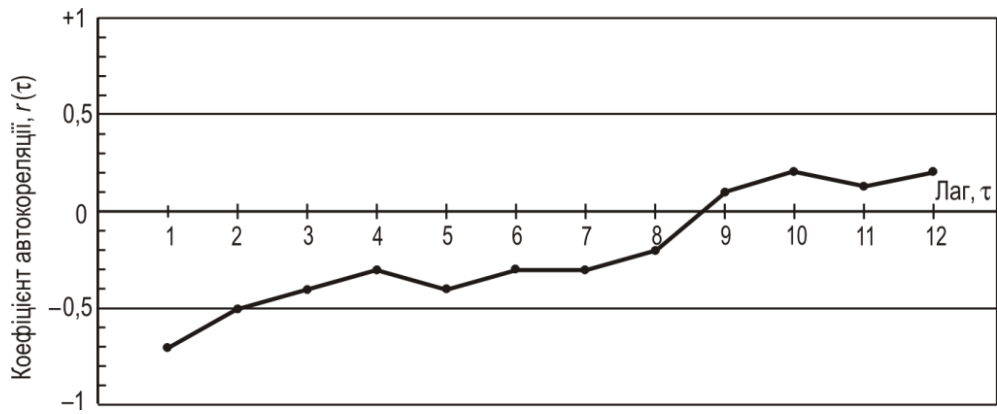


Рис. 3.4. Корелограма часового ряду з лінійно-адитивним трендом

Корелограму часового ряду із сезонною компонентою наведено на рис. 3.5.

Для сезонних явищ спостерігається щільна кореляційна залежність між рівнями часового ряду, що відповідають одному і тому ж періоду часу. Найбільші за абсолютною величиною значення набувають коефіцієнти автокореляції з лагами 6, 12, 18 і 24. Вони є статистично значущими, і можна зробити висновок про наявність у часовому ряді сезонних коливань з періодом в один рік.

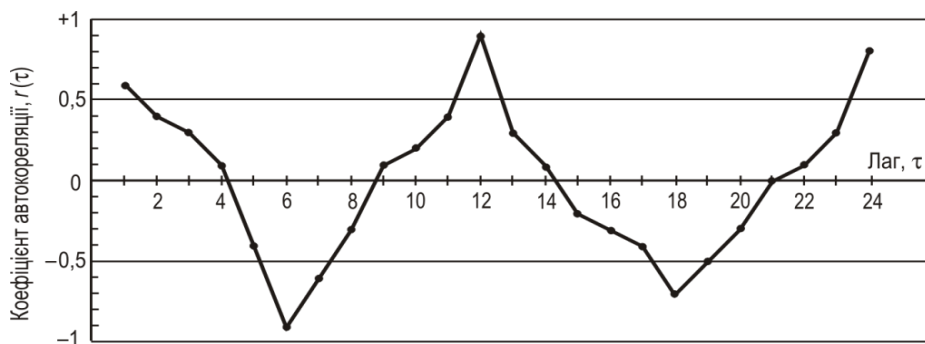


Рис. 3.5. Корелограма часового ряду із сезонною компонентою

У разі стаціонарності часового ряду зі зростанням лага τ взаємозв'язок між рівнями послаблюється і автокореляційна функція $r(\tau)$ спадає (за абсолютним значенням). Якщо формально побудована корелограма не є

функцією, що швидко спадає, то це, як правило, свідчить про нестационарність ряду.

Тренд часового ряду в переважній більшості випадків прихований дією сезонних, циклічних і випадкових коливань. З метою виділення тренда («очищення» рівнів часового ряду від впливу інших складових) виконують процедуру, яку прийнято називати *згладжуванням* (*вирівнюванням*, *фільтрацією*). Завдання процедури згладжування – заміна фактичних значень рівнів часового ряду розрахунковими. Останні менш схильні до коливань, що сприяє чіткішому прояву основної тенденції.

Інструменти згладжування часових рядів умовно поділяють на дві групи:

алгоритмічні – методи, які ґрунтуються на усередненні декількох сусідніх рівнів часового ряду;

аналітичні – методи, які ґрунтуються на представленні тренда у вигляді деякої відомої з точністю до параметрів функції.

В основу алгоритмічних методів згладжування покладено обчислення *плинних середніх* – послідовних серій середніх значень із встановленим періодом згладжування.

Розрізняють *просту*, *зважену* та *експоненційно зважену плинні середні*.

Розглянемо часовий ряд y_1, y_2, \dots, y_n і виберемо інтервал згладжування довжиною m . Довжина інтервалу згладжування – це кількість рівнів, які беруть участь в усередненні часового ряду.

Сукупність спостережень (рівнів часового ряду), які використовують для розрахунку плинних середніх значень, називають *активним інтервалом згладжування*.

Довжина інтервалу згладжування може виражатися парним ($m = 2p$) або непарним ($m = 2p + 1$) числом. Якщо m – непарне число, то згладжене значення відповідає конкретному періоду (моменту) часу. Якщо ж m є

парним числом, то плинна середня відповідає точці між двома моментами часу i , як правило, пари плинних середніх центрують (усереднюють).

Значення простої плинної середньої (\bar{y}_t) для моменту часу t обчислюють за формулою

$$\bar{y}_t = \frac{y_{t-p} + y_{t-p+1} + \dots + y_t + \dots + y_{t+p}}{m}. \quad (3.69)$$

Вибір довжини інтервалу згладжування (вікна згладжування) залежить від задач і цілей дослідження. Насамперед керуються намаганням усунути або зменшити вплив випадкових і сезонних чинників. Чим більша довжина інтервалу згладжування, тим краще усувається (елімінується) вплив цих чинників, тобто якісніше представляється тренд часового ряду. Разом з тим потрібно враховувати, що ряд плинних середніх коротший від первісного часового ряду на $(m - 1)$ рівнів.

У процесі згладжування часового ряду, який містить сезонну компоненту, необхідно враховувати період сезонності. Для часових рядів, які не містять сезонної компоненти, розмір вікна згладжування найчастіше приймають рівним 3 або 5.

Обчислюючи просту плинну середню, ваги (частки впливу) всіх рівнів часового ряду вважають однаковими. У зв'язку з цим просту плинну середню можна використовувати лише тоді, коли графік динаміки часового ряду близький до прямої лінії. Якщо ж тренд має перегин, то проста плинна середня може істотно спотворювати розвиток процесу (одночасно вирівнюються опуклі та вігнуті інтервали, ігнорується зміна знаків лінії).

Зважене значення плинної середньої для t -го рівня часового ряду обчислюють за такою формулою:

$$\bar{y}_t = \sum_{i=-p}^p w_i y_{t+i}, \quad (t = \overline{p+1, m-p}, \quad m = 2p + 1), \quad (3.70)$$

де w_i – вага для i -го рівня в інтервалі згладжування ($w_i \geq 0$; $\sum w_i = 1$).

Встановлені на ваги (w_i) умови не забезпечують однозначного вибору вагових коефіцієнтів. Розглядають декілька модифікацій зваженої плинної

середньої, які використовують різні варіанти розрахунку ваг. Найчастіше вибір значень w_i ґрунтується на припущенні, що довільну гладку функцію на обмеженому інтервалі можна представити поліномом деякого степеня. Значення рівня часового ряду (y_t) в момент t замінюють значенням полінома в цій точці, яке залежить від степеня полінома (q) та довжини інтервалу згладжування (m). Вибираючи степінь полінома, слід враховувати, що з його збільшенням зростає точність апроксимації, але гірше елімінуються впливи випадкових і сезонних коливань, що ускладнює виявлення основної тенденції часового ряду.

Якщо представити оцінку t -го рівня часового ряду (\hat{y}_t) поліномом q -го степеня

$$\hat{y}_t = b_0 + b_1 t + b_2 t^2 + \dots + b_q t^q \quad (3.71)$$

то невідомі коефіцієнти b_0, b_1, \dots, b_q можна визначити на основі критерію мінімізації суми квадратів відхилень значень полінома від фактичних значень рівнів часового ряду.

Припустимо, що довжина інтервалу згладжування дорівнює 5, і значення рівнів часового ряду апроксимуються поліномом 2-го степеня ($m = 5; q = 2$). Перенесемо початок відліку часу (початок координат) в середину активного інтервалу і розглянемо моменти часу $t: -2, -1, 0, 1, 2$. Тоді невідомі параметри b_0, b_1 і b_2 можна оцінити на підставі мінімізації функціоналу

$$Q = \sum_{t=-2}^2 (y_t - b_0 - b_1 t - b_2 t^2) \rightarrow \min. \quad (3.72)$$

Прирівнявши частинні похідні виразу (3.72) до нуля і розв'язавши відповідну систему рівнянь, отримуємо:

$$b_0 = \frac{1}{35} (-3y_{-2} + 12y_{-1} + 17y_0 + 12y_1 - 3y_2);$$

$$b_1 = \frac{1}{10} (-2y_{-2} - y_{-1} + y_1 + 2y_2);$$

$$b_2 = \frac{1}{14} (2y_{-2} - y_{-1} + 2y_0 - y_1 + 2y_2).$$

Згладжене значення рівня, що відповідає середині активного інтервалу, приймається рівним значенню параметра b_0 у зв'язку з перенесенням початку координат. Тому заново обчислювати вагові коефіцієнти для рівнів, що входять до наступного активного інтервалу згладжування, немає потреби (вони будуть однаковими для кожного активного інтервалу).

Отже, оцінка згладженого значення в центральній точці активного інтервалу визначається як зважена середня арифметична п'яти рівнів цього інтервалу з такими ваговими коефіцієнтами (w_i):

$$-\frac{3}{35}; \frac{12}{35}; \frac{17}{35}; \frac{12}{35}; -\frac{3}{35}.$$

Враховуючи симетричність відносно центрального елемента, користуються таким скороченим символічним записом значень вагових коефіцієнтів:

$$b_0 = \frac{1}{35} [-3; 12; 17].$$

У загальному випадку для кожного активного інтервалу довжиною $m = 2p + 1$ коефіцієнти полінома порядку q розраховують за критерієм

$$Q = \sum_{t=-p}^p (y_t - b_0 - b_1 t - b_2 t^2 - \dots - b_q t^q)^2 \rightarrow \min. \quad (3.73)$$

У цьому разі вагові коефіцієнти, обчислені на основі полінома степеня $q = 2l + 1$, будуть такими ж, як і в разі використання полінома степеня $q = 2l$.

Вагові коефіцієнти, які обчислені вищезазначеним методом, мають такі властивості: сума ваг дорівнює 1; ваги симетричні відносно середини інтервалу згладжування.

Вагові коефіцієнти, які використовуються для обчислення зважених плинних середніх, подано в табл. 3.4.

Використання плинних середніх (як простої, так і зваженої) не забезпечує згладжування перших і останніх рівнів часового ряду. Проте відсутність перших рівнів не є такою важливою, як відсутність останніх, особливо тоді, коли метою дослідження є прогнозування. До методів

згладжування, використання яких дає можливість згладити перші та останні рівні, відносять *метод експоненційного згладжування*.

Таблиця 3.4

Вагові коефіцієнти для зваженої плинної середньої ($q = 2; 3$)

Довжини інтервалів згладжування	Вагові коефіцієнти
$m = 3$	$\frac{1}{4}, \frac{1}{2}, \frac{1}{4}$
$m = 5$	$\frac{1}{35}[-3; 12; 17]$
$m = 7$	$\frac{1}{21}[-2; 3; 6; 7]$
$m = 9$	$\frac{1}{231}[-21; 14; 39; 54; 59]$
$m = 11$	$\frac{1}{429}[-36; 9; 44; 69; 84; 89]$

Особливість експоненційного згладжування полягає в тому, що у процесі вирівнювання часового ряду використовують лише попередні рівні з певною вагою.

Згладжене значення рівня часового ряду (S_t) на момент часу t обчислюють за формулою

$$S_t = \alpha y_t + (1 - \alpha)S_{t-1}, \quad (3.74)$$

де α – параметр згладжування ($0 < \alpha < 1$), який слугує для встановлення ваг рівнів часового ряду.

Згідно (3.74) згладжене значення ряду (S_t) на момент часу t подають як лінійну комбінацію фактичного рівня (y_t) для цього ж моменту часу і згладженого рівня для попереднього моменту часу (S_{t-1}).

Значення S_{t-1} відповідно рівне

$$S_{t-1} = \alpha y_{t-1} + (1 - \alpha)S_{t-2}.$$

Продовживши процес такого розгляду, отримаємо:

$$S_t = \alpha y_t + \alpha(1 - \alpha)y_{t-1} + \alpha(1 - \alpha)^2 y_{t-2} + \dots + \alpha(1 - \alpha)^{t-1} y_1 + \alpha(1 - \alpha)^t y_0$$

або

$$S_t = \alpha \sum_{i=0}^{t-1} (1 - \alpha)^i y_{t-i} + \alpha(1 - \alpha)^t S_0, \quad (3.75)$$

де $\alpha(1 - \alpha)^i$ – вага $(t-i)$ -го рівня часового ряду.

Послідовність значень $\alpha, \alpha(1 - \alpha), \alpha(1 - \alpha)^2, \dots, \alpha(1 - \alpha)^n$ експоненційно спадає, оскільки $0 < \alpha < 1$, тобто вагова функція формується з урахуванням новизни інформації.

Розрахунок експоненційних середніх вимагає зазначення початкових умов, а саме значень S_0 і α .

Значення початкового рівня S_0 можна вибрати рівним:

- середньому арифметичному даних минулих спостережень або їхньої частини (якщо є такі дані);
- середньому арифметичному значень декількох перших рівнів початкового ряду;
- значенню першого рівня початкового ряду.

Правильний вибір початкового рівня S_0 відіграє важливу роль у згладжуванні часових рядів із невеликою кількістю спостережень.

Математичні сподівання початкового часового ряду та згладженого за допомогою експоненційної зваженої середньої є рівними між собою. Дисперсію (δ^2) згладженого ряду обчислюють за формулою:

$$\delta^2 = \frac{\alpha}{2 - \alpha} \sigma^2, \quad (3.76)$$

де σ^2 – дисперсія початкового часового ряду.

З (3.76) можна зробити висновок, що дисперсія початкового часового ряду більша за дисперсію згладженого часового ряду.

Вибираючи значення параметра згладжування α , необхідно враховувати, що від нього залежить «чутливість» і «стійкість» експоненційної середньої. Чим більше значення α , тим «чутливішою» є експоненційна середня до останніх рівнів часового ряду. Чим менше значення α , тим експоненційна середня є «стійкішою». На практиці підбір значення параметра згладжування α рекомендують проводити емпіричним шляхом, тобто ітеративно перебираючи його можливі значення і керуючись

критерієм мінімізації відхилень фактичних і згладжених значень рівнів часового ряду.

Досконалішим методом згладжування часових рядів є аналітичне вирівнювання, яке ґрунтується на тому, що зміну рівнів ряду можна наближено представити деякою математичною функцією. У процесі аналітичного вирівнювання фактичні рівні часового ряду замінюють теоретичними, розрахованими, як правило, на підставі рівняння парної регресії, в якому час розглядають як незалежну змінну, а рівні ряду – як функцію цієї змінної.

Аналітичні методи згладжування часових рядів ґрунтуються на такій моделі:

$$y_t = f(t) + \varepsilon_t, \quad (3.77)$$

де $f(t)$ – функція, параметри якої необхідно оцінити (детермінована складова); ε_t – випадкова складова, яка акумулює вплив множини випадкових чинників.

Випадкова складова, на відміну від детермінованої, не пов'язана зі зміною часу. Її аналіз слугує основою перевірки гіпотези про адекватність вибраної моделі. За умови, що модель вибрана правильно, випадкова складова є нормально розподіленою величиною з нульовим математичним сподіванням і дисперсією

$$\sigma_\varepsilon^2 = \frac{\sum_{t=1}^n (y_t - f(t))^2}{n-m}, \quad (3.78)$$

де m – кількість параметрів функції, n – кількість рівнів ряду.

Параметри функції $f(t)$ оцінюють на основі критерію мінімізації суми квадратів відхилень фактичних рівнів часового ряду (y_i) від обчислених за рівнянням регресії (\hat{y}_i):

$$Z = \sum_{i=1}^n (y_i - \hat{y}_i)^2 \rightarrow \min. \quad (3.79)$$

На практиці переважно використовують функції, параметри яких мають змістовну інтерпретацію (абсолютну швидкість зміни рівнів ряду, приріст абсолютної швидкості тощо).

Рівняння лінійного тренда має вигляд:

$$\hat{y} = b_0 + b_1 t. \quad (3.80)$$

Лінійний тренд використовують для згладжування часових рядів, які характеризуються рівномірним збільшенням (зменшенням) рівнів, коли абсолютні ланцюгові прирости близькі між собою (b_1 – постійний приріст за початкового рівня b_0).

Функції, які використовують для представлення нелінійних трендів соціально-економічних процесів, залежно від динаміки розвитку умовно поділяють на групи функцій, які описують необмежені та обмежені монотонні процеси.

Типовим прикладом застосування функцій першої групи є представлення динаміки показників виробництва продукції, а другої – середньодушового споживання харчових продуктів.

Функції другої групи називають *кривими насичення*. Криві насичення, які мають точку перегину, називають *S-подібними кривими*. Ці криві описують два послідовні процеси розвитку: один з прискоренням розвитку; інший – із сповільненням.

Поліном другого порядку (параболу)

$$\hat{y} = b_0 + b_1 t + b_2 t^2 \quad (3.81)$$

використовують для опису процесів, характерною особливістю яких є рівноприскорене зростання або спадання рівнів ряду. Тоді параметр b_1 характеризує постійну швидкість зростання, а параметр b_2 – постійну швидкість приросту (пришвидшення зростання).

Тренд часового ряду з порівняно стабільними темпами приросту доцільно подавати у вигляді показникової (3.82) або модифікованої показникової (3.83) функції:

$$\hat{y} = ab^t; \quad (3.82)$$

$$\hat{y} = ab^t + c. \quad (3.83)$$

Характерною ознакою S -подібних кривих є різна швидкість зміни функції в області її визначення: на початковому проміжку часу ($0 < t < t_0$) спостерігається повільне зростання функції; в околі точки перегину $M(t_0; y_0)$ – прискорене зростання; на проміжку часу ($t > t_0$) функція уповільнено зростає, асимптотично наближаючись до лінії насичення (рис. 3.6).

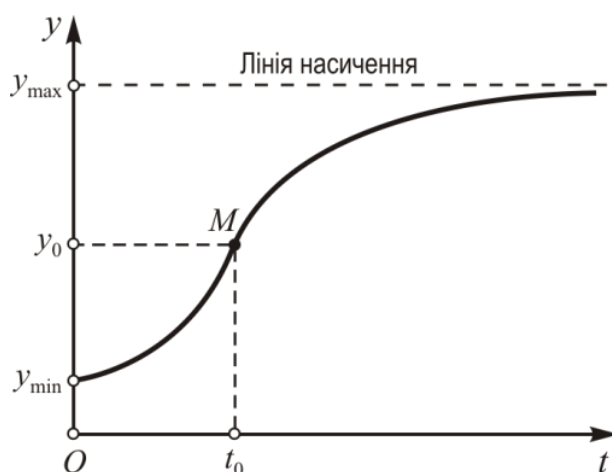


Рис. 3.6. S -подібна крива

З-поміж відомих S -подібних кривих як тренд найчастіше використовують такі:

$$y_t = ca^{bt} - \text{крива Гомперця}; \quad (3.84)$$

$$y_t = \frac{c}{1+be^{-at}} - \text{крива Перла-Ріда (логістична крива)}; \quad (3.85)$$

$$y_t = a + \frac{(b-a) \cdot t^d}{c+t^d} - \text{крива степеневого вигляду}. \quad (3.86)$$

На графіках S -подібних кривих можна виділити чотири інтервали:

- на першому спостерігається незначний приріст функції;
- на другому приріст збільшується;
- на третьому темпи приросту сповільнюються і є приблизно постійними;

- на четвертому спостерігається уповільнення темпів приросту, і функція необмежено наближається до асимптоти лінії насичення.

Моделі (3.81) – (3.86) належать до класу парних нелінійних регресій.

Модифікована показникова функція та *S*-подібні криві мають асимптоти.

Оцінюючи їхні параметри, розглядають два випадки:

- значення асимптоти відоме (наприклад, коефіцієнт використання обладнання не може перевищувати одиниці або виробничі потужності не можуть забезпечити обсяг виробництва, що перевищує деякий рівень) або ж його визначають експертним шляхом;

- значення асимптоти невідоме.

Якщо припустити, що значення асимптоти у модифікованій показниковій функції відоме (c^*), то інші параметри можна визначити за допомогою 1МНК, попередньо провівши перетворення:

$$y_t - c^* = ab^t; \quad \ln(y_t - c^*) = \ln a + t \ln b.$$

Аналогічно можна оцінити параметри кривої (3.85):

$$\frac{c^*}{y_t} - 1 = be^{-at}; \quad \ln\left(\frac{c^*}{y_t} - 1\right) = \ln b - at.$$

Оцінюючи параметри кривої (3.85), виконують перетворення

$$\frac{y_t}{c^*} = a^{b^t}$$

і, двічі прологарифмувавши, отримують:

$$\ln(\ln \frac{y_t}{c^*}) = \ln(\ln a) + t \ln b.$$

У випадку, коли значення асимптоти наперед невідоме, для знаходження параметрів розглянутих вище функцій використовують наближені методи, з якими пропонуємо читачеві ознайомитися самостійно.

Моделювання трендів можна здійснити з використанням деяких інших кривих, вигляд яких вибирають на підставі теоретичного аналізу змісту досліджуваного процесу. Але треба зауважити, що для аналітичного вирівнювання часових рядів недоцільно вибирати функції з великою кількістю параметрів (особливо за малих обсягів вибірки), оскільки в такому

разі трендове рівняння може відображати не основну тенденцію процесу, а випадкові коливання.

Трендову модель вважають адекватною, якщо вона правильно відображає основну тенденцію часового ряду. Ця вимога еквівалентна вимогам, які стосуються випадкової компоненти, а саме:

відхилення фактичних рівнів часового ряду від рівнів, обчислених на підставі рівняння тренда, є випадковими;

відхилення підпорядковуються нормальному закону розподілу;

математичне сподівання відхилень дорівнює нулю;

відхилення незалежні між собою (відсутність автокореляції).

Для вибору «найкращого» тренда найчастіше порівнюють оцінки середніх квадратичних відхилень фактичних рівнів часового ряду від рівнів, обчислених на основі рівняння регресії. Що меншою є оцінка середнього квадратичного відхилення, тим краще рівняння регресії описує тренд.

Багатьом соціально-економічним процесам властиві періодичні коливання. Причини таких коливань різноманітні і залежать від часу та показників, які досліджують.

Періодичні коливання, що тривають один рік, називають *сезонними*. Проявом сезонних коливань слугує наявність інтервалів зростання та спадання рівнів часового ряду всередині року, які регулярно повторюються впродовж декількох років.

Сезонність зумовлюють природно-кліматичні, календарні та соціальні причини. Сезонні коливання виникають у виробництві сільськогосподарської продукції, завантаженні транспорту, формуванні попиту на окремі види продукції та ін. Врахування сезонності є актуальним для вироблення управлінських рішень щодо ціноутворення в умовах нерівномірного попиту, оцінювання наявних потужностей і резервів підприємства, потреби в робочій силі в періоди пікових навантажень тощо.

Виявити сезонні коливання в часовому ряді можна на основі аналізу коефіцієнтів автокореляції. Якщо з-поміж коефіцієнтів автокореляції максимального значення набуває значущий коефіцієнт автокореляції порядку $\tau > 3$, то часовий ряд містить сезонні коливання з періодичністю τ .

Показниками інтенсивності сезонних коливань рівнів часового ряду в окремі періоди слугують *індекси сезонності*, сукупність яких утворює *сезону хвилю*. Індекси сезонності обчислюють як відношення фактичних рівнів часового ряду до середньорічного рівня, розрахованого на основі фактичних або згладжених даних спостережень.

Якщо часовий ряд є стаціонарним (відсутній тренд), то індекс сезонності (I_t^c) знаходять на основі емпіричних даних без їх попереднього згладжування:

$$I_t^c = \frac{\bar{y}_t}{\bar{y}} \cdot 100\%, \quad (3.87)$$

де \bar{y} – середній рівень часового ряду за рік; \bar{y}_t – середній рівень часового ряду за інтервал часу t (місяць або квартал).

Що більше значення індексу сезонності, тим більша інтенсивність (амплітуда) коливань відповідного рівня ряду відносно середнього рівня.

З метою отримання стійких оцінок розміру сезонних коливань у випадку нестационарних часових рядів (наявності тренда) рекомендують дотримуватися такої послідовності дій:

- згладжують рівні часового ряду за допомогою плинної середньої або аналітичних методів вирівнювання;
- знаходять індекси сезонності для окремих інтервалів часу;
- усереднюють індекс сезонності за всі роки.

Індекси сезонності використовують у методах моделювання сезонних процесів, зокрема, у методі *сезонної декомпозиції* для розкладання часового ряду на окремі складові з метою їх подальшого аналізу. На практиці найчастіше використовують два види моделей взаємодії компонент часового ряду: *адитивну* та *мультиплікативну моделі* (моделі (3.56) і (3.57)).

У багатьох літературних джерелах трендову компоненту називають «тренд-циклічною компонентою». Це пояснюється тим, що під час декомпозиції циклічні коливання (наприклад, пов'язані з періодами економічного розвитку, періодами зростання і спадання економічної активності) не відокремлюють від початкового часового ряду, а вважають такими, що входять до трендової складової.

Модель (3.56) застосовують у тих випадках, коли із зміною рівнів ряду амплітуда сезонних коливань не змінюється (рис. 3.7а). Якщо із зростанням або спаданням рівнів ряду амплітуда змінюється (рис. 3.7б), то використовують мультиплікативну модель.

Користуючись адитивною моделлю, припускають, що випадкова складова розподілена нормально з нульовим математичним сподіванням і постійною дисперсією: $\varepsilon \sim N(0; \sigma_{\varepsilon_t}^2)$.

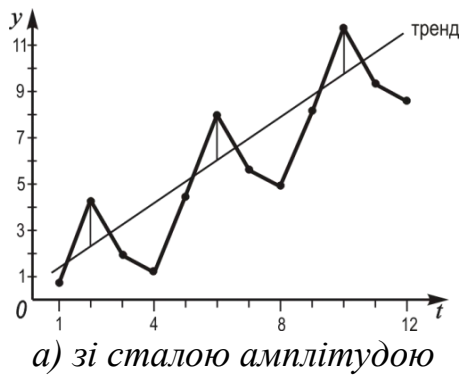
У моделі (3.57) випадкові відхилення (ε_t) враховують мультиплікативно, а не адитивно. Якщо прологарифмувати ліву і праву частини виразу (3.57), то модель буде представлено в адитивному вигляді:

$$\ln y_t = \ln f(t) + \ln s_t + \ln \varepsilon_t. \quad (3.88)$$

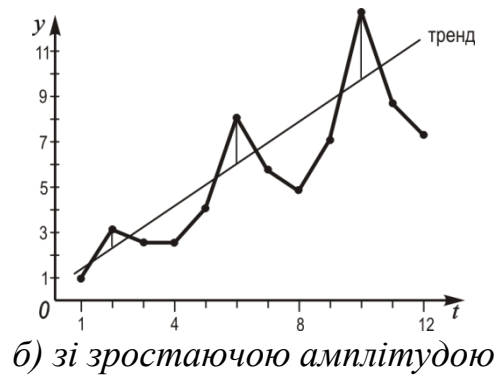
Невірна ідентифікація типу сезонності в часовому ряді призводить до отримання завищеного значення випадкової компоненти, що впливає на зростання величини довірчого інтервалу. Одним із критеріїв оцінювання правильності вибору типу моделі може слугувати нормальність розподілу випадкової компоненти:

$$\varepsilon_t = y_t - (f(t) \cdot s_t) - \text{для адитивної моделі}; \quad (3.89)$$

$$\varepsilon_t = \frac{y_t}{f(t) \cdot s_t} - \text{для мультиплікативної моделі}. \quad (3.90)$$



коливань



коливань

Рис. 3.7. Часові ряди з сезонними коливаннями

Послідовність побудови і реалізації моделей сезонної декомпозиції представлено на рис. 3.8.

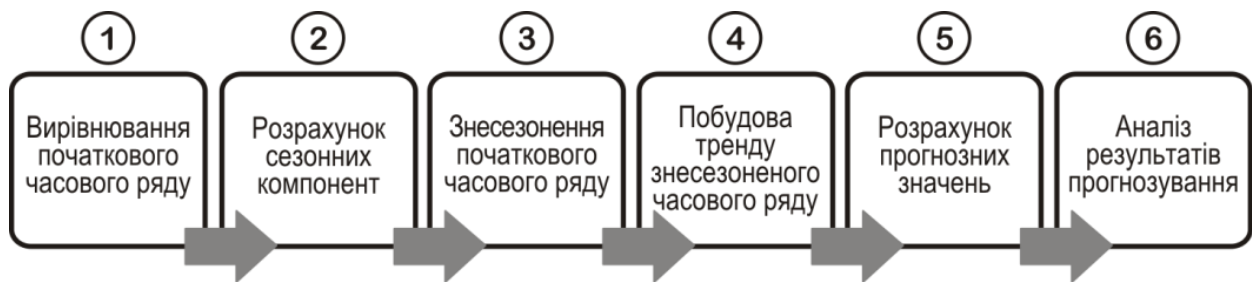


Рис. 3.8. Послідовність побудови та реалізації моделей сезонної декомпозиції

На першому етапі готують дані для знаходження сезонних компонент, а саме – «очищають» початковий часовий ряд від сезонних і випадкових коливань за методами плинних середніх або експоненційного згладжування.

Інтервал («ширину вікна») згладжування вибирають рівним періоду сезонності. Якщо інтервал згладжування є парним числом (наприклад, прогнозування квартальних показників на основі місячних даних за декілька років), то оцінки є точками, що лежать між фактичними рівнями, хоча повинні відповідати фактичним. Тому слід виконати *центрування* – розрахувати центровані плинні середні (середні з пари сусідніх значень). У результаті отримують згладжений ряд (\tilde{y}_t).

На другому етапі розраховують сезонні компоненти:

$$s_t = y_t - \tilde{y}_t - \text{для адитивної моделі}; \quad (3.91)$$

$$s_t = \frac{y_t}{\tilde{y}_t} - \text{для мультиплікативної моделі}. \quad (3.92)$$

Очевидно, що ці компоненти містять випадкові відхилення. Для того, щоб їх позбутися, розраховують середні сезонні оцінки (\bar{s}_t). Середні сезонні оцінки, як правило, коригують таким чином, щоб їхня сума дорівнювала нулю для адитивної моделі, а для мультиплікативної моделі – добуток дорівнював одиниці. У результаті отримують набір «чистих» сезонних компонент.

На третьому етапі середні сезонні оцінки вилучають з початкового часового ряду і отримують знесезоннений ряд (Y_t):

$$Y_t = y_t - \bar{s}_t - \text{для адитивної моделі}; \quad (3.93)$$

$$Y_t = \frac{y_t}{\bar{s}_t} - \text{для мультиплікативної моделі}. \quad (3.94)$$

На четвертому етапі згладжують знесезоннений ряд і на його основі будують рівняння тренда ($\hat{f}(t)$).

Користуючись побудованим рівнянням тренда і середніми значеннями сезонних оцінок, можна розраховувати прогнозні значення для моментів часу t ($y_t^{\text{прогн}}$):

$$y_t^{\text{прогн}} = \hat{f}(t) + \bar{s}_t - \text{для адитивної моделі}; \quad (3.95)$$

$$y_t^{\text{прогн}} = \hat{f}(t) \cdot \bar{s}_t - \text{для мультиплікативної моделі}. \quad (3.96)$$

На завершальному етапі обчислюють помилки прогнозування та на їхній основі аналізують якість побудованої моделі.

Розглянутий вище класичний метод сезонної декомпозиції не позбавлений певних недоліків, а саме:

– у процесі згладжування втрачаються перші та останні рівні початкового часового ряду, що є актуальним за незначної кількості спостережень;

- не передбачаються зміни сезонних компонент упродовж тривалого періоду часу, що на практиці часто не виконується;
- у випадку наявності аномальних рівнів («викидів») сезонні компоненти можуть істотно спотворюватися.

Номер, який отримує одиниця спостереження (об'єкт) в упорядкованій за певним правилом статистичній сукупності, називається *рангом*, а процедура переходу від сукупності спостережень до послідовності їх рангів – *ранжуванням*. Ранжування одиниць статистичної сукупності згідно з рангами дає змогу визначати перевагу однієї одиниці перед іншою, встановлювати їхню належність до певної групи. Ранжування використовують для порівняння різних інвестиційних проектів, оцінювання привабливості об'єктів, які виставляються на продаж, побудови систем сертифікації і якості продукції тощо.

У тому разі, коли кожна одиниця статистичної сукупності відображена лише однією ознакою, упорядкування можна виконати, замінивши значення ознаки відповідними рангами. Якщо одиниця сукупності характеризується декількома ознаками, то виникає потреба в обчисленні *інтегральної оцінки* (інтегрального показника) одиниці сукупності комплексного показника, який характеризує властивість одиниці сукупності в найзагальнішій формі її прояву, враховуючи сумарне значення різних за своєю суттю ознак.

Інтегральна оцінка повинна відповідати таким вимогам:

- відображати мету її обчислення, піддаватися простій змістовній інтерпретації;
- максимально враховувати інформативність її складових, забезпечуючи при цьому стиснення надлишкової інформації;
- бути інваріантною (незмінною, постійною) стосовно одиниць вимірювання її складових;
- представляти достатню роздільну спроможність щодо одиниць статистичної сукупності;

- відтворювати варіацію її складових.

Якщо сукупність складається з n одиниць, кожна з яких описується значеннями m ознак, а x_{ij} – значення j -ї ознаки для i -ї одиниці сукупності, то значення ознак i -ї одиниці сукупності можна задати вектором-рядком $X_i = (x_{i1}, x_{i2}, \dots, x_{im})$. Перетворення векторних величин на скалярні (інтегральні оцінки Q_i) називають *агрегуванням (скаляризацією)*. Агрегування ознак ґрунтується на теорії «адитивної цінності», згідно з якою цінність цілого еквівалентна (дорівнює) сумі цінностей його складових.

Визначення інтегральної оцінки передбачає виконання таких процедур:

- формування множини ознак;
- вибір методу стандартизації значень ознак;
- вибір методу агрегування стандартизованих значень ознак.

Формування множини ознак полягає у виборі ознак, значення яких відіграють вагому роль у визначенні загальної властивості одиниці сукупності. Воно повинно відбуватися на підставі якісного аналізу статистичної сукупності або номенклатури ознак загальноприйнятих систем класифікації техніко-економічної інформації, соціальних і демографічних явищ тощо. Наприклад, щоб визначити інтегральну оцінку рівня якості роботизованої системи для порівняння її з базовою моделлю, множина ознак може містити такі виокремлені показники: річну продуктивність у разі безвідмовної роботи (тис.год); частку простою через відмову (%); ціну системи (тис. грош. од.); річні витрати на ремонт (тис. грош. од.); інші річні експлуатаційні витрати (тис. грош. од.); термін служби (роки).

Ознаки (показники), більші значення яких призводять до підвищення загальної оцінки якості об'єкта, називають *стимуляторами* (продуктивність праці, фондвіддача, рентабельність продажу, власний капітал підприємства та ін.).

Ознаки, зростання значень яких призводить до погіршення загальної оцінки об'єкта, називають *дестимуляторами* (термін окупності проекту, коефіцієнт зносу основних засобів, коефіцієнт левериджу та ін.).

Стимулятори (x_{st}) прямо впливають на інтегральну оцінку (Q_i) i -го об'єкта, а дестимулятори (x_{dst}) – зворотно.

Здебільшого ознаки статистичної сукупності мають різні одиниці вимірювання, представлені в різних масштабах і змінюються в різних діапазонах, у зв'язку з чим виникає необхідність у *стандартизації* (*нормалізації*) значень ознак, тобто приведення їх до єдиної основи із збереженням співвідношень між окремими значеннями ознак.

Оскільки процедура стандартизації безпосередньо впливає на кінцевий результат агрегування, то вона повинна максимально враховувати інформацію, яку містить ознакова множина.

За результатами процедури стандартизації вектор-рядок первісних значень ознак i -ї одиниці сукупності $\mathbf{X}_i = (x_{i1}, x_{i2}, \dots, x_{im})$ замінюють на вектор стандартизованих значень $\mathbf{Z}_i = (z_{i1}, z_{i2}, \dots, z_{im})$.

Існують різні методи стандартизації, але всі вони передбачають відносне зіставлення емпіричних значень показника x_{ij} з деякими величинами a і q :

$$z_{ij} = \frac{x_{ij}}{a}, \quad (3.97)$$

або

$$z_{ij} = \frac{x_{ij} - a}{q}. \quad (3.98)$$

Якщо для показників, які характеризують властивості одиниці статистичної сукупності, встановлені стандарти або інші нормативні значення (x_{oj}), то упорядкування сукупності можна провести за допомогою інтегральної оцінки рівня відхилення від еталону. Тоді стандартизовані значення, розраховують за формулою

$$z_{ij} = \left| \frac{x_{ij} - x_{oj}}{x_{oj}} \right|. \quad (3.99)$$

Порівняльний (рейтинговий) аналіз часто здійснюють на підставі інтегральної оцінки, зіставляючи значення ознак із середніми ($z_{ij} = x_{ij}/\bar{x}_j$). У такому випадку очевидно, що при $Q_i > 1$ інтегральна оцінка рівня розвитку (інтенсивності) явища для i -ї одиниці сукупності перевищує середній у сукупності рівень, а при $Q_i < 1$ є нижчою за середній рівень.

Рейтинг окремих одиниць сукупності можна встановити також на підставі стандартизованих значень, які визначають за формулою

$$z_{ij} = \frac{x_{ij}}{\sum_{i=1}^n x_{ij}}, \quad z_{ij} \in (0; 1), \quad (3.100)$$

де $\sum_{i=1}^n x_{ij}$ – загальний обсяг значень j -ї ознаки в сукупності.

Стандартизовані значення можна також розрахувати на підставі відхилень значень ознаки від мінімального і максимального значень показника в сукупності за формулами:

• для стимуляторів:

$$z_{ij} = \frac{x_{ij} - x_{min}}{x_{max} - x_{min}}, \quad (3.101)$$

• для дестимуляторів:

$$z_{ij} = \frac{x_{max} - x_{ij}}{x_{max} - x_{min}}. \quad (3.102)$$

Значення z_{ij} характеризує позицію i -ї одиниці сукупності в діапазоні варіації за j -ю ознакою. Чим більше значення z_{ij} , тим більший внесок j -ї ознаки у величину інтегральної оцінки i -ї одиниці сукупності.

За умови нормального розподілу початкових даних для стандартизації значень ознак найчастіше користуються формулою

$$z_{ij} = \frac{x_{ij} - \bar{x}_j}{\sigma_{x_j}}, \quad (3.103)$$

де \bar{x}_j – середня арифметична j -ї ознаки, а σ_{x_j} – її стандартне відхилення.

У результаті формується стандартизована шкала, яка відображає місце довільного значення ознаки в наборі даних, вимірюючи його відхилення від середньої арифметичної в одиницях стандартного відхилення. Стандартизація за формулою (7.7) забезпечує приведення початкового набору даних до набору даних з розподілом, який характеризується середньою арифметичною, рівною 0, і стандартним відхиленням, рівним 1.

Інтегральну оцінку Q_i для ознак з однаковою вагомістю визначають за формулою простої арифметичної середньої:

$$Q_i = \frac{1}{m} \sum_{j=1}^m z_{ij}, \quad (i = \overline{1, n}). \quad (3.104)$$

Якщо вплив окремих ознак на формування інтегральної оцінки неоднаковий, то використовують формулу зваженої арифметичної середньої:

$$Q_i = \sum_{j=1}^m w_j z_{ij}, \quad (3.105)$$

де w_j – ваговий коефіцієнт j -ї ознаки, $\sum_{j=1}^m w_j = 1, w_j > 0$

Зазвичай вагові коефіцієнти представляються в частках одиниці, тобто їх сума рівна 1, що дозволяє легко інтерпретувати значимість ознак.

Наявність великої кількості ознак у статистичній сукупності ускладнює вибір вагових коефіцієнтів і вважається у значній мірі суб'єктивною процедурою. У переважній більшості випадків вагові коефіцієнти визначають на основі експертних суджень.

Упорядкування багатовимірних статистичних сукупностей можна виконати засобами *таксономічного аналізу*. *Таксономія* – це наука про правила упорядкування і класифікації багатовимірних об'єктів різної природи. Характерною особливістю моделей і методів таксономії вважаються невелика кількість припущень щодо властивостей об'єктів і умов аналізу та відносно нескладний математичний апарат.

Сутність таксономічного ранжування багатовимірних одиниць статистичної сукупності полягає в реалізації таких процедур:

- визначення «ідеального» (з точки зору мети аналізу) багатовимірного об'єкта;

- знаходження відстаней від кожного реального об'єкта до «ідеального»;

- упорядкування багатовимірних об'єктів (точок) на основі ступеня їх наближення до «ідеального» об'єкта і вибір найкращого об'єкта за критерієм мінімальної відстані до «ідеального» об'єкта.

В основу обчислення таксономічних оцінок покладено поняття *таксономічної відстані* – відстані між двома одиницями статистичної сукупності як точками у багатовимірному просторі. Чим менша відстань між двома точками, тим подібнішими вважаються відповідні об'єкти. Отже, для оцінювання ступеня подібності (відмінності) об'єктів можна скористатися відстанню між об'єктами у просторі координат об'єктів. Знайдені відстані між об'єктами дають змогу визначити положення об'єктів (точок) відносно інших і виконати їх упорядкування.

Послідовність етапів упорядкування багатовимірної статистичної сукупності на основі таксономічного аналізу представлена на рис. 3.9.

Початковим етапом для реалізації необхідних таксономічних процедур є формування ознакової множини і матриці спостережень (X). З точки зору мети дослідження вибрані ознаки необхідно розділити на стимулятори і дестимулятори.

Елементи ознакової множини представляють різні властивості об'єктів, а тому можуть мати різні одиниці виміру і змінюватися в різних діапазонах. Тому наступним етапом таксономічного аналізу є стандартизація значень початкових даних. В результаті отримують матрицю стандартизованих значень ознак (Z).

На основі матриці стандартизованих значень ознак проводиться пошук еталонного вектора (Z_0), координати якого відповідають «ідеальному» об'єкту. Критерієм пошуку еталонного вектора слугують максимальні (для ознак-стимуляторів) і мінімальні (для ознак-дестимуляторів) значення відповідних ознак досліджуваних об'єктів. Одиниця сукупності, що

відповідає «ідеальному» об'єкту, реально не існує, проте з точки зору мети аналізу вважається «еталоном розвитку».



Рис. 3.9. Послідовність етапів таксономічного аналізу

Після того, як обчислені координати еталонного вектора, знаходять відстані від реальних об'єктів (одиниць статистичної сукупності) до

«ідеального» об'єкта (C_{io}). Очевидно, що значення C_{io} не може дати однозначної відповіді щодо ступеня наближення i -ї одиниці статистичної сукупності до еталонної точки.

Інформативнішим є таксономічний показник рівня розвитку (d_i), який обчислюється як відношення C_{io} до максимально можливої середньої відстані одиниць статистичної сукупності до еталонного вектора C :

$$d_i = \frac{C_{io}}{c}. \quad (3.106)$$

Величину C як оцінку максимальної відстані знаходять за «правилом двох (або трьох) сигм», для чого попередньо потрібно розрахувати середню відстань до еталонного вектора (C_0) і середнє квадратичне відхилення (σ_0).

Чим меншим є значення таксономічного показника розвитку ($0 \leq d_i \leq 1$) для i -ї одиниці сукупності, тим ближче вона знаходиться до еталону.

Модифікацією формули (3.106) слугує:

$$d_i^* = 1 - \frac{C_{io}}{c}. \quad (3.107)$$

У випадку використання формули (3.107) інтерпретація таксономічного показника рівня розвитку є такою: i -та одиниця сукупності характеризується тим більшим рівнем розвитку, чим ближчим до одиниці є значення d_i^* .

За допомогою таксономічних показників рівня розвитку можна здійснювати ранжування багатовимірних об'єктів на основі порівнянь двох типів:

- порівняння різних об'єктів (на певну дату або за певний період функціонування) – просторове ранжування ;
- порівняння рівнів розвитку об'єкта у часі – часове ранжування.

Кластер (англ. cluster – скупчення) – це об'єднання декількох однорідних елементів, яке може розглядатися як самостійна одиниця, що характеризується певними властивостями. У математиці під кластером прийнято розуміти групу подібних елементів статистичної сукупності.

Кластерний аналіз – це багатовимірна статистична процедура, завдання якої полягає у розділенні заданої вибірки об'єктів на підмножини (кластери) таким чином, щоб до кожного кластера входили подібні об'єкти, а об'єкти різних кластерів суттєво відрізнялися.

Застосування кластерного аналізу вважається ефективним інструментом вирішення таких задач:

- організування даних в наочні структури, побудова ієрархії множини об'єктів (задачі таксономії);
- спрощення даних для подальшої обробки за рахунок розчленування вибірки на подібні групи, з якими можна працювати окремо (задачі прогнозування, задачі кореляційно-регресійного аналізу);
- виявлення нетипових об'єктів, які не вдається приєднати до жодного з кластерів (задачі встановлення належності об'єкта до єдиної вибраної групи).

Процес класифікації полягає у віднесенні об'єктів до наперед сформованих груп, а кластеризації – у поділі множини об'єктів на групи, які визначаються лише її результатом.

У загальному випадку методи кластерного аналізу (автоматичної класифікації, розпізнавання образів без навчаючої вибірки, як їх ще називають) призначені для розв'язування задач оптимального поділу початкової множини статистичних спостережень $X = \{x_i; i = \overline{1, n}\}$ на k підмножин (кластерів, класів) $X_r = \{x_{rl}; l = \overline{1, n_k}\}$, ($r = \overline{1, k}$) згідно з вибраним критерієм якості.

Кластерний аналіз знаходить широке застосування в різних сферах наукової і практичної діяльності. В маркетингу кластерний аналіз є ефективним інструментом вирішення задач сегментації ринку. Типовими прикладами використання кластерного аналізу в менеджменті слугують задачі класифікації постачальників, виявлення подібних виробничих

ситуацій, поділ персоналу на різні за мотивацією групи. В соціології – це поділ респондентів на однорідні групи.

На відміну від багатьох математико-статистичних методів кластерний аналіз не вимагає строгих припущень щодо об'єктів, які розглядаються, і дає змогу досліджувати множину початкових даних практично довільної природи.

Вирішення задачі кластеризації є неоднозначним, оскільки:

- відсутня точна постановка задачі кластеризації;
- існує декілька критеріїв якості кластеризації;
- існує ряд евристичних методів кластеризації;
- кількість кластерів, як правило, заздалегідь невідома;
- результати кластеризації суттєво залежать від вибраної метрики, яку дослідник вказує суб'єктивно.

В якості міри подібності у процесі кластерного аналізу використовують метричні простори (відстані між об'єктами).

Очевидним є той факт, що чим менша відстань між об'єктами, тим вони подібніші. Обчислення відстані між об'єктами базується на певній метриці.

Якщо ознаки об'єктів є кількісними, то найчастіше користуються *евклідовою відстанню*, згідно з якою відстань між i -им і j -им об'єктами (d_{ij}^E) дорівнює

$$d_{ij}^E = \sqrt{\sum_{l=1}^m w_l (x_{il} - x_{jl})^2}, \quad (3.108)$$

де w_l – ваговий коефіцієнт l -ї ознаки; x_{il} і x_{jl} – відповідно значення l -ї ознаки для i -го і j -го об'єктів у m -вимірному просторі.

Для кластеризації також можна використати манхеттенську відстань (d_{ij}^H) між i -им і j -им об'єктами, яку обчислюють так:

$$d_{ij}^H = \sum_{l=1}^m |x_{il} - x_{jl}|. \quad (3.109)$$

У порівнянні з евклідовою відстанню вплив окремих великих значень ознак зменшується (оскільки різниці не підносять до квадрату). Якщо ознаки

набувають лише значень 0 і 1, то манхеттенську відстань називають *хеммінговою відстанню*.

Узагальненням евклідової відстані у випадку залежності (корельованості) ознак слугує *відстань Махаланобіса* – міра відмінності між двома випадковими векторами \mathbf{X}_i та \mathbf{X}_j (d_{ij}^M):

$$d_{ij}^M = \sqrt{(\mathbf{X}_i - \mathbf{X}_j)^T \mathbf{C}^{-1} (\mathbf{X}_i - \mathbf{X}_j)}. \quad (3.110)$$

де \mathbf{C} – коваріаційна матриця.

За відстань між об'єктами, які представляються категоризованими ознаками, можуть прийматися коефіцієнти рангової кореляції, взаємної спряженості, лінійні коефіцієнти кореляції.

Важливу роль при розрахунку відстаней між об'єктами відіграє правильний вибір масштабу і одиниць вимірювання ознак. Неоднорідність вимірювань призводить до некоректного порівняння відстаней між об'єктами. Наприклад, якщо значення ознаки x_1 знаходиться в діапазоні від 0 до 1, а значення ознаки x_2 – в діапазоні від 500 до 1000, то у процесі обчислення відстані між об'єктами, яка відображає положення об'єктів у просторі їх властивостей, ознака x_2 буде домінувати над ознакою x_1 . Ця проблема вирішується шляхом стандартизації початкових даних.

Застосування кластерного аналізу в загальному випадку передбачає послідовне виконання таких процедур:

- визначення множини змінних (ознак), за якими будуть оцінюватися об'єкти;
- стандартизація значень змінних (при потребі);
- обчислення значень міри подібності (відстаней) між об'єктами;
- вибір алгоритму кластеризації і його реалізація;
- представлення та інтерпретація результатів аналізу.

У результаті виконання перших трьох процедур формується інформаційна база кластерного аналізу – матриця відстаней між

монокластерами (кластерами, до складу яких входить лише один об'єкт). У подальшому вона трансформується у матрицю відстаней між кластерами. Кластери, яким відповідає мінімальний елемент у матриці відстаней, називають найближчими кластерами.

На даний момент часу розроблено багато алгоритмів кластерного аналізу, які прийнято ділити на дві категорії: *ієрархічні* і *неієрархічні*.

Сукупність алгоритмів упорядкування даних, призначених для створення ієрархії (дерева) вкладених кластерів, називають ієрархічною кластеризацією. Зміст ієрархічної кластеризації зводиться до послідовного об'єднання менших кластерів у більші або розчленування більших кластерів на менші. Кінцева кількість кластерів, яка буде сформована, наперед невідома і визначається у процесі аналізу на підставі оптимізації деякого критерію. Неієрархічні алгоритми кластеризації поділяють множину об'єктів на апріорно визначену кількість кластерів, а нові кластери формуються до тих пір, коли не буде досягнута умова зупинки алгоритму.

Розрізняють *чіткі* і *нечіткі* алгоритми кластерного аналізу. Чіткі алгоритми кожному об'єкту вибірки ставлять у відповідність номер кластера, тобто кожний об'єкт належить лише одному кластеру. Нечіткі алгоритми кожному об'єкту ставлять у відповідність набір дійсних значень, які вказують ступінь відношення об'єкта до кластерів, тобто об'єкт відноситься до кожного кластера з деякою ймовірністю.

Використовуючи різні методи кластеризації для одних і тих же даних, дослідник може отримати різні результати, що вважається типовим явищем у кластерному аналізі.

Сутність ієрархічної кластеризації полягає у послідовному об'єднанні менших кластерів у більші (*агломеративні методи*) або у розділенні більших кластерів на менші (*дивізімні методи*).

Згідно з ієрархічними агломеративними методами на першому кроці об'єднують два кластери, відстань між якими є мінімальною. В результаті,

замість n кластерів утворюється один кластер, що містить два об'єкти, і $(n - 2)$ монокластерів. На другому і наступних кроках монокластер об'єднується з іншим монокластером або приєднується до кластера, що складається із двох і більше об'єктів. На кожному кроці відбувається перерахунок значень матриці відстаней з пониженням її розмірності. Об'єднання може продовжуватися до тих пір, поки всі об'єкти не будуть утворювати один кластер.

Дивізімні методи є логічною протилежністю агломеративних методів. На початковому кроці реалізації алгоритму дивізімної кластеризації всі об'єкти належать до одного кластера. На наступних кроках відбувається розчленування більших кластерів на менші. Приклад покрокової процедури об'єднання кластерів наведено на рис. 3.10.

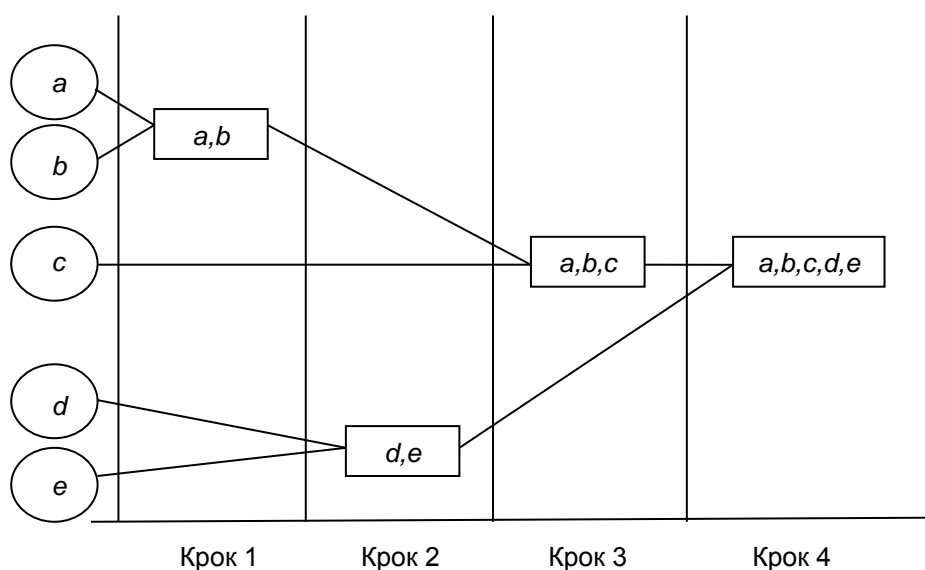


Рис. 3.10. Приклад покрокової процедури об'єднання кластерів

Результати ієрархічної кластеризації зручно представляти у вигляді *деревоподібних діаграм (дендрограм)*, які репрезентують розподіл одиниць сукупності на групи під час окремих кроків виконання процедури кластерного аналізу. На одній осі дендрограми відкладають номери одиниць статистичної сукупності, а на другій – відстані, за якими їх об'єднують у кластери.

Дендрограма як засіб візуалізації послідовності утворення кластерів наочно показує ступінь подібності окремих об'єктів. Кількість рівнів дендрограми відповідає кількості кроків укрупнення або розчленування кластерів. На дендрограмі об'єкти можуть розташовуватися вертикально або горизонтально, тобто розрізняють вертикальні і горизонтальні дендрограми.

Приклад вертикальної дендрограми агломеративної кластеризації наведено на рис. 3.11.

На першому кроці об'єднуються об'єкти O_1 і O_5 , утворюючи кластер $\langle O_1, O_5 \rangle$. При цьому відстань між об'єктами O_1 і O_5 складає 0,5 од. у вибраній метриці оцінювання. На другому кроці об'єкти O_2 і O_4 об'єднуються у кластер $\langle O_2, O_4 \rangle$ (відстань між цими об'єктами рівна 1,0). Наступний крок полягає у приєднанні об'єкта O_3 до кластера $\langle O_2, O_4 \rangle$ (відстань між об'єктом O_3 і кластером $\langle O_2, O_4 \rangle$ рівна 1,5), в результаті чого формується кластер $\langle O_2, O_4, O_3 \rangle$. І, нарешті, кластери $\langle O_1, O_5 \rangle$ і $\langle O_2, O_4, O_3 \rangle$ групуються на найвищому рівні ієрархії кластерів.

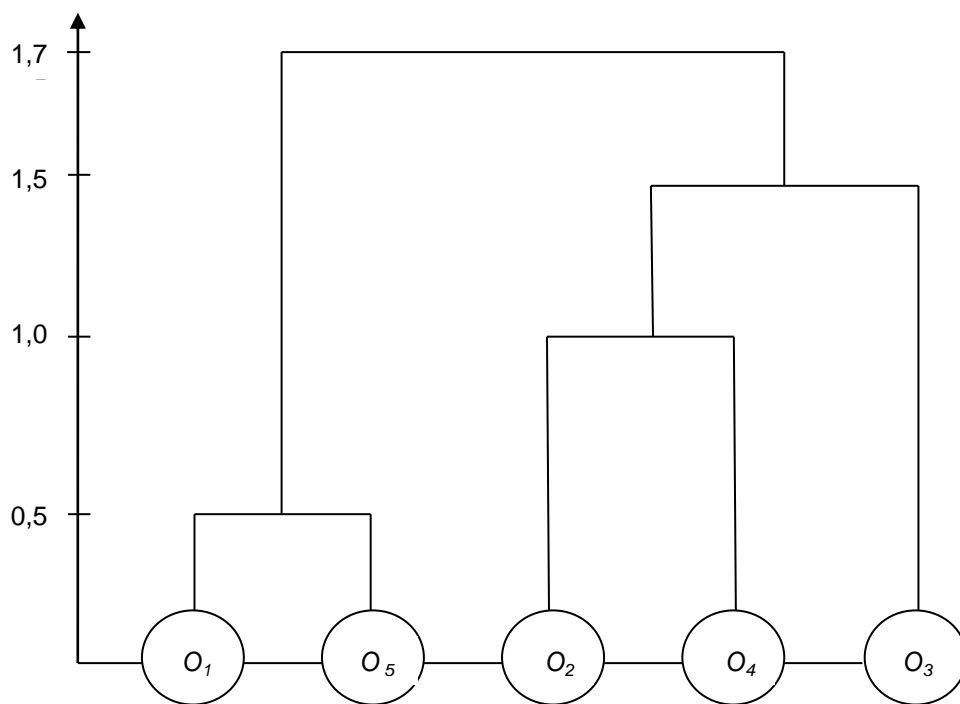


Рис. 3.11. Приклад вертикальної дендрограми

Практика застосування методів ієрархічного кластерного аналізу віддає перевагу агломеративним процедурам, які відрізняються правилами (способами, методами) об'єднання об'єктів і кластерів. Проте всі правила базуються на єдиному алгоритмі послідовної кластеризації.

Найчастіше використовують такі правила об'єднання:

- *одиночного зв'язку* («найближчого сусіда»);
- *повного зв'язку* («найдалшого сусіда»);
- *середнього зв'язку*;
- *Уорда*;
- *парного середнього*.

У випадку одиночного зв'язку вимогою включення об'єкта до кластера слугує максимальна подібність лише з одним елементом кластера, тобто за відстань між кластерами беруть відстань між найменш віддаленими елементами кластерів. Від цього походить назва алгоритму, так як для приєднання об'єкта до кластера вимагається тільки один зв'язок (зв'язок нового об'єкта з кластером визначається на основі одного з елементів кластера). Перед початком роботи алгоритму формується матриця відстаней між об'єктами. На кожному кроці в матриці відстаней знаходять мінімальне значення, яке відповідає відстані між найближчими кластерами. Знайдені кластери об'єднуються і утворюють новий кластер, а значення елементів матриці відстаней, розмірність якої понижується на одиницю, перераховуються.

Відповідно до правила повного зв'язку відстанню між кластерами слугує відстань між найвіддаленішими елементами кластерів. Два об'єкти об'єднуються в один кластер, якщо їх коефіцієнт подібності менший за деяке порогове значення. В термінах евклідової відстані це означає, що відстань між двома об'єктами кластера не повинна перевищувати деякого порогового значення, яке визначається максимально допустимим діаметром підмножини, що утворює кластер.

Правило середнього зв'язку ґрунтується на тому, що відстань між двома кластерами визначається як відстань між їх *центрами тяжіння* (*центроїдами*). На кожному кроці об'єднують два кластери, які мають мінімальну відстань між центроїдами. У випадку об'єднання двох кластерів, в одному з яких кількість об'єктів значно перевищує кількість об'єктів в іншому, характеристики кластера з меншою кількістю об'єктів практично ігноруються. З метою врахування різниці між розмірами кластерів необхідно використати вагові коефіцієнти.

Правило Уорда ґрунтується на дисперсійному аналізі для оцінювання відстані між кластерами. Критерієм включення об'єкта до кластера слугує мінімізація приросту суми квадратів відстаней об'єктів до центрів кластерів. На кожному кроці відбувається об'єднання таких двох кластерів, які забезпечують мінімальний приріст внутрішньогрупової суми квадратів відстаней. Метод спрямований на об'єднання кластерів, які розташовані близько один від одного, і характеризується приблизно рівними розмірами.

Відповідно до правила попарного середнього відстань між кластерами обчислюється як середня відстань між всіма парами об'єктів, які їх утворюють. Можливі два варіанти використання попарних середніх – просте попарне середнє і зважене попарне середнє (в якості вагового коефіцієнти вибирається розмір кластера).

Ієрархічні методи кластеризації рекомендують використовувати при невеликій кількості спостережень, а їх основна перевага – наочність.

При великих обсягах спостережень використовують неієрархічні методи, сутність яких полягає у поділенні набору об'єктів на певну (наперед вказану) кількість окремих кластерів. Існує два підходи до такого поділу. Перший підхід полягає у визначенні меж кластерів найщільніших областей у багатовимірному просторі даних (визначенні кластерів там, де має місце велике «згущення» точок), а другий – в мінімізації міри відмінностей об'єктів.

Неієрархічні методи кластеризації – це ітераційні методи поділу початкової сукупності об'єктів, які ґрунтуються на оптимізації цільової функції, що визначає найкраще у певному розумінні розбиття множини об'єктів на кластери.

Неієрархічні методи кластерного аналізу різняться між собою за:

- вибором початкових умов кластеризації;
- правилами формування нових кластерів;
- умовами зупинки роботи алгоритму.

Типовим прикладом неієрархічних методів кластерного аналізу є метод *k-середніх*, в основу якого покладена мінімізація суми квадратів відстаней між кожним об'єктом спостереження і центром його кластера.

Процедура методу *k-середніх* складається з двох етапів:

- початкове закріплення об'єктів за кластерами;
- ітераційний перерозподіл об'єктів між кластерами.

На першому етапі із сукупності, що містить n об'єктів, на підставі апріорних міркувань або випадковим чином вибирають k об'єктів, які приймаються за еталони. Кожному еталону присвоюють порядковий номер, який одночасно вважається номером кластера, а його центр тяжіння кластера – координати об'єкта. Після цього розраховують відстані від $(n - k)$ об'єктів, які залишилися, до еталонів. Об'єкт приєднують до одного із k кластерів-еталонів на підставі критерію мінімальної відстані згідно з вибраною метрикою. Якщо існує дві або більше мінімальних відстаней, об'єкт приєднують до кластера з найменшим порядковим номером. Еталон з урахуванням приєднання кластерів замінюють новим кластером з обов'язковим перерахунком центру тяжіння. У результаті через $(n - k)$ кроків всі об'єкти будуть віднесені до одного із k кластерів. На цьому початковий етап завершується, причому початкові центри кластерів-еталонів будуть замінені кластерними середніми.

Ітераційний перерозподіл об'єктів між кластерами полягає у віднесенні кожного із n об'єктів до одного із k кластерів на підставі мінімальної відстані об'єкта до центрів тяжіння кластерів. Існують дві модифікації методу k -середніх. Перша полягає у перерахунку центра тяжіння кластера після кожної його зміни, а друга – після завершення перегляду всіх об'єктів. Ітераційний процес, як правило, продовжують доти, поки зміни координат кластерних центрів не стануть мінімальними.

Результати неієрархічного кластерного аналізу можна подавати у вигляді таблиць (матриць), які містять основні характеристики статистичних ознак у кластерах.

Метод k -середніх у порівнянні з ієрархічними методами вважається стійкішим до вибраної метрики подібності, використання у процесі кластеризації незначущих ознак, аномальних даних.

Основними проблемами, які виникають у процесі застосування методу k -середніх, є:

- чутливість до вибору початкового наближення;
- необхідність вказання кількості кластерів, на які розбивається множина об'єктів.

Вказані проблеми часто вирішуються шляхом розгляду декількох варіантів початкових умов методу і поступового нарощування кількості кластерів з оцінюванням результатів на підставі вибраного функціоналу якості кластеризації.

Результатом кластерного аналізу є набір кластерів, що містять об'єкти початкової множини. Необхідно оцінити, наскільки якісно виконаний розподіл об'єктів за кластерами, тобто оцінити якість отриманих результатів кластеризації.

Розрізняють два підходи до аналізу якості результатів кластеризації – *неформалізований* і *формалізований*.

Неформалізований підхід полягає у змістовній інтерпретації кластерів у термінах задачі конкретної предметної області, а також інтерпретації середніх значень ознак об'єктів, що утворюють кластери. Аналіз середніх значень ознак може бути підставою для іменування кластерів (присвоєння їм умовних назв) та їх характеристик.

Формалізований підхід до оцінки якості результатів кластеризації ґрунтується на критеріях, обчислених на основі кількісних характеристик, отриманих згідно використаних методів кластерного аналізу.

Використання різних методів кластерного аналізу для однієї і тієї ж сукупності об'єктів, як правило, зумовлює отримання різних результатів (різна кількість кластерів, різний склад кластерів, різні ступені подібності). На характеристики кластерної структури впливає низка чинників, найважливішими серед яких вважаються вибір:

- сукупності ознак, на основі яких проводиться кластеризація;
- міри подібності об'єктів;
- алгоритму кластеризації.

Застосувавши той чи інший метод кластерного аналізу, потрібно виконати процедуру перевірки якості отриманої кластерної структури. Так як основна задача кластеризації полягає у поділі множини об'єктів на подібні (однорідні) підмножини, то якість кластеризації визначається властивостями:

- *компактності* – елементи одного кластера розташовані максимально близько один до одного;
- *відокремленості* – елементи різних кластерів розташовані якомога далі один від одного;
- *концентрації* – елементи кластерів розташовані навколо їх центроїдів.

Показники компактності, відокремленості і концентрації поділяються на:

- *внутрішні* – побудовані на основі порівняння отриманих результатів з апіорі відомим розподілом на кластери (з тестовою множиною з відомою структурою кластерів);
- *зовнішні* – побудовані лише на підставі результатів кластеризації (апіорна інформація про структуру кластерів невідома);
- *відносні* – побудовані на порівнянні декількох кластерних структур (без апіорної інформації).

Існує близько 50 показників (індексів), за якими оцінюють якість результатів кластерного аналізу. З ними можна ознайомитися у спеціалізованих літературних джерелах.

Задачу кластеризації можна сформулювати як пошук кластерів, оптимальних відносно вибраного критерію (значення показника, необхідного для прийняття рішення). У практиці кластерного аналізу найчастіше користуються такими критеріями (функціоналами):

- сума квадратів відстаней до центрів кластерів:

$$f_1 = \sum_{l=1}^k \sum_{i \in S_l} d^2(X_i, \bar{X}_l), \quad (3.111)$$

де l – номер кластера ($l = \overline{1, k}$); X_i – вектор значень ознаки для i -го об'єкта; \bar{X}_l – центр l -го кластера; $d(X_i, \bar{X}_l)$ – відстань між i -им об'єктом і центром l -го кластера;

- сума внутрішньокластерних відстаней між об'єктами:

$$f_2 = \sum_{l=1}^k \sum_{i, j \in S_l} d_{ij}^2, \quad (3.112)$$

де d_{ij}^2 – евклідова відстань між i -им та j -им об'єктами в кластері;

- сумарна внутрішньокластерна дисперсія:

$$f_3 = \sum_{j=1}^k \sigma_j^2, \quad (3.113)$$

де σ_j^2 – внутрішньокластерна дисперсія j -го кластера.

Оптимальною вважається кластеризація з мінімальними значеннями функціоналів f_1, f_2 і f_3 .

3.3. Методи багатовимірного статистичного аналізу

Дискримінантний аналіз — розділ багатовимірного статистичного аналізу, представлений сукупністю методів для дослідження відмінностей між двома і більше групами об'єктів на підставі декількох ознак одночасно.

Процес дискримінантного аналізу ґрунтується на реалізації декількох пов'язаних статистичних процедур, які забезпечують *інтерпретацію міжгрупових відмінностей і класифікацію об'єктів*.

Процедури інтерпретації міжгрупових відмінностей спрямовані на вирішення таких задач:

- перевірки існування між групами істотних відмінностей з точки зору незалежних змінних;
- можливості на підставі заданого набору ознак відрізнити одну групу від іншої;
- пошуку функції, яка найкраще розділяє групи;
- визначення ознак, які забезпечують найбільший внесок у міжгрупові відмінності.

Процедури класифікації спрямовані на:

- оцінювання точності поділу об'єктів на групи;
- пошук груп, до яких можуть бути віднесені нові об'єкти (спостереження) за принципом максимальної подібності.

На відміну від кластерного аналізу, дискримінантний аналіз не призначений для утворення нових груп, а допомагає виявити різницю між існуючими групами і сформувані правило віднесення нових об'єктів до одного із сформованих раніше кластерів.

Дискримінантний аналіз – один з ефективних інструментів вирішення задач сегментації ринку, зокрема, виявлення відмінностей між окремими групами споживачів, конкурентів, продукції, брендів тощо.

Існує два підходи до задач сегментації ринку товарів і послуг – апріорний та апостеріорний. Апріорний підхід передбачає, що ознаки сегментації, кількість груп і їх характеристики наперед відомі, тобто сегментні групи сформовані. У випадку апостеріорної сегментації будь-які групи відсутні і задача полягає у тому, щоб виявити групи для вироблення маркетингових рішень і заходів. Дискримінантний аналіз використовується для вирішення задач апріорної сегментації ринку, на відміну від кластерного аналізу, який використовується для вирішення задач апостеріорної сегментації.

Можна навести багато прикладів застосування дискримінантного аналізу для вирішення задач дослідження ринку товарів і послуг. Зокрема, за допомогою дискримінантного аналізу можна отримати відповіді на запитання:

- чим з точки зору демографічних характеристик відрізняються покупці дешевих (базових) марок деякого товару від покупців дорогих (унікальних) марок?
- яка відмінність між споживачами темного і світлого шоколаду?
- на який товарний ринок поставляти певний вид продукції, щоб вона була конкурентоспроможною?
- чим відрізняються покупці, які віддають перевагу товарам фірми А, від покупців товарів-аналогів фірми Б?
- чи різняться між собою окремі сегменти ринку за відношенням до рекламних компаній?
- чим відрізняються позичальники, які повернули товарний кредит в термін, від тих, що не повернули?

Дискримінантний аналіз вважається альтернативою множинного регресійного аналізу на випадок, коли залежна змінна є категоріальною (вимірюється у номінальній або порядковій шкалі), і вирішує такі задачі:

- передбачення значень (категорій) залежної змінної;

▪ встановлення незалежних змінних, які найкраще підходять для такого передбачення.

Залежні змінні в дискримінантному аналізі прийнято називати *групувальною змінною*. Вона слугує ознакою розділення номінальної або порядкової шкали. Змінні, якими оперують у процесі дискримінації, називають *дискримінантними змінними* або *пре дикторами*.

У загальному випадку кількість дискримінантних змінних не обмежена, але загальна кількість об'єктів (спостережень) повинна перевищувати кількість дискримінантних змінних по крайній мірі на дві. Крім того, передумовами застосування дискримінантного аналізу є такі вимоги:

- дискримінантні змінні виміряні в інтервальній шкалі або шкалі відношень;
- до множини дискримінантних змінних не включають функціонально зв'язані і висококорельовані ознаки;
- дискримінантні змінні підпорядковуються багатовимірному нормальному закону розподілу;
- відмінність між коваріаційними матрицями груп є статистично незначущою (коваріаційні матриці груп приблизно рівні).

Якщо вказані вище передумови не виконуються, то статистичні висновки щодо застосування результатів дискримінантного аналізу не можна вважати надійними.

Об'єкти дискримінантного аналізу належать до однієї з двох або більше груп (класів), причому групи сформовані таким чином, що кожний об'єкт (спостереження) відноситься лише до однієї групи.

Якщо групувальна змінна представлена лише двома категоріями, то метод дослідження відмінностей називають *дискримінантним аналізом для двох груп*. Якщо групувальна змінна представлена трьома або більше категоріями, то метод дослідження відмінностей називають *множинним дискримінантним аналізом*.

Дискримінантні змінні можна розглядати як осі багатовимірного евклідового простору, а кожний об'єкт як точку цього простору. Групи представляються як згущення точок в окремих областях простору. Розташування групи об'єктів у просторі характеризується *центроїдом групи* – уявною точкою, координатами якої є середні значення змінних у групі.

Розгляд окремих дискримінантних змінних не може забезпечити багатовимірний аналіз групових відмінностей. З цією метою використовують центроїди груп. Для того, щоб відрізнити відносно положення центроїдів різних груп, достатньо обмежитися розмірністю простору, на одиницю меншою за кількість груп.

Розглянемо випадок, коли об'єкти, які характеризуються двома ознаками (x_1 і x_2), розділені на дві групи (k_1 і k_2) (рис. 3.12). Проекції об'єктів на осі x_1 і x_2 перетинаються, тобто по кожній змінній окремо деякі об'єкти обох груп мають подібні характеристики. Для того, щоб найкраще розділити ці групи, потрібно побудувати відповідну лінійну комбінацію змінних x_1 і x_2 , що означає перехід до нової системи координат. Нові осі координат v і u повинні бути розташовані так, щоб:

а) проекції об'єктів, що належать різним групам, на вісь v були максимально розділені;

б) вісь u перпендикулярна до осі v і розділяє групи таким чином, що вони знаходяться по різні боки від неї.

Функція $f(x) = a_1x_1 + a_2x_2$ є лінійною комбінацією змінних x_1 та x_2 і називається *дискримінантною функцією двох змінних*.

Коефіцієнти дискримінантної функції двох змінних визначаються на основі сформульованих вище вимог щодо перетворення простору.

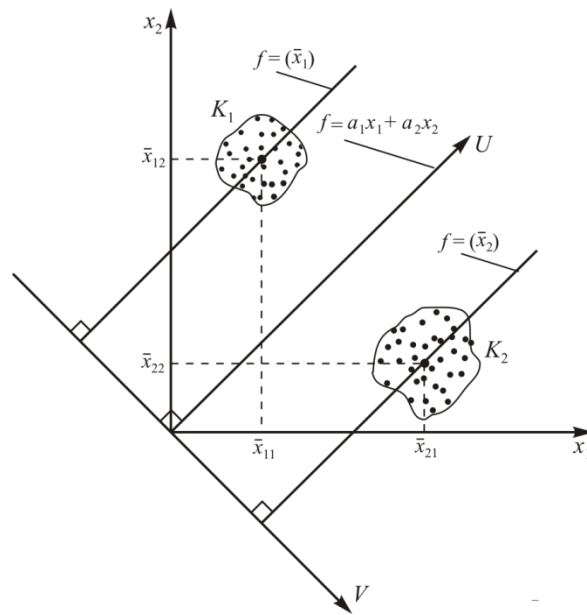


Рис. 3.12. Геометрична інтерпретація дискримінантної функції для двох груп

Середнє значення дискримінантної функції для групи K_1 дорівнює

$$\bar{f}_1 = a_1 \bar{x}_{11} + a_2 \bar{x}_{12}, \quad (3.114)$$

а для групи K_2 —

$$\bar{f}_2 = a_1 \bar{x}_{21} + a_2 \bar{x}_{22}. \quad (3.115)$$

Якщо n_1 і n_2 відповідно обсяги вибірок K_1 і K_2 , то коефіцієнти дискримінантної функції a_1 і a_2 визначаються згідно умови максимальної відмінності між \bar{f}_1 і \bar{f}_2 , тобто

$$|f(\bar{x}_1) - f(\bar{x}_2)| \Rightarrow \max. \quad (3.116)$$

Для віднесення нових спостережень до однієї із двох груп потрібно визначити *межу (константу) дискримінації (M)*, яка розділяє ці групи. Логічно, що в даному випадку такою межею може бути значення, рівновіддалене від \bar{f}_1 і \bar{f}_2 :

$$M = \frac{\bar{f}_1 + \bar{f}_2}{2}. \quad (3.117)$$

Якщо для значень ознак x_{01} і x_{02} деякого нового об'єкта (спостереження) виконується умова

$$a_1x_{01} + a_2x_{02} \geq M, \quad (3.118)$$

то об'єкт зараховують до групи K_1 , інакше — до групи K_2 .

У випадку трьох і більше груп для дослідження відмінностей між ними використовують *множинний дискримінантний аналіз*. Основна мета множинного дискримінантного аналізу - пошук і дослідження дискримінантних функцій, які розділяють (дискримінують) групи, що розглядаються.

Якщо є декілька груп, то можна оцінити більше, чим одну дискримінантну функцію. Наприклад, якщо є три групи, то першу функцію можна використати для дискримінації 1-ї і 2-ї груп, другу — для дискримінації 1-ї і 3-ї груп, третю — для дискримінації 1-ї і 2-ї та 3-ї груп разом і т.д. Проте не має потреби вказувати можливі комбінації груп для формування окремих дискримінантних функцій. На основі аналізу *канонічних кореляцій* (кореляцій між лінійними функціями двох множин) можна визначити оптимальні комбінації дискримінантних змінних, для яких перша функція забезпечить максимальну дискримінацію груп, друга буде найкращою після першої і т.д. Отримані за допомогою канонічного аналізу дискримінантні функції будуть ортогональними (незалежними), а їх внески у дискримінацію груп не перетинаються.

Лінійна комбінація дискримінантних змінних

$$f_{lk} = a_0 + a_1x_{1lk} + a_2x_{2lk} + \dots + a_mx_{mlk}, \quad (3.119)$$

отримана на основі аналізу канонічних кореляцій, називається *канонічною дискримінантною* функцією. Використані у (3.119) позначення мають такий зміст: m – кількість дискримінантних змінних; a_i ($i = \overline{1, m}$) – дискримінантні множники; a_0 – вільний член; f_{lk} – значення функції для l -го об'єкта в k -й групі.

Коефіцієнти першої канонічної дискримінантної функції вибираються таким чином, щоб її середні значення для різних груп якомога більше

відрізнялися між собою. Коефіцієнти другої дискримінантої функції також вибираються на основі критерію максимізації відмінностей між середніми значеннями груп, але при цьому накладається обмеження – значення другої і першої дискримінантних функцій не корелюють між собою. Аналогічно третя дискримінантна функція не повинна корелювати з двома першими.

Розглядаючи групи як згущення точок в деяких областях евклідового простору, для їх розділення доцільно використати центроїди. Точка, яка визначається середніми значеннями сукупності об'єктів, називається *головним центроїдом*.

Відносно головного центроїда можна проводити різноманітну орієнтацію осей за умови, що вони належать простору, “натягнутому на центроїди”. Найважливішу вісь можна спрямувати під кутом, для якого середні значення груп різняться у більшій мірі, чим для будь-якого іншого напрямку. Вважаючи, що є декілька груп, другу вісь можна зорієнтувати таким чином, щоб забезпечити максимальне розрізнення груп, але виконується обмеження: друга вісь ортогональна до першої. Аналогічно спрямовуються інші осі.

Описаний вище спосіб розташування осей слугує критерієм для побудови канонічних дискримінантних функцій.

Залежність (3.119) вказує математичне перетворення m -вимірного простору дискримінантних змінних в q -вимірний простір канонічних дискримінантних функцій (q – максимальна кількість функцій). Кожній осі q -вимірного простору відповідає своя дискримінантна функція, а для конкретного об'єкта значення дискримінантної функції інтерпретується як його координата в просторі дискримінантних функцій. У випадку, коли кількість дискримінантних змінних (m) менша за кількість груп (k), то перехід від простору із більшою розмірністю до простору із меншою розмірністю не відбувається (проводиться тільки заміна координат, яка задовільняє певний критерій).

Для розгляду основних теоретичних положень множинного дискримінантного аналізу введемо такі позначення:

N – загальна кількість об’єктів у всіх групах (обсяг навчальної вибірки);

K – кількість груп;

n_k – кількість об’єктів в k -й групі $\left(\sum_{k=1}^K n_k = N\right)$;

m – кількість ознак, якими характеризується об’єкт (дискримінантних змінних);

x_{ilk} ($i = \overline{1, m}$; $l = \overline{1, n_k}$; $k = \overline{1, K}$) – значення i -ї ознаки для l -го об’єкта в k -й групі;

\bar{x}_{ik} – середнє значення i -ї ознаки в k -й групі;

\bar{x}_i – середнє значення i -ї ознаки в сукупності (по всіх групах).

Для вимірювання відмінностей між об’єктами досліджуваної сукупності використовують симетричну матрицю значень загальних сум квадратів відхилень і попарних добутоків відхилень від загальних середніх (T), елементи якої обчислюються за допомогою співвідношення (3.120)

$$t_{ij} = \sum_{k=1}^K \sum_{l=1}^{n_k} (x_{ilk} - \bar{x}_i)(x_{jlk} - \bar{x}_j). \quad (3.120)$$

Якщо кожний елемент матриці T поділити на $(N-1)$, то отримаємо коваріаційну матрицю дискримінантних змінних.

Залежність між двома дискримінантними змінними характеризується значенням коефіцієнта парної кореляції між ними. Поділивши кожний елемент матриці T на квадратний корінь із добутку відповідних діагональних елементів, можна знайти матрицю коефіцієнтів парної кореляції між дискримінантними змінними.

Вимірювання розсіювання об’єктів всередині груп проводиться за допомогою матриці (W), елементи якої дорівнюють

$$w_{ij} = \sum_{k=1}^K \sum_{l=1}^{n_k} (x_{ilk} - \bar{x}_{ik})(x_{jlk} - \bar{x}_{jk}) \quad (i, j = \overline{1, m}). \quad (3.121)$$

В результаті ділення кожного елемента матриці W на $(N - K)$ отримують відповідні елементи внутрішньогрупової коваріаційної матриці.

Якщо центроїди груп різняться, то елементи матриці W будуть меншими за відповідні елементи матриці T .

Матриця $B = T - W$ називається *матрицею міжгрупової суми квадратів і попарних добутків відхилень*, елементи якої дорівнюють

$$b_{ij} = t_{ij} - w_{ij} \quad (i, j = \overline{1, m}). \quad (3.122)$$

Співвідношення між значеннями відповідних елементів матриць B і W слугують мірою відмінності між групами.

Очевидно, що чим меншим є розсіювання точок відносно центроїда всередині групи і чим більша відстань між центроїдами груп, тим кращою буде дискримінація груп. Тому в основу знаходження коефіцієнтів канонічної дискримінантної функції покладають критерій максимізації відношення міжгрупової варіації до внутрішньогрупової.

Опускаючи математичні аспекти проблеми побудови канонічних дискримінантних функцій, відзначимо, що знаходження їх коефіцієнтів зводиться до пошуку розв'язків системи рівнянь, яка у матричній формі має вигляд:

$$(B - \lambda W)V = 0, \quad (3.123)$$

де λ і V – відповідно власні значення і власні вектори.

Кожний розв'язок системи рівнянь (3.123) (власне значення λ_i і власний вектор V_i) відповідає одній дискримінантній функції.

Компоненти власного вектора (v_i) інтерпретуються як коефіцієнти i -ї дискримінантної функції, але тоді початок координат не буде співпадати з

головним центроїдом. Для того, щоб початок координат співпадав з головним центроїдом, проводять нормування компонент власного вектора:

$$a_i = \sqrt{N - K} \cdot V_i; \quad (3.124)$$

$$a_0 = \sum_{i=1}^m a_i \bar{x}_i. \quad (3.125)$$

При цьому дискримінантні значення вимірюються в одиницях стандартного відхилення.

Для фахівця в галузі дослідження ринку тварів і послуг важливим вважається не стільки знання алгоритму пошуку коефіцієнтів дискримінантної функції, скільки особливості змістовної інтерпретації результатів, отриманих у процесі дискримінантного аналізу. Сучасні пакети прикладних програм, призначені для статистичної обробки даних, забезпечують користувача потужним інструментарієм, за допомогою якого, не вникаючи глибоко у сутність математичного апарату і не переймаючись обчислювальною складністю методів, можна ефективно вирішувати задачі дискримінантного аналізу.

Перш за все у процесі інтерпретації результатів дискримінантного аналізу необхідно вирішити, чи всі розраховані дискримінантні функції корисні з точки зору опису міжгрупових відмінностей. З цією метою використовуються власні значення системи рівнянь (3.123).

Найбільшому власному значенню λ^* відповідає власний вектор V^* , компоненти якого є коефіцієнтами першої канонічної дискримінантної функції. Кожна наступна побудована канонічна дискримінантна функція забезпечує максимальний внесок у міжгрупову варіацію, яка не пояснена попередніми функціями. Максимальна кількість канонічних дискримінантних функцій на одиницю менша за кількість груп або рівна кількості дискримінантних змінних (залежно від того, яка із цих величин є меншою).

Для оцінювання статистичної значущості і корисності кожної із побудованих канонічних дискримінантних функцій використовують декілька критеріїв, кожний з яких надає особливий зміст процесу дискримінації.

Перш за все рівень дискримінації груп залежить від значення власних чисел, а саме: чим більшим є власне число, тим сильніша дискримінація. Найбільшу роздільну спроможність має дискримінантна функція, якій відповідає розв'язок із максимальним значенням власного числа λ , - перша дискримінантна функція.

Значення $\lambda_i > 0$, яке відповідає i -й дискримінантній функції, не надає повну інформацію про неї, оскільки не піддається безпосередній інтерпретації. Якщо існує декілька дискримінантних функцій, то потрібний показник для порівняння їх роздільної (дискримінантної) спроможності. Роздільну спроможність i -ї дискримінантної функції оцінюють за допомогою показника відсоткового вмісту власного числа λ_i (відношення величини власного числа λ_i до суми всіх власних чисел у %). Цей показник характеризує корисність відповідної функції у порівнянні з іншими, але нічого не додає до визначення відмінностей між групами.

Іншою характеристикою корисності дискримінантної функції є *коефіцієнт канонічної кореляції*. У загальному випадку *канонічна кореляція* - це міра зв'язку між двома множинами змінних. У випадку дискримінантного аналізу - це ступінь залежності між групами і дискримінантною функцією. Коефіцієнт канонічної кореляції для i -ї дискримінантної функції (r_i^*) дорівнює

$$r_i^* = \sqrt{\frac{\lambda_i}{1 + \lambda_i}}. \quad (3.126)$$

Чим більшим є значення коефіцієнта канонічної кореляції (змінюється в межах від 0 до 1), тим кращою вважається роздільна спроможність відповідної дискримінантної функції.

Як правило, статистичну значущість дискримінантної функції перевіряють непрямым способом. Замість перевірки спроможності самої дискримінантної функції розглядається *залишкова дискримінантна спроможність функції* — спроможність змінних розрізняти групи за умови вилучення інформації, отриманої за допомогою раніше обчислених функцій. Якщо залишкова дискримінація виявляється дуже малою, то не має потреби у побудові чергової дискримінантної функції.

Оцінювання залишкової дискримінантної спроможності i -ї дискримінантної функції проводиться за допомогою *статистики лямбда Уїлкса* (статистики Уїлкса)

$$\Lambda_i = \prod_{j=i+1}^K \frac{1}{1 + \lambda_j}, \quad (3.127)$$

де i — кількість побудованих дискримінантних функцій.

Лямбда Уїлкса — це відношення міри внутрішньогрупової мінливості до міри загальної мінливості. Чим меншим є значення лямбда Уїлкса, тим більш високою роздільною спроможністю характеризується відповідна дискримінантна функція, тим кращим виявляється поділ на групи при дискримінантному аналізі (максимальне значення дорівнює 1). Одиничного значення лямбда набуває у випадку, коли спостережні середні значення груп рівні.

На основі лямбда-Уїлкса розраховують статистику χ^2 -квадрат із $V = (m - i)(K - i - 1)$ ступенями вільності:

$$\chi_i^2 = - \left[N - \left(\frac{m - K}{2} \right) - 1 \right] \ln \Lambda_i. \quad (3.128)$$

Якщо розрахункове значення вказаної вище статистики перевищує критичне, то i -та дискримінантна функція вважається статистично значущою.

На відносний внесок окремих дискримінантних змінних у значення кожної дискримінантної функції вказують стандартизовані коефіцієнти (чим більше значення стандартизованого коефіцієнта, тим більший внесок відповідної змінної). Внесок стандартизованого коефіцієнта в дискримінантну функцію є пропорціональним до його величини. Тому стандартизовані коефіцієнти можна використовувати з метою зменшення розмірності початкового ознакового простору. Якщо абсолютна величина стандартизованого коефіцієнта дискримінантної функції є малою, то відповідну дискримінантну змінну можна виключити з моделі.

Близькими за інформаційним змістом до стандартизованих коефіцієнтів дискримінантної функції є структурні коефіцієнти, які характеризують кореляцію між окремою дискримінантною змінною і дискримінантною функцією. Структурні коефіцієнти (змінюються в межах від 0 до 1) інтерпретуються як факторні навантаження змінних на дискримінантну функцію. Близьке до 1 значення структурного коефіцієнта вказує на те, що практично вся інформація про дискримінантну функцію міститься у відповідній змінній. Якщо ж значення структурного коефіцієнта близьке до 0, то кореляція між дискримінантною змінною і дискримінантною функцією є слабкою.

На відміну від стандартизованих коефіцієнтів, які враховують одночасний вплив всіх змінних на дискримінантну функцію, структурні коефіцієнти є парними кореляціями, а тому на них не впливають взаємозв'язки між змінними.

Структурні коефіцієнти можуть розраховуватися як для всієї сукупності об'єктів, так і для кожної групи окремо. Зокрема, якщо необхідно встановити, як дискримінантні функції пов'язані з дискримінантними змінними в межах окремої групи, то використовують внутрішньогрупові структурні коефіцієнти.

Важливою задачею, яка виникає у процесі дискримінантного аналізу, є задача відбору інформативних дискримінантних змінних, які забезпечують розділення апріорі заданих груп. Можливі два підходи до вирішення цієї задачі – суб'єктивний і алгоритмічний.

Суб'єктивний підхід ґрунтується виключно на судженні дослідника про роль конкретної змінної, сформованого на підставі зіставлення абсолютних значень стандартизованих коефіцієнтів статистично значущих дискримінантних функцій.

В рамках алгоритмічного підходу список інформативних змінних формується автоматично у відповідності з тим чи іншим алгоритмом покрокового дискримінантного аналізу.

Покроковий дискримінантний аналіз буває двох видів:

- *покроковий аналіз із включенням;*
- *покроковий аналіз із виключенням.*

На першому кроці аналізу із включенням визначається дискримінантна змінна, для якої середні значення в апріорно заданих групах відрізняються найбільше. Ця змінна включається до моделі. На кожному наступному кроці визначається нова змінна, яка найкраще розрізняє групи. Процес завершується тоді, коли жодна із дискримінантних змінних, що залишилися, не забезпечує істотного внеску у розрізнення груп.

Покроковий аналіз із виключенням проводиться у зворотному напрямі. Спочатку всі дискримінантні змінні включаються до моделі, а потім на кожному наступному кроці вилучаються змінні, які не мають істотного впливу на дискримінацію груп.

Покрокова процедура дискримінантного аналізу реалізується на підставі часткових F -критеріїв. Значення F -статистики для дискримінантної змінної вказує на її статистичну значущість при дискримінації груп, тобто є мірою внеску змінної у дискримінацію сукупності.

Критерієм оцінювання доцільності включення до моделі дискримінантної змінної може слугувати значення часткової лямбда Уїлкса — показника того, наскільки окрема змінна сама спроможна виконувати дискримінацію. Відбирається та змінна, для якої часткова статистика Уїлкса набуває найменшого значення.

Необхідну точність обчислень забезпечує *тест толерантності*. Якщо деяка змінна є лінійною комбінацією або приблизно дорівнює лінійній комбінації інших змінних, то її толерантність рівна нулю або близька до нуля. Така змінна не несе нової інформації, а лише створює обчислювальні проблеми. Толерантність ще не відібраної змінної рівна одиниці мінус квадрат множинної кореляції цієї змінної з усіма вже відібраними змінними.

Викладений вище матеріал стосується побудови канонічних дискримінантних функцій за відомої належності об'єктів до певної групи та спроможності цих функцій пояснити відмінності між групами. Проте не менш актуальним є встановлення груп, до яких найімовірніше можуть бути віднесені нові випадково відібрані об'єкти (спостереження), тобто класифікація об'єктів. Необхідно відзначити, що процеси класифікації за допомогою кластерного і дискримінантного аналізу принципово різні. Кластерний аналіз передбачає класифікацію об'єктів на основі їх відмінностей без будь-якої попередньої інформації про склад і кількість груп. При використанні дискримінантного аналізу спочатку задаються склад і кількість груп, після того вирішується задача, яка полягає у визначенні, наскільки точно можна передбачити належність об'єктів до груп за допомогою набору дискримінантних змінних.

Процедура класифікації може ґрунтуватися як на самих дискримінантних змінних, так і на канонічних дискримінантних функціях.

Розрізняють три підходи до класифікації об'єктів за результатами дискримінації:

- на основі функцій класифікації — для кожної групи будується лінійна класифікаційна функція, а об'єкт відноситься до k -ї групи, якщо значення k -ї функції для нього є максимальним;

- на основі відстані — об'єкт відноситься до тієї групи, відстань до центроїда якої є найменшою (як правило, використовується відстань Махаланобіса);

- на основі максимальної правдоподібності (апостеріорної ймовірності) - об'єкт відноситься до групи, для якої ймовірність належності об'єкта є максимальною.

За першим підходом у процесі класифікації використовують лише дискримінантні змінні. На основі критерію максимального розрізнення груп будують лише функції класифікації, а процедуру дослідження міжгрупових відмінностей і статистичної значущості розрізнення груп не проводять.

Функція класифікації як лінійна комбінація дискримінантних змінних, призначена для визначення групи, до якої найімовірніше може бути віднесений новий об'єкт:

$$h_k = b_{k0} + b_{k1}x_1 + b_{k2}x_2 + \dots + b_{km}x_m \quad (k = \overline{1, K}), \quad (3.129)$$

де h_k – функції класифікації для k -ї групи.

Параметри функції класифікації (b_{ki}) визначають так:

$$b_{ki} = (N - K) \sum_{j=1}^m (w_{ij})^{-1} x_{jk} \quad (i = \overline{1, m}, k = \overline{1, K}),$$

$$b_{k0} = -0,5 \sum_{j=1}^m b_{kj} x_{jk}, \quad (3.130)$$

де $(w_{ij})^{-1}$ – елемент матриці, оберненої до внутрішньогрупової матриці суми квадратів і попарних добутків відхилень (W).

Для кожної групи будується окрема функція класифікації. На підставі значень ознак, які характеризують об'єкт, що підлягає класифікації, обчислюють значення функцій класифікації, після чого об'єкт відносять до тієї групи, для якої значення функції класифікації є найбільшим.

Оскільки коефіцієнти функції класифікації не стандартизовані і для кожної групи будується окрема функція, то вони не підлягають змістовній інтерпретації.

Задачу класифікації можна розв'язувати у просторі канонічних дискримінантних функцій на підставі відстані від об'єкта до центроїдів кожної групи. Об'єкт, визначений як точка у просторі канонічних дискримінантних функцій, відносять до тієї групи, до центроїда якої він розташований найближче.

У випадках, коли ознаки, що характеризують багатовимірні об'єкти, представлені в різних одиницях вимірювання, мають різні стандартні відхилення і корелюють між собою, найкращою мірою відстані вважається *відстань Махаланобіса*.

Для кожної групи обчислюється відстань від її центроїда до об'єкта, який потрібно класифікувати, після чого об'єкт відносять до групи, для якої ця відстань є мінімальною. Використовуючи для класифікації квадрат відстані Махаланобіса, можна обчислити ймовірності належності об'єкта до кожної із сформованих груп.

Класифікація об'єктів за допомогою методу максимальної правдоподібності (правила Байєса) ґрунтується на тому, що задача сформована у термінах теорії ймовірності і відомі:

- апіорні ймовірності $P(G_k)$ — ймовірності належності об'єкта до k -ї групи;

умовні ймовірності $P(X/G_k)$ належності об'єкта до кожної із груп (умовні розподіли вектора ознак X за умови, що об'єкт належить до k -ї групи).

Неоднакова кількість спостережень у різних групах може бути результатом:

- реального розподілу випадкових величин у генеральній сукупності;
- випадкового відбору.

У першому випадку апіорні ймовірності належності об'єктів до груп приймаються пропорційними чисельностям груп, а в другому – однаковими.

За формулою Байєса ймовірність того, що об'єкт X належить до k -ї групи, дорівнює

$$P(G_k / X) = \frac{P(X / G_k) P(G_k)}{\sum_{i=1}^K P(X / G_i) P(G_i)} \quad (k = \overline{1, K}). \quad (3.131)$$

Об'єкт класифікується до тієї групи, для якої апостеріорна ймовірність $P(G_k / X)$ є максимальною, що відповідає найменшій відстані Махаланобіса (за винятком тих випадків, коли ймовірності апіорної класифікації занадто малі).

Мірою оцінювання якості правила класифікації об'єктів слугує *класифікаційна матриця (матриця передбачень)*, яка формується в результаті застосування правила класифікації до об'єктів навчальної вибірки (множини об'єктів з відомою належністю до груп) і містить ряд правильно і помилково класифікованих випадків. Частка правильно класифікованих об'єктів називається *коефіцієнтом результативності*. Вона вважається найкращим показником інформативності відібраних дискримінантних змінних і корисності застосування дискримінантних функцій для інтерпретації міжгрупових відмінностей і вказує на відповідність дискримінантної моделі емпіричним даним. З практичної точки зору коефіцієнт результативності вважається задовільним показником класифікації, якщо він перевищує 0,70.

Ефективність дії побудованого правила класифікації у значній мірі залежить від коректності навчальної вибірки. Застосувавши правило класифікації до навчальної вибірки, можна визначити об'єкти, які за своїми показниками не забезпечують однорідність відповідної групи, тобто віднесені до групи помилково. З метою покращення якості навчальної вибірки такі об'єкти, як правило, вилучають із подальшого розгляду.

Процедура вилучення є покроковою — на кожному кроці вилучається один об'єкт, який найбільше не підходить до групи. Процедуру вилучення об'єктів продовжують до тих пір, поки загальний відсоток коректності у класифікаційній матриці не стане рівним 100%. Вилучаючи об'єкт слід пам'ятати, що при цьому змінюється центроїд групи, а це може призвести до гірших результатів, ніж були отримані для початкової навчальної вибірки, оскільки можуть з'явитися нові невірно класифіковані об'єкти.

Висновки, зроблені на основі коефіцієнта результативності класифікації вважаються переоціненими, оскільки вони ґрунтуються на одній і тій самій навчальній вибірці, яка використана для побудови класифікаційних функцій. Якщо вибірка є великою, то одну частину (наприклад, половину) випадковим чином можна вибрати в якості навчальної, а другу використати для тестування. Процедура оцінки точності класифікації за допомогою тестової вибірки шляхом порівняння точності класифікації з тією, яка досягнута на навчальній вибірці, називається *крос-перевіркою* (модель будується на основі навчальної вибірки, а точність класифікації перевіряється за тестовою вибіркою).

Послідовність етапів множинного дискримінантного аналізу представлена на рис. 3.13.

Ринок товарів і послуг описується достатньо великою кількістю ознак (показників, змінних), більшість з яких взаємопов'язані. Для зручності опрацювання і подальшого використання кількість ознак доцільно скорочувати до рівня, за якого втрати інформації є мінімальними і не перевищують деяке задане значення. Сукупність методів багатовимірного статистичного аналізу, призначених для скорочення кількості змінних і їх узагальнення, називають *факторним аналізом*.



Рис. 3.13. Послідовність етапів множинного дискримінантного аналізу

У результаті факторного аналізу зв'язки між корельованими змінними подаються за допомогою нових змінних, які називають *факторами*. Фактор - це гіпотетична, прихована (латентна) змінна, яку не можна безпосередньо виміряти, але вона в тій чи іншій мірі пов'язана з ознаками, які підлягають вимірюванню. Кількість факторів є меншою у порівнянні із початковою множиною ознак. Фактори інтерпретують як причини сумісної зміни декількох ознак. Ідентифікацію факторів через початкові ознаки називають *інтерпретацією факторів*. Вона полягає у присвоєнні фактору імені, яке за змістом є узагальненням імен тих ознак, що входять до фактора.

Існують три передумови доцільності переходу від великої кількості початкових ознак до суттєво меншої кількості найінформативніших факторів:

- дублювання інформації за рахунок щільного зв'язку між ознаками;
- неінформативність ознак, які мало варіюють у різних спостереженнях;
- можливість агрегування окремих ознак.

Методи і моделі факторного аналізу застосовують для вирішення таких задач:

- виявлення факторів (латентних змінних), які пояснюють структуру кореляцій всередині набору спостережуваних ознак;
- стиснення (зниження) розмірності простору ознак об'єкта за рахунок зведення взаємозалежних спостережуваних ознак до деяких інтегрованих неспостережуваних показників (*редукція даних*);
- класифікації об'єктів на основі стиснутого простору ознак;
- заміни корельованого набору змінних на некорельований для подальшого регресійного аналізу (вилучення мультиколінеарності).

Таким чином, сутність факторного аналізу полягає у переході від представлення об'єкта великою кількістю ознак, які піддаються безпосередньому вимірюванню, до представлення об'єкта меншою кількістю узагальнених (інтегрованих) показників (факторів), які відображають найсуттєвіші властивості об'єкта. Узагальнення часто дозволяє по-новому подивитися на початкові дані, виявити зв'язки між початковими змінними, які раніше не були очевидними, і перейти на вищий рівень пізнання сутності досліджуваного об'єкта.

Факторний аналіз може використовуватися як *розвідувальний (експлораторний)* і як *перевірочний (конфірматорний)* метод аналізу даних. Розвідувальний факторний аналіз слугує інструментом структуризації емпіричних даних і формулювання гіпотез про причини існування виявлених зв'язків між ними. Перевірочний факторний аналіз застосовують на

завершальній стадії дослідження як засіб перевірки відповідності сформульованих гіпотез емпіричним даним.

У дослідженні ринку товарів і послуг факторний аналіз найчастіше використовується для вирішення таких задач:

- визначення латентних змінних (факторів) для групування споживачів при сегментації ринку;
- визначення характеристик торгової марки, що впливають на вибір споживачів, при розробленні товарної стратегії;
- визначення характеристик споживачів, чутливих до зміни цін, при розробленні стратегії ціноутворення;
- визначення характеру телевізійних передач, яким віддають перевагу споживачі цільового ринку, при розробленні рекламної стратегії;
- визначення факторів для оцінювання конкурентоспроможності продукції і підприємств.

На відміну від регресійного і дисперсійного аналізу, в яких змінні поділяють на залежні і незалежні, у факторному аналізі такого поділу немає.

У процесі регресійного аналізу увага акцентується на визначенні кількісної оцінки впливу незалежних змінних на результатний показник (залежну змінну). Дисперсійний аналіз спрямований на вивчення впливу якісних незалежних змінних на залежні кількісні змінні. Факторний аналіз призначений для комплексного представлення досліджуваного явища або об'єкта, яке виражається у взаємозв'язках і взаємообумовленості окремих змінних, і належить до розвідувальних статистичних інструментів.

За наявності великої кількості змінних факторний аналіз як інструмент редукції даних може бути використаний в якості передумови кластерного аналізу. По-перше, скорочення кількості змінних, які приймають участь у процесі кластеризації, забезпечує спрощення і чіткість інтерпретації кластерних рішень. Очевидно, що рішення, отримане на основі декількох факторів, легше піддається змістовній інтерпретації, чим рішення, отримане

на основі десяти і більше змінних. По-друге, застосування різних методів кластеризації за наявності високої кореляції між окремими змінними може істотно впливати на відмінність результатів, що ускладнює вироблення загального висновку.

Більшість методів факторного аналізу ґрунтуються на аналізі головних компонент. Ідея застосування головних компонент полягає у переході до нового ортогонального базису, осі якого орієнтовані за напрямками максимальної дисперсії початкових даних. Вздовж першої осі нового базису дисперсія буде максимальною. Друга вісь максимізує дисперсію за умови ортогональності до першої і т.д. Дисперсія вздовж останньої осі є найменшою. Вказане перетворення забезпечує заміну набору корельованих початкових змінних новим набором некорельованих змінних.

З геометричної точки зору задача факторного аналізу полягає у пошуку базису найменшої розмірності, який із заданою точністю представляє досліджуваний об'єкт.

Розглянемо геометричну інтерпретацію головних компонент на прикладі додатно корельованих змінних x_1 і x_2 .

Графік двовимірного розсіювання об'єктів можна подати у вигляді еліпса, оскільки більшим значенням змінної x_1 будуть відповідати більші значення змінної x_2 і навпаки (рис. 3.13). Якщо поставлена задача представлення об'єктів в термінах однієї змінної, то головна вісь еліпса (m_1) підходить найкраще, оскільки уздовж неї об'єкти різняться більше (дисперсії більші), чим уздовж будь-якої іншої прямої. Відбувається перехід від координат кожного об'єкта по двох осях x_1 і x_2 до координат по одній осі m_1 — головній компоненті. Нова змінна (фактор) є лінійною комбінацією двох початкових змінних. Зауважимо, що при переході від двох змінних до однієї головної компоненти втрачається частина початкової інформації, причому, чим щільніший зв'язок між змінними, тим менша втрата інформації.

Якщо змінні x_1 і x_2 не корелюють, то осі m_1 і m_2 з точки зору інформативності слід вважати рівнозначними, а тому визначити одну з них як головну неможливо.

У випадку, коли кількість корельованих змінних більша двох, принцип визначення головних компонент аналогічний. У просторі трьох і більше змінних графік розсіювання об'єктів представляється еліпсоїдом, кількість осей якого вважається рівною кількості змінних. Перша вісь пройде по найбільшому діаметру. Кожна наступна вісь ортогональна до попередньої і в напрямку кожної наступної осі розкид спостережень буде меншим, чим у попередньої.

Аналіз головних компонент як засіб скорочення змінних за умови збереження максимальної частки дисперсії спостереження є початковою процедурою багатьох методів факторного аналізу і може розглядатися як їх спрощений аналог.

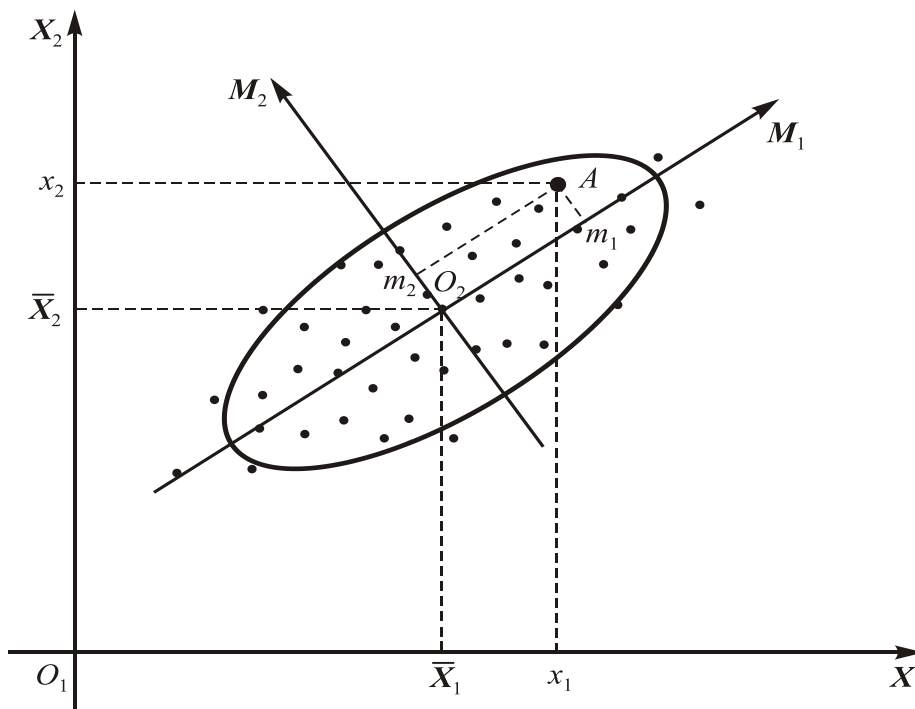


Рис. 3.14. Перехід від двовимірного розподілу ознак x_1 і x_2 до одновимірного Z_1

Практичне застосування факторного аналізу ґрунтується на таких умовах:

- всі змінні виражені у метричній шкалі;
- вибірка однорідна;
- обсяг вибірки щонайменше у два рази більший за кількість змінних.

Нехай сформована вибірка спостережень за деяким об'єктом (явищем)

$$\mathbf{X} = [\mathbf{X}_1, \mathbf{X}_2, \dots, \mathbf{X}_m] = \begin{bmatrix} x_{11} & x_{12} & \dots & x_{1m} \\ x_{21} & x_{22} & \dots & x_{2m} \\ \dots & \dots & \dots & \dots \\ x_{n1} & x_{n2} & \dots & x_{nm} \end{bmatrix},$$

де n – кількість спостережень; m – кількість спостережуваних ознак.

Необхідним етапом факторного аналізу є процедура стандартизації змінних, тому представимо результати спостережень у стандартизованому вигляді:

$$\mathbf{Z} = [\mathbf{Z}_1, \mathbf{Z}_2, \dots, \mathbf{Z}_m] = \begin{bmatrix} z_{11} & z_{12} & \dots & z_{1m} \\ z_{21} & z_{22} & \dots & z_{2m} \\ \dots & \dots & \dots & \dots \\ z_{n1} & z_{n2} & \dots & z_{nm} \end{bmatrix}.$$

Припустимо, що кількість спостережуваних ознак є надлишковою і може бути охарактеризована меншим набором із p ($p < m$) визначальних ознак (факторів).

Завдання факторного аналізу - представлення стандартизованих змінних Z_j у термінах факторів, які поділяють на *загальні (головні)* і *характерні*. Загальні фактори F_k ($k = \overline{1, p}$) беруть участь у формуванні кожної змінної X_j ($j = \overline{1, m}$) і виявляють внутрішні властивості досліджуваного об'єкта. Характерні фактори F_j ($j = \overline{1, m}$) відображають ту частину змінної X_j , яку неможливо пояснити загальними факторами. Для кожної змінної X_j існує свій характерний фактор, який виражає лише її властивості. Характерний фактор можна інтерпретувати як деяке відхилення, пов'язане із тим, що p -

вимірний базис із головних факторів є неповним для представлення початкових змінних, або початкові змінні не можуть бути виражені лінійно через головні фактори.

Графічна інтерпретація моделі факторного аналізу представлена на рис.3.15.

Лінійна модель класичного факторного аналізу має вигляд:

$$Z_j = \sum_{k=1}^p a_{kj} F_k + \alpha_j F_j^* \quad (j = \overline{1, m}), \quad (3.132)$$

де a_{kj} і α_j - відповідно стандартизовані коефіцієнти множинної лінійної регресії j -ї змінної по k -му загальному фактору та j -му характерному фактору.

Загальні фактори лінійно незалежні і не корелюють із характерними факторами, а характерні фактори не корелюють між собою.

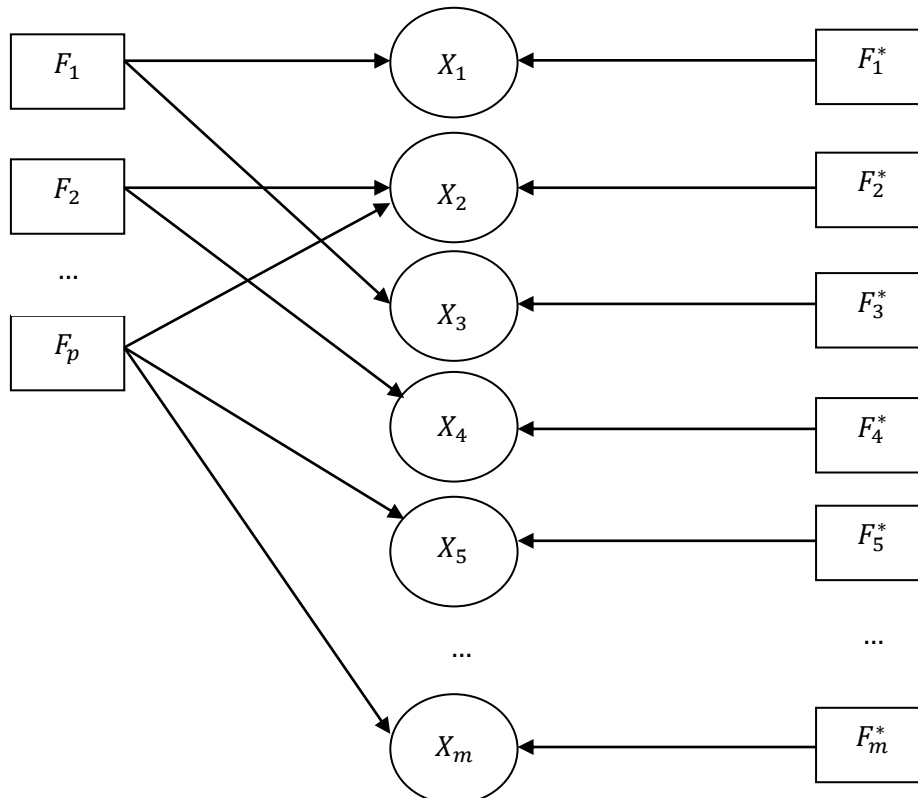


Рис. 3.15. Схема моделі факторного аналізу

Характерний фактор пов'язаний лише з однією змінною. Загальні фактори пов'язуються ваговими коефіцієнтами більше, ніж з однією змінною. Якщо істотними вважаються всі вагові коефіцієнти загального фактора, то він називається *генеральним фактором*.

Значення коефіцієнтів $a_{jk} (j = \overline{1, m}; k = \overline{1, p})$ і α_j в моделі (3.132) відповідно інтерпретуються як значення коефіцієнта парної кореляції між j -ю змінною та k -им загальним фактором і значення коефіцієнта парної кореляції між j -ю змінною та j -им характерним фактором і називаються *факторними навантаженнями* (або наповненням змінної фактором).

Матриця M , яка рівна сумі матриць навантажень загальних (A) і характерних факторів (α), називається *повною факторною матрицею*

$$M = A + \alpha, \quad (3.133)$$

де

$$A = \begin{bmatrix} a_{11} & a_{12} & \dots & a_{1p} \\ a_{21} & a_{22} & \dots & a_{2p} \\ \dots & \dots & \dots & \dots \\ a_{m1} & a_{m2} & \dots & a_{mp} \end{bmatrix} \quad \text{і} \quad \alpha = \begin{bmatrix} \alpha_1 & 0 & \dots & 0 \\ 0 & \alpha_2 & \dots & 0 \\ \dots & \dots & \dots & \dots \\ 0 & 0 & \dots & \alpha_m \end{bmatrix}.$$

Квадрати факторних навантажень відображають внески факторів у дисперсію змінної $S_j^2 (j = \overline{1, m})$. Оскільки значення змінних стандартизовані, то дисперсія кожної з них дорівнює одиниці. Та частина дисперсії змінної, яка пояснена впливом загальних факторів (h_j^2), називається *спільністю*, а та, яка обумовлена специфікою змінної і помилками вимірювання (α_j^2), — *характерністю*. Має місце рівність:

$$S_j^2 = h_j^2 + \alpha_j^2 = 1 \quad (j = \overline{1, m}), \quad (3.134)$$

де $h_j^2 = \sum_{k=1}^p a_{kj}^2$.

Характерність можна подати у вигляді:

$$\alpha_j^2 = b_j^2 + e_j^2 \quad (j = \overline{1, m}), \quad (3.135)$$

де b_j^2 – специфічна дисперсія j -ї змінної, пов’язана з характерним для неї фактором;

e_j^2 – дисперсія помилки j -ї змінної, зумовлена точністю вимірювання, впливом випадкових чинників.

Сумарна дисперсія всіх змінних, зумовлена дією окремого фактора, дорівнює сумі часток дисперсій по цьому фактору. Якщо її розділити на кількість змінних, то отримаємо показник, який називається *інформативністю* або *потужністю фактора*. Потужність k -го фактора (V_k) рівна

$$V_k = \frac{1}{p} \sum_{j=1}^m h_j^2 \quad (k = \overline{1, p}) \quad (3.136)$$

Величина

$$V = \sum_{k=1}^p V_k \quad (3.137)$$

називається *повною факторизацією*. Чим більше значення V , тим якіснішим вважається результат факторного аналізу (чітких статистичних критеріїв повноти факторизації нема).

Усі розрахунки, пов’язані з факторним аналізом, ґрунтуються на матриці кореляцій (кореляційній матриці) R , елементи якої r_{ij} є коефіцієнтами кореляції між i -ю та j -ю змінними:

$$R = \begin{bmatrix} r_{11} & r_{12} & \dots & r_{1m} \\ r_{21} & r_{22} & \dots & r_{2m} \\ \dots & \dots & \dots & \dots \\ r_m & r_{n2} & \dots & r_{nm} \end{bmatrix}.$$

Якщо всі значення елементів головної діагоналі матриці R рівні 1 (кореляції кожної змінної самої з собою), то враховується повна дисперсія кожної змінної. Тим самим до уваги приймається сумарний вплив загальних і специфічних факторів.

Матриця кореляцій, в якій елементи головної діагоналі рівні 1, називається *повною матрицею кореляцій* (R_1). Якщо ж на головній діагоналі кореляційної матриці знаходяться значення спільностей (h_j^2), то враховується лише вплив загальних факторів, тобто вплив характерностей елімінується. Матриця кореляцій, в якій елементи головної діагоналі відповідають спільностям, називається *редукованою матрицею* (R_h). Має місце рівність:

$$R = R_h + \alpha \quad (3.138)$$

Матриця A називається *факторним відображенням* і не враховує навантаження характерних факторів. Тому вона є *редукованою матрицею факторних навантажень*.

Перехід від редукованої матриці кореляцій змінних до редукованої матриці факторних навантажень покладено в основу факторного аналізу і переслідує дві цілі:

- визначення навантажень загальних факторів на змінні;
- визначення кількості загальних факторів, необхідних для відображення кореляційних зв'язків між змінними для певного рівня інформативності.

Так як кореляція між факторами відсутня, то має місце така рівність:

$$r_{ij} = \sum_{k=1}^p a_{ik} a_{kj}, \quad (i, j = \overline{1, n}) \quad (3.139)$$

або у матричній формі -

$$R_h = AA^T, \quad (3.140)$$

яка називається *фундаментальною теоремою факторного аналізу*.

Фундаментальна теорема стверджує, що кореляційна матриця може бути відтворена за допомогою факторного відображення і кореляції між факторами.

Відновлені лише по загальних факторах коефіцієнти кореляції будуть меншими за абсолютною величиною від обчислених безпосередньо за

даними спостереження, а елементи головної діагоналі відновленої кореляційної матриці рівні відповідним спільностям (h_j^2).

Максимально можливе відтворення коефіцієнтів кореляції між змінними за допомогою факторів досягається за рахунок варіації кількості факторів і варіації діагональних елементів (спільностей) редукованої кореляційної матриці.

Як правило, наперед невідомо, скільки факторів потрібно для представлення набору змінних. Сама ж процедура факторного аналізу передбачає попереднє зазначення кількості факторів. Для вирішення цієї проблеми на першому етапі факторного аналізу досліджують головні компоненти на основі власних значень.

Головна задача факторного аналізу полягає у визначенні факторних навантажень, а потім і самих факторів. Існує декілька методів розв'язання цієї задачі, серед яких найбільше розповсюдження отримав *метод головних компонент* (англ. — *Principal Component Analysis (PCA)*).

Модель аналізу головних компонент записують так:

$$Z_j = \sum_{k=1}^m a_{kj} F_k \quad (j = \overline{1, m}) \quad (3.141)$$

або у матричній формі -

$$Z = AF.$$

Метод головних компонент – це єдиний статистично обґрунтований метод з поміж тих, що використовуються для реалізації задач факторного аналізу.

Алгоритм методу головних компонент полягає у послідовному виборі компонент, які упорядковуються за спаданням частки поясненої дисперсії початкових змінних.

Кожний наступний фактор визначається на підставі критерію максимізації дисперсії, що залишилася непоясненою попередніми факторами, тому фактори є некорельованими (ортогональними). Максимально можлива кількість некорельованих компонент рівна кількості

початкових змінних, тобто класичний компонентний аналіз зберігає розмірність простору початкових змінних.

Кожній головній компоненті ставиться у відповідність власне значення, яке характеризує частку загальної варіації змінних X_1, X_2, \dots, X_m , що пояснюється цією компонентою. Знаходження власних значень повної матриці кореляції (R_1) покладено в основу методу головних компонент.

Власні значення $\lambda_1, \lambda_2, \dots, \lambda_m$ матриці \mathbf{R} є коренями характеристичного рівняння

$$|\mathbf{R} - \lambda \mathbf{I}| = \begin{vmatrix} 1-\lambda & r_{12} & \dots & r_{1m} \\ r_{12} & 1-\lambda & \dots & r_{2m} \\ \dots & \dots & \dots & \dots \\ r_{m1} & r_{m2} & \dots & 1-\lambda \end{vmatrix} = 0. \quad (3.142)$$

На підставі розв'язків характеристичного рівняння (3.142) отримують матрицю власних значень Λ і матрицю квадратних коренів із власних значень матриці $\Lambda^{\frac{1}{2}} \sqrt{\Lambda}$:

$$\Lambda = \begin{bmatrix} \lambda_1 & 0 & \dots & 0 \\ 0 & \lambda_2 & \dots & 0 \\ \dots & \dots & \dots & \dots \\ 0 & 0 & \dots & \lambda_m \end{bmatrix}; \quad \sqrt{\Lambda} = \begin{bmatrix} \sqrt{\lambda_1} & 0 & \dots & 0 \\ 0 & \sqrt{\lambda_2} & \dots & 0 \\ \dots & \dots & \dots & \dots \\ 0 & 0 & \dots & \sqrt{\lambda_m} \end{bmatrix},$$

де $\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_m$; $\sum_{k=1}^m \lambda_k = m$.

Так як значення змінних представлені у стандартизованій формі, то дисперсія кожної змінної рівна 1, а сумарна дисперсія всіх змінних – кількості змінних, тобто m . Метод головних компонент передбачає збереження всієї сумарної дисперсії. Тому кількість виділених компонент і сума всіх власних значень дорівнює кількості початкових змінних. В результаті застосування методу головних компонент факторна структура спотворюється у бік перебільшення абсолютних значень величин факторних навантажень.

Власне значення λ_k , поділене на кількість змінних – це внесок k -ї змінної у сумарну дисперсію змінних.

Кожному власному значенню λ_k відповідає власний вектор \mathbf{V}_k , який знаходять в результаті розв'язування такого рівняння

$$(\mathbf{R} - \lambda_k \mathbf{I}) \mathbf{V}_k = 0. \quad (3.143)$$

Після цього формують матрицю \mathbf{U} , яка складається із нормованих власних векторів:

$$\mathbf{U} = \begin{bmatrix} u_{11} & u_{12} & \dots & u_{1m} \\ u_{21} & u_{22} & \dots & u_{2m} \\ \dots & \dots & \dots & \dots \\ u_{m1} & u_{m2} & \dots & u_{mm} \end{bmatrix},$$

де $u_k = \frac{\mathbf{V}_k}{|\mathbf{V}_k|}$ ($k = \overline{1, m}$).

На підставі матриць \mathbf{U} і $\sqrt{\Lambda}$ знаходять матрицю значень факторних навантажень \mathbf{A}

$$\mathbf{A} = \mathbf{U} \sqrt{\Lambda} = \begin{bmatrix} a_{11} & a_{12} & \dots & a_{1m} \\ a_{21} & a_{22} & \dots & a_{2m} \\ \dots & \dots & \dots & \dots \\ a_{m1} & a_{m2} & \dots & a_{mm} \end{bmatrix}. \quad (3.144)$$

Елемент a_{ij} матриці \mathbf{A} як коефіцієнт парної кореляції визначає навантаження j -ї головної компоненти на i -ту ознаку. Сума квадратів всіх компонентних навантажень (за стовпцем) дорівнює власному значенню даної компоненти:

$$\lambda_j = \sum_{i=1}^m a_{ij}^2 \quad (j = \overline{1, m}). \quad (3.145)$$

Матриця значень факторних навантажень \mathbf{A} використовується для обчислення матриці значень головних компонент \mathbf{F} :

$$\mathbf{F} = \mathbf{A}^{-1} \mathbf{Z}. \quad (3.146)$$

Матричне рівняння (3.146) – це обернене лінійне перетворення змінних Z_1, Z_2, \dots, Z_m у змінні F_1, F_2, \dots, F_p , яке дає змогу безпосередньо кількісно оцінити досліджувані латентні змінні.

Декомпозицію сумарної дисперсії ознакової множини X і факторної множини F , отриманої за методом головних компонент, можна подати у вигляді табл. 3.5.

Таблиця 3.5

Декомпозицію сумарної дисперсії ознакової множини

Ознакова множина Z	Факторна множина F				Дисперсія Z_i
	F_1	F_2	...	F_m	
Z_1	a_{11}^2	a_{12}^2	...	a_{1m}^2	1
Z_2	a_{21}^2	a_{22}^2	...	a_{2m}^2	1
...	1
Z_m	a_{m1}^2	a_{m2}^2	...	a_{mm}^2	1
Дисперсія F_j	λ_1	λ_2	...	λ_m	m

Крім переходу до повного набору некорельованих змінних, метод головних компонент дає змогу:

- візуалізувати складний набір змінних;
- побачити найінформативніші змінні;
- побачити “особливі” спостереження.

Незважаючи на те, що за допомогою розглянутого вище методу на основі m змінних може бути виділена така ж кількість головних компонент (факторів), внески більшості з них у сумарну варіацію виявляється незначним. Велика частка сумарної варіації припадає на декілька перших компонент (саме вони називаються головними).

Кількість головних компонент, які виділяються у процесі факторного аналізу, у значній мірі визначається суб’єктивно на підставі розуміння того, яка величина сумарної дисперсії сприймається як випадкова мінливість (вплив неконтрольованих чинників, помилки вимірювань тощо).

Існує декілька статистичних критеріїв визначення кількості факторів.

Маючи власні значення головних компонент, для визначення кількості факторів можна скористатися критеріями Кайзера і Кеттелла.

За *критерієм Кайзера* кількість факторів, які залишаються для аналізу, дорівнює кількості компонент, власні значення яких більші за одиницю.

Критерій Кеттелла — це графічний метод, який полягає у візуальному аналізі залежності власних значень від кількості факторів у порядку їх спадання. Кількість факторів визначається наближено по точці перегину на графіку власних значень (до його виходу на пологую пряму після різкого спаду).

Метод головних компонент розглядають як спрощений варіант факторного аналізу. В моделі головних компонент (3.141) дисперсія кожної змінної пояснюється лише головними компонентами, тоді як у моделі факторного аналізу (3.132) дисперсія кожної змінної поділена на дві частини: дисперсію, зумовлену загальними факторами (спільність), і дисперсію, зумовлену варіацією змінної (характерність).

Проведення факторного аналізу передбачає виконання таких етапів:

- формування кореляційної матриці;
- виділення (вибір) факторів;
- інтерпретація результатів.

Сукупність способів вибору факторів (отримання факторної структури) прийнято називати *методами факторного аналізу*. У практиці факторного аналізу використовують декілька методів, які різняться вирішенням проблеми спільності.

Для дослідників, які не є фахівцями в галузі математичної статистики, але використовують у своїй роботі факторний аналіз, важливим є основний зміст процедури факторного аналізу, що полягає у переході від матриці кореляцій до матриці факторних навантажень. Обмежимося лише перерахуванням окремих методів, доповнивши його короткими коментарями.

З погляду логіки реалізації до методу головних компонент подібні методи *головних осей (головних факторів)* і *факторизації образів (аналізу образів)*.

Метод факторизації образів – це метод головних компонент, що застосовується до редукованої матриці (R_h), діагональні елементи якої оцінюються як квадрати коефіцієнтів множинної кореляції кожної змінної з усіма іншими змінними. На думку теоретиків і практиків факторного аналізу, така оцінка забезпечує точніші результати, ніж аналіз головних компонент. Проте значення спільностей вважаються недооціненими, що призводить до певного спотворення факторної структури, хоча і меншого, ніж при використанні методу головних компонент (при використанні методу головних компонент факторна структура спотворюється в бік завищення абсолютних значень величин факторних навантажень).

Точніші оцінки забезпечує метод головних осей (головних факторів). Для знаходження оцінок факторних навантажень і спільностей застосовують ітераційний алгоритм. На першому кроці за допомогою методу головних компонент знаходять оцінки факторних навантажень і на їх основі – оцінки спільностей. На кожному наступному кроці власні значення і факторні навантаження обчислюються на основі оцінок попередніх значень спільностей. Ітераційний процес завершується при виконанні вказаної кількості ітерацій або при досягненні мінімальних відмінностей на двох сусідніх кроках.

На зменшення різниці початкових і обчислених (відтворених) кореляцій спрямовані ітераційні процедури, покладені в основу *методів найменших квадратів, узагальнених найменших квадратів і максимальної правдоподібності*.

Мета використання методу найменших квадратів – мінімізація помилки факторної структури при фіксованій кількості факторів.

Перший крок методу найменших квадратів полягає в оцінюванні спільностей через квадрати коефіцієнтів множинної кореляції між змінними і розрахунку факторної структури. Після цього відтворюються (обчислюються) коефіцієнти кореляції. На другому кроці на основі нових значень спільностей (обчислених за значеннями факторної структури, отриманої на першому кроці) розраховують нову факторну структуру і перевіряють різницю між початковими і відтвореними коефіцієнтами кореляції. Ітераційний процес повторюється до досягнення мінімально можливої різниці між початковими і обчисленими кореляціями при заданій кількості факторів.

Узагальнений метод найменших квадратів відрізняється від попереднього тим, що змінним ставляться у відповідність вагові коефіцієнти. Чим більша вага надається змінній, тим у більшій мірі вона впливає на факторну структуру.

Метод максимальної правдоподібності забезпечує отримання статистичної оцінки повноти факторизації. Мірою повноти (якості) факторизації слугує оцінка відмінності між початковими і відтвореними коефіцієнтами кореляції за критерієм χ^2 , значущість якого визначається в залежності від кількості змінних і кількості факторів. Розрахункове значення критерію визначають за формулою

$$\chi_{\text{розр}}^2 = \left[n - \frac{1}{6}(2p + 5) - \frac{2}{3}m \right] \ln \frac{|AA^T|}{|R|}.$$

Якщо розрахункове значення критерію перевищує критичне (кількість ступенів вільності рівна $V = 0,5[(p - m)^2 - p - m]$), то нульову гіпотезу про те, що m загальних факторів достатньо для пояснення кореляційної матриці відхиляють і переходять до моделі з $(m + 1)$ загальними факторами.

Зауважимо, що однозначних рекомендацій щодо вибору методу факторного аналізу нема. Тому у кожному конкретному випадку потрібно порівнювати результати застосування різних методів, зосереджуючи особливу увагу на величині частки кожної дисперсії, кількості виділених

факторів і отриманні максимально простої інтерпретації факторної структури.

Як правило, рішення, отримане за допомогою вищевказаних методів, дозволяє відразу змістовно охарактеризувати фактори. Разом з тим, цінність результату факторного аналізу визначається перш за все можливістю ідентифікації факторів як причини сумісної зміни декількох ознак. У зв'язку з цим застосовують процедуру обертання факторних осей у багатовимірному просторі з метою отримання простої структури матриці факторних навантажень.

Критеріями простої структури матриці факторних навантажень вважаються:

- наявність у кожній стрічці хоча б одного елемента зі значенням, рівним 0 або близьким до 0;
- для кожної пари факторів (стовпців матриці) існують ознаки, факторні навантаження яких мають максимальні навантаження на один фактор і близькі до 0 на другий фактор, а також ознаки, які мають невеликі навантаження на обидва фактори;
- кожна змінна отримує більше навантаження лише за одним фактором, а за іншими її навантаження близькі до 0.

Система рівнянь (3.132) не має однозначного розв'язку. Розрахунок факторних навантажень виконується з точністю до довільного лінійного перетворення у її правій частині, що еквівалентно можливості довільного повороту факторних осей навколо векторів-змінних.

Геометрична інтерпретація процедури обертання факторів представлена на рис. 3.16. Змінні зображені у вигляді векторів у просторі загальних факторів, а факторні навантаження – це координати векторів. Коефіцієнти кореляції між кожною парою змінних рівні косинусу кута між відповідними векторами. Вказані співвідношення між змінними з математичної точки зору залишаються такими самими, якщо повернути осі на

довільний кут (штрих-пунктирні лінії). Змінилися лише величини факторних навантажень.

Фактори можна повертати відносно змінних будь-яким чином, але повинна зберігатися умова їх ортогональності. При цьому намагаються робити такий поворот осей, щоб кожна змінна отримала максимальне навантаження по одному фактору і мінімальне – по всіх інших. Тоді змінна буде віднесена лише до одного фактора, що є важливим для інтерпретації факторної структури.

В сучасних пакетах прикладних програм, призначених для реалізації задач факторного аналізу, використовуються аналітичні способи обертання факторів: кожна пара факторів повертається відносно змінних до тих пір, поки не буде досягнута максимально проста структура.

Розрізняють два способи обертання факторів:

- *ортогональне обертання* – при повороті осей кут між факторами залишається прямим (зберігається припущення про некорельованість факторів);
- *косокутне обертання* – обмеження про некорельованість факторів не ставиться.

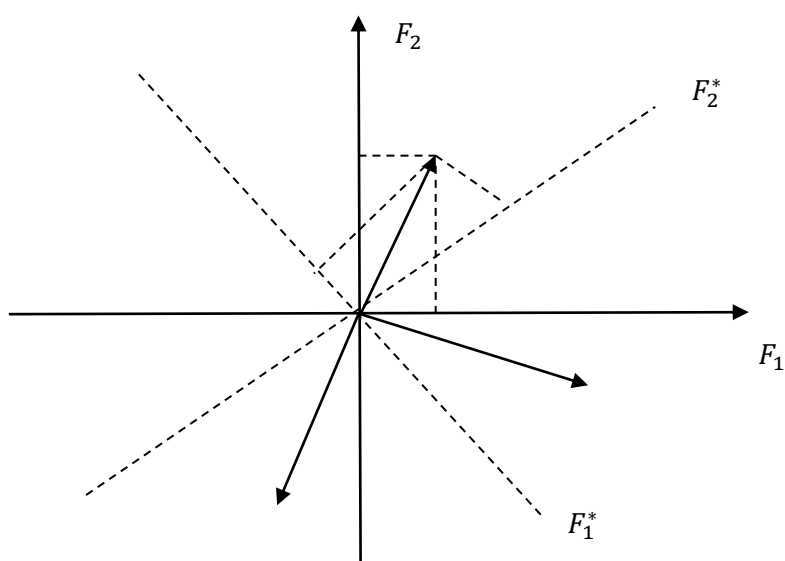


Рис. 3.16. Геометрична інтерпретація обертання осей

З поміж способів ортогонального обертання виділяють:

- *варимакс*;
- *квартимакс*;
- *еквімакс*.

Ціль використання способу варимакс – спрощення стовпців матриці факторних навантажень, а квартимакс – стрічок. Еквімакс займає проміжне становище – використання даного способу спрямовується на спрощення як стовпців, так і стрічок факторної матриці.

Потрібно зауважити, що обертання факторів не завжди полегшує інтерпретацію отриманих залежностей, а в деяких випадках навіть призводить до гірших результатів. Погіршення результатів після обертання свідчить про непридатність вибраної факторної моделі.

Дисперсійний аналіз (від лат. *dispersio* – розсіювання) - це статистичний метод дослідження впливу однієї або декількох номінальних або порядкових змінних на одну кількісну змінну.

У дисперсійному аналізі одні змінні розглядаються як причини (незалежні змінні, фактори), а інші — як наслідки (залежні змінні). Незалежні змінні (предиктори, фактори) є категоризованими, а залежні — метричними.

Дисперсійний аналіз ґрунтується на розкладанні загальної дисперсії на окремі компоненти, зумовлені впливом конкретних факторів, і перевірці гіпотези про значущість впливу факторів на залежну змінну. Порівняння компонент загальної дисперсії за *F*-критерієм дає змогу встановити, яка частка загальної варіації залежної змінної викликана дією факторних змінних.

За кількістю незалежних змінних (регульованих змінних) дисперсійний аналіз поділяють на:

- *однофакторний* — досліджується вплив одного фактора на результати спостереження (*one-way ANOVA*);

▪ *багатофакторний* — досліджується як вплив декількох факторів на результати спостереження окремо, так і їх взаємодія (*n-way ANOVA*).

Дисперсійний аналіз проводять лише за умови, коли розподіл залежної змінної є нормальним.

Типовими прикладами застосування дисперсійного аналізу в маркетингу є дослідження:

- впливу упаковки товару на його сприйняття покупцем;
- зміни обсягу продажу при різних варіантах рекламної компанії;
- впливу соціальних характеристик покупців на кількість постійних клієнтів;
- впливу різних рівнів ціни і реклами на обсяг реалізації товару;
- впливу регіону на обсяг реалізації товару;
- залежності ціни на товар від місця знаходження пункту продажу і пори року;
- впливу рівня освіти на вибір споживачем торгової марки.

Особливу роль дисперсійний аналіз відіграє у процесі дослідження результатів експерименту, так як дає змогу визначити характер зміни дисперсії результатів при зміні рівнів факторів, що вивчаються. Якщо дисперсії будуть відрізнятися значущо, то необхідно зробити висновок про вплив фактора на середнє значення спостережуваної випадкової величини.

З метою вивчення впливу факторів на декілька змінних (багатовимірну залежну змінну) використовують *багатовимірний дисперсійний аналіз*. Багатовимірний дисперсійний аналіз, крім перевірки гіпотези про вплив факторів на кожну залежну змінну окремо, забезпечує перевірку гіпотези про вплив факторів на всю сукупність змінних як одну багатовимірну змінну.

Прикладом застосування багатовимірного дисперсійного аналізу у дослідженні ринку товарів і послуг може слугувати вивчення впливу інтенсивності реклами (високий, середній і низький рівень) і наявності

акційних заходів (наявність, відсутність) одночасно на обсяги продажів товару і кількість відвідувачів супермаркету.

Початковим матеріалом для дисперсійного аналізу слугують дані дослідження трьох і більше вибірок, які можуть бути як рівними, так і нерівними за чисельністю, як зв'язаними, так і незв'язаними.

Припустимо, що проведено спостереження за деякою випадковою величиною X , на значення якої впливає один фактор, що має m рівнів градації, і сформовано вибірку (табл. 3.6).

Таблиця 3.6

Початкові дані для однофакторного дисперсійного аналізу

Рівень фактора	Значення вибірки	Обсяг вибірки
1	$x_{11}, x_{12}, \dots, x_{1j}, \dots, x_{1n_1}$	n_1
2	$x_{21}, x_{22}, \dots, x_{2j}, \dots, x_{2n_2}$	n_2
.....
i	$x_{i1}, x_{i2}, \dots, x_{ij}, \dots, x_{in_i}$	n_i
.....
m	$x_{m1}, x_{m2}, \dots, x_{mj}, \dots, x_{mn_m}$	n_m

Позначення: x_{ij} – j -те значення залежної змінної для i -го рівня незалежної змінної (i -го рівня фактора);
 n_i ($i = \overline{1, m}$) – кількість спостережень для i -го рівня;
 $n_1 + n_2 + \dots + n_m = N$ – загальний обсяг вибірки.

На основі даних табл. 3.6 можна обчислити \bar{x}_i — середнє значення залежної змінної в i -й групі (для i -го рівня фактора) та \bar{x} — загальновибіркову середню:

Тоді

$$\bar{x}_i = \frac{1}{n_i} \sum_{j=1}^{n_i} x_{ij} \quad (i = \overline{1, m}); \quad (3.147)$$

$$\bar{x} = \frac{1}{N} \sum_{i=1}^m \sum_{j=1}^{n_i} x_{ij}. \quad (3.148)$$

Основна мета дисперсійного аналізу полягає у дослідженні статистичної значущості відмінностей між груповими середніми за допомогою порівняння дисперсій.

Нехай:

- SST — сума квадратів відхилень залежної змінної від загально-вибіркової середньої (загальна варіація залежної змінної);
- SSB — міжгрупова сума квадратів відхилень (частина варіації залежної змінної, обумовлена відмінністю середніх у групах незалежної змінної);
- SSW — внутрішньогрупова сума квадратів відхилень (частина варіації залежної змінної, обумовлена мінливістю всередині груп);.

Можна показати, що

$$SST = SSB + SSW, \quad (3.149)$$

де

$$SST = \sum_{i=1}^m \sum_{j=1}^{n_i} (x_{ij} - \bar{x})^2; \quad (3.150)$$

$$SSB = \sum_{i=1}^{n_i} n_i (\bar{x}_i - \bar{x})^2; \quad (3.151)$$

$$SSW = \sum_{i=1}^m \sum_{j=1}^{n_i} (x_{ij} - \bar{x}_i)^2. \quad (3.152)$$

В основу процедури дисперсійного аналізу покладено обчислення відношення міжгрупової (систематичної) дисперсії до внутрішньогрупової (випадкової) дисперсії.

Міжгрупова дисперсія (MSB) вимірює систематичну варіацію залежної змінної, зумовлену впливом фактора, покладеного в основу групування, і дорівнює середньому квадрату відхилень групових середніх від загальної середньої:

$$MSB = \frac{SSB}{m-1}. \quad (3.153)$$

Внутрішньогрупова дисперсія (MSV) відображає випадкову варіацію — частину загальної варіації, яка спричинена впливом неврахованих чинників і не залежить від фактора, покладеного в основу групування:

$$MSW = \frac{SSW}{N-m}. \quad (3.154)$$

Загальна дисперсія залежної змінної MST дорівнює

$$\frac{SST}{N-1}. \quad (3.155)$$

Якщо фактор не впливає на значення залежної змінної, то внутрішньогрупова і міжгрупова дисперсії повинні бути приблизно рівними і для перевірки гіпотези про рівність групових середніх застосовують F -критерій

$$F_{\text{розр}} = \frac{MSB}{MSW}. \quad (3.156)$$

Для вибраного рівня значущості α і $V_1 = m-1$ та $V_2 = N-m$ ступенів вільності знаходять критичне значення статистики Фішера ($F_{\text{крит}}$). Якщо $F_{\text{розр}} > F_{\text{крит}}$, то гіпотеза про рівність групових середніх відхиляється.

Величина

$$\eta^2 = \frac{SSB}{SST} \quad (3.157)$$

характеризує силу впливу (ефект) незалежної змінної на залежну змінну і називається *кореляційним відношенням*. Кореляційне відношення змінюється в межах від 0 (всі групові середні рівні) до 1 (всередині кожної групи незалежної змінної залежна змінна не варіює).

Умовами коректного застосування дисперсійного аналізу вважаються:

- кожна група (градація, рівень) є незалежною випадковою вибіркою з нормально розподіленої сукупності;
- дисперсії однакові на різних рівнях досліджуваного фактора (умова гомоскедастичності).

Стосовно першої умови потрібно зауважити, що однофакторний дисперсійний аналіз відносно мало чутливий до її порушення.

Якщо обсяги вибірки є однаковими, то порушення умови про рівність дисперсій несуттєво впливає на висновок, зроблений на підставі F -критерію. Проте, якщо обсяги вибірок різняться, порушення умови про рівність дисперсій може серйозно спотворити результати дисперсійного аналізу. Для перевірки того, що вибірки мають рівні дисперсії, найчастіше користуються критеріями Левене і Бартлета, з якими читач може ознайомитися у спеціалізованих джерелах.

Вище відзначалося, що важливою умовою застосування дисперсійного аналізу є нормальність розподілу залежної змінної і рівність дисперсій фактора у порівнюваних групах. У випадку суттєвого порушення цих умов рекомендують *непараметричний дискримінантний аналіз*. Непараметричний дискримінантний аналіз не передбачає припущення про нормальність вибірки.

Альтернативою класичного однофакторного аналізу на випадок, коли залежна змінна представлена у порядковій шкалі вимірювання, є *дисперсійний аналіз Краскела-Уолліса*. Він ґрунтується на критерію, який призначений для перевірки рівності медіан (Me) декількох вибірок.

Нульова і альтернативна гіпотези формулюються так:

H_0 : кожна група характеризується однаковим розподілом;

H_0 : не всі групи мають однаковий розподіл.

Для розрахунку критерію Краскела-Уолліса потрібно замінити значення спостережень за залежною змінною рангами. При цьому ранг 1 присвоюється найменшому значенню спостереження, а N – найбільшому.

Якщо окремі значення повторюються, то їм присвоюють відповідне середнє значення рангів.

Статистика, яка використовується у критерію Краскела-Уолліса, аналогічна величині SSB (міжгруповій варіації), на основі якої обчислюється F -статистика в однофакторному дисперсійному аналізі. Замість порівняння середніх значень всіх груп із загальною середньою, порівнюються середні ранги всіх груп із середнім рангом, обчисленим на основі всіх спостережень.

Для розрахунку статистики Краскела-Уолліса користуються формулою

$$H = \frac{12}{N(N+1)} \sum_{i=1}^m \frac{R_i^2}{n_i} - 3(N+1), \quad (3.158)$$

де R_i – сума рангів i -ї вибірки.

H -статистика апроксимується χ^2 -розподілом з $(m-1)$ ступенем вільності. Якщо розрахункове значення H -статистики перевищує критичне значення χ^2 -статистики, то нульова гіпотеза відхиляється, тобто рівень ознаки при переході від групи до групи змінюється.

Варіація залежних змінних, які є синтетичними показниками, зумовлена одночасною дією декількох факторів, причому фактори мають різну силу впливу. Тому використання однофакторного дисперсійного аналізу для дослідження відмінностей між груповими середніми у вказаному випадку є недостатнім.

Для дослідження одночасного впливу декількох факторів на залежну змінну використовують багатофакторний дисперсійний аналіз. Проведення багатофакторного дисперсійного аналізу передбачає наявність однієї нормально розподіленої залежної змінної (представлена у метричній шкалі вимірювань) і двох або більше факторів (категоризованих незалежних змінних).

Характерна особливість багатофакторного дисперсійного аналізу — можливість оцінити не тільки вплив кожного фактора на залежну змінну, а і їхню взаємодію. У зв'язку з цим використання багатофакторного

дисперсійного аналізу забезпечує отримання результатів, які виявити за допомогою однофакторного дисперсійного аналізу не реально.

Процедура багатфакторного дисперсійного аналізу аналогічна процедурі однофакторного дисперсійного аналізу і ґрунтується на розкладанні загальної суми квадратів відхилень (загальної варіації) залежної змінної на окремі складові.

Розглянемо теоретичні основи двохфакторного дисперсійного аналізу.

Припустимо, що необхідно дослідити вплив факторів A і B на залежну змінну, причому фактор A характеризується m рівнями, а фактор B — n рівнями. Загальна варіація залежної змінної (SST) може бути представлена так:

$$SST = SS_A + SS_B + SS_{AB} + SS_W, \quad (3.159)$$

де SS_A і SS_B — відповідно суми квадратів відхилень, що спричинені дією факторів A і B (окремо кожного);

SS_{AB} — сума квадратів відхилень, спричинена взаємодією факторів A і B ;

SS_W — сума квадратів відхилень, спричинена дією неврахованих чинників.

Кожній сумі квадратів відхилень відповідає певний ступінь вільності:

$$\begin{aligned} F^T &= N-1, \text{ де } N - \text{обсяг вибірки;} \\ F^A &= m-1; \quad F^B = n-1; \quad F^{AB} = (m-1)(n-1); \quad F^W = N-mn. \end{aligned} \quad (3.160)$$

Ступені вільності утворюють рівність

$$F^T = F^A + F^B + F^{AB} + F^W. \quad (3.161)$$

Якщо кожен суму квадратів відхилень поділити на відповідний ступінь вільності, то отримуємо такі дисперсії:

$$MS = \frac{SST}{N-1} \text{ — загальна дисперсія;}$$

$$MS_A = \frac{SS_A}{m-1} \text{ — дисперсія, спричинена рівнями фактора } A;$$

$$MS_B = \frac{SS_B}{n-1} \text{ — дисперсія, спричинена рівнями фактора } B;$$

$MS_{AB} = \frac{SS_{AB}}{(m-1)(n-1)}$ — дисперсія, спричинена спільною дією (взаємодією)

факторів;

$MS_W = \frac{SS_W}{N - mn}$ — залишкова дисперсія.

Взаємодія між факторами описується у вигляді зміни одного фактора під впливом іншого.

Чим більший вплив фактора на залежну змінну, тим більшою є відповідна йому сума квадратів у виразі (3.160). Якщо фактори не залежать один від одного, то $SS_{AB} = 0$.

Об'єднаний вплив обох факторів називають *повним ефектом* (множинним кореляційним відношенням):

$$\eta^2 = \frac{SS_A + SS_B + SS_{AB}}{SST} \quad (3.162)$$

Виявлення впливу факторів A і B на загальну змінну проводять з використанням трьох незалежних нульових гіпотез:

H_A : фактор A не впливає на залежну змінну;

H_B : фактор B не впливає на залежну змінну;

H_{AB} : взаємодія факторів не впливає на залежну змінну.

Перевірку вказаних вище гіпотез проводять за допомогою критерію Фішера (табл. 3.7).

Таблиця 3.7

Статистики для перевірки гіпотез двохфакторного дисперсійного аналізу

Джерело варіації	Кількість ступенів вільності	Сума квадратів	Дисперсія	$F_{\text{розр}}$
A	$m - 1$	SS_A	MS_A	$F^A = MS_A / MS_W$
B	$n - 1$	SS_B	MS_B	$F^B = MS_B / MS_W$
AB	$(m - 1) \cdot (n - 1)$	SS_{AB}	MS_{AB}	$F^{AB} = MS_{AB} / MS_W$
Залишок	$N - mn$	SS_W	X	X
Разом	$N - 1$	SST	X	X

Кожна з трьох нульових гіпотез відхиляється, якщо для заданого рівня значущості відповідне розрахункове значення F -статистики перевищує критичне.

З метою дослідження залежності нормально розподіленої величини від факторів A , B і C використовують трифакторний дисперсійний аналіз. Загальна сума квадратів відхилень у цьому випадку матиме такий вигляд:

$$SST = SS_A + SS_B + SS_C + SS_{AB} + SS_{AC} + SS_{BC} + SS_W \quad (3.163)$$

Можна розглядати і більшу кількість факторів, але потрібно розуміти, що аналіз великої кількості факторів пов'язаний із двома проблемами:

- збором і представленням даних;
- інтерпретацією результатів.

Якщо мають місце повторні вимірювання одних і тих же змінних (у різний час або за різних умов), то говорять про наявність *факторів повторних вимірювань*. Фактори повторних вимірювань також називають *внутрішньогруповими факторами*, так як для їх оцінки використовується внутрішньогрупова сума квадратів. У випадку, коли різним градаціям фактора відповідає одна і та ж вибірка (залежні вибірки), використовують різновид дисперсійного аналізу, який називається *дисперсійним аналізом із повторними вимірюваннями*. Дисперсійний аналіз із повторними вимірюваннями (*RM-ANOVA*) – це метод багатовимірною статистичного аналізу, за допомогою якого одні і ті ж самі об'єкти (респонденти) досліджуються за різних комбінацій рівнів фактора (за різних умов експерименту) з повторними вимірюваннями одних і тих самих змінних. В англійській літературі цей варіант дисперсійного аналізу отримав назву *Repeated-Measures Design* – експериментальний план з повторними вимірюваннями.

Типовим прикладом слугують панельні дослідження, в яких респонденти відповідають на одне і те саме питання через певні проміжки часу. Мета дисперсійного аналізу у даному випадку полягає в оцінюванні

впливу на відповіді респондентів фактора часу (зміни уподобань споживачів у часі).

Використання повторних вимірювань призводить до порушення припущення дисперсійного аналізу про незалежність вибірок.

У випадку повторних спостережень існує зв'язок між значеннями вимірювань для кожного респондента (значення ознаки, отримані від одного респондента, будуть ближчими між собою, чим значення ознаки, отримані від різних респондентів), а отже дисперсія значень за умови повторних вимірювань буде меншою.

У процесі звичайного дискримінантного аналізу (*ANOVA*) розглядається лише мінливість об'єкта. У процесі дисперсійного аналізу з повторними вимірюваннями (*RM-ANOVA*) виділяють два джерела варіації залежної змінної – рівень вимірювань і сам об'єкт, що забезпечує вилучення із загальної дисперсії даних частини дисперсії, зумовленої індивідуальними відмінностями у рівнях залежної змінної. Тому дисперсійний аналіз із повторними вимірюваннями стає чутливішим до впливу факторів, тобто підвищується статистична потужність аналізу.

Дисперсійний аналіз із повторними вимірюваннями ґрунтується на представленні загальної дисперсії у вигляді таких трьох складових:

- дисперсії, зумовленої відмінностями між об'єктами (респондентами) вимірювання;
- дисперсії, зумовленої градацією (рівнями) факторів повторних вимірювань;
- дисперсії залишків (випадкової компоненти).

Розрізняють два підходи до реалізації задач однофакторного дисперсійного аналізу з повторними вимірюваннями – одновимірний і багатовимірний.

Застосування одновимірного підходу передбачає використання додаткового припущення – дисперсії залежної змінної для різних рівнів

внутрішньогрупового фактора приблизно однакові і має місце кореляція між повторними вимірюваннями (умова сферичності). Умова сферичності є необхідною і достатньою умовою для обґрунтованого використання F -критерію. Перевірка гіпотези про те, що набір вибірок задовольняє умову сферичності, проводиться за допомогою тесту Маучлі.

Для дослідження значущості одновимірних факторів повторних вимірювань, які мають більше, ніж два рівні, можна застосувати методи багатовимірної дисперсійної аналізу.

Багатовимірний підхід менше чутливий до коливань залежної змінної, але його застосування не передбачає припущення про корельованість вимірювань залежної змінної. Замість F -критерію (як в одновимірному підході) використовуються багатовимірні тести (Лямбда Уїлкса, слід Піллая).

3.4. Статистичний аналіз і моделювання з використанням нейронних мереж

Штучна нейронна мережа (нейронна мережа, НМ) – це математична модель об'єкта або процесу разом з її програмним та апаратним забезпеченням, побудована на основі принципів організації функціонування біологічних нейронних мереж.

Зазвичай НМ застосовують у процесі статистичного аналізу і моделювання тоді, коли початково не відомий точний вид зв'язку між змінними, а лінійна апроксимація моделі за умови існування нелінійних зв'язків не може слугувати надійним засобом отримання достовірних результатів дослідження.

Спроможність моделювати нелінійні залежності, узагальнювати і виділяти приховані залежності між вхідними і вихідними даними, визначати неінформативні для аналізу дані та відсіювати їх, адаптуватися до мінливих

зовнішніх впливів, працювати з неповними даними тощо є головною перевагою НМ порівняно з традиційними статистичними методами, що використовують для розв'язання багатьох задач. Типовими задачами статистичного аналізу і моделювання соціально-економічних об'єктів та процесів, які ефективно вирішуються за допомогою використання НМ, є задачі дослідження динаміки функціонування економічних систем, кластеризації, передбачення рейтингів, прогнозування часових рядів.

Останнім часом на основі НМ розроблено ряд програмних продуктів для прогнозування на фондовому ринку, оцінювання ризиків неповернення кредиту в банківській сфері, оцінювання фінансового стану підприємств та ін.

Одиницею обробки інформації в НМ є *штучний нейрон* (далі – нейрон) – аналог біологічного нейрона (рис. 3.17). Він має декілька каналів вводу інформації (*дендритів*) та один канал виводу інформації (*аксон*). Аксон нейрона з'єднаний з іншими нейронами зв'язками, які називають *синапсами*. Синапс характеризується одним параметром – ваговим коефіцієнтом (w), значення якого впливає на зміни інформації, коли вона передається від одного нейрона до іншого. Додатне значення вагового коефіцієнта відповідає збудним сигналам, а від'ємне – гальмуючим сигналам.

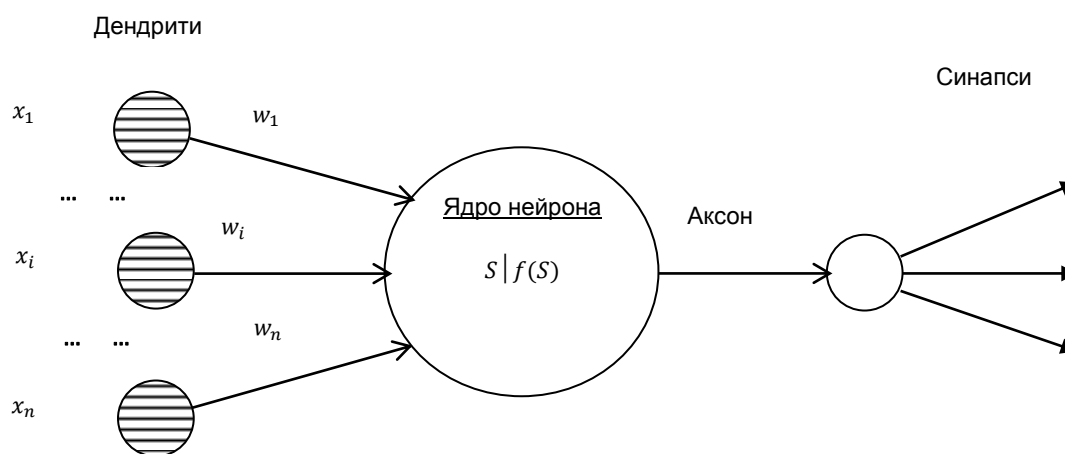


Рис. 3.17. Схема штучного нейрона

Математично штучний нейрон подають у вигляді деякої нелінійної функції, аргументом якої є лінійна комбінація всіх його вхідних сигналів, а значенням – вихід нейрона (аксон). У теорії НМ таку функцію називають *функцією активації* (активаційною функцією):

$$y = f(S), \quad (1.164)$$

$$S = \sum_{i=1}^n w_i \cdot x_i, \quad (1.165)$$

де n – кількість входів нейрона; x_i ($i = \overline{1, n}$) – значення i -го входу нейрона; w_i ($i = \overline{1, n}$) – вага i -го входу нейрона; S – зважена сума входів нейрона;

$f(S)$ – активаційна функція; y – значення вихідного сигналу нейрона (активаційної функції).

Функція активації як засіб перетворення зваженого входу нейрона у вихідний сигнал застосовується на останньому етапі обробки сигналу нейроном з метою обмеження амплітуди вихідного сигналу. Від вибраної функції активації та її параметрів залежить якість навчання НМ на конкретних прикладах. Найчастіше вживані активаційні функції представлено в табл. 3.8.

Таблиця 3.8

Активаційні функції

Назви функцій	Функціональні залежності	Області значень
Ступінчаста порогова	$y = \begin{cases} 0, \text{ якщо } S < S^* \\ 1, \text{ якщо } S \geq S^* \end{cases}$	0; 1
Тотожна (лінійна)	$y = S$	$(-\infty; \infty)$
Напівлінійна	$y = \begin{cases} 0, \text{ якщо } S < 0 \\ S, \text{ якщо } S \geq 0 \end{cases}$	$[0; \infty)$
Напівлінійна з посиленням	$y = \begin{cases} 0, \text{ якщо } S \leq 0 \\ S, \text{ якщо } 0 < S < 1 \\ 1, \text{ якщо } S \geq 1 \end{cases}$	$[0; 1]$
Лінійна з посиленням	$y = \begin{cases} -1, \text{ якщо } S \leq -1 \\ S, \text{ якщо } -1 < S < 1 \\ 1, \text{ якщо } S \geq 1 \end{cases}$	$[-1; 1]$

Сигмоїдна (логістична)	$y = \frac{1}{1+e^{-as}}$ де a – параметр, що визначає нахил функції	(0; 1)
Гіперболічний тангенс	$y = \frac{e^s - e^{-s}}{e^s + e^{-s}}$	(-1; 1)
Радіальна базисна (Гаусса)	$y = e^{-s^2/2}$	(0; 1)

Задачі, які вирішуються за допомогою НМ, зводяться до апроксимації багатовимірних функцій – побудови відображення $F: x \rightarrow y$, яке ставить у відповідність кожному елементу множини x єдиний елемент множини y . Якщо використовуються порогові активаційні функції (будь-який із виходів мережі дорівнює 0 або 1), то НМ слугують інструментом класифікації. Використання неперервних активаційних функцій доцільне для задач регресійного аналізу.

Нейронна мережа – це сукупність нейронів і зв'язків між ними (рис. 3.18).

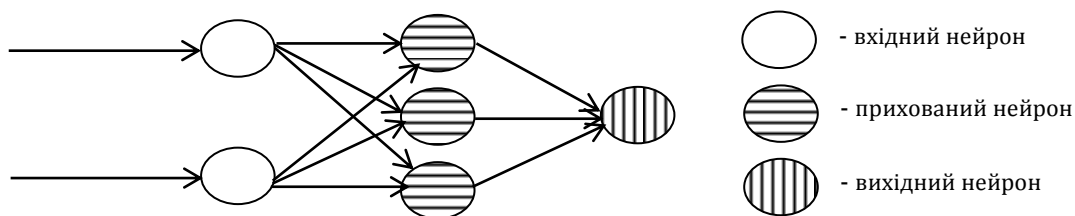


Рис. 3.18. Схема простої нейронної мережі

За розташуванням в топології мережі і виконуваної функції нейрони поділяють на:

- *вхідні нейрони* – нейрони, які отримують вхідні сигнали і передають їх на наступні рівні (як правило, не виконують обчислювальних операцій);
- *вихідні нейрони* – виходи мережі (можуть виконувати обчислювальні операції);
- *приховані нейрони* – нейрони, які виконують основні обчислювальні операції.

Сукупність нейронів, об'єднаних за особливостями їх функціонування (за функціями призначення), утворюють шар нейронів. У загальному випадку штучна НМ містить три шари нейронів, а саме: *вхідний шар, приховані шари, вихідний шар*.

Вхідний шар – це перший шар НМ, який призначений для прийому вхідних сигналів і передачі їх прихованому шару (крім вхідних нейронів може містити один нейрон для зміщення – зсуву графіка функції активації вправо або вліво). Приховані шари розташовані між вхідним і вихідним шарами, виконують різні перетворення стосовно вхідних сигналів. Усі нейрони у прихованому шарі зв'язані з кожним нейроном у наступному шарі, а кількість прихованих шарів визначається задачею, для вирішення якої створюється НМ. Вихідний шар – це шар, призначений для видачі результатів моделювання.

Розрізняють *одношарові* і *багатошарові НМ*. В одношарових НМ нема прихованих шарів і сигнали з вхідного шару відразу передаються вихідному шару, тобто нейрони вхідного шару вирішують лише задачу розподілення сигналів і не виконують обчислювальних операцій. Необхідні обчислення і видачу результатів забезпечує вихідний шар нейронів.

Багатошарові НМ, крім вхідного і вихідного шарів, містять один або декілька *прихованих шарів*, які розташовані між ними і призначені для обробки інформації. Нейрони вхідного шару приймають виміряні значення характеристик досліджуваних об'єктів (процесів) і передають їх на перший прихований шар без перетворення. Приховані шари і вихідний шар відображають специфіку знань та перетворюють вхідні дані. Вихідний шар остаточно генерує розв'язок задачі у вигляді скалярних або векторних значень. Багатошарова НМ спроможна моделювати ситуації, яким притаманний високий рівень складності і невизначеності. Встановлення кількості прихованих шарів та кількості елементів у них є важливим проблемним питанням побудови НМ.

У той час, коли кількість нейронів у вхідному та вихідному шарах визначається досить легко на основі постановки задачі, визначення кількості нейронів у прихованих шарах є неоднозначною задачею. Нині немає строгих правил для визначення кількості нейронів у прихованих шарах. Тому на практиці часто використовують доволі трудомісткий процес – починають з одного нейрона і додають по одному нейрону в прихований шар до тих пір, поки вихідний результат не буде достатньо точним.

З-поміж багатошарових НМ виділяють:

- *мережі прямого поширення* (без зворотних зв'язків);
- *рекурентні мережі* (зі зворотними зв'язками);
- *мережі радіальних базисних функцій (RBF-мережі)*.

У мережах *прямого поширення* сигнали передаються строго в напрямку від вхідного шару до прихованих шарів і від прихованих шарів до вихідного шару. У зворотному напрямку сигнали не передаються, і нейрони одного шару не пов'язані між собою. Такі мережі успішно використовують для вирішення задач прогнозування.

НМ, у яких виходи нейронів наступних шарів мають семантичні зв'язки з нейронами попередніх шарів, називають *рекурентними*. Виділяють декілька різновидів рекурентних НМ. Прикладом рекурентної НМ є *мережа Елмана* (рис. 3.19).

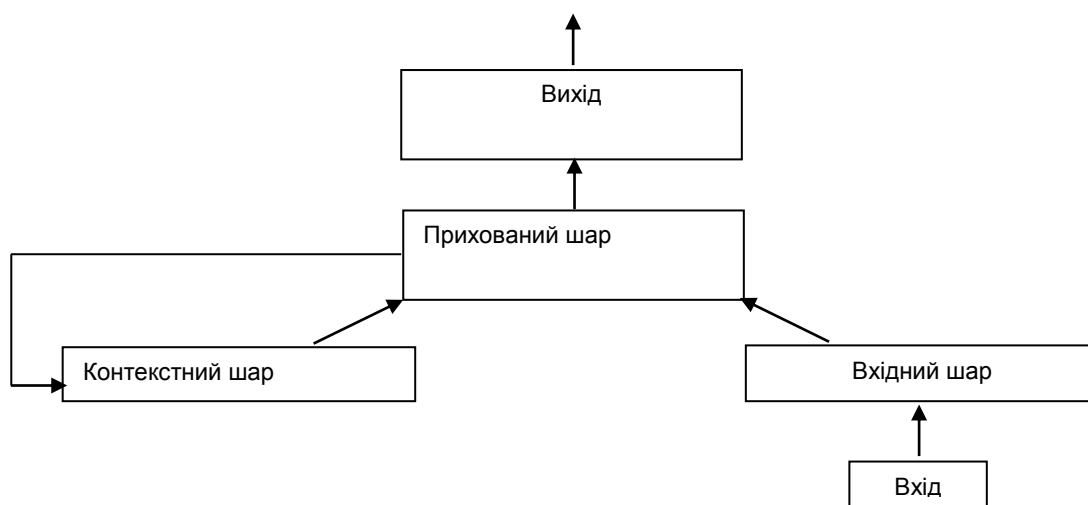


Рис. 3.19. Рекурентна мережа Елмана

Ця мережа побудована таким чином, що вихідні сигнали прихованого шару передаються не тільки нейронам вихідного шару, а й додатковим нейронам, які називають *контекстними*. Вони виконують функцію зберігання інформації про попередній стан нейронів прихованого шару. Така архітектура НМ є ефективною у випадку відтворення часових рядів.

RBF-мережі – це двошарові нейронні мережі прямого поширення, які будують на основі використання радіальних базисних функцій. Активаційна функція кожного нейрона прихованого шару – це радіальна базисна функція. Оскільки радіальна базисна функція є нелінійною, то для моделювання довільної функції достатньо використовувати лише один шар прихованих нейронів.

Історично першими були створені одношарові НМ, які називають *перцептронами* (з лат. *perception* – сприйняття). Їхнє призначення – класифікація лінійно розділених об'єктів. До складу такої мережі входять елементи трьох типів: А-елементи, S-елементи та R-елементи (рис. 3.20).

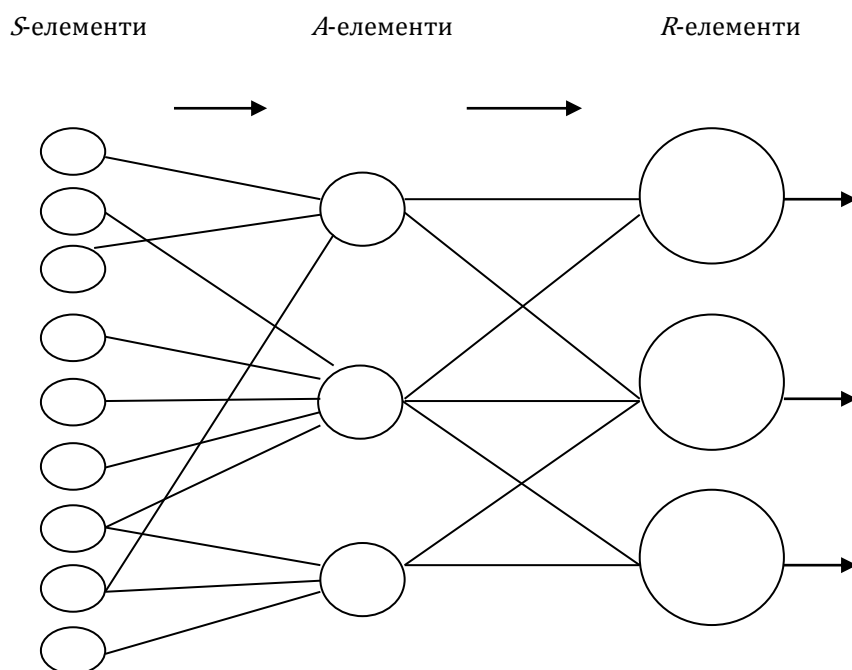


Рис. 3.20. Логічна схема перцептрона з трьома виходами

А-елементи (асоціативні елементи) – це шар нейронів, які з'єднані з множиною входів (S-елементів, рецепторів, сенсорів). Кожен S-елемент може перебувати в одному з двох станів – спокою або збудження. В останньому випадку відбувається передача одиничного сигналу асоціативним елементам. Асоціативному елементу відповідає певний набір (асоціація) S-елементів. Асоціативні елементи активізуються, якщо кількість сигналів від S-елементів на їхньому вході перевищує деяку величину. Сигнали від збуджених А-елементів передаються суматору (R-елементам). Причому, сигнал від кожного асоціативного елемента має свою вагу. Реагуючі елементи на основі порогових активаційних функцій формують вихідний сигнал.

Надалі були розроблені багат шарові перцептрони, які знайшли успішне застосування для вирішення багатьох складних задач, зокрема, дослідження нелінійних тенденцій.

Наявність нелінійної функції активації, одного або декількох прихованих шарів нейронів, високий рівень семантичних з'єднань разом зі спроможністю до навчання забезпечують обчислювальну потужність багат шарового перцептрона.

Теоретичною основою для побудови НМ слугує твердження: для довільної множини пар вхідного і вихідного векторів існує двошарова однорідна НМ з послідовними зв'язками, кінцевою кількістю нейронів і сигмоїдальними функціями активації.

Однією з головних переваг НМ перед традиційними алгоритмами є спроможність навчатися, що з технічної точки зору зводиться до пошуку адекватних коефіцієнтів зв'язку між нейронами. У результаті навчання виявляються і узагальнюються закономірності між вхідними і вихідними даними. За умови успішного навчання НМ буде формувати правильні відповіді на підставі нових даних (навіть у випадку неповних і частково спотворених даних).

Набір нейронів і структура семантичних зв'язків, як правило, формуються один раз на початку створення НМ і не змінюються в процесі її експлуатації. Процес налаштування семантичних ваг називають *навчанням НМ*. Залежно від способів навчання розрізняють:

- *навчання з учителем;*
- *навчання без учителя;*
- *навчання з підкріпленням.*

Навчання з учителем передбачає наявність навчальних прикладів (навчальної множини, даних спостереження), кожен з яких є парою: <вхід - бажаний вихід>.

Кожен приклад подається на вхід НМ і підлягає обробці, результатом якої є вихідний сигнал. Його значення порівнюють з відповідним значенням бажаного виходу.

Навчання з учителем (з математичної точки зору) можна сприймати як аналог задачі інтерполяції або апроксимації деякої таблично заданої функції.

Опрацьовуючи під час навчання вхідні та відповідні їм вихідні значення даних, НМ знаходить деякі залежності між ними. В ідеальному випадку після навчання вона має видавати бажаний вихідний сигнал у разі надходження на вхід довільного сигналу з навчального набору даних.

Різницю між бажаним (очікуваним, цільовим) значенням виходу моделі (y) і фактичним значенням виходу моделі (y^*) називають *помилкою навчання* (E). Якщо НМ має лише один вихідний нейрон, то $E = y - y^*$. У випадку n вихідних нейронів помилку навчання визначають як середній квадрат помилок на кожному виході, тобто:

$$E = \frac{1}{n} \sum_{i=1}^n (y_i - y_i^*)^2. \quad (3.166)$$

Помилка навчання є показником точності налаштування НМ на навчальній множині і може бути використана як умова зупинки навчання. Проте вона не може слугувати оцінкою точності функціонування моделі з даними, які не приймали участі у навчанні, тобто узагальненою оцінкою

спроможності мережі. Контроль якості навчання НМ можна виконати на тестовій множині – на наборі прикладів, які не пред’являлися у процесі навчання. Помилка виходу НМ стосовно прикладів тестової множини називається *помилкою узагальнення*.

Якість функціонування НМ залежить не тільки від її архітектури (кількості шарів і нейронів у кожному шарі), але й від того, наскільки добре вона була навчена.

Спосіб навчання НМ, за якого не вказують правильні відповіді на вхідні сигнали, називають навчанням без учителя. Алгоритм навчання без учителя застосовують у випадках, коли відомі лише вхідні сигнали. Переважно параметри НМ налаштовують так, щоб вона видавала однакові результати для достатньо близьких значень вхідних сигналів.

Навчання НМ з підкріпленням є частковим випадком навчання з учителем. У процесі такого способу мережа навчається, взаємодіючи з деяким середовищем. Такий спосіб навчання використовують у тому випадку, коли необхідно вибрати кращий варіант з доступної множини рішень.

Властивість НМ навчатися робить їх більш привабливими порівняно з системами, які функціонують за наперед заданими правилами. Виділяють два класи методів навчання НМ:

- *детерміновані* – коригування параметрів мережі проводять на основі їхніх поточних значень, величин входів, фактичних і бажаних виходів (типовий приклад – метод зворотного розповсюдження помилки);
- *стохастичні* – параметри змінюються випадковим чином, причому зберігаються лише ті зміни, які зумовили покращання якості моделі.

Навчання НМ ґрунтується на прямій і зворотній передачі інформації про помилки навчання (рис. 3.21).



Рис. 3.21. Схема процесу навчання НМ

Вхідні сигнали послідовно надходять у НМ, яка їх обробляє і формує відповіді. За результатами сформованої відповіді обчислюється помилка навчання. Якщо помилка навчання задовільна, то мережа переходить до опрацювання наступної групи вхідних сигналів, інакше – проводиться процедура мінімізації помилки навчання.

Процедури 2-6 повторюються для кожної пари навчальної множини, допоки помилка навчання на всій навчальній множині не досягає прийняттого рівня.

Процес навчання НМ є ітеративним, при цьому необхідно розрізнити такі поняття:

- *тренувальний сет* – послідовність даних, якими оперує мережа (наприклад, у випадку використання операції логічного множення розглядаються 4 тренувальних сеті: $0 L 0 = 0$; $0 L 1 = 0$; $1 L 0 = 0$; $1 L 1 = 1$);
- *ітерація* – кількість тренувальних сетів, пройдених мережею (лічильник, який збільшується кожний раз при проходженні одного сету);
- *epoca* – один перегляд всіх прикладів з навчальної множини (для кожної епохи обчислюється помилка навчання).

Після того, як сформовано навчальну вибірку та обрано спосіб обчислення функції помилки, навчання НМ набуває характеру задачі багатовимірної оптимізації, яка зводиться до пошуку вагових коефіцієнтів синапсів, які мінімізують розбіжність між отриманим і бажаним результатом. Для вирішення цієї задачі використовують:

- алгоритми локальної оптимізації (метод градієнтного спуску, метод спряжених градієнтів, метод Ньютона);
- стохастичні алгоритми оптимізації (метод пошуку у випадковому напрямку, метод Монте-Карло);
- алгоритми глобальної оптимізації (методи перебору значень змінних).

Метод *градієнтного спуску* вважається одним із найпростіших і логічно обґрунтованих методів пошуку мінімального значення функції багатьох змінних. На першому етапі цього методу вибирають деяку початкову точку M_0 та обчислюють у ній градієнт (вектор, який задає напрямок найшвидшого зростання функції). Далі роблять крок в антиградієнтному напрямку. У результаті отримують нову точку M_1 , значення функції в якій зазвичай не

більше за значення функції в точці M_0 . Якщо значення функції збільшилося, то потрібно зменшити крок. Після цього в новій точці M_1 обчислюють градієнт і рухаються в антиградієнтному напрямку. Момент завершення ітераційного процесу настає тоді, коли рух з отриманої точки за малого кроку призводить до збільшення значення функції або градієнт у цій точці рівний нулеві (точка локального мінімуму).

Якщо для мінімізації функції помилок використовується метод градієнтного спуску, то для обчислення семантичних ваг користуються такою залежністю:

$$W(t + 1) = W(t) - \eta V, \quad (3.167)$$

де $W(t)$ – вектор вагових коефіцієнтів для епохи t ; $V = \frac{\partial E(t)}{\partial W(t)}$ – градієнт функції помилки навчання для епохи t ; η – коефіцієнт швидкості навчання.

Коефіцієнт швидкості навчання – це параметр, який керує величиною коригування ваг на кожній ітерації і змінюється в діапазоні $[0; 1]$. Вибір значення η є проблемною задачею:

- великі значення η відповідають великому значенню кроку корекції; у такому разі для знаходження мінімуму функції помилки буде потрібно менше ітерацій, але потенційно збільшується помилка навчання;
- малі значення η відповідають меншому кроку корекції семантичних ваг, а тому збільшується точність налаштування алгоритму мінімізації функції помилки, але також збільшується кількість епох навчання.

У процесі навчання НМ можливе явище *перенавчання*, коли навчена модель добре розпізнає приклади з навчальної множини («зазубрила» пред'явлені їй спостереження), однак не розпізнає або погано розпізнає інші приклади, які не були використані у процесі навчання. Перенавчання є результатом надмірного підлаштування параметрів моделі до залежностей, які є в навчальній множині, тобто підлаштування під особливості окремих прикладів, які можуть містити аналогічні значення та помилки.

Одним із способів запобігання перенавчанню НМ є паралельне спостереження за зміною помилки мережі на навчальній і тестовій множинах. За деяку кількість кроків навчання помилки зменшуються на обох множинах. Проте на якомусь кроці помилка на тестовій множині починає збільшуватися, а на навчальній множині продовжує зменшуватися.

Уся інформація, яку НМ використовує для розв'язання поставленої задачі, зосереджена в навчальній множині (в навчальних прикладах). Тому якість навчання НМ насамперед залежить від кількості прикладів у навчальній множині, а також від того, наскільки добре ці приклади характеризують задачу.

Для зупинки процесу навчання використовують такі критерії:

- виконання заданої кількості ітерацій;
- досягнення функцією якості навчання встановленого мінімального значення;
- стабілізація функції якості навчання;
- функція якості навчання на тестовій вибірці починає зростати, а на навчальній – спадати.

Життєвий цикл НМ охоплює такі основні етапи:

Етап 1. Постановка задачі та вибір типу НМ:

- формулювання задачі;
- обґрунтування застосування нейромережевого підходу.

Етап 2. Формування інформаційної бази:

- визначення змінних моделі НМ;
- формування вибіркової множини даних;
- попередня обробка вибіркової множини даних;
- розділення вибіркової множини даних на навчальну і тестову.

Етап 3. Визначення структури НМ:

- визначення кількості прихованих шарів;
- визначення кількості нейронів у кожному шарі.

Етап 4. Налаштування параметрів НМ та алгоритму навчання:

- вибір активаційних функцій;
- вибір способу навчання НМ;
- задавання умов завершення навчання.

Етап 5. Додаткове навчання НМ (вхідні і вихідні дані, отримані в процесі експлуатації, можуть бути використані для подальшого перерахунку семантичних ваг).

НМ достатньо проста у використанні. Користувачеві необхідно мати лише мінімальний рівень знань про те, як вибирати архітектуру мережі, формувати тестову і навчальну множини, інтерпретувати результати процесу моделювання.

У наш час розроблено багато програмних пакетів, які реалізують нейромережеві технології. Частина з них до певної міри є універсальними, інші – вузькоспеціалізовані.

Основні положення та вказані моделі використовувалися при побудові:

- моделей динаміки фінансових ресурсів, які дозволяють описати процеси еволюції власного капіталу банку залежно від динаміки залучених ресурсів і реалізованої політики нагромадження. Запропоновано підхід до адаптації традиційних моделей дослідження виробничих підприємств і організацій до аналізу діяльності банківських структур.

Благу́н І. С., Буртняк І. В. Моделі аналізу та оцінки діяльності банків в умовах нестабільного економічного середовища. Вісник Хмельницького національного університету. Серія: Економічні науки. 2020. № 4. С. 41–45.

- моделей взаємозв'язку між нерівністю доходів і економічним зростанням в контексті відносного споживання економічними агентами. Розроблено автором просту модель ендогенного зростання, в якій агенти мають різні початкові рівні доходів і діють у відповідності до каскадних

переваг, тобто порівнюють свій власний рівень споживання з рівнем споживання найближчих за величиною початкового доходу економічних агентів.

Благун І. С., Дмитришин Л. І. Моделювання взаємозв'язку «нерівність доходів–економічне зростання» за допомогою теорії каскадних переваг. Проблеми економіки. 2012. № 4. С. 216–221.

- моделей регресії для оцінки кореляції підприємств усіх секцій (розділів) економічної діяльності, які здійснюють експортні операції. На основі кореляційно-регресійного аналізу досліджено існування взаємозв'язку між рівнем інтернаціоналізації та фінансовими результатами діяльності підприємств. Побудовано моделі множинної регресії для підприємств різних секцій (розділів) економічної діяльності, які описують вплив рівня інтернаціоналізації на фінансові результати, що дозволило зробити висновок про існування взаємозв'язку рівня інтернаціоналізації та фінансових результатів діяльності підприємств.

Благун І. С., Ільчук П. Г. Моделі взаємозв'язку рівня інтернаціоналізації та фінансових результатів діяльності українських підприємств. Проблеми економіки. 2015. № 1. С. 259–264.

- DEA-модель для оцінки якості управління комерційними банками. Це дозволило уникнути проблему гетероскедастичності, яка виникає при параметричному моделюванні. Удосконалено алгоритм отримання інформації із спостережень, що дозволило побудувати Парето-оптимальну межу всіх рішень, враховуючи кожне окреме спостереження. Розроблено автором модель вимірювання ефективності діяльності комерційних банків, яка полягає у знаходженні мінімального значення коефіцієнта ефективності, що дозволяє зменшити вхідні дані, але так, щоб рівень вихідних даних залишався на тій самій позиції.

Благуи I. С. Оцінка ефективності комерційних банків за допомогою методу DEA. Бізнес Інформ. 2020. № 11. С. 192–197.

- чутливості показників ефективності інвестиційних проєктів, що дозволяє визначити критичні фактори впливу на реалізацію проєкту та його ефективність. Розроблено автором коефіцієнт чутливості на основі критичних змінних для кожної пояснювальної змінної моделі. Визначено поняття екстремальної змінної в процесі реалізації методу оцінки ефективності проєкту. Таким чином, на основі використання запропонованого підходу вдалося виявити найбільш значущі показники стійкості інвестиційного проєкту та сформувати базу для використання більш складних методів оцінки ризику та розробки показників ефективності інвестиційних проєктів.

Blahun I. S., Leshuk H. V. Sensitivity analysis of indicators of efficiency of investment projects in the economic space of a region. The European Journal of Economics and Management. 2017. Vol. 3. P. 39–45.

РОЗДІЛ 4. ПРОГНОЗУВАННЯ СОЦІАЛЬНО-ЕКОНОМІЧНИХ СИСТЕМ

4.1. Експертні методи прогнозування

Важливу роль у передбаченні тенденцій і параметрів економічних явищ і процесів відграють методи експертних оцінок — методи, які поєднують інтуїтивно-аналітичний аналіз проблеми висококваліфікованими фахівцями з кількісним формальним опрацюванням експертних висновків. Найчастіше методи експертних оцінок використовують у таких випадках:

- неможливості побудови формалізованої моделі об'єкта прогнозування;
- відсутності або неповноти інформації про об'єкт прогнозування;
- наявності альтернативних варіантів розвитку об'єкта прогнозування.

Основна сфера застосування методів експертних оцінок — передбачення якісно нових процесів та їх характеристик, нетипових для економічної практики ситуацій. Глибоке розуміння експертами суті та актуальності досліджуваної проблеми, а також їх практичний досвід дають змогу сформулювати множину елементів дослідження (події, показники і критерії оцінювання, альтернативні варіанти розвитку об'єкта прогнозування тощо) та оцінити характеристики об'єкта прогнозування (ймовірність і час настання окремих подій, значення прогнозованих показників, корисність різних варіантів рішень тощо).

Процес прогнозування на основі експертних методів розглядають як сукупність впорядкованих етапів, що охоплюють підготовку, опрацювання й верифікацію нової інформації (рис. 4.1).

Документ, яким керуються в процесі проведення експертизи, має містити такі положення:

- мету розроблення прогнозу та обґрунтування методу експертного оцінювання;

- перелік та терміни виконання окремих етапів і завдань експертного оцінювання;
- джерела фінансового і матеріального забезпечення;
- склад, завдання, обов'язки і права учасників організаційної та експертної групи.

•



Рис. 4.1. Послідовність процесу прогнозування на основі експертних методів

Перед проведенням експертизи для вирішення поставленої задачі необхідно сформувати колектив експертів – експертну групу. Кількісний склад експертної групи залежить від складності проблеми, яку необхідно розв'язати. Очевидно, що кількість експертів повинна бути достатньою для врахування суттєвих аспектів проблеми та її вирішення з певною точністю. Разом з тим, потрібно зважити на те, що у разі надто великої експертної групи експертні оцінки можуть виявитися неузгодженими (як правило, через розбіжність оцінок окремих експертів), а також з'явитися додаткові труднощі, пов'язані з організуванням та проведенням експертизи. Тому

вирішення проблеми вибору кількісного складу експертної групи перш за все залежить від складності поставленої задачі прогнозування і спроможності експертів адекватно оцінювати об'єкт прогнозування (компетентності експертів).

Після визначення кількісного складу експертної групи, приступають до підбору експертів. Існують різні методики персонального формування експертної групи. Зокрема, список кандидатів можна визначити за методом «снігової кулі». Спочатку складають список із декількох фахівців з проблематики, що підлягає дослідженню. Кожному з фахівців, зазначених у списку, пропонують доповнити список іншими кваліфікованими, на їхню думку, фахівцями. Процес поповнення списку відбувається декілька разів (як правило, поки з'являються нові імена).

Основна вимога, яка ставиться до формування експертної групи, — забезпечити ефективність вирішення проблеми, тобто отримати прогноз з високим рівнем достовірності та обмежених витратах на проведення експертного оцінювання. Достовірність експертного оцінювання обумовлюється складом і компетентністю експертної групи. Високий рівень компетентності — основна умова залучення експерта до участі у розробці прогнозу. Найбільшого розповсюдження отримали такі способи оцінювання компетентності експертів:

- самооцінка — оцінювання експертом рівня власної компетенції за певною шкалою;
- тестування — оцінювання компетентності за результатами відповідей на запитання спеціально розробленої тест-анкети;
- ретроспективне оцінювання за результатами попередньої участі в експертизах;
- групове оцінювання кожним експертом компетентності інших експертів, які увійшли до експертної групи;

• зовнішнє оцінювання компетентності учасників експертної групи іншими експертами.

Оцінювання компетентності експерта іншими фахівцями полягає у тому, що ряду фахівців пропонують визначитися щодо доцільності залучення пропонованих осіб до експертної групи. Якщо оцінювання компетентності n експертів, які відібрані для включення до експертної групи, проводять m фахівців, то можна сформувати матрицю рекомендацій X розмірності $(m \times n)$, елементи якої x_{li} визначають за формулою:

$$x_{li} = \begin{cases} 1, & \text{якщо } l - \text{й фахівець рекомендує } i - \text{го експерта,} \\ 0, & \text{якщо } l - \text{й фахівець не рекомендує } i - \text{го експерта.} \end{cases}$$

З урахуванням матриці рекомендацій коефіцієнт компетентності (K_i) i -го експерта розраховують за формулою:

$$K_i = \frac{\sum_{l=1}^m x_{li}}{\sum_{l=1}^m \sum_{i=1}^n x_{li}}, \quad i = \overline{1, n}, \quad (4.1)$$

До основних завдань, які ставлять перед сформованою експертною групою, відносять:

- *кількісне оцінювання* майбутніх параметрів економічної системи;
- *попарне порівняння* прогнозних оцінок для виявлення кращої з двох наявних оцінок;
- *ранжування* – впорядкування передбачень за їхньою пріоритетністю;
- *шкалювання* – визначення важливості прогнозних альтернатив із зазначенням рівня переваги однієї альтернативи над іншою.

Судження експерта або експертної групи стосовно поставленої задачі називають *експертною оцінкою*. У першому випадку використовують термін «індивідуальні експертні оцінки», а в другому – «колективні експертні оцінки».

Залежно від поставленої мети прогнозування після опитування експертів виникають задачі:

побудова узагальненої оцінки на основі індивідуальних прогнозних значень, вказаних кожним експертом;

визначення вагових коефіцієнтів параметрів об'єкта та оцінювання ймовірностей перебування параметрів у певних межах;

побудова узагальненої оцінки на основі парного порівняння об'єктів кожним з експертів;

побудова узагальненої рейтингової оцінки;

визначення рівня узгодженості висновків експертів;

оцінювання надійності результатів експертного оцінювання.

У разі вимірювання параметрів об'єкта прогнозування в кількісній шкалі побудова узагальненої експертної оцінки зводиться до визначення описових статистик (середнього арифметичного, моди, медіани тощо). Якщо оцінки параметрів виражені у порядковій шкалі (ранжування, попарне порівняння), то опрацювання індивідуальних експертних оцінок має на меті впорядкування елементів певної множини (об'єктів, параметрів тощо).

Результати оцінювання, які представлені кожним із експертів і виражені в кількісній шкалі вимірювання, слід розглядати як реалізацію деякої випадкової величини з множини допустимих оцінок і застосовувати методи математичної статистики для їхнього узагальнення. Це забезпечує кількісне представлення узагальненого висновку, визначення рівня узгодженості думок експертів та статистичної значущості результатів експертизи.

Якщо експертизу проводять в один тур за допомогою n незалежних експертів, то за узагальнену прогнозну оцінку приймають зважене середнє арифметичне значення (\bar{a}) індивідуальних експертних оцінок (з урахуванням компетентності експертів):

$$\bar{a} = \frac{\sum_{i=1}^n K_i a_i}{\sum_{i=1}^n K_i}, \quad (4.2)$$

де a_i – оцінка i -го експерта; K_i – коефіцієнт компетентності i -го експерта.

У випадку, коли вагові коефіцієнти K_i невідомі (або вважають, що усі експерти однаково компетентні), їм привласнюють значення $K_i = 1$.

Дисперсія (σ^2) прогнозних оцінок (a_i):

$$\sigma^2 = \frac{\sum_{i=1}^n (a_i - \bar{a})^2 K_i}{\sum_{i=1}^n K_i} \quad (4.3)$$

слугує мірою узгодженості індивідуальних оцінок експертів. Більші значення дисперсії свідчать про більші відхилення індивідуальних оцінок від середнього значення, тобто про більшу розбіжність думок експертів.

Вважаючи, що оцінки експертів розподілені нормально з математичним сподіванням (\bar{a}) і дисперсією (σ^2), можемо визначити довірчий інтервал узагальненої оцінки (\tilde{a}):

$$\bar{a} - t^{\text{крит}} \mu_a \leq \tilde{a} \leq \bar{a} + t^{\text{крит}} \mu_a, \quad (4.4)$$

де значення $t^{\text{крит}}$ вибирається з таблиці критичних значень розподілу Стьюдента з урахуванням рівня значущості α і числа ступенів вільності $V = n - 1$, а стандартну помилку μ_a обчислюють залежно від кількості експертів за формулою:

$$\mu_a = \begin{cases} \sqrt{\frac{\sigma^2}{n}}, & \text{якщо } n > 30, \\ \sqrt{\frac{\sigma^2}{n-1}}, & \text{якщо } n \leq 30. \end{cases}$$

З метою підвищення точності прогнозу експертам пропонують наводити три можливі оцінки об'єкта прогнозування: оптимістичну — $a^{(1)}$; правдоподібну (поміркувану) — $a^{(2)}$ та песимістичну — $a^{(3)}$.

Тоді узагальнена прогнозна оцінка \bar{a} буде такою:

$$\bar{a} = \frac{\sum_{i=1}^n K_i \bar{a}_i}{\sum_{i=1}^n K_i}, \quad (4.5)$$

де \bar{a}_i — середня прогнозна оцінка i -го експерта.

Міру узгодженості думок експертів (σ^2) визначають за формулою:

$$\sigma^2 = \frac{\sum_{i=1}^n K_i \sigma_i^2}{\sum_{i=1}^n K_i} + \frac{\sum_{i=1}^n K_i (\bar{a}_i - \bar{a})^2}{\sum_{i=1}^n K_i}, \quad (4.6)$$

де σ_i^2 — дисперсія прогновної оцінки i -го експерта.

У формулах (4.5) і (4.6) середню оцінку i -го експерта (\bar{a}_i) і дисперсію (σ_i^2) визначають за формулами:

$$\bar{a}_i = \frac{w^{(1)}a_i^{(1)} + w^{(2)}a_i^{(2)} + w^{(3)}a_i^{(3)}}{w^{(1)} + w^{(2)} + w^{(3)}}; \quad (4.7)$$

$$\sigma_i^2 = \frac{a_i^{(3)} - a_i^{(1)}}{w^{(4)}}, \quad (4.8)$$

де $w^{(1)}$, $w^{(2)}$, $w^{(3)}$ – вагові коефіцієнти оптимістичної, правдоподібної (реалістичної) та песимістичної оцінок i -го експерта, відповідно; $w^{(4)}$ – ступінь невпевненості експерта.

Значення вагових коефіцієнтів $w^{(1)}$, $w^{(2)}$, $w^{(3)}$ та $w^{(4)}$ можна встановити за однією із двох методик:

$$\text{М1: } w^{(1)} = 1; w^{(2)} = 1; w^{(3)} = 1; w^{(4)} = 36;$$

$$\text{М2: } w^{(1)} = 3; w^{(2)} = 0; w^{(3)} = 2; w^{(4)} = 25.$$

Важливими завданнями, що постають перед експертами в процесі передбачення розвитку складних об'єктів, є встановлення вагових коефіцієнтів параметрів та оцінювання ймовірностей перебування значень цих параметрів у певних межах.

Припустимо, що n експертів оцінюють значення m параметрів (показників) деякого об'єкта прогнозування. Нехай a_{ij} — оцінка j -го параметра i -м експертом. Тоді узагальнена оцінка j -го параметра (\bar{a}_j) становитиме:

$$\bar{a}_j = \frac{\sum_{i=1}^n K_i a_{ij}}{\sum_{i=1}^n K_i}, \quad j = \overline{1, m}. \quad (4.9)$$

За значеннями узагальнених оцінок окремих параметрів (\bar{a}_j) можна обчислити їхні вагові коефіцієнти у структурі об'єкта (k_j):

$$k_j = \frac{\bar{a}_j}{\sum_{j=1}^m \bar{a}_j}, \quad j = \overline{1, m}. \quad (4.10)$$

Перед експертами може постати завдання оцінювання ймовірності потрапляння досліджуваного параметра в деякий інтервал значень. Для вирішення цього завдання організатори експертизи розбивають інтервал

можливих значень оцінюваного параметра на l проміжків. Після цього кожному експерту пропонують оцінити ймовірність потрапляння значення параметра в кожний із проміжків, у результаті чого формують матрицю ймовірностей P :

$$P = \begin{bmatrix} p_{11} & p_{12} & \dots & p_{1l} \\ p_{21} & p_{22} & \dots & p_{2l} \\ \dots & \dots & \dots & \dots \\ p_{n1} & p_{n2} & \dots & p_{nl} \end{bmatrix},$$

де p_{ij} — оцінка i -го експерта ймовірності потрапляння значення параметра в j -й проміжок.

Узагальнену оцінку потрапляння значення параметра в j -й проміжок (P_j) з урахуванням компетентності експертів (K_i , причому $\sum_{i=1}^n K_i = 1$) розраховують за такою формулою:

$$P_j = \sum_{i=1}^n K_i p_{ij}, \quad j = \overline{1, l}. \quad (4.11)$$

Очевидно, що $\sum_{j=1}^l P_j = 1$. На підставі значень P_j будують ряд розподілу, а за узагальнену оцінку приймають математичне сподівання, моду, медіану та інші оцінки ймовірностей.

Важливим завданням економічного прогнозування, яке ефективно вирішують експертними методами, є ранжування (упорядкування) об'єктів (показників, альтернатив, чинників впливу та ін.) за їхньою пріоритетністю.

Позначимо через r_{ij} — ранг, присвоєний i -м експертом j -му об'єкту. Результати опитування експертів зручно представляти в табличній формі (табл. 4.1). В i -му рядку таблиці вказують ранги, які надав i -й експерт об'єктам ранжування, — числа натурального ряду (кількість рангів відповідає кількості об'єктів ранжування). Надалі будемо вважати, що найменш пріоритетному об'єкту присвоюють найбільший ранг. У разі виконання припущення про існування рівноцінних об'єктів наявне *нестроге ранжування*, у протилежному випадку — *строге ранжування*. У випадку нестрогого ранжування (k об'єктів обійняли ранги r_1, r_2, \dots, r_k), здійснюють стандартизацію рангів — присвоєння рівноцінним об'єктам

рангу, який дорівнює середньому арифметичному рангів, поділених між ними.

Таблиця 4.1

Форма представлення результатів ранжування
прогнозних альтернатив

Експерти	Ранги об'єктів прогнозування			
	O_1	O_2	...	O_m
1	r_{11}	r_{12}	...	r_{1m}
2	r_{21}	r_{22}	...	r_{2m}
...
n	r_{n1}	r_{n2}	...	r_{nm}
Сума рангів, $R_j = \sum_{i=1}^n r_{ij}$	R_1	R_2	...	R_m

В останньому $(n + 1)$ -му рядку табл. 4.1 записують суму рангів j -го об'єкта:

$$R_j = \sum_{i=1}^n r_{ij}, \quad j = \overline{1, m}. \quad (4.12)$$

Оскільки кількість рангів дорівнює кількості об'єктів ранжування (m) , то виконується умова:

$$\sum_{j=1}^m R_j = m(m + 1)/2.$$

Обчислення сумарного рангу j -го об'єкта за формулою (4.12) передбачає однакову компетентність експертів. Якщо ж компетентність експертів є різною, то ця формула потребує незначного уточнення. Нехай компетентність експертів оцінено значеннями K_i ($i = \overline{1, n}$), причому $\sum_{i=1}^n K_i = 1$. Тоді суму рангів (R_j) j -го об'єкта розраховують так:

$$R_j = \sum_{i=1}^n K_i r_{ij}, \quad j = \overline{1, m}. \quad (4.13)$$

Найпростіший спосіб отримання узагальненої оцінки ранжування – це впорядкування об’єктів за ланцюгом нерівностей:

$$R_k < R_l < \dots < R_q,$$

де $R_k = \min_j R_j$; $R_l = \min_j R_j$ ($j \neq l$); $R_q = \max_j R_j$.

У результаті отримуємо таку послідовність об’єктів за пріоритетами:

$$O_k > O_l > \dots > O_q.$$

Розміщення об’єктів у порядку зростання величин R_j приймають як остаточне ранжування, в якому пріоритетною вважають прогнозу альтернативу з мінімальною сумою рангів.

Мірою узгодженості ранжування, виконаного експертами, є коефіцієнт конкордації, який у випадку строгого ранжування розраховують за формулою:

$$W = \frac{12 \sum_{j=1}^m (R_j - \bar{R})^2}{n^2(m^3 - m)}, \quad (4.14)$$

де $\bar{R} = n(m + 1)/2$ – середня сума рангів; n , m – кількість експертів та об’єктів оцінювання відповідно.

Коефіцієнт конкордації W знаходиться в межах $0 \leq W \leq 1$ (вищому рівню узгодженості відповідає більше значення W). Для перевірки істотності коефіцієнта конкордації використовують критерій χ^2 з $(m - 1)$ ступенями вільності. Розрахункове значення критеріальної статистики $\chi_{\text{розр}}^2 = Wn(m - 1)$ порівнюють з табличним значенням для вибраного рівня значущості α . Якщо розрахункове значення перевищує критичне (табличне), то узгодженість думок експертів підтверджується.

У випадку нестроного ранжування коефіцієнт конкордації обчислюють за формулою:

$$W = \frac{12 \sum_{j=1}^m (R_j - \bar{R})^2}{n^2(m^3 - m) - n \sum_{i=1}^n \sum_{j=1}^{m_i} (t_{ij}^3 - t_{ij})}, \quad (4.15)$$

де m_i – кількість груп рівних рангів, визначених i -м експертом; t_{ij} – кількість усереднених рангів у j -й групі рівних рангів, визначених i -м експертом.

Якщо кількість експертів дорівнює $n = 2$, а кількість об'єктів порівняння — m , то статистичний аналіз узгодженості думок експертів проводять за коефіцієнтом рангової кореляції Спірмена (ρ), який має вигляд:

$$\rho = 1 - \frac{6 \sum_{i=1}^m d_i^2}{m^3 - m}, \quad \rho \in [-1; +1], \quad (4.16)$$

де d_i — різниця рангів, виставлених експертами i -му об'єкту.

Знак коефіцієнта рангової кореляції Спірмена вказує на напрям зв'язку («плюс» — прямий, «мінус» — зворотний). Для оцінювання статистичної значущості цього коефіцієнта розраховане значення ρ порівнюють з критичним (вибирають з таблиці критичних значень коефіцієнта рангової кореляції Спірмена). Якщо $\rho > \rho_{\text{крит}}$, то коефіцієнт є статистично значущим.

У практиці експертного оцінювання часто використовують бальні оцінки — умовні градації альтернатив порівняння. У такому разі необхідно дотримуватися певних правил градації (шкали вимірювання, критеріїв присвоєння балів тощо). Якщо результати експертного оцінювання представлено у вигляді бальних оцінок, то задача ранжування об'єктів передбачає переведення бальних оцінок у рейтингові.

Для ранжування прогнозних альтернатив можна використати *метод нормування*, який ґрунтується на тому, що кожний експерт розподіляє наперед визначену суму балів A між альтернативами:

$$\sum_{j=1}^m a_{ij} = A, \quad i = \overline{1, n},$$

де a_{ij} — оцінка в балах j -ї альтернативи, заданої i -м експертом.

Зведена оцінка j -ї прогнозної альтернативи рівна:

$$A_j = \sum_{i=1}^n a_{ij}, \quad j = \overline{1, m}. \quad (4.17)$$

Тоді на основі її значень можна розрахувати відносні вагові коефіцієнти альтернатив:

$$\alpha_j = \frac{A_j}{\sum_{j=1}^m A_j} = \frac{A_j}{nA}, \quad j = \overline{1, m}. \quad (4.18)$$

Ранжування прогнозних альтернатив проводять за зменшенням значення відносного вагового коефіцієнта α_i — більшому значенню коефіцієнта відповідає вищий ранг.

Ранжування на підставі одночасного розгляду великої кількості об'єктів часто викликає певні труднощі, що впливає на якість прогнозу. У таких випадках ранжування можна проводити за допомогою методу *парних порівнянь*.

Кожний експерт порівнює кожен об'єкт з іншими і заповнює табл. 4.2.

Таблиця 4.2

Таблиця парних порівнянь

Об'єкти (параметри)	1	2	...	k	...	n
1	–					
2	–	–				
...	–	–	–			
k	–	–	–	–		
...	–	–	–	–	–	
n	–	–	–	–	–	–

Експерти заповнюють лише клітинки, розташовані праворуч від діагоналі (клітинки $(i, j), i < j$). У клітинку (i, j) записують номер об'єкта (i або j), якому експерт віддає перевагу або знак «~», якщо об'єкти є рівноцінними.

На підставі заповнених таблиць аналітик формує матриці індивідуальних переваг $A^{(i)}$ розміром $t * t$, де t – кількість об'єктів порівнянь. За *строгого ранжування* елементи матриці переваг рівні:

$$a_{rs}^{(i)} =$$

$\begin{cases} 1, & \text{якщо } i - \text{й експерт віддає перевагу } r - \text{му об'єкту перед } s - \text{м;} \\ 0 & - \text{у протилежному випадку.} \end{cases}$

Якщо має місце *нестроге ранжування*, то:

$$a_{rs}^{(i)} =$$

$\begin{cases} 1, & \text{якщо } i - \text{й експерт віддає перевагу } r - \text{му об'єкту перед } s - \text{м;} \\ 0,5, & \text{якщо } i - \text{й експерт вважає об'єкти рівноцінними;} \\ 0, & \text{якщо } i - \text{й експерт віддає перевагу } s - \text{му об'єкту перед } r - \text{м.} \end{cases}$

На основі матриць $A^{(i)}$ формують матрицю сумарних переваг A :

$$A = \sum_{i=1}^n A^{(i)}, \quad (4.19)$$

або $A = \sum_{i=1}^n K_i A^{(i)}$, якщо задані коефіцієнти компетентності K_i експертів.

За результат упорядкування приймають розміщення об'єктів у порядку зростання сум елементів рядків матриці A , тобто елементів a_r :

$$a_r = \sum_{s=1}^m a_{rs}, \quad r = \overline{1, m}. \quad (4.20)$$

Інформаційною базою для формалізованих методів прогнозування слугують *динамічні ряди (ряди динаміки)* – послідовно розташовані у хронологічному порядку статистичні дані, які відображають розвиток процесу (зміни явища) у часі. Оскільки ознакою, за якою впорядковують значення економічних показників, є час, то динамічні ряди також називають *часовими*.

Часові ряди характеризують двома ознаками:

- періодами (проміжками) або моментами часу (на початок або на кінець періоду) (t);
- рівнями ряду — значеннями економічних показників за певний період часу або на певний момент часу (y_t).

Рівні часового ряду можна представити абсолютними, відносними або середніми величинами.

За кількістю показників, для яких визначають рівні, часові ряди поділяють на *одновимірні* та *багатовимірні (векторні)*.

За способом формування розрізняють такі види часових рядів:

інтервальні — сукупність показників, які характеризують розвиток об'єкта дослідження за певний період часу (наприклад, обсяг виробленої продукції за рік);

моментні — сукупність показників, які характеризують стан об'єкта дослідження на певну дату (наприклад, оборотні активи підприємства на початок та кінець року).

Для кількісного аналізування динаміки економічних процесів найчастіше використовують такі аналітичні показники: *абсолютний приріст*, що характеризує абсолютну швидкість зростання або спадання рівнів ряду; *коефіцієнт зростання (або темп зростання)*, який характеризує інтенсивність зміни рівнів ряду; *темп приросту*, що характеризує відносну величину приросту рівнів ряду.

Кожний із зазначених показників може бути *базисним, ланцюговим* або *середнім*.

Показник називають *базисним*, якщо його порівнюють з фіксованим рівнем ряду, який прийнято за базу порівняння. Як правило, за базу порівняння вибирають початковий рівень ряду (y_1) або рівень, з якого починається новий етап розвитку процесу (y_6). Якщо кожний наступний рівень ряду порівнюють із попереднім (порівняння зі змінною базою), то отримані показники називають *ланцюговими*.

Формули для розрахунку основних показників динаміки представлено в табл. 4.3.

Показники динаміки часового ряду

Показники	Абсолютний приріст	Коефіцієнт зростання	Темп зростання, %	Темп приросту, %
Базисний	$\Delta y_t^{\bar{b}} = y_t - y_0$	$k_t^{\bar{b}} = \frac{y_t}{y_0}$	$T_{зр\ t}^{\bar{b}} = k_t^{\bar{b}} \cdot 100$	$T_{пр\ t}^{\bar{b}} = T_{зр\ t}^{\bar{b}} - 100$
Ланцюговий	$\Delta y_t^{\bar{л}} = y_t - y_{t-1}$	$k_t^{\bar{л}} = \frac{y_t}{y_{t-1}}$	$T_{зр\ t}^{\bar{л}} = k_t^{\bar{л}} \cdot 100$	$T_{пр\ t}^{\bar{л}} = T_{зр\ t}^{\bar{л}} - 100$

Середній абсолютний приріст ($\Delta \bar{y}$), середній коефіцієнт зростання (\bar{k}), середній темп зростання ($\bar{T}_{зр}$) і середній темп приросту ($\bar{T}_{пр}$) обчислюють, відповідно, за такими формулами:

$$\Delta \bar{y} = \frac{y_n - y_1}{n-1} = \frac{1}{n-1} \sum_{t=1}^{n-1} \Delta y_t^{\bar{л}}, \quad (4.21)$$

$$\bar{k} = \sqrt[n-1]{\frac{y_n}{y_1}} = \sqrt[n-1]{k_2^{\bar{л}} \cdot k_3^{\bar{л}} \cdots k_n^{\bar{л}}}, \quad \bar{T}_{зр} = \bar{y} \cdot 100\%, \quad (4.22)$$

$$\bar{T}_{пр} = \bar{T}_{зр} - 100\%. \quad (4.23)$$

Середній абсолютний приріст і середній коефіцієнт зростання (або середній темп зростання) як узагальнюючі характеристики, відповідно, швидкості та інтенсивності зміни рівнів часового ряду можна використовувати для побудови найпростіших прогнозів.

У випадку, коли тенденція розвитку процесу є близькою до лінійної, тобто значення ланцюгових абсолютних приростів є приблизно однаковими, то за точковий прогноз можна прийняти:

$$y_{n+\tau}^{\text{прогн}} = y_n + \tau \cdot \Delta \bar{y}, \quad (4.24)$$

де τ – період випередження.

Якщо рівні часового ряду наближено відповідають характеру зміни показникової або експоненційної функції, то прогнозне значення можна отримати за формулою:

$$y_{n+\tau}^{\text{прогн}} = \bar{k}y_n. \quad (4.25)$$

Основний недолік прогнозування на основі середніх показників часового ряду — врахування лише початкового та кінцевого рівнів. Проте такі наближені методи є достатньо поширеними в практиці прогнозування та слугують базою для більш глибокого кількісного і якісного аналізу з метою побудови складніших моделей прогнозування.

Важливу роль у процесі дослідження тенденції часових рядів відіграють оцінки середнього арифметичного значення, дисперсії та коваріації. Для інтервального ряду з рівновіддаленими у часі рівнями їх обчислюють за такими формулами:

$$\bar{y} = \frac{1}{n} \sum_{t=1}^n y_t - \text{середнє арифметичне значення}; \quad (4.26)$$

$$\hat{\sigma}^2 = \frac{1}{n-1} \sum_{t=1}^n (y_t - \bar{y})^2 - \text{дисперсія}; \quad (4.27)$$

$$\text{cov}(y_t; y_{t+k}) = \frac{1}{n-1} \sum_{t=k+1}^n (y_t - \bar{y})(y_{t+k} - \bar{y}) - \text{коваріація } k\text{-го порядку}. \quad (4.28)$$

Для отримання якісного прогнозу необхідна повна і достовірна інформація про рівні часового ряду. Часовий ряд об'єктивно відображає тенденцію розвитку економічного процесу лише за умови можливості зіставлення (порівняння) його рівнів. Основними вимогами щодо зіставлення часових рядів вважають такі:

- рівність періодів (моментів часу), яким відповідають рівні ряду;
- однакова повнота охоплення досліджуваних частин процесу (явища);
- збіг просторових меж процесу (явища);
- спільномірність рівнів ряду (однаковий масштаб виміру);
- єдине тлумачення одиниці об'єкта спостереження.

Рівні часового ряду можуть складатися зі значень, які сформувалися під впливом характерних для досліджуваного процесу чинників, а також отриманих під впливом інших чинників, які не є типовими для цього процесу. Такі значення різко виділяються і, як наслідок, використання методології статистичного аналізу, яка не враховує аномальні значення, може призвести до істотних помилок прогнозування. Через це рівні, які суттєво виділяються на фоні інших, потребують додаткового вивчення.

Для виявлення аномальних рівнів найчастіше використовують метод Ірвіна, що ґрунтується на такій λ -статистиці:

$$\lambda_t = \frac{|y_t - y_{t-1}|}{\sigma_y}, \quad t = \overline{2, n}, \quad (4.29)$$

де σ_y – середнє квадратичне відхилення рівнів часового ряду.

Розраховані значення статистики λ_t порівнюють з критичними $\lambda_{\text{крит}}$ (табл. 4.4). Якщо $\lambda_t > \lambda_{\text{крит}}$, то відповідний рівень часового ряду y_t вважають аномальним.

Таблиця 4.4

Критичні значення критерію Ірвіна ($\alpha = 0,05$)

n	10	20	30	50	100
$\lambda_{\text{крит}}$	1,5	1,3	1,2	1,1	1,0

Після виявлення аномальних рівнів часового ряду аналізують причини їхнього виникнення. Якщо аномальні значення зумовлені причинами випадкового характеру, то їх переважно вилучають або замінюють апроксимованими значеннями чи середніми арифметичними значеннями двох сусідніх рівнів. Аномальні значення, пов'язані з дією чинників, що мають об'єктивний характер, заміні не підлягають.

Часовий ряд можна інтерпретувати як спостереження над неперервним випадковим процесом у певні моменти часу. Під *випадковим процесом* розуміють функцію $Y(t)$ з невідповідним аргументом t , яка за довільного значення t є випадковою величиною. Послідовність спостережень $y_{t_1}, y_{t_2},$

..., Y_{t_n} за випадковим процесом $Y(t)$ у статистиці називають *реалізацією випадкового процесу* або *часовим рядом*.

Часовий ряд називають *строго стаціонарним* (або *стаціонарним у вузькому розумінні*), якщо спільний розподіл імовірностей n спостережень $Y_{t_1}, Y_{t_2}, \dots, Y_{t_n}$ відповідає розподілу ймовірностей $Y_{t_1+\tau}, Y_{t_2+\tau}, \dots, Y_{t_n+\tau}$.

Фахівця, який розробляє прогноз, як правило, цікавить не спільний розподіл спостережень, а середнє значення, дисперсія і коваріація ряду. Тому на практиці частіше користуються поняттям стаціонарності у широкому розумінні («слабкої» стаціонарності).

Часовий ряд називають *стаціонарним у широкому розумінні*, якщо він характеризується сталими математичним сподіванням, дисперсією та автокореляцією будь-якого порядку в усіх проміжках часу. Якщо ці умови не виконуються, то ряд відносять до *нестационарного*.

Стаціонарні часові ряди представляють процеси, які характеризуються тим, що не зазнають якісних змін, тобто не змінюють свою структуру, силу і напрям взаємозв'язків між елементами.

Окремим випадком стаціонарних часових рядів є «білий шум», під яким розуміють послідовність незалежних між собою однаково розподілених величин із нульовим математичним сподіванням і постійною дисперсією.

У загальному випадку часовий ряд можна розкласти на такі складові:

- *тренд* ($f(t)$) — характеризує основну тенденцію розвитку процесу (явища) і є результатом комплексної дії чинників упродовж тривалого періоду часу;

- *сезонна компонента* (s_t) — характеризує вплив періодичних або близьких до них коливань сезонного характеру (вплив пори року на виробництво та реалізацію певних видів товарів і надання послуг — продаж зимового спортивного інвентарю, споживання безалкогольних напоїв, морозива, банківські послуги, транспортні перевезення тощо);

• *циклічна компонента* (c_t) — характеризує вплив циклів економічного розвитку (наприклад, пов'язаних з демографічними процесами або процесами технічного переозброєння виробництва);

• *випадкова (нерегулярна) компонента* (ε_t) — характеризує дію випадкових чинників.

Трендову, сезонну і циклічну компоненти називають *систематичними (регулярними)* складовими рівнів часового ряду. Вони характеризують вплив постійно діючих (основних) чинників на рівні часового ряду.

Випадкова компонента формується під впливом чинників двох видів: раптової (неочікуваної) та (або) поточної дії. Чинники першого виду (різкі зміни курсу валют, стихійні лиха тощо) зумовлюють значні стрибки рівнів часового ряду. Чинники другого виду – випадкові коливання, які є результатом дії певної кількості побічних причин. Вплив кожної із причин є незначним, проте відчувається їхня сумісна дія.

Необхідно зазначити, що компоненти часового ряду не піддаються спостереженню і є теоретичними величинами.

Основна мета статистичного аналізу часових рядів полягає у встановленні співвідношення між закономірністю та випадковістю в процесі формування значень економічних показників. Закономірність пояснює динаміку показників у минулому і використовується для прогнозування їхніх значень у майбутньому. Дослідження випадковості дає змогу визначити ймовірність і величину відхилення показників від значень, обумовлених закономірністю розвитку.

Часовий ряд представляють різними типами моделей.

Модель часового ряду, яку представлено у вигляді суми тренду, сезонної, циклічної і випадкової компонент, називають *адитивною* моделлю:

$$y_t = f(t) + s_t + c_t + \varepsilon_t, \quad (4.30)$$

Модель часового ряду, яку представлено у вигляді добутку тренду, сезонної, циклічної і випадкової компонент — *це мультиплікативна* модель:

$$y_t = f(t) \cdot s_t \cdot c_t \cdot \varepsilon_t, \quad (4.31)$$

Модель, яка має вигляд:

$$y_t = f(t) \cdot s_t \cdot c_t + \varepsilon_t, \quad (4.32)$$

називають *комбінованою* моделлю часового ряду.

Використані в моделях компоненти не обов'язково притаманні будь-якому часовому ряду. Існують ряди, в яких відсутні як тренд, так сезонні та циклічні коливання. У цьому випадку рівні ряду є функцією випадкової компоненти і коливаються навколо середнього рівня, що характерно для стаціонарних часових рядів.

На основі моделей (4.30)-(4.32) шляхом елімінування (виключення впливу) окремих компонент досліджують інші компоненти, зокрема елімінуючі сезонні, циклічні і випадкові коливання, а також *основну тенденцію ряду* – тренд.

Побудова прогнозів на основі часових рядів вимагає дослідження тенденції (порівняно стійкого напрямку зміни):

- основної тенденції (тренда);
- дисперсії;
- автокореляції.

Тренд, як загальний напрям розвитку процесу (явища), переважно подають у вигляді гладкої кривої (траєкторії), що є функцією від часу. У цьому разі передбачають, що вплив усіх основних чинників на процес дослідження можна виразити через функцію часу. У зв'язку з цим під трендом часто розуміють рівняння регресії $\hat{y} = f(t)$. Зазначимо, що представлення тренду лише одним параметром t не дає змоги повністю описати характер розвитку процесу, і його не можна розглядати як єдине джерело формування закономірності процесу.

Перші припущення щодо характеру зміни тренда можна зробити в результаті візуального аналізу графіка часового ряду.

Якщо графічний аналіз не забезпечує чіткого уявлення про характер часового ряду, то проводять його тестування за статистичними критеріями. У випадку, коли в практичних дослідженнях результати одного тесту не дають чіткі підстави для прийняття або відхилення висунутої гіпотези, виникає необхідність використання інших тестів. Сукупність тестів для перевірки основної тенденції часового ряду поділять на дві основні групи:

- непараметричні — не передбачають наявності будь-яких припущень щодо закону розподілу часового ряду та його параметрів;
- параметричні — ґрунтуються на відносно строгих передумовах щодо закону розподілу часового ряду та його параметрів (параметричні тести застосовують лише у випадку нормально розподілених даних).

Для перевірки гіпотези про наявність тренда в часовому ряді можна використати метод перевірки різниць середніх – параметричний тест на основі t -критерію (критерію Стьюдента). Вигляд математичного виразу для обчислення критеріальної статистики залежить від прийнятого припущення щодо дисперсії часового ряду. Розглянемо послідовність тестування за одним із варіантів методу різниць середніх.

На першому етапі часовий ряд ділять на дві послідовні приблизно однакові за кількістю рівнів частини, кожен з яких розглядають як окрему нормально розподілену вибірку. Якщо розбіжність між середніми двох сукупностей є неістотною (випадковою), то можна стверджувати про відсутність основної тенденції у часовому ряді, в протилежному випадку — про наявність тренду. Тому перевірка наявності основної тенденції зводиться до перевірки гіпотези про рівність середніх двох нормально розподілених сукупностей.

Прийmemo такі позначення: n_1, n_2 — обсяги першої і другої частини сукупності відповідно; \bar{y}_1, \bar{y}_2 — середні значення для першої і другої частини сукупності відповідно.

На другому етапі оцінюють дисперсії першої і другої частини часового ряду (\hat{S}_1^2 і \hat{S}_2^2):

$$\hat{S}_1^2 = \frac{1}{n_1-1} \sum_{t=1}^{n_1} (y_t - \bar{y}_1)^2; \quad \hat{S}_2^2 = \frac{1}{n_2-1} \sum_{t=n_1+1}^{n_2} (y_t - \bar{y}_2)^2.$$

Метод різниць середніх ґрунтується на припущенні про рівність дисперсій рівнів першої і другої частин ($S_1^2 = S_2^2$), тому на третьому етапі перевіряють гіпотезу про рівність дисперсій за F -критерієм (критерієм Фішера). Розрахункове значення критеріальної F -статистики обчислюють за формулою:

$$F^{\text{розр}} = \frac{\max(\hat{S}_1^2; \hat{S}_2^2)}{\min(\hat{S}_1^2; \hat{S}_2^2)}. \quad (4.33)$$

Якщо для вибраного рівня значущості α та V_1 і V_2 ступенів вільності виконується умова:

$$F^{\text{крит}} \left(1 - \frac{\alpha}{2}; V_1; V_2 \right) < F^{\text{розр}} < F^{\text{крит}} \left(\frac{\alpha}{2}; V_1; V_2 \right), \quad (4.34)$$

то гіпотезу про рівність дисперсій ($H_0: S_1^2 = S_2^2$) не відхиляють, у протилежному випадку — приймають альтернативну ($H_1: S_1^2 \neq S_2^2$) гіпотезу. Якщо $\hat{S}_1^2 > \hat{S}_2^2$, то $V_1 = n_1 - 1$ і $V_2 = n_2 - 1$, інакше — $V_1 = n_2 - 1$ і $V_2 = n_1 - 1$.

У випадку прийняття гіпотези про рівність дисперсій приступають до четвертого етапу — перевірки гіпотез про рівність середніх рівнів двох сукупностей ($H_0: \bar{y}_1 = \bar{y}_2; H_1: \bar{y}_1 \neq \bar{y}_2$) за t -критерієм (критерієм Стьюдента). Для цього обчислюють значення t -статистики:

$$t^{\text{розр}} = \frac{|\bar{y}_1 - \bar{y}_2|}{\hat{S}^2 \cdot \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}}. \quad (4.35)$$

У формулі (4.35) \hat{S}^2 — це загальна оцінка дисперсії, обчислена за оцінками дисперсій \hat{S}_1^2 першої і \hat{S}_2^2 другої частин:

$$\hat{S}^2 = \frac{1}{n_1+n_2-2} [(n_1-1)\hat{S}_1^2 + (n_2-1)\hat{S}_2^2]. \quad (4.36)$$

З таблиці розподілу Стьюдента для вибраного рівня значущості α та $V = n_1 + n_2 - 2$ ступенів вільності знаходять критичне значення $t^{\text{крит}}$ і

порівнюють з розрахованою t -статистикою ($t^{\text{розр}}$). Якщо $t^{\text{розр}} < t^{\text{крит}}$, то нульову гіпотезу H_0 про рівність середніх не відхиляють, тобто розбіжність між середніми рівнями вважають незначною, що свідчить про відсутність тренду в часовому ряді.

Якщо гіпотезу про рівність дисперсій відхиляють, то t -критерій (4.35) застосовувати не можна.

Розглянутий вище метод дає надійні результати тоді, коли часовий ряд змінюється монотонно. Якщо ж для ряду властива тенденція зміни загального напрямку розвитку, то перевірка гіпотези може не виявити наявності основної тенденції.

Надійнішим методом перевірки наявності тренда в часовому ряді, порівняно із методом перевірки різниць середніх, є застосування непараметричного *методу Фостера–Стюарта*. Крім перевірки наявності тренда, цей метод дає змогу встановити наявність тенденції дисперсії часового ряду. Відсутність тенденції дисперсії означає, що варіація рівнів часового ряду є постійною.

Реалізацію методу Фостера–Стюарта розпочинають з порівняння кожного рівня часового ряду (починаючи з другого, $t = \overline{2, n}$) з усіма попередніми рівнями. У результаті порівняння формують дві числові послідовності:

$$U_t = \begin{cases} 1, \text{ якщо значення } y_t \text{ більше за всі попередні рівні,} \\ 0 \text{ у протилежному випадку;} \end{cases} \quad (4.37)$$

$$W_t = \begin{cases} 1, \text{ якщо значення } y_t \text{ менше за всі попередні рівні,} \\ 0 \text{ у протилежному випадку.} \end{cases} \quad (4.38)$$

Після цього обчислюють такі значення:

$$m_t = U_t + W_t; \quad (4.39)$$

$$k_t = U_t - W_t; \quad (4.40)$$

$$M = \sum_{t=2}^n m_t; \quad (4.41)$$

$$K = \sum_{t=2}^n k_t. \quad (4.42)$$

Статистика M набуває значень від 0 (ряд монотонно спадає) до $(n - 1)$ (ряд монотонно зростає), і використовується для перевірки наявності тенденції дисперсії.

Статистика K набуває значень від мінус $(n - 1)$ (ряд монотонно спадає) до $(n - 1)$ (ряд монотонно зростає або має місце чергування зростання і зменшення рівнів). Її використовують для перевірки наявності тренду.

Для перевірки наявності основної тенденції та тенденції дисперсії використовують критерій Стьюдента:

$$t_1^{\text{розр}} = \frac{M - \mu}{\sigma_1}; \quad (4.43)$$

$$t_2^{\text{розр}} = \frac{K - o}{\sigma_2}. \quad (4.44)$$

де μ – математичне сподівання M ; σ_1 – середнє квадратичне відхилення M ; σ_2 – середнє квадратичне відхилення K .

Значення μ , σ_1 та σ_2 табульовані для різних обсягів спостережень (табл. 4.5). Величини $t_1^{\text{розр}}$ і $t_2^{\text{розр}}$ порівнюють з критичними значеннями $t^{\text{крит}}$ для вибраного рівня значущості α і $V = n - 1$ ступенів вільності. Якщо розрахункове значення перевищує критичне, то відповідну гіпотезу про відсутність тенденції відхиляють.

Таблиця 4.5

Табульовані значення μ , σ_1 та σ_2

n	μ	σ_1	σ_2
10	3,858	1,288	1,964
15	4,636	1,521	2,158
20	5,195	1,677	2,279
25	5,632	1,791	2,373
30	5,990	1,882	2,447
35	6,294	1,956	2,509
40	6,557	2,019	2,561
45	6,790	2,072	2,606
50	6,998	2,121	2,645

З поміж інших методів і критеріїв тестування тенденції часового ряду виділяють методи Манна–Уїтні та Кокса–Стюарта, метод серій, критерії Кокрена, Бартлетта, Дікі–Фуллера та ін.

Кореляційну залежність між послідовними рівнями часового ряду називають *автокореляцією*.

Кількісною мірою щільності лінійного зв'язку (залежності) між послідовними рівнями часового ряду слугує *коефіцієнт автокореляції* $r(\tau)$ з часовим лагом τ (*коефіцієнт автокореляції порядку* τ), який розраховують за формулою:

$$r(\tau) = \frac{(n-\tau) \sum_{t=1}^{n-\tau} y_t y_{t+\tau} - \sum_{t=1}^{n-\tau} y_t \sum_{t=1}^{n-\tau} y_{t+\tau}}{\sqrt{(n-\tau) \sum_{t=1}^{n-\tau} y_t^2 - (\sum_{t=1}^{n-\tau} y_t)^2} \cdot \sqrt{(n-\tau) \sum_{t=1}^{n-\tau} y_{t+\tau}^2 - (\sum_{t=1}^{n-\tau} y_{t+\tau})^2}}. \quad (4.45)$$

Коефіцієнт автокореляції $r(\tau)$ характеризує щільність зв'язку між послідовними рівнями часового ряду y_1, y_2, \dots, y_n та $y_{1+\tau}, y_{2+\tau}, \dots, y_{n+\tau}$ (зміщеними один відносно іншого на τ одиниць) і змінюється в межах $[-1; +1]$.

Близьке до одиниці за абсолютною величиною значення коефіцієнта автокореляції вказує на щільну лінійну залежність між початковими і зсуненими рівнями часового ряду, а близьке до нуля – на їхню незалежність.

Формулу (4.45) використовують лише для оцінювання щільності лінійного зв'язку між рівнями часового ряду, а тому її застосування у випадку чітко вираженої нелінійної тенденції є неправомірним.

На підставі значення коефіцієнта автокореляції можна зробити висновки лише про наявність чи відсутність лінійної (або близької до неї) тенденції, але за знаком коефіцієнта автокореляції не можна зробити висновок про зростаючу чи спадну тенденції в рівнях ряду.

Лаг τ (зсув у часі) визначає порядок автокореляції. Якщо $\tau = 1$, то наявна автокореляція 1-го порядку, якщо $\tau = 2$, то — 2-го порядку і т.д. Послідовність значень коефіцієнтів автокореляції першого, другого та інших порядків утворює *автокореляційну функцію* — функцію оцінювання

автокореляції залежно від часового лага. Графік автокореляційної функції називають *корелограмою*. Аналіз автокореляційної функції і корелограми дає змогу визначити лаг, за якого зв'язок між поточним та попередніми рівнями ряду є найщільнішим.

Обчисливши декілька коефіцієнтів автокореляції, можна визначити лаг, за якого автокореляція є найбільшою, і встановити таким чином структуру часового ряду:

- якщо коефіцієнт автокореляції 1-го порядку є статистично значущим і характеризується максимальним значенням серед інших, то ряд містить лише лінійний тренд;

- якщо максимального значення набуває статистично значущий коефіцієнт автокореляції p -го порядку, то часовий ряд, крім тенденції, містить коливання з періодичністю в p моментів часу, причому ці коливання можуть бути як сезонними, так і циклічними;

- якщо жоден із коефіцієнтів автокореляції не є статистично значущим, то ряд характеризується випадковою структурою або сильною нелінійною тенденцією, виявлення якої вимагає додаткових досліджень.

Зі збільшенням лага (τ) кількість пар $(n - \tau)$ спостережень $(y_t, y_{t+\tau})$ зменшується, тому лаг τ необхідно вибирати таким, щоб величина $(n - \tau)$ була достатньою для знаходження $r(\tau)$. На практиці користуються правилом, за яким максимальне значення τ не повинно перевищувати величину $\frac{n}{4}$, де n — кількість рівнів часового ряду.

З практичного погляду ($\alpha = 0,05$) можна стверджувати, що коефіцієнт автокореляції $r(\tau)$ буде статистично значущим, якщо його абсолютне значення перевищує $\frac{2}{\sqrt{n-\tau}}$.

Корелограма наочно показує, як зміни попередніх рівнів впливають на наступні рівні часового ряду, що слугує основою для виявлення структури часового ряду. На рис. 4.2 наведено типову корелограму стаціонарного ряду,

для якого індивідуальні значення коливаються навколо середнього значення, не виявляючи при цьому ні суттєвої тенденції зміни середнього, ні сезонних особливостей. У зміні значень коефіцієнтів автокореляції відсутня певна закономірність, і коефіцієнти автокореляції статистично незначущі (значення коефіцієнтів автокореляції розташовані між паралельними прямими, які визначають межі значущості).

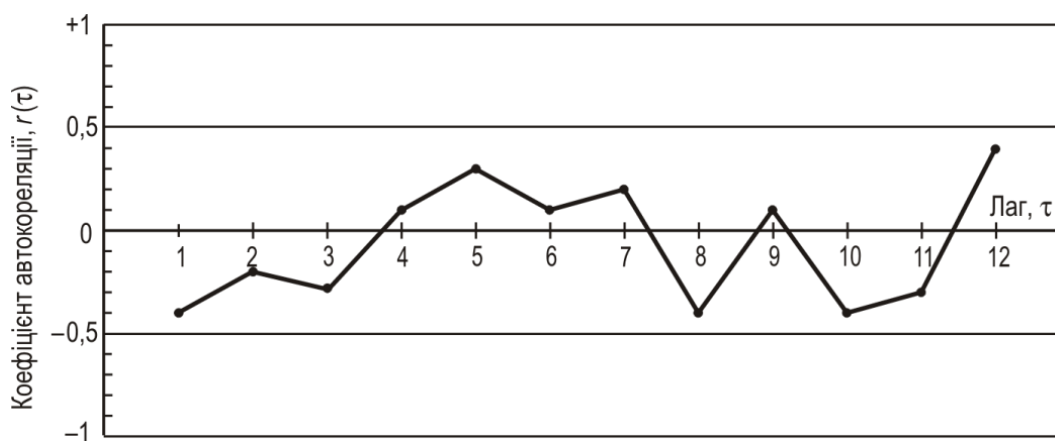


Рис. 4.2. Корелограма стаціонарного часового ряду

Корелограму часового ряду з лінійно-адитивним трендом (зростання середнього значення відбувається приблизно на однакову величину для кожного моменту часу) наведено на рис. 4.3. Візуальний аналіз цієї корелограми вказує на залежність значень коефіцієнта автокореляції від величини лага. Зі зростанням лага абсолютні значення коефіцієнтів автокореляції зменшуються, а максимальне значення (за абсолютною величиною) відповідає лагу $\tau = 1$, причому значущим є лише коефіцієнт автокореляції $r(1)$.

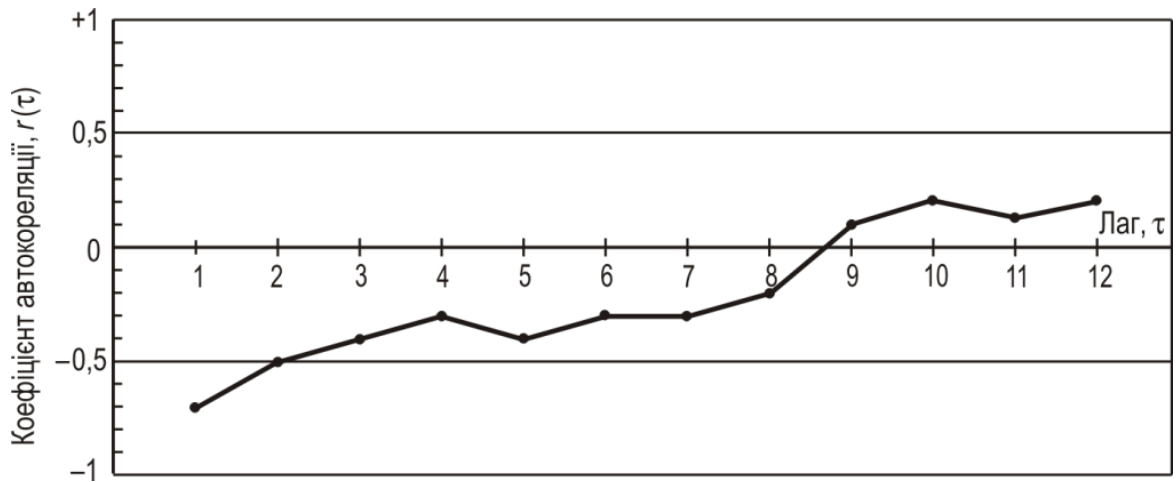


Рис. 4.3. Корелограма часового ряду з лінійно-адитивним трендом

Корелограму часового ряду із сезонною компонентою наведено на рис. 4.4. Для сезонних явищ спостерігається щільна кореляційна залежність між рівнями часового ряду, що відповідають одному і тому ж періоду часу. Найбільші за абсолютною величиною значення набувають коефіцієнти автокореляції з лагами 6, 12, 18 і 24. Вони є статистично значущими, і можна зробити висновок про наявність у часовому ряді сезонних коливань з періодом в один рік.

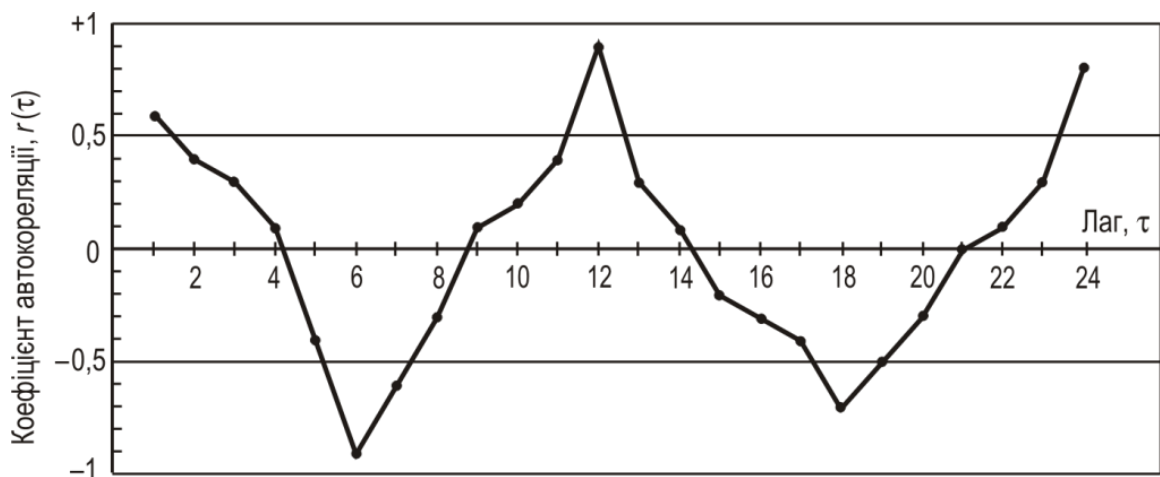


Рис. 4.4. Корелограма часового ряду із сезонною компонентою

Дослідження випадкової складової часового ряду пов'язують з дослідженням стаціонарності залишків (різниці між рівнями часового ряду і

систематичними компонентами). У разі стаціонарності часового ряду зі зростанням лага τ взаємозв'язок між рівнями послаблюється, і автокореляційна функція $r(\tau)$ спадає (за абсолютним значенням). Якщо формально побудована корелограма не є функцією, що швидко спадає, то це, як правило, свідчить про нестаціонарність ряду.

4.2. Методи адаптивного прогнозування

Методи прогнозування на основі аналітичного вирівнювання часових рядів ґрунтуються на однаковій інформаційній цінності всіх рівнів часового ряду. Представлення економічних процесів з урахуванням однакової інформаційної цінності рівнів часового ряду не завжди є виправданим, особливо, коли досліджують не лише тенденції «в середньому», а й розвиток тенденції. Попередні в часі дані часто виявляються малозначущими і можуть спотворювати результати прогнозування.

Ефективнішими методами прогнозування на основі обчисленої вибірки вважають методи, які враховують різну інформативність рівнів часового ряду – *методи адаптивного прогнозування*.

До адаптивних методів прогнозування відносять методи, які ґрунтуються на моделях, що відображають умови прогнозованого показника в часі з урахуванням інформативної цінності різних рівнів часового ряду. Процедуру коригування моделі на підставі надходження нової інформації називають *адаптацією моделі*.

Основне призначення адаптивних методів — короткострокове прогнозування. Поняття «короткострокове прогнозування» в цьому випадку потрібно сприймати як прогноз на один або незначну кількість періодів випередження, а не в розумінні його тривалості менше одного року, як це прийнято в економічній термінології.

Адаптивні моделі постійно пристосовуються до нових даних і в кінцевому результаті відображають тенденцію, що склалася на кінець періоду спостережень. Динаміка прогнозованого показника в кінці періоду спостережень є важливішою, ніж тенденція його розвитку в минулому. Алгоритм реалізації методів адаптивного прогнозування передбачає виконання таких кроків:

1. На підставі декількох початкових рівнів часового ряду оцінюють параметри вибраної моделі прогнозування та обчислюють прогнозні значення показника.
2. Обчислюють помилку прогнозування.
3. Проводять «навчання моделі». Якщо помилка прогнозування задовільна, то роблять прогноз на один крок вперед, інакше – коригують модель, розраховують прогнозні значення і переходять до 2-го кроку.

«Навчання моделі» проводять до тих пір, поки не будуть розглянуті всі спостереження. Новий прогноз є результатом коригування попереднього прогнозу на основі його помилки. Коефіцієнти адаптивної моделі оцінюють, як правило, на основі рекурентного підходу. Тому немає потреби повторювати весь обсяг обчислень за наявності нових даних.

«Навчання» (самокоригування) моделі полягає у виборі найкращого параметра адаптації, який визначають за результатами пробного моделювання на основі фактичних даних. Параметр адаптації характеризує швидкість реакції моделі на зміни в процесі функціонування об'єкта. Висновок про те, наскільки добре модель піддається «навчанню», роблять за результатами аналізу спроможності моделі адекватно відтворювати закономірності часового ряду. Після того, як визначено параметр адаптації, «навчання» моделі відбувається у процесі опрацювання нових даних.

Незважаючи на закладену в адаптивних моделях спроможність пристосовуватися до мінливих у часі умов, їх не можна вважати універсальними. Вибираючи конкретну адаптивну модель для практичної

реалізації, порівнюючи недоліки і сильні сторони окремих моделей, слід користуватися певним критерієм якості (точності) моделі. У переважній більшості випадків таким критерієм слугує критерій мінімізації середнього квадрата помилок прогнозування. Будуючи конкретні адаптивні моделі, потрібно враховувати ймовірнісні закономірності розвитку реального процесу, зіставляти динамічні властивості часового ряду з можливостями моделі до їхнього відтворення.

У модель слід закладати такі адаптивні властивості, які дадуть змогу своєчасно виявляти моменти зміни тенденції у часовому ряді та швидко коригувати параметри. Проблема покращення адаптивних властивостей прогновної моделі ефективно вирішується за допомогою використання спеціальних контрольних сигналів (індикаторів), які для заданого рівня статистичної значущості характеризують міру невідповідності прогнозів фактичним даним (неадекватності моделі).

Прикладом такого контрольного сигналу слугує індикатор Тригга (θ_t), який обчислюють як відношення експоненційно зваженої середньої помилки прогнозу (\bar{e}_t) до експоненційно зваженої середньої абсолютної помилки (MAD_t) з урахуванням параметра згладжування помилок (γ):

$$\theta_t = \frac{\bar{e}_t}{MAD_t}, \quad (4.46)$$

де $\bar{e}_t = \gamma e_t + (1 - \gamma)\bar{e}_{t-1}$; $MAD_t = \gamma |e_t| + (1 - \gamma)MAD_{t-1}$.

Оскільки $MAD_t \geq \bar{e}_t$, то контрольний сигнал перебуває в межах від -1 до +1. Параметр γ рекомендують вибирати таким чином, щоб виконувалася умова $\gamma \leq \alpha$, де α – параметр згладжування рівнів часового ряду. Ситуація, за якої $MAD_t \approx |\bar{e}_t|$, характерна для неадекватних моделей.

У табл. 4.6 наведено критичні значення контрольного сигналу Тригга ($\theta_t^{кр}$) для лінійної моделі Р. Брауна. Якщо в певний момент часу абсолютне значення сигналу перевищує критичне, то для вибраного рівня значущості модель прогнозування стає неадекватною реальному процесу.

Критичні значення індикатора Трігга для лінійної моделі Р. Брауна

Імовірність	$\gamma = 0,1$	$\gamma = 0,2$	$\gamma = 0,3$	$\gamma = 0,4$	$\gamma = 0,5$
0,90	0,32	0,46	0,57	0,67	0,76
0,95	0,39	0,54	0,67	0,76	0,83
0,99	0,46	0,65	0,76	0,86	0,90

Будь-яка адаптивна модель ґрунтується на одній із схем:

- *плинної середньої* — оцінкою поточного рівня часового ряду слугує зважена середня всіх попередніх рівнів, причому вагомість (інформаційна цінність) рівнів зменшується з їх віддаленням від поточного рівня, а реакцію на помилку прогнозу визначають за допомогою *параметрів згладжування (адаптації)*;

- *авторегресії* — оцінкою поточного рівня часового ряду слугує зважена сума декількох попередніх рівнів, вагомість яких визначається не їхньою віддаленістю від поточного рівня, а щільністю зв'язку між ними.

Використання простої і зваженої плинних середніх не забезпечує згладжування перших і останніх рівнів часового ряду. Відсутність згладжених останніх рівнів особливо актуальна в тому разі, коли завданням дослідження є прогнозування. Згладжені значення всіх рівнів часового ряду можна отримати за допомогою методів експоненційного згладжування.

Ідея методів експоненційного згладжування полягає у вирівнюванні рівнів часового ряду за допомогою зваженої плинної середньої з експоненційно розподіленими вагами. Що більше віддалена точка, для якої розраховують зважену плинну середню, від кінцевого рівня часового ряду, тим меншу «участь» вона приймає у побудові прогнозу.

Для часового ряду y_1, y_2, \dots, y_n прогнозне значення за *методом простого експоненційного згладжування* розраховують за такими рекурентними співвідношеннями:

$$S_t = \alpha y_t + (1 - \alpha)S_{t-1}, \quad (4.47)$$

де α ($0 < \alpha < 1$) – параметр згладжування;

S_{t-1} – значення експоненційно згладженої середньої для моменту $t - 1$ (прогнозне значення для моменту $t - 1$);

S_t – прогнозне значення для моменту t .

Від величини параметра згладжування α залежить швидкість зміни впливу попередніх спостережень на значення прогнозованого показника. Що більше α , тим менше попередні спостереження впливають на прогнозне значення. Вибір α близьким до одиниці призводить до врахування лише останніх спостережень у формуванні прогнозу. Якщо ж α є близьким до нуля, то в обчисленні прогнозу враховуються всі (або майже всі) попередні спостереження.

На практиці параметр α , як правило, визначають двома способами:

- експериментальним способом на сітці – можливі варіанти значення α розподіляють сіткою з певним кроком (наприклад, розглядають сітку значень від $\alpha=0,1$ до $\alpha=0,9$ з кроком $h=0,1$);
- за допомогою квазіньютонівської процедури мінімізації середнього квадрата помилок (або абсолютної середньої помилки).

Початкове значення S_0 , як правило, вибирають рівним значенню y_1 або середній декількох перших рівнів часового ряду.

Довірчий інтервал для вибраного рівня значущості α розраховують за формулою:

$$[S_t - t_\alpha \Delta_t; S_t + t_\alpha \Delta_t], \quad (4.48)$$

де t_α – критичне значення критерію Стюдента з кількістю ступенів вільності $V = n - 2$;

$$\Delta_t^2 = \frac{2}{2-\alpha} \sigma_e^2 - \text{дисперсія індивідуального прогнозу};$$

$$\sigma_e^2 = \frac{\sum_{t=1}^n (y_t - S_t)^2}{n-1} - \text{дисперсія помилок}.$$

Припустимо, що часовий ряд можна описати поліномом порядку q :

$$y_t = a_0 + a_1 t + \frac{a_2}{2!} t^2 + \dots + \frac{a_q}{q!} t^q + \varepsilon_t. \quad (4.49)$$

У такому разі прогнозне значення $y_{t+\tau}^{\text{прогн}}$ в момент $t + \tau$ можна знайти, розкладаючи його в ряд Тейлора:

$$y_{t+\tau}^{\text{прогн}} = y_t^{(0)} + \tau y_t^{(1)} + \frac{\tau^2}{2!} y_t^{(2)} + \dots + \frac{\tau^q}{q!} \cdot y_t^{(q)}, \quad (4.50)$$

де $y_t^{(i)}$ – i -та похідна в момент часу t .

Р. Брауном та Р. Майером доведено, що будь-яку i -ту похідну виразу (4.50) можна обчислити як лінійну комбінацію експоненційних середніх першого і вищих порядків (залежно від порядку полінома, вибраного для апроксимації значень рівнів часового ряду).

Експоненційною середньою першого порядку $S_t^{[1]}$ часового ряду для моменту t називають величину

$$S_t^{[1]} = \alpha \sum_{i=0}^n (1 - \alpha)^i y_{t-i} = \alpha y_t + (1 - \alpha) S_{t-1}^{[1]}, \quad (4.51)$$

де α – параметр згладжування ($0 < \alpha < 1$).

Експоненційні середні вищих порядків $S_t^{[k]}$ обчислюють за такою рекурентною формулою:

$$S_t^{[k]} = \alpha \cdot S_t^{[k-1]} + (1 - \alpha) \cdot S_{t-1}^{[k]}. \quad (4.52)$$

Як видно з формули (4.52), експоненційна середня $S_t^{[k]}$ — це лінійна комбінація попередніх спостережень з вагами, які зменшуються за геометричною прогресією стосовно минулих рівнів.

На практиці, зазвичай, для обчислення прогнозних значень користуються моделями Р. Брауна першого або другого порядку:

$$y_{t+\tau}^{\text{прогн}} = \hat{a}_0 + \hat{a}_1 \tau; \quad (4.53)$$

$$y_{t+\tau}^{\text{прогн}} = \hat{b}_0 + \hat{b}_1 \tau + \frac{1}{2} \hat{b}_2 \tau^2, \quad (4.54)$$

де $\hat{a}_0, \hat{a}_1, \hat{b}_0, \hat{b}_1, \hat{b}_2$ – коефіцієнти, що зв'язують експоненційні середні з відповідними коефіцієнтами рівняння регресії, яке описує тренд часового ряду; τ – горизонт прогнозування.

На першому етапі реалізації методу Р. Брауна задають початкові умови, а саме: експоненційні середні нульового порядку і параметр згладжування.

Якщо для прогнозування використовують модель першого порядку, то будують лінійне рівняння регресії:

$$\hat{y}_t = a_0 + a_1 t, \quad (4.55)$$

а початкові значення експоненційних середніх вибирають такими:

$$S_0^{[1]} = a_0 - \frac{1-\alpha}{\alpha} a_1; \quad (4.56)$$

$$S_0^{[2]} = a_0 - \frac{2(1-\alpha)}{\alpha} a_1. \quad (4.57)$$

У випадку використання моделі другого порядку будують квадратичне рівняння регресії:

$$\hat{y}_t = b_0 + b_1 t + b_2 t^2, \quad (4.58)$$

а початкові значення експоненційних середніх обчислюють за такими формулами:

$$S_0^{[1]} = b_0 - \frac{(1-\alpha)}{\alpha} b_1 + \frac{(1-\alpha)(2-\alpha)}{2\alpha^2} b_2; \quad (4.59)$$

$$S_0^{[2]} = b_0 - \frac{2(1-\alpha)}{\alpha} b_1 + \frac{(1-\alpha)(3-2\alpha)}{\alpha^2} b_2; \quad (4.60)$$

$$S_0^{[3]} = b_0 - \frac{3(1-\alpha)}{\alpha} b_1 + \frac{3(1-\alpha)(4-3\alpha)}{2\alpha^2} b_2. \quad (4.61)$$

Чітких рекомендацій щодо знаходження оптимального значення коефіцієнта згладжування α не встановлено. Однак треба зважати на те, що чим більшим буде значення α , тим істотнішим буде вплив останніх спостережень для розрахунку прогнозних значень.

Значення α , зокрема, можна вибирати з урахуванням довжини інтервалу згладжування за формулою

$$\alpha = \frac{2}{n+1}, \quad (4.62)$$

де n – кількість спостережень, що входять в інтервал згладжування.

Для пошуку коефіцієнта згладжування за великого обсягу спостережень часовий ряд можна розділити на дві частини. За даними першої частини для різних значень α розробляють прогноз на період, що відповідає другій частини. Після цього обчислюють дисперсії помилок прогнозів і для подальшої роботи коефіцієнт α , для якого дисперсія є мінімальною.

На другому етапі за рекурентними співвідношеннями обчислюють експоненційні середні та оцінюють коефіцієнти моделі, вибраної для прогнозування.

Якщо тренд описується лінійною моделлю (4.54), то коефіцієнти розраховують за формулами:

$$\hat{a}_0 = 2S_t^{[1]} - S_t^{[2]}; \quad (4.63)$$

$$\hat{a}_1 = \frac{\alpha}{1-\alpha} [S_t^{[1]} - S_t^{[2]}]. \quad (4.64)$$

У разі квадратичної моделі тренду (4.54) коефіцієнти встановлюють за такими формулами:

$$\hat{b}_0 = 3 [S_t^{[1]} - S_t^{[2]}] + S_t^{[3]}; \quad (4.65)$$

$$\hat{b}_1 = \frac{\alpha}{2(1-\alpha)^2} [(6 - 5\alpha)S_t^{[1]} - 2(5 - 4\alpha)S_t^{[2]} + (4 - 3\alpha)S_t^{[3]}]; \quad (4.66)$$

$$\hat{b}_2 = \frac{\alpha^2}{(1-\alpha)^2} [S_t^{[1]} - 2S_t^{[2]} + S_t^{[3]}]. \quad (4.67)$$

На третьому етапі розраховують точкові та інтервальні прогнози. Точковий прогноз отримують в результаті підстановки обчислених вище коефіцієнтів і довжини періоду випередження у формули (4.53) і (4.54).

Вважаючи, що випадкова величина ε є нормально розподіленою з нульовим математичним сподіванням і дисперсією σ_ε^2 , інтервальний прогноз для вибраного рівня значущості α отримаємо таким:

$$y_{t+\tau}^{\text{прогн}} - t_\alpha \sigma_{y_{t+\tau}} \leq y_{t+\tau} \leq y_{t+\tau}^{\text{прогн}} + t_\alpha \sigma_{y_{t+\tau}}, \quad (4.68)$$

де $\sigma_{y_{t+\tau}}$ – середня квадратична помилка прогнозу; t_α – критична точка розподілу Стьюдента (з $V = n - 2$ ступенями вільності для лінійної моделі і $V = n - 3$ – для квадратичної моделі).

Середню квадратичну помилку прогнозу для лінійної і квадратичної моделі визначають відповідно за формулами:

$$\sigma_{y_{t+\tau}} = \sigma_{\varepsilon_t} \sqrt{\frac{\alpha}{(2-\alpha)} [1 + 4(1-\alpha) + 5(1-\alpha)^2 + 2\alpha(4-3\alpha)\tau + 2\alpha^2\tau^2]}; \quad (4.69)$$

$$\sigma_{y_{t+\tau}} = \sigma_{\varepsilon_t} \sqrt{2\alpha + 3\alpha^2 + 3\alpha^3\tau^2}, \quad (4.70)$$

де σ_{ε_t} – середня квадратична помилка відхилення фактичних даних від тренда.

Середня квадратична помилка σ_{ε_t} відхилення для лінійної моделі рівна

$$\sigma_{\varepsilon_t} = \sqrt{\frac{\sum_{t=2}^n (y_t - y_t^{\text{прогн}})^2}{n-3}}, \quad (4.71)$$

а для квадратичної моделі:

$$\sigma_{\varepsilon_t} = \sqrt{\frac{\sum_{t=2}^n (y_t - y_t^{\text{прогн}})^2}{n-4}}. \quad (4.72)$$

Наведені моделі Р. Брауна характеризуються такими позитивними властивостями: чіткою та зрозумілою концепцією побудови; можливістю емпіричного пошуку оптимального значення лише одного параметра згладжування; спільним оцінюванням коефіцієнтів моделі, що зменшує автокореляцію залишків.

Недолік моделей Р. Брауна полягає в тому, що часовий ряд розглядають ізольовано від інших процесів. Навіть у випадку, коли наявна додаткова інформація, то її можна використати лише шляхом регулювання швидкості адаптації. Крім того, вся специфіка часового ряду відображається єдиним параметром згладжування α .

Метод Р. Брауна вважають ефективним інструментом прогнозування процесів, у яких відсутня чітко виражена динаміка зростання або спадання показників (нестійких рядів зі змінною тенденцією). Для прогнозування процесів, які характеризуються чітко вираженою тенденцією зростання або спадання, розроблено низку модифікацій цього методу, зокрема метод Ч. Хольта.

Коефіцієнти полінома у методі Р. Брауна виражають через експоненційні середні різних порядків, які в свою чергу отримують шляхом згладжування рівнів часового ряду за допомогою одного параметра. За методом Ч. Хольта (*подвійного експоненційного згладжування*) використовують два параметри згладжування (α і β). У цьому випадку точковий прогноз на момент часу $(t + \tau)$ розраховують за формулою:

$$y_{t+\tau}^{\text{прогн}} = a_t + \tau b_t, \quad (4.73)$$

де a_t – розрахункове значення коефіцієнта, який характеризує зміну середнього рівня; b_t – розрахункове значення коефіцієнта, який характеризує приріст рівня за одиницю часу; τ – період випередження.

Коефіцієнти a_t і b_t оцінюють за допомогою двох плинних середніх, які мають незалежні параметри згладжування, за такими рекурентними співвідношеннями:

$$a_t = \alpha y_t + (1 - \alpha)(a_{t-1} + b_{t-1}); \quad (4.74)$$

$$b_t = \beta(a_t - a_{t-1}) + (1 - \alpha)b_{t-1}. \quad (4.75)$$

Реалізація методу передбачає встановлення початкових значень a_0 і b_0 та значень параметрів згладжування α і β . Чітких рекомендацій щодо їх вибору, як і для методу Р. Брауна, немає. Значення a_0 і b_0 можна прийняти рівними відповідним параметрам лінійного рівняння регресії, побудованого на основі декількох початкових рівнів часового ряду, або a_0 можна прийняти рівним середньому значенню декількох перших рівнів часового ряду, а b_0 — середньому значенню декількох перших різниць ряду.

Для розрахунку коефіцієнта a використовують значення коефіцієнта b , а для розрахунку коефіцієнта b — значення коефіцієнта a . Оскільки кожний із коефіцієнтів розраховують за допомогою значення власного параметра згладжування, то параметр α впливає на параметр β і навпаки. Тому використання класичних меж зміни параметрів згладжування (від 0 до 1) в методі Ч. Хольта є необґрунтованим. Обмеження області зміни параметрів згладжування α і β означає об'єднання їх властивостей.

Значення параметрів згладжування α і β у практиці прогнозування вибирають з інтервалів:

$$0 < \alpha \leq 0,30; \quad 0,01 \leq \beta \leq 0,25.$$

Для оптимального вибору значень α і β можна виконати експериментальні дослідження, користуючись критерієм мінімізації середнього квадрата помилок прогнозування.

Інтервальний прогноз розраховують за формулою (4.68), а наближене значення середньої квадратичної помилки прогнозу становитиме

$$\sigma_{y_{t+\tau}} \approx \sigma_{\varepsilon_t} \sqrt{(1 - \alpha\beta)[1,25 + (1 - \alpha\beta)\tau]}. \quad (4.76)$$

Суттєвим недоліком методу Ч. Хольта є припущення щодо лінійної тенденції зростання або спадання рівнів часового ряду. Якщо ділянки лінійної тенденції прогнозованого показника почергово змінюються ділянками нелінійної тенденції, то метод Ч. Хольта дає посередні результати.

В основу методу гармонійних ваг покладено ідею використання плинного тренда замість плинної середньої, тобто екстраполяцію рівнів часового ряду проводять за допомогою плинного тренда (ламаної лінії). У цьому випадку окремі точки плинного тренда зважують за допомогою гармонійних ваг, що забезпечує присвоєння пізнішим спостереженням вищої вагомості.

На першому кроці методу гармонійних ваг формують фази. До фази входять k рівнів початкового часового ряду y_1, y_2, \dots, y_n (практика прогнозування рекомендує вибирати k від 3 до 5). Першу фазу утворюють рівні y_1, y_2, \dots, y_k , другу — y_1, y_2, \dots, y_{k+1} , останню — $y_{n-k+1}, y_{n-k+2}, \dots, y_n$. Очевидно, що кількість утворених фаз K буде дорівнювати $K = n - k + 1$.

Після того, як сформовані фази, для кожної з них за МНК знаходять параметри лінійних трендів

$$\hat{y}_i(t) = b_{0i} + b_{1i}t \quad i = \overline{1, K}; \quad t = \overline{i, i + k - 1}. \quad (4.77)$$

За формулою (4.77) обчислюють середні значення плинного тренда в точках $t = \overline{1, n}$. Наприклад, якщо $k = 3$, то для знаходження середніх значень використовуємо співвідношення:

$$\left\{ \begin{array}{l} \bar{y}(1) = \hat{y}_1(1); \\ \bar{y}(2) = \frac{\hat{y}_1(2) + \hat{y}_2(2)}{2}; \\ \bar{y}(3) = \frac{\hat{y}_1(3) + \hat{y}_2(3) + \hat{y}_3(3)}{3}; \\ \dots \dots \dots \dots \dots \dots \dots; \\ \bar{y}(n-1) = \frac{\hat{y}_{n-3}(n-1) + \hat{y}_{n-2}(n-1)}{2}; \\ \bar{y}(n) = \hat{y}_{n-2}(n). \end{array} \right.$$

Для виконання наступних кроків необхідно перевірити припущення про те, що відхилення від плинного тренду мають випадковий характер і утворюють стаціонарний процес. Якщо припущення виконуються, то наступним кроком є обчислення ланцюгових приростів (W_{t+1}) і середнього ланцюгового приросту (\bar{W}):

$$W_{t+1} = \bar{y}(t+1) - \bar{y}(t); \quad (4.78)$$

$$\bar{W} = \sum_{t=1}^{n-1} C_{t+1} \cdot W_{t+1}, \quad (4.79)$$

де C_{t+1} — гармонійні коефіцієнти, які знаходять на основі гармонійних ваг m_{t+1} .

Для визначення гармонійних ваг користуються одним із співвідношень:

$$\left\{ \begin{array}{l} m_{t+1} = \sum_{i=1}^t \frac{1}{n-i}, \\ m_{t+1} = m_t + \frac{1}{n-t} \quad (t = \overline{1, n-1}). \end{array} \right. \quad (4.80)$$

При цьому самій попередній інформації надається вага $m_2 = 1/(n-1)$, а сума гармонійних ваг дорівнюватиме

$$\sum_{t=1}^{n-1} m_{t+1} = n-1. \quad (4.81)$$

Гармонійні коефіцієнти на підставі формули (4.81) визначають за формулою

$$C_{t+1} = \frac{m_{t+1}}{n-1}. \quad (4.82)$$

Очевидно, що гармонійні коефіцієнти відповідають таким умовам:

$$C_{t+1} > 0, \quad t = \overline{1, n-1}; \quad \sum_{t=1}^{n-1} C_{t+1} = 1.$$

З описаної вище схеми розрахунку гармонійних коефіцієнтів кожний ланцюговий приріст має більшу вагомість, ніж попередній.

Точковий прогноз обчислюють за формулою

$$y_{t+\tau}^{\text{прогн}} = \bar{y}(t) + \tau \bar{W}. \quad (4.83)$$

Побудова *довірчого інтервалу* для прогнозного значення показника ґрунтується на нерівності Чебишева:

$$P\{|W_{t+1} - E(W)| > a\sigma_W\} < \frac{1}{a^2}, \quad (4.84)$$

де a – задане ціле додатне число; σ_W – середнє квадратичне відхилення випадкової величини W .

Оцінку середнього квадратичного відхилення $\hat{\sigma}_W$ випадкової величини W визначають за формулою

$$\hat{\sigma}_W = \sqrt{\sum_{t=1}^{n-1} C_{t+1} (W_{t+1} - \bar{W})^2}, \quad (4.85)$$

а довірчі межі для прогнозного значення показника будуть становити

$$y_{n+\tau}^{\text{прогн}} - A(\tau)\hat{\sigma}_W \leq y_{\text{інт}}^{\text{прогн}} \leq y_{n+\tau}^{\text{прогн}} + A(\tau)\hat{\sigma}_W; \quad (4.86)$$

$$A(\tau) = a \sum_{t=1}^{\tau+1} C_{n-t+1}, \quad (4.87)$$

де τ – горизонт прогнозування.

Багатьом економічним процесам властиві періодичні коливання. Причини таких коливань різноманітні і залежать від часу та показників, які досліджують. Якщо у часових рядах спостерігаються періодичні коливання, то їх згладжування за допомогою традиційних методів не дає позитивних результатів.

Одним із методів дослідження часових рядів, для яких характерні періодичні коливання, слугує *гармонійний (спектральний) аналіз*. Він забезпечує розпізнавання коливань різної довжини і побудову математичної моделі часових рядів, рівні яких подаються як сума періодичного тренда $f(t)$ і випадкової компоненти ε_t . Періодичність тренда полягає у точному і регулярному повторенні своєї поведінки через певні проміжки часу, тобто, якщо період дорівнює φ , то $f(t + \varphi) = f(t)$. Тренд вважають деякою функцією від часу, а вплив на нього випадкової компоненти — відсутнім.

Основними характеристиками стаціонарного часового ряду з періодичними коливаннями є:

- *період коливання (тривалість циклу, T)* — інтервал часу, через який рівні ряду повторюються;

- *частота (f)* — кількість циклів за одиницю часу, тобто $f = 1/T$;

- *амплітуда (A)* — відстань від середнього рівня до максимального (або мінімального) рівня часового ряду;

- *фаза (F)* — відстань між початком відліку часу ($t = 0$) і найближчим моментом часу, у якому рівень ряду досягає максимального або мінімального значення;

- *кутова частота коливань (ω)* — швидкість зміни фази коливань, вимірюється в радіанах за одиницю часу і дорівнює $\omega = 2\pi/T$ ($0 \leq \omega \leq 2\pi$).

У моделях часових рядів, які представляють динаміку економічних процесів, за одиницю часу прийнято вважати одне спостереження. Якщо ряд містить місячні дані за декілька років, то відповідна періодичність визначає річний цикл. Таким чином, 12 спостережень складає один цикл ($T = 12$), а кожне спостереження — 0,083 циклу ($f = 1/12$). На практиці часові ряди містять скінченну кількість рівнів (n). Якщо інтервали часу між спостереженнями рівні (наприклад, місяць), то найбільший період буде дорівнювати n місяцям, що відповідає кутовій частоті $\omega = 2\pi/n$, а найменший — двом місяцям.

Крім вказаних вище характеристик для задання стаціонарного часового ряду з періодичними коливаннями необхідно вказати середнє значення (\bar{y}), навколо якого відбуваються коливання.

Згідно з теоретичними основами гармонійного аналізу стаціонарний часовий ряд з періодичними коливаннями можна подати у вигляді ряду Фур'є — суми середнього рівня та функцій синуса і косинуса (*гармонік*):

$$y_t = a_0 + \sum_{k=1}^m (a_k \cos \omega_k t + b_k \sin \omega_k t), \quad (4.88)$$

де a_0 – середній рівень часового ряду; a_k, b_k – коефіцієнти біля гармонік; m – кількість гармонік; k – порядковий номер гармоніки; ω_k – кутова частота k -ї гармоніки.

Гармонійний аналіз — це метод вирівнювання деякої періодичної функції за допомогою гармонік різного порядку. Більшість методів, розроблених для прогнозування періодичних коливань, ґрунтуються на припущенні, що довжина періодичної компоненти наперед відома або виявляється візуально на основі графіка часового ряду, після чого включається в теоретичну модель. На відміну від них гармонійний аналіз дає змогу детальніше розпізнати періодичні коливання, знайти декілька циклів різної довжини, які повторюються і на перший погляд сприймаються як випадкові коливання.

У результаті застосування МНК до виразу (4.88) отримують:

$$\left\{ \begin{array}{l} a_0 = \frac{1}{n} \sum_{t=1}^n y_t; \\ a_k = \frac{2}{n} \sum_{t=1}^n y_t \cos(\omega_k t), \quad k = \overline{1, m-1}; \\ b_k = \frac{2}{n} \sum_{t=1}^n y_t \sin(\omega_k t), \quad k = \overline{1, m-1}; \\ a_m = \frac{1}{n} \sum_{i=1}^n y_t \cos(\omega_k t); \\ b_m = 0. \end{array} \right. \quad (4.89)$$

Коефіцієнти біля синусів і косинусів (*коефіцієнти Фур'є*) характеризують рівень кореляції відповідних функцій з даними спостережень. Висока кореляція (велике значення коефіцієнта при певній функції) показує, що на відповідній частоті існує строга періодичність в даних.

Коефіцієнти a_k і b_k пов'язані з амплітудою коливань співвідношеннями:

$$A_k^2 = a_k^2 + b_k^2. \quad k = \overline{1, m}. \quad (4.90)$$

Кутову частоту k -ї гармоніки розраховують за формулою:

$$\omega_k = \frac{2\pi k}{n}. \quad (4.91)$$

Кожний член суми (4.88) є гармонікою з визначеним періодом. Перша має період, що дорівнює довжині ряду (n). Далі в порядку нумерації гармонік періоди зменшуються: друга має період, що дорівнює $\frac{n}{2}$, третя — $\frac{n}{3}$ і т.д.

Максимальна кількість гармонік, які використовують для моделювання часового ряду, не повинна перевищувати половини довжини часового ряду, тобто $m \leq \frac{n}{2}$. Якщо кількість спостережень непарна, то останній рівень ряду з розгляду вилучають. З практичної точки зору модель періодичних коливань добре описується декількома першими гармоніками. Переважно розраховують до чотирьох гармонік, а потім визначають, з якою кількістю з них ряд Фур'є найкраще відображає зміни рівнів часового ряду.

Дисперсія k -ї гармоніки (δ_k^2) дорівнює $\frac{A_k^2}{2}$. Оцінкою її вкладу в загальну дисперсію процесу (σ_y^2) слугує дисперсійне відношення (R_k^2):

$$R_k^2 = \frac{\delta_k^2}{\sigma_y^2}. \quad (4.92)$$

Оскільки жодні дві гармоніки через їх ортогональність не корелюють між собою, то дисперсії, які враховуються різними гармоніками, додаються і в результаті їхня сума дорівнює загальній дисперсії часового ряду.

Виявлення наявності періодичних коливань і визначення періоду проводять за допомогою корелограм і періодограм.

Періодограма — це послідовність значень

$$P_k = \frac{T}{2} \cdot A_k^2, \quad (k = \overline{1, m}). \quad (4.93)$$

Значення періодограми інтерпретують як дисперсію даних на відповідній частоті.

Графічне зображення періодограми слугує засобом візуального виявлення скритої періодичності та попереднього вибору гармонік, які

будуть включені до моделі. Проявом періодичних коливань вважають наявність гострих вузьких піків у періодограмі на відповідних частотах.

Представлений вище підхід до моделювання періодичних часових рядів застосовують лише до стаціонарних рядів. Для часових рядів з тенденцією ряд Фур'є зводять до стаціонарного вигляду. Для цього знаходять лінійний тренд ($\hat{y}_t = a + bt$) і застосовують гармонійний аналіз до відхилень від тренда ($e_t = y_t - \hat{y}_t$) або визначають абсолютні ланцюгові прирости ($\Delta y_t = y_t - y_{t-1}$), які в подальшому використовують як інформаційну базу для побудови ряду Фур'є.

Якщо керуватися припущенням, що в майбутньому періоді збережеться амплітуда коливання, то гармонійну модель можна використати для оцінювання досліджуваного показника в перспективі. Прогнозне значення для певного моменту часу у випадку стаціонарності часового ряду розраховують підстановкою відповідних параметрів у рівняння (4.88). У випадку нестаціонарності ряду спочатку будують прогноз на підставі тенденції розвитку, а потім до нього додають прогноз за рядом Фур'є, що побудований за даними відхилень від тренда (або ланцюгових приростів).

Періодичні коливання, що тривають один рік, називають *сезонними*. Проявом сезонних коливань слугує наявність інтервалів зростання та спадання рівнів часового ряду всередині року, які регулярно повторюються впродовж декількох років.

Сезонність зумовлюють природно-кліматичні, календарні та соціальні причини. Сезонні коливання виникають у виробництві сільськогосподарської продукції, завантаженні транспорту, формуванні попиту на окремі види продукції та ін. Врахування сезонності є актуальним для вироблення управлінських рішень щодо ціноутворення в умовах нерівномірного попиту, оцінювання наявних потужностей і резервів підприємства, потреби в робочій силі в періоди пікових навантажень тощо.

Виявити сезонні коливання в часовому ряді можна на основі аналізу коефіцієнтів автокореляції і періодограми. Якщо з поміж коефіцієнтів автокореляції максимального значення набуває значущий коефіцієнт автокореляції порядку $\tau > 3$, то часовий ряд містить сезонні коливання з періодичністю в τ моментів часу. Проявом сезонних коливань на періодограмі є наявність гострих вузьких піків на відповідних частотах (періодах спостережень).

Показниками інтенсивності сезонних коливань рівнів часового ряду в окремі періоди слугують *індекси сезонності*, сукупність яких утворює *сезону хвилю*. Індекси сезонності обчислюють як відношення фактичних рівнів часового ряду до середньорічного рівня, розрахованого на основі фактичних або згладжених даних спостережень.

Якщо часовий ряд є стаціонарним (відсутній тренд), то індекс сезонності (I_t^c) знаходять на основі емпіричних даних без їх попереднього згладжування:

$$I_t^c = \frac{\bar{y}_t}{\bar{y}} \cdot 100\%, \quad (4.94)$$

де \bar{y} – середній рівень часового ряду за рік; \bar{y}_t – середній рівень часового ряду за інтервал часу t (місяць або квартал).

Що більше значення індексу сезонності, тим більша інтенсивність (амплітуда) коливань відповідного рівня ряду відносно середнього рівня.

З метою отримання стійких оцінок розміру сезонних коливань у випадку нестационарних часових рядів (наявності тренда) рекомендують дотримуватися такої послідовності дій:

- згладжують рівні часового ряду за допомогою плинної середньої або аналітичних методів вирівнювання;
- знаходять індекси сезонності для окремих інтервалів часу;
- усереднюють індекс сезонності за всі роки.

Індекси сезонності використовують у методах прогнозування сезонних процесів, зокрема, у методі *сезонної декомпозиції* для розкладання часового ряду на окремі складові з метою їх подальшого аналізу. На практиці

найчастіше використовують два види моделей взаємодії компонент часового ряду: *адитивну* та *мультиплікативну* моделі.

В адитивній моделі значення рівнів часового ряду (y_t) подають як суму трендової ($f(t)$), сезонної (s_t) та випадкової (ε_t) компонент:

$$y_t = f(t) + s_t + \varepsilon_t, \quad t = \overline{1, n}, \quad (4.95)$$

де n – кількість рівнів часового ряду.

Мультиплікативна модель має такий вигляд:

$$y_t = f(t) \cdot s_t \cdot \varepsilon_t, \quad t = \overline{1, n}. \quad (4.96)$$

У багатьох літературних джерелах трендову компоненту називають «тренд-циклічною компонентою». Це пояснюється тим, що під час декомпозиції циклічні коливання (наприклад, пов'язані з періодами економічного розвитку, періодами зростання і спадання економічної активності) не відокремлюють від початкового часового ряду, а вважають такими, що входять до трендової складової.

Модель (4.95) застосовують у тих випадках, коли із зміною рівнів ряду амплітуда сезонних коливань не змінюється (рис. 4.5а). Якщо із зростанням або спаданням рівнів ряду амплітуда змінюється (рис. 4.5б), то використовують мультиплікативну модель.

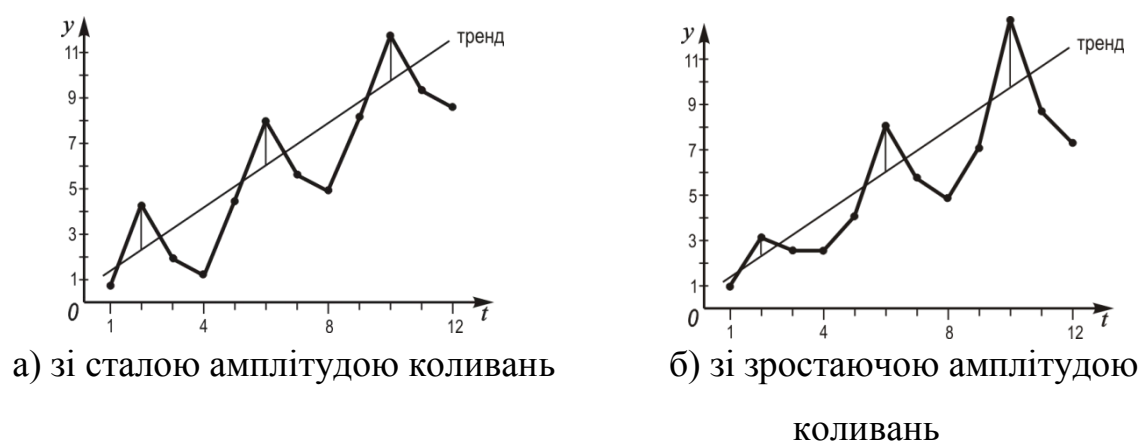


Рис. 4.5. Часові ряди з сезонними коливаннями

Користуючись адитивною моделлю, припускають, що випадкова складова розподілена нормально з нульовим математичним сподіванням і постійною дисперсією $\varepsilon \sim N(0; \sigma_{\varepsilon_t}^2)$.

У моделі (4.95) випадкові відхилення (ε_t) враховують мультиплікативно, а не адитивно. Якщо злогарифмувати ліву і праву частини виразу (4.96), то модель буде представлено в адитивному вигляді

$$\ln y_t = \ln f(t) + \ln s_t + \ln \varepsilon_t. \quad (4.97)$$

Подання мультиплікативної моделі у вигляді (4.97) дає певне припущення щодо розподілу випадкової компоненти, а саме, логічно припустити, що випадкова компонента розподілена нормально з відмінним від нуля математичним сподіванням і постійною дисперсією. А це свідчить про те, що у випадку використання мультиплікативних моделей отримують *зміщені оцінки*.

Невірна ідентифікація типу сезонності в часовому ряді призводить до отримання завищеного значення випадкової компоненти, що впливає на зростання величини довірчого інтервалу. Одним із критеріїв оцінювання правильності вибору типу моделі може слугувати нормальність розподілу випадкової компоненти:

$$\varepsilon_t = y_t - (f(t) - s_t) \text{ — для адитивної моделі;} \quad (4.98)$$

$$\varepsilon_t = \frac{y_t}{f(t) \cdot s_t} \text{ — для мультиплікативної моделі.} \quad (4.99)$$

Послідовність побудови і реалізації моделей сезонної декомпозиції представлено на рис. 4.6.

На першому етапі готують дані для знаходження сезонних компонент, а саме — «очищають» початковий часовий ряд від сезонних і випадкових коливань за методами плинних середніх або експоненційного згладжування.

Інтервал («ширину вікна») згладжування вибирають рівним періоду сезонності. Якщо інтервал згладжування є парним числом (наприклад, прогнозуючи квартальні показники на основі місячних даних за декілька років), то оцінки є точками, що лежать між фактичними рівнями, хоча повинні відповідати фактичним. Тому слід виконати *центрування* —

розрахувати центровані плинні середні (середні з пари сусідніх значень). У результаті отримують згладжений ряд (\tilde{y}_t).

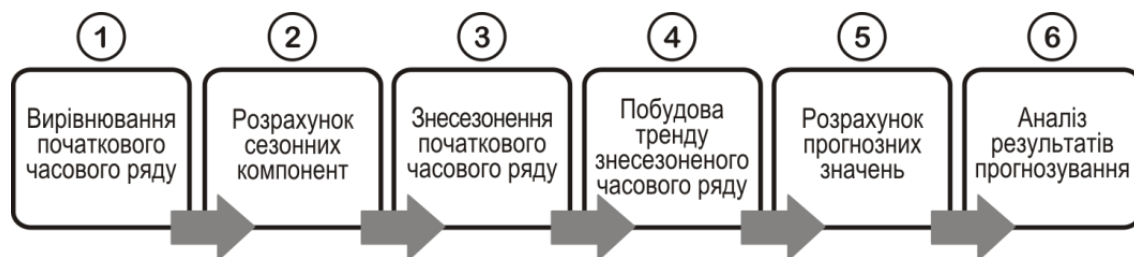


Рис. 4.6. Послідовність побудови та реалізації моделей сезонної декомпозиції

На другому етапі розраховують сезонні компоненти:

$$s_t = y_t - \tilde{y}_t \text{ — для адитивної моделі;} \quad (4.100)$$

$$s_t = \frac{y_t}{\tilde{y}_t} \text{ — для мультиплікативної моделі.} \quad (4.101)$$

Очевидно, що ці компоненти містять випадкові відхилення. Для того, щоб їх позбутися, розраховують середні сезонні оцінки (\bar{s}_t). Середні сезонні оцінки, як правило, коригують таким чином, щоб їхня сума дорівнювала нулю для адитивної моделі, а добуток — одиниці для мультиплікативної моделі. У результаті отримують набір «чистих» сезонних компонент.

На третьому етапі середні сезонні оцінки вилучають з початкового часового ряду і отримують знесезонений ряд (Y_t):

$$Y_t = y_t - \bar{s}_t \text{ — для адитивної моделі;} \quad (4.102)$$

$$Y_t = \frac{y_t}{\bar{s}_t} \text{ — для мультиплікативної моделі.} \quad (4.103)$$

На четвертому етапі згладжують знесезонений ряд і на його основі будують рівняння тренда ($\hat{f}(t)$).

На наступному етапі, користуючись побудованим рівнянням тренда і середніми значеннями сезонних оцінок, розраховують прогнозні значення для моментів часу t ($y_t^{\text{прогн}}$):

$$y_t^{\text{прогн}} = \hat{f}(t) + \bar{s}_t \text{ — для адитивної моделі;} \quad (4.104)$$

$$y_t^{\text{прогн}} = \hat{f}(t) \cdot \bar{s}_t \text{ — для мультиплікативної моделі.} \quad (4.105)$$

На завершальному етапі обчислюють помилки прогнозування та на їхній основі аналізують якість побудованої моделі.

Розглянутий вище класичний метод сезонної декомпозиції не позбавлений певних недоліків, а саме:

- у процесі згладжування втрачаються перші та останні рівні початкового часового ряду, що є актуальним за незначної кількості спостережень;

- не передбачаються зміни сезонних компонент впродовж тривалого періоду часу, що на практиці часто не виконується;

- у випадку наявності аномальних рівнів («викидів») сезонні компоненти можуть істотно спотворюватися.

Нами було розглянуто модель Ч. Хольта для адаптивного прогнозування процесів із чітко вираженою лінійною тенденцією. П. Вінтерс удосконалив її на випадок з лінійною тенденцією та мультиплікативним сезонним ефектом. Побудована модель, яка є об'єднанням двопараметричної моделі лінійного зростання і моделі сезонності, отримала назву *моделі Хольта–Вінтерса* (або трипараметричної моделі згладжування).

Точковий прогноз з періодом випередження τ за моделлю Хольта–Вінтерса дорівнює

$$y_{t+\tau}^{\text{прогн}} = (a_t + \tau b_t) \cdot c_{t-l+\tau}, \quad (4.106)$$

де a_t – експоненційна середня рівнів часового ряду; b_t – експоненційна середня приросту коефіцієнта a_t ; $c_{t-l+\tau}$ – коефіцієнт сезонності (коефіцієнт, що враховує сезонний ефект); l – період сезонності (для квартальних даних $l = 4$, для щомісячних $l = 12$).

Коефіцієнти моделі обчислюють з використанням таких рекурентних співвідношень:

$$a_t = \alpha \frac{y_t}{c_{t-l}} + (1 - \alpha)(a_{t-1} + b_{t-1}), \quad (4.107)$$

$$b_t = \beta(a_t - a_{t-l}) + (1 - \beta)b_{t-l}, \quad (4.108)$$

$$c_t = \gamma \frac{y_t}{a_t} + (1 - \gamma)c_{t-l}, \quad (4.109)$$

де α , β , γ – параметри згладжування ($\alpha, \beta, \gamma \in (0; 1)$).

Алгоритм реалізації моделі Хольта–Вінтерса вимагає встановлення початкових умов. Спочатку потрібно розрахувати коефіцієнти a_0 і b_0 лінійного тренда на основі частини або всього часового ряду. Після цього на деякій ділянці часового ряду, яка містить l спостережень, обчислюють коефіцієнти сезонності:

$$c_t = \frac{y_t}{a_0 + b_0 t}.$$

Параметри експоненційного згладжування (α, β, γ), які використовують у моделі Хольта–Вінтерса, знаходяться в інтервалі $(0; 1)$, і їх вибирають незалежно. Оскільки чітких правил щодо їхнього вибору немає, то на практиці необхідно експериментувати для мінімізації середньоквадратичної помилки прогнозу. У такому разі слід зважати на те, що збільшення значень кожного з параметрів згладжування призводить до зростання ваги (інформаційної цінності) останніх спостережень, а зменшення — до покращення згладжування випадкових відхилень.

Г. Тейл і С. Вейдж запропонували модель для прогнозування процесів, яким притаманна лінійна тенденція і адитивний сезонний ефект. Точковий прогноз з періодом випередження τ за моделлю Тейла–Вейджа розраховують за формулою:

$$y_{t+\tau}^{\text{прогн}} = a_t + \tau b_t + c_{t-l+\tau}. \quad (4.110)$$

Змістовна інтерпретація коефіцієнтів моделі (4.109) така ж, як і для моделі Хольта–Вінтерса.

Коефіцієнти моделі Тейла–Вейджа оновлюють за такими співвідношеннями:

$$a_t = \alpha(y_t - c_{t-l}) + (1 - \alpha)(a_{t-1} + b_{t-1}); \quad (4.111)$$

$$b_t = \beta(a_t - a_{t-l}) + (1 - \beta)b_{t-1}; \quad (4.112)$$

$$c_t = \gamma(y_t - a_t) + (1 - \gamma)c_{t-1}, \quad (4.113)$$

де α , β , γ – параметри згладжування ($\alpha, \beta, \gamma \in (0; 1)$).

Коефіцієнти a_0 і b_0 обчислюють аналогічно до моделі Хольта–Вінтерса, а значення $c_t = y_t - a_0 - b_0 t$.

Модель з експоненційною тенденцією та мультиплікативною сезонністю можна звести до моделі з лінійною тенденцією і адитивною сезонністю. У результаті заміни значень початкового часового ряду їхніми логарифмами експоненційну тенденцію перетворюють у лінійну. Одночасно мультиплікативна модель стає адитивною.

У багатьох випадках економічні показники мають настільки складну структуру, що їхнє прогнозування на основі традиційних моделей не гарантує отримання задовільних результатів. Ефективним інструментом вирішення цієї проблеми слугує використання методів адаптивного прогнозування за схемою авторегресії.

Ідея, яка покладена в основу авторегресійних моделей, ґрунтується на логічному припущенні, що значення прогнозованого показника в який-небудь момент часу t впливає на його значення в наступні моменти часу, тобто рівні одного й того ж ряду є взаємозалежними.

Моделі часових рядів, в яких поточний рівень представляють у вигляді лінійної комбінації попередніх рівнів і випадкової величини, якій притаманна властивість «білого шуму», називають *моделями авторегресії (AR-моделями — з англ. autoregressive model)*. Моделі авторегресії використовують для моделювання як стаціонарних, так і нестаціонарних часових рядів. У представленому матеріалі цього розділу основну увагу зосереджено на часових рядах, з рівнів яких вилучено невинуваті компоненти, тобто на стаціонарних часових рядах.

Найпростіша авторегресійна модель — це модель авторегресії першого порядку ($AR(1)$):

$$y_t = \alpha_1 y_{t-1} + \varepsilon_t, \quad (t = \overline{1, n}). \quad (4.114)$$

де ε_t – випадкова величина з нульовим математичним сподіванням ($M\{\varepsilon_t\} = 0$) і постійною дисперсією ($D\{\varepsilon_t\} = \sigma_0^2$), причому величини ε_i та ε_j не корелюють між собою для всіх $i \neq j$; α – параметр моделі.

Модель (4.114) представляє дискретний випадок *марківського процесу* — випадкового процесу, значення якого для довільного моменту часу $t + 1$ залежить від значення в момент часу t , але не залежить від значень в моменти $t - 1, t - 2$ і т.д. У термінах кореляційного аналізу часових рядів це означає, що існує кореляційний зв'язок початкового часового ряду з рядом, зміщеним на один часовий інтервал, і відсутній зв'язок з рядами, зміщеними на два і більше часових інтервалів. Нагадаємо, що щільність зв'язку між послідовними рівнями часового ряду і зміщеними один відносно одного на τ одиниць характеризує коефіцієнт автокореляції порядку $r(\tau)$.

Коефіцієнт α моделі (4.114) дорівнює коефіцієнту автокореляції першого порядку, тобто $\alpha_1 = r(1)$. Умовою стаціонарності процесу, який описується моделлю (4.114), є $|\alpha_1| < 1$.

Авторегресію другого порядку ($AR(2)$) називають *процесом Юла*. Модель $AR(2)$ має вигляд:

$$y_t = \alpha_1 y_{t-1} + \alpha_2 y_{t-2} + \varepsilon_t, \quad (4.115)$$

де відносно ε_t зберігаються зазначені вище умови для ($AR(1)$).

Параметри моделі (4.115) розраховують за формулами:

$$\alpha_1 = \frac{r(1)(1-r(2))}{1-r(1)^2}, \quad (4.116)$$

$$\alpha_2 = \frac{r(2)-r(1)^2}{1-r(1)^2}. \quad (4.117)$$

Умовами стаціонарності процесу Юла є така система нерівностей:

$$\begin{cases} |\alpha_1| < 2 \\ |\alpha_2| < 1 - |\alpha_1|. \end{cases} \quad (4.118)$$

У загальному випадку модель авторегресії порядку p ($AR(p)$) має вигляд:

$$y_t = \alpha_1 y_{t-1} + \alpha_2 y_{t-2} + \dots + \alpha_p y_{t-p} + \varepsilon_t. \quad (4.119)$$

Позначимо

$$\alpha = \begin{bmatrix} \alpha_1 \\ \alpha_2 \\ \alpha_3 \\ \dots \\ \alpha_p \end{bmatrix}; \quad r = \begin{bmatrix} r(1) \\ r(2) \\ r(3) \\ \dots \\ r(p) \end{bmatrix}; \quad R = \begin{bmatrix} 1 & r(1) & r(2) & \dots & r(p-1) \\ r(1) & 1 & r(1) & \dots & r(p-2) \\ r(2) & r(1) & 1 & \dots & r(p-3) \\ \dots & \dots & \dots & \dots & \dots \\ r(p-1) & r(p-2) & r(p-3) & \dots & 1 \end{bmatrix}.$$

Тоді коефіцієнти моделі (4.119) можуть бути представлені таким матричним виразом:

$$\alpha = r \cdot R^{-1}. \quad (4.120)$$

На практиці теоретичні значення коефіцієнтів автокореляції не відомі, тому їх замінюють оцінками, які обчислюють за даними вибірки. Підставивши оцінки коефіцієнтів автокореляції в (4.120), отримують оцінки коефіцієнтів AR-моделі. Зазначимо, що оцінювання параметрів AR-моделі за формулою (4.120) рекомендують використовувати лише як перше наближення для подальшого дослідження.

Модель (4.119) можна узагальнити на випадок, коли $M\{y_t\} = \mu \neq 0$:

$$y_t = \alpha_0 + \alpha_1 y_{t-1} + \alpha_2 y_{t-2} + \dots + \alpha_p y_{t-p} + \varepsilon_t. \quad (4.121)$$

Якщо середнє значення рівнів часового ряду відмінне від нуля, то можна або відняти це значення від рівнів і будувати модель без вільного члена, або просто будувати модель з вільним членом α_0 , що пов'язаний із середнім значенням часового ряду таким співвідношенням:

$$\alpha_0 = \mu(1 - \alpha_1 - \alpha_2 - \dots - \alpha_p). \quad (4.122)$$

Для оцінювання коефіцієнтів моделі $AR(p)$ можна використати звичайний МНК, оскільки регресори відносяться до попередніх моментів часу, а ε_t утворюють «білий шум» і не корелюють з регресорами. Отже, задача оцінювання параметрів моделі $AR(p)$ за рівнянням (4.120) аналогічна

оцінюванню її параметрів за МНК, а отримані оцінки вважають незміщеними, ґрунтовними та ефективними.

Актуальною проблемою побудови моделі $AR(p)$ вважають її ідентифікацію, яка полягає у визначення порядку авторегресії. Для цього використовують властивості *автокореляційної* та *частинної автокореляційної функції*.

Частинний коефіцієнт автокореляції характеризує статистичну залежність (кореляцію) між спостереженнями y_t і y_{t-k} за умови, що вплив проміжних членів ряду $y_{t-1}, y_{t-2}, \dots, y_{t-k+1}$ на y_t вилучений. Лаг (зсув у часі) визначає порядок частинного коефіцієнта кореляції.

Частинний коефіцієнт автокореляції першого порядку $r^*(1)$ дорівнює коефіцієнту автокореляції першого порядку ($r(1)$), оскільки немає проміжних лагів. Частинні коефіцієнти автокореляції другого $r^*(2)$ і третього $r^*(3)$ порядків відповідно дорівнюють:

$$r^*(2) = \frac{r(2)-r(1)^2}{1-r(1)^2},$$

$$r^*(3) = \frac{r(3)-r(1)^2r(3)-2r(1)r(2)+r(1)r(2)^2}{1-2r(1)^2+2r(1)^2r(2)-r(2)^2}.$$

У загальному випадку частинну автокореляційну функцію (ЧАКФ) можна обчислювати за такою рекурентною формулою:

$$r(k) = \begin{cases} r(1), & \text{якщо } k = 1; \\ \frac{r^*(k) - \sum_{j=1}^{k-1} r^*(j) \cdot r_{k-j}}{1 - \sum_{j=1}^{k-1} r^*(j) \cdot r_{k-j}}, & k = 2, 3, \dots \end{cases} \quad (4.123)$$

Для вибору порядку авторегресійної моделі $AR(p)$ користуються такими властивостями автокореляційної та частинної автокореляційної функцій:

- значення автокореляційної функції експоненційно спадають, монотонно, або по чергово змінюючи знак;
- коефіцієнт останнього члена авторегресійної моделі p -го порядку співпадає з p -м значенням частинної автокореляційної функції;

• значення частинної автокореляційної функції є статистично значущими (відмінними від нуля) до лагу p включно.

Як попередній порядок моделі $AR(p)$ розглядають число p , починаючи з якого всі наступні оцінки вибіркової частинної автокореляційної функції відхиляються від нуля не більше, ніж на величину $\left| \frac{2}{\sqrt{n}} \right|$ для всіх $k > p$.

AR -моделі передбачають стаціонарність часових рядів і добре описують коливання нестійких показників. Тому в більш широкому розумінні ідентифікація таких моделей включає також вибір способу трансформації початкового часового ряду з метою приведення його до стаціонарності.

Стаціонарність процесу, який генерується авторегресійною моделлю, можна визначати на основі розв'язків такого характеристичного рівняння:

$$1 - \alpha_1 \lambda - \alpha_2 \lambda^2 - \dots - \alpha_p \lambda^p = 0. \quad (4.124)$$

Підставивши в рівняння (4.124) значення коефіцієнтів моделі (4.120), отримаємо p коренів, які можуть бути як дійсними, так і комплексними числами. Якщо всі отримані корені знаходяться за межами одиничного кола (перевищують за модулем 1), то модель авторегресії описує стаціонарний процес. Таким чином, умову стаціонарності для моделей авторегресії записують як $|\lambda| > 1$. Процеси, для яких $|\lambda| = 1$, називають процесами одиничного кореня. Їх відносять до нестаціонарних процесів.

Кінцевий вибір порядку авторегресійної моделі ґрунтується на оцінюванні статистичної значущості її коефіцієнтів та дослідженні залишків, отриманих шляхом віднімання від рівнів початкового часового ряду відповідних значень, встановлених за моделлю $AR(p)$. Якщо для залишків характерні властивості «білого шуму», то підбір моделі завершується, у протилежному разі потрібно змінювати порядок моделі або використовувати більш складні моделі.

У процесі прогнозування намагаються уникати моделей, порядок авторегресії яких більший, ніж два. Така рекомендація зумовлюється тим, що поява елементів третього і четвертого порядку зазвичай вказує на наявність сезонності. Для прогнозування сезонних процесів за допомогою

авторегресій розроблено модифікації розглянутих вище авторегресійних моделей.

Точковий прогноз, який роблять у момент часу t на τ кроків уперед, позначимо через $y_{t+\tau}^{\text{прогн}}$. Тоді прогнози на один, два і τ кроків для моделі $AR(1)$ відповідно розраховують:

$$y_{t+1}^{\text{прогн}} = \alpha y_t, \quad (4.125)$$

$$y_{t+2}^{\text{прогн}} = \alpha y_{t+1}^{\text{прогн}} = \alpha^2 y_t, \quad (4.126)$$

$$y_{t+\tau}^{\text{прогн}} = \alpha^\tau y_t. \quad (4.127)$$

Оскільки в моделі $AR(2)$ рівень часового ряду в момент часу t залежить від двох попередніх значень, то точкові прогнози на один, два і τ кроків обчислюють відповідно за такими рекурентними співвідношеннями:

$$y_{t+1}^{\text{прогн}} = \alpha_1 y_t + \alpha_2 y_{t-1}, \quad (4.128)$$

$$y_{t+2}^{\text{прогн}} = \alpha_1 y_{t+1}^{\text{прогн}} + \alpha_2 y_t, \quad (4.129)$$

$$y_{t+\tau}^{\text{прогн}} = \alpha_1 y_{t+\tau-1}^{\text{прогн}} + \alpha_2 y_{t+\tau-2}^{\text{прогн}}. \quad (4.130)$$

У загальному випадку для моделі $AR(p)$ точковий прогноз на τ ($\tau > p$) кроків обчислюють так:

$$y_{t+\tau}^{\text{прогн}} = \alpha_1 y_{t+\tau-1}^{\text{прогн}} + \alpha_2 y_{t+\tau-2}^{\text{прогн}} + \dots + \alpha_p y_{t+\tau-p}^{\text{прогн}}. \quad (4.131)$$

Доведено, що оцінка прогнозу є незміщеною, а дисперсія прогнозу збігається з дисперсією часового ряду.

Довірчий інтервал прогнозу рівний:

$$y_{t+\tau}^{\text{прогн}} - t_\alpha \hat{\sigma}_\varepsilon \leq y_{t+\tau}^{\text{інт}} \leq y_{t+\tau}^{\text{прогн}} + t_\alpha \hat{\sigma}_\varepsilon, \quad (4.132)$$

де t_α – критичне значення критерію Стюдента для рівня значущості α та $V = n - 2$ ступенів вільності;

$\hat{\sigma}_\varepsilon$ – оцінка середнього квадратичного відхилення випадкової величини.

Наведені вище рекомендації щодо вибору порядку авторегресійних моделей з практичної точки зору є доволі жорсткими. Часто корелограма не дає чіткої відповіді на поставлене питання. Тому дослідник має бути готовим до експериментування і, вибираючи найкращу з поміж адекватних моделей,

користуватися правилом: що меншою є кількість змінних у моделі, тим вона простіша і краща.

Розвиток ідеї щодо залежності поточного рівня часового ряду від попередніх рівнів ґрунтується на припущенні про те, що значення ряду можуть залежати від випадкових помилок попередніх спостережень. Представлення стаціонарних часових рядів у вигляді лінійної комбінації поточної і попередніх помилок називають *моделлю плинної середньої* (МА-моделлю, від англ. *Moving Average*). МА-моделі є авторегресійними моделями, але їх будують не на фактичних значеннях рівнів часових рядів, а на поточних і лагових значеннях випадкової складової.

МА-моделі розглядають як автономні засоби дослідження часових рядів або як доповнення до моделей авторегресії (МА-моделі не слід ототожнювати з моделями адаптивного прогнозування за схемою плинної середньої).

Найпростішою є модель плинної середньої першого порядку $MA(1)$:

$$y_t = \varepsilon_t - \beta_1 \varepsilon_{t-1}, \quad t = \overline{1, n}. \quad (4.133)$$

Середнє значення і дисперсія процесу, який описується моделлю $MA(1)$, відповідно дорівнюють:

$$M\{y_t\} = 0; \quad D\{y_t\} = \sigma^2(1 + \beta_1^2).$$

Теоретично автокореляційна функція пов'язана з параметром моделі $MA(1)$ таким співвідношенням:

$$r(k) = \begin{cases} \frac{-\beta_1}{1+\beta_1^2}, & \text{якщо } k = 1; \\ 0, & \text{якщо } k > 1. \end{cases} \quad (4.134)$$

Відповідно оцінювання параметра β_1 зводиться до розв'язування квадратного рівняння

$$\beta_1^2 + \frac{\beta_1}{r(1)} + 1 = 0, \quad (4.135)$$

причому, з двох коренів рівняння за оцінку параметра β_1 вибирають той, який за абсолютною величиною менший від одиниці, що пов'язано з умовами стаціонарності часового ряду.

Умови стаціонарностей виконуються для довільного процесу $MA(1)$, тобто не залежать від значення параметра β_1 . У випадку $|\beta_1| < 1$ має місце «двоїстість», яка означає, що модель плинної середньої можна перетворити в модель авторегресії і виразити значення поточного рівня не через поточне і попереднє значення випадкової складової, а через поточне значення випадкової компоненти та лагові значення рівнів часового ряду:

$$y_t = \varepsilon_t + \beta_1 y_{t-1} + \beta_1^2 y_{t-2} + \beta_1^3 y_{t-3} + \dots$$

Наприклад, якщо $\beta_1 = 0,5$, то авторегресійне перетворення має вигляд:

$$y_t = \varepsilon_t + 0,5y_{t-1} + 0,25y_{t-2} + \dots$$

Модель плинної середньої другого порядку $MA(2)$ має такий вигляд:

$$y_t = \varepsilon_t - \beta_1 \varepsilon_{t-1} - \beta_2 \varepsilon_{t-2}, \quad t = \overline{1, n}. \quad (4.136)$$

Зв'язок параметрів моделі $MA(2)$ з автокореляційною функцією задають таким співвідношенням:

$$r(k) = \begin{cases} \frac{-\beta_1(1-\beta_1)}{1+\beta_1^2+\beta_2^2}, & \text{якщо } k = 1; \\ \frac{-\beta_2}{1+\beta_1^2+\beta_2^2}, & \text{якщо } k = 2; \\ 0, & \text{якщо } k > 2. \end{cases} \quad (4.137)$$

У загальному випадку модель плинної середньої порядку q $MA(q)$ записують так:

$$y_t = \varepsilon_t - \beta_1 \varepsilon_{t-1} - \beta_2 \varepsilon_{t-2} - \dots - \beta_q \varepsilon_{t-q}. \quad (4.138)$$

Коефіцієнти автокореляції пов'язані з параметрами моделі плинної середньої таким співвідношенням:

$$r(k) = \begin{cases} \frac{-\beta_k + \sum_{i=1}^{q-k} \beta_i \beta_{i+k}}{1 + \sum_{k=1}^q \beta_k^2}, & \text{якщо } k \leq q. \\ 0, & \text{якщо } k > q. \end{cases} \quad (4.139)$$

Оскільки значення y_t в моделі (4.138) залежать від попередніх випадкових помилок, то застосувати МНК для оцінювання параметрів моделі плинної середньої не можна, відтак у процесі оцінювання використовують інші числові методи.

Оцінювання параметрів моделі $MA(q)$ полягає у послідовній реалізації таких дій:

- обчислення на основі даних спостережень вибірових оцінок автокореляцій ($\hat{r}(k)$) для всіх $k = \overline{1, q}$;
- побудова системи рівнянь (4.139) для визначення невідомих параметрів моделі, в якій замість теоретичних кореляцій $r(k)$ використовуються їхні оцінки $\hat{r}(k)$;
- розв'язування побудованої системи рівнянь.

Алгоритм обчислення параметрів моделі плинної середньої чутливий до вибраного порядку моделі (особливо в сторону збільшення). Тому на початковому етапі дослідження не рекомендують розглядати моделі з великою кількістю параметрів.

Для вибору порядку моделі $MA(q)$ використовують таке правило: якщо всі значення частинної автокореляційної функції послідовно змінюють знак і, починаючи з порядку k , рівні нулеві (статистично незначучо відрізняються від нуля), то часовий ряд описують моделлю $MA(k-1)$.

Для моделі плинної середньої $MA(q)$ характеристичне рівняння має такий вигляд:

$$1 + \beta_1 \lambda + \beta_2 \lambda^2 + \dots + \beta_q \lambda^q = 0 \quad (4.140)$$

Процес, який описується моделлю $MA(q)$, завжди є стаціонарним. Крім того, модель авторегресії можна представити моделлю плинної середньої, а модель плинної середньої — авторегресійною моделлю (інакше кажучи, моделі вважають оборотними). Умовою оборотності є виконання нерівності $|\lambda_j| > 1$ ($j = \overline{1, q}$), де λ_j – корені характеристичного рівняння (4.140).

Модель $MA(q)$ як і модель $AR(p)$ можна узагальнити на випадок ненульового математичного сподівання:

$$y_t = \varepsilon_t + \beta_0 + \beta_1 \varepsilon_{t-1} + \dots + \beta_q \varepsilon_{t-q}.$$

Із врахуванням того, що випадкова величина для прогнозованих моментів часу невідома, точкові прогнози для моделі (4.138) на один, два і τ кроків розраховують за такими співвідношеннями:

$$y_{t+1}^{\text{прогн}} = -\beta_1 \varepsilon_t - \beta_2 \varepsilon_{t-1} - \dots - \beta_q \varepsilon_{t-q+1}; \quad (4.141)$$

$$y_{t+2}^{\text{прогн}} = -\beta_2 \varepsilon_t - \beta_3 \varepsilon_{t-1} - \dots - \beta_q \varepsilon_{t-q+2}; \quad (4.142)$$

$$y_{t+\tau}^{\text{прогн}} = -\beta_q \varepsilon_t, \quad \tau \leq q. \quad (4.143)$$

Випадки, коли економічні процеси розвиваються як авторегресійні в чистому вигляді зустрічаються не часто, однак моделі плинної середньої у поєднанні з моделями авторегресії дають змогу адекватно описати реальні економічні процеси.

Часові ряди, що характеризують зміну економічних явищ і процесів, як правило, є нестационарними. Ознаками цього слугують наявність тенденції і сезонної складової, а також зміна взаємозалежностей між рівнями ряду.

Нестационарний часовий ряд можна перетворити у стационарний після вилучення тренда та сезонної компоненти або в результаті віднімання сусідніх рівнів ряду. Процес переходу від нестационарного часового ряду до стационарного за допомогою операторів послідовних різниць (шляхом віднімання сусідніх рівнів), називають *диференціюванням ряду*:

$$\begin{cases} \Delta^1 y_t = y_t - y_{t-1} \\ \Delta^2 y_t = \Delta^1 y_t - \Delta^1 y_{t-1} \\ \dots \\ \Delta^d y_t = \Delta^{d-1} y_t - \Delta^{d-1} y_{t-1}. \end{cases} \quad (4.144)$$

Початковий часовий ряд називають *інтегрованим рядом першого порядку*, якщо його перші різниці утворюють стационарний ряд. Якщо ж для отримання стационарного ряду необхідно обчислити другі різниці, то його називають *інтегрованим рядом другого порядку*, а в загальному випадку — *інтегрованим рядом порядку d* . Оператор послідовних різниць застосовують до того часу, поки не отримають стационарний інтегрований ряд з параметром інтеграції (d).

Нестационарні процеси бувають двох типів:

- процеси, стаціонарні в кінцевих різницях, які представляють моделями трендів, а також процеси з різкою зміною рівнів ряду (із переходом показника на якісно новий рівень);

- процеси, не стаціонарні в кінцевих різницях (всі інші нестационарні процеси).

Якщо нестационарний ряд описується лінійним трендом $y_t = a_0 + a_1 t + \varepsilon_t$, то в результаті застосування оператора перших різниць він стане стаціонарним, оскільки отриманий ряд має постійне математичне сподівання та дисперсію:

$$\Delta y_t = a_0 + a_1 t + \varepsilon_t - (a_0 + a_1(t-1) + \varepsilon_{t-1}) = a_1 + \varepsilon_t - \varepsilon_{t-1}. \quad (4.145)$$

Як і у випадку з лінійним трендом, можна показати, що нестационарний процес, який описується параболічною функцією $y_t = b_0 + b_1 t + b_2 t^2 + \varepsilon_t$, є стаціонарним у других різницях.

Зазначимо, що процеси з різкою зміною рівнів ряду після застосування оператора перших різниць, стануть стаціонарними (рис. 4.7).

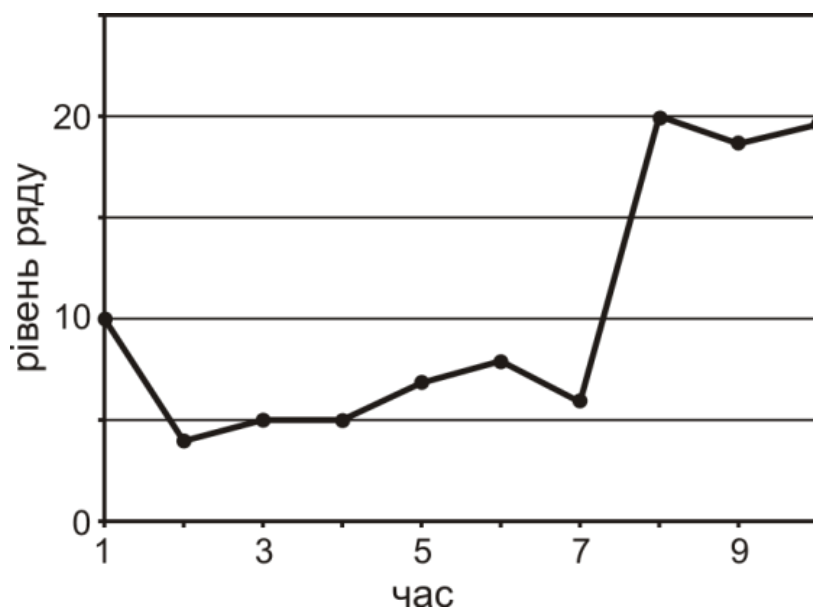


Рис. 4.7. Нестационарний процес із різкою зміною рівнів ряду

Якщо початковий часовий ряд є стаціонарним, то і будь-які його різниці утворюють стаціонарний ряд.

Сезонність вилучають за допомогою сезонних різницевих операторів:

- для річної сезонності: $\Delta y_t^{cp} = y_{t+12} - y_t$;
- для квартальної сезонності: $\Delta y_t^{ckB} = y_{t+4} - y_t$.

Звичайні і сезонні різницеві оператори можна застосовувати до нестаціонарного часового ряду в довільному порядку. Проте, якщо ряд має чітко виражену сезонну тенденцію, то процедуру його диференціювання краще розпочинати з використання сезонних різницевих операторів, оскільки після цього ряд одразу може виявитися стаціонарним.

Використовуючи різницеві оператори, необхідно пам'ятати: що менше разів продиференційовано ряд, тим менша дисперсія підсумкового прогнозу.

Необхідно враховувати, що використання послідовних різниць з метою отримання стаціонарного ряду не завжди покращує точність прогнозу. А для процесів, яким характерні постійні зміни всіх статистичних параметрів, підхід отримання стаціонарних рядів на основі послідовних різниць може призвести до помилкових результатів прогнозування. Це пояснюється тим, що ряд стає стаціонарним у статистичному розумінні і не представляє реальний процес, оскільки тенденція розвитку процесу на періоді прогнозування може змінитися.

Нестаціонарний ряд можна також звести до стаціонарного вигляду за допомогою інших процедур, зокрема:

- логарифмуванням рівнів ряду: $x_t = \ln y_t, \quad t = \overline{1, n}$;
- логарифмуванням ланцюгових індексів: $x_t = \ln(y_t/y_{t-1}), \quad t = \overline{2, n}$;
- заміною рівнів ряду темпами зростання: $x_t = \frac{y_t}{y_{t-1}}, \quad t = \overline{2, n}$.

Вибираючи процедури перетворення, слід враховувати вигляд графіка, який представляє нестаціонарний початковий ряд.

Висновок щодо стаціонарності чи нестаціонарності часового ряду можна зробити на підставі дослідження автокореляційної та частинної

автокореляційної функції. Для отримання більш адекватної оцінки стаціонарності часових рядів розроблено тести, які ґрунтуються на перевірці статистичних гіпотез. Прикладом формального критерію для перевірки рядів на стаціонарність і визначення порядку інтеграції ряду є *тест Дікі-Фуллера (DF-тест)*.

Нехай початковий часовий ряд описується авторегресійною моделлю $AR(1)$. Якщо коефіцієнти цієї моделі є рівними або більшими за одиницю, то ряд є нестаціонарним. Цю властивість покладено в основу DF -тесту.

Віднімаючи y_{t-1} від лівої і правої частини моделі (4.114), отримаємо таку модель:

$$\Delta y_t = b y_{t-1} + \varepsilon_t, \quad (4.146)$$

де вжито такі позначення: $\Delta y_t = y_t - y_{t-1}$; $b = \alpha_1 - 1$.

Враховуючи зв'язок між коефіцієнтами початкової і трансформованої моделі, отримуємо:

- якщо $b = 0$, то $\alpha_1 = 1$, і ряд не є стаціонарним (інтегрований ряд першого порядку);
- якщо $b < 0$, то $\alpha_1 < 1$, і ряд є стаціонарним.

Після оцінювання параметрів рівняння регресії (4.146) за допомогою МНК формують нульову (H_0) та альтернативну (H_1) гіпотези:

- $H_0: b = 0$ — часовий ряд є нестаціонарним;
- $H_1: b < 0$ — часовий ряд є стаціонарним.

DF -статистика — це статистика для перевірки значущості коефіцієнта лінійної регресії (відношення значення оцінюваного параметра до його середнього квадратичного відхилення):

$$DF = \frac{\hat{b}}{\hat{\sigma}_b}, \quad (4.147)$$

де \hat{b} — оцінка коефіцієнта регресії (4.146) за МНК; $\hat{\sigma}_b$ — оцінка стандартної помилки коефіцієнта регресії.

Розподіл DF -статистики відрізняється від класичного розподілу Стьюдента. Для різних рівнів значущості розроблено таблиці критичних значень тесту залежно від вигляду рівняння регресії та обсягу спостережень. Якщо розрахункове значення DF -статистики менше від критичного значення (критичні значення є від'ємними) за заданого рівня значущості, тобто знаходиться лівіше від критичного значення, то нульову гіпотезу відхиляють.

Крім моделі (4.146), перевірку на стаціонарність можна проводити для моделей:

- з константою, але без тренда: $\Delta y_t = a_0 + by_{t-1} + \varepsilon_t$; (4.148)

- з константою та лінійним трендом: $\Delta y_t = a_0 + a_1t + by_{t-1} + \varepsilon_t$. (4.149)

Тестування моделей (4.148) і (4.149) дає додаткову інформацію про початковий часовий ряд, наприклад, статистична значущість параметра біля змінної t вказує на наявність лінійного тренда.

Для кожної з трьох тестових регресій існують свої критичні значення DF -статистики (табл. 4.7).

Таблиця 4.7

Критичні значення DF -статистики для 5% рівня значущості

Види моделей	Обсяг вибірки					
	25	50	100	250	500	∞
$\Delta y_t = by_{t-1} + \varepsilon_t$	-1,95	-1,95	-1,95	-1,95	-1,95	-1,95
$\Delta y_t = a_0 + by_{t-1} + \varepsilon_t$	-3,00	-2,93	-2,89	-2,88	-2,87	-2,86
$\Delta y_t = a_0 + a_1t + by_{t-1} + \varepsilon_t$	-3,60	-3,50	-3,45	-3,43	-3,42	-3,41

Якщо нульову гіпотезу не відхиляють, то ряд є нестаціонарним, і можливі два варіанти:

- 1) часовий ряд має вищий порядок інтеграції;
- 2) часовий ряд не можна звести до стаціонарного вигляду за допомогою операторів послідовних різниць.

DF-тест можна застосовувати лише за умови некорельованості випадкових величин ε_t . Якщо ця умова не виконується, то до моделей, які використовують для тестування, необхідно включати лагові значення послідовних різниць, тобто використовувати *розширений тест DF* (*ADF*-тест).

З метою більш повного врахування структури часового ряду в основу *ADF*-тесту покладено регресію:

$$\Delta y_t = b y_{t-1} + c_1 \Delta y_{t-1} + c_2 \Delta y_{t-2} + \dots + c_p \Delta y_{t-p} + \varepsilon_t, \quad (4.150)$$

тобто до базової моделі (4.146) включено p лагових значень залежної змінної у різницях. Як правило, *ADF*-тест використовують з кількістю лагів, що не перевищує 10 % від кількості спостережень. Як один із можливих варіантів вирішення проблеми вибору лага є побудова $(p + 1)$ моделей вигляду (4.150) і вибір із них моделі з найменшим значенням інформаційного критерію.

До моделі (4.150) можна додатково включити константу та лінійний тренд. Процедура тестування часового ряду на стаціонарність за допомогою *ADF*-тесту ідентична процедурі *DF*-тесту.

Для того, щоб пересвідчитися в достовірності результатів тестування часових рядів на стаціонарність, доцільно використовувати декілька тестів, підкріплюючи отримані висновки графічними процедурами.

Об'єднання авторегресійного процесу з процесом плинної середньої породжує ще один клас моделей часових рядів — *моделі ARMA* (від англ. *Auto Regressive Moving Average model* — *авторегресії та плинної середньої*).

У загальному випадку модель авторегресії та плинної середньої порядків p (кількість лагів стосовно авторегресії) і q (кількість лагів стосовно плинної середньої) (модель *ARMA*(p, q)) має такий вигляд:

$$y_t = \alpha_1 y_{t-1} + \alpha_2 y_{t-2} + \dots + \alpha_p y_{t-p} + \varepsilon_t - \beta_1 \varepsilon_{t-1} - \beta_2 \varepsilon_{t-2} - \dots - \beta_q \varepsilon_{t-q}. \quad (4.151)$$

За допомогою моделі $ARMA(p, q)$ можна компактно і доволі точно представити будь-який стаціонарний процес за наявності достатнього масиву даних, проте коефіцієнти моделі не піддаються змістовній інтерпретації.

Модель $ARMA(p, q)$ забезпечує розроблення прогнозу, який залежить як від поточного і попереднього значень рівнів часового ряду, так і від поточних і попередніх значень випадкової величини.

Використання $ARMA$ -моделей переважно не вимагає економіко-теоретичного обґрунтування їхньої специфікації. Необхідною умовою їхнього практичного застосування є стаціонарність процесу, який розглядають. Процес, який описується за допомогою моделі $ARMA(p, q)$, є стаціонарним, якщо всі корені характеристичного рівняння (4.124) задовольняють умову: $|\lambda_i| > 1, i = \overline{1, p}$.

Переважна більшість часових рядів, які представляють економічні процеси і явища, є нестаціонарними, але за допомогою оператора послідовних різниць їх приводять до стаціонарних.

Узагальненням моделі $ARMA(p, q)$ на випадок використання оператора послідовних різниць є модель $ARIMA(p, d, q)$ — *інтегрована модель авторегресії – плинної середньої* (від англ. *Auto Regressive Integrated Moving Average model*). Зміст параметрів моделі $ARIMA(p, d, q)$ є таким:

- p — порядок авторегресії;
- d — порядок інтегрування часового ряду;
- q — порядок плинної середньої.

Сукупність процедур визначення, коригування і перевірки $ARMA$ -моделей на адекватність на основі даних часових рядів прийнято називати *методологією прогнозування Бокса–Дженкінса* (від імені авторів Дж. Бокса та Г. Дженкінса).

За методологією Бокса–Дженкінса задача побудови моделі $ARIMA$ передбачає виконання 4-х етапів:

- *ідентифікація моделі;*

- оцінювання параметрів моделі;
- перевірка моделі на адекватність;
- прогнозування на основі побудованої моделі.

Ідентифікація моделі $ARIMA(p, d, q)$ полягає у визначенні таких величин: p, d, q .

Оскільки $ARMA$ -моделі застосовують лише для стаціонарних часових рядів, то першою стадією етапу ідентифікації є перевірка часового ряду на стаціонарність. У випадку нестаціонарності часового ряду застосовують оператор послідовних різниць і встановлюють порядок (ступінь) інтегрованості часового ряду d . На другій стадії ідентифікації моделі визначають оптимальні лаги стосовно авторегресії (p) і плинної середньої (q).

Будуючи модель часового ряду, намагаються мінімізувати кількість її параметрів і, як наслідок, порядок моделі. Такий підхід до моделювання пояснюється тим, що параметри моделі оцінюють на основі коефіцієнтів автокореляції. Зі збільшенням порядку моделі виникає потреба в збільшенні вибірових коефіцієнтів автокореляції, що супроводжується зниженням точності оцінювання параметрів моделі. У зв'язку з цим виникає необхідність пошуку моделі з найменшою кількістю параметрів порівняно з іншими можливими варіантами і найбільш адекватної реальному процесу. Для практичних цілей найчастіше користуються закономірностями, які пов'язують параметри моделі з поведінкою її автокореляційної і частинної автокореляційної функцій.

У випадку, коли вибірові коефіцієнти автокореляції зі зростанням лагу експоненційно прямують до нуля, а коефіцієнти частинної автокореляції швидко відсікаються, модель повинна включати авторегресійні складові.

У тому разі, коли вибірові коефіцієнти автокореляції швидко відсікаються, а коефіцієнти частинної автокореляції прямують до нуля, в моделі повинні бути складові плинної середньої.

Якщо як вибірковий коефіцієнт автокореляції, так і вибірковий коефіцієнт частинної автокореляції плавно прямують до нуля, то це вказує на необхідність включення в модель складових обох типів.

У випадку великої кількості спостережень (n) коефіцієнти автокореляції та частинної автокореляції вважають статистично значущими, якщо вони відхиляються від нуля не більше, ніж на величину $|2/\sqrt{n}|$.

Методологія Бокса–Дженкінса не передбачає якої-небудь особливої структури рівнів часового ряду, а ґрунтується на ітеративному підході до виявлення оптимальної (або прийнятної) моделі з поміж загального класу *ARIMA*-моделей.

Якщо досліджувана модель не забезпечує досягнення поставленої мети, то переходять до вибору та аналізу іншої моделі. Ітеративну процедуру перебору й аналізу моделей повторюють до тих пір, поки не буде знайдено модель, яка задовольняє дослідника (рис. 4.8).

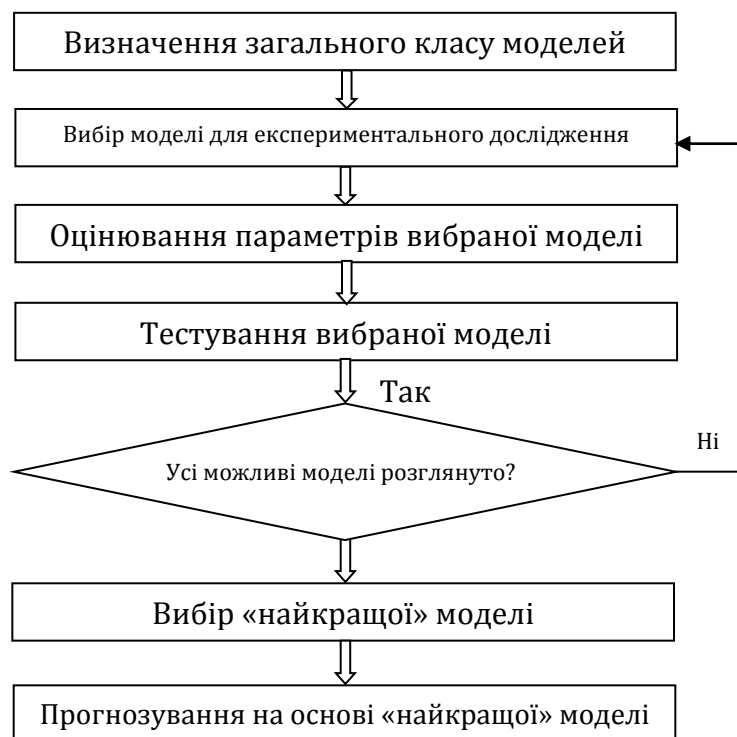


Рис. 4.8. Прогнозування за методологією Бокса–Дженкінса

Модель $ARIMA(p, d, q)$ (модель Бокса–Дженкінса) призначена для опису часових рядів, в яких не випадкова складова $f(t)$ входить до часового ряду адитивно і має вигляд алгебраїчного полінома.

Характеристики автокореляційної та частинної автокореляційної функцій для найпростіших $ARMA$ -моделей представлено в табл. 4.8.

Таблиця 4.8

Властивості автокореляційної і частинної автокореляційної функцій

Моделі	Функції	
	АКФ	ЧАКФ
$ARMA(1, 0)$	Експоненційно спадає, монотонно або по чергово змінюючи знак. За від'ємного значення авторегресії змінює знак, за додатного – не змінює	Має викид (пік, різко виділене значення) на лагу 1. Для інших лагів кореляція відсутня
$ARMA(2, 0)$	Експоненційно або синусоїдально спадає. За від'ємного значення авторегресії змінює знак, за додатного – не змінює	Має викиди на лагах 1 і 2. Для інших лагів кореляція відсутня
$ARMA(0, 1)$	Має викид на лагу 1. Для інших лагів кореляція відсутня	Експоненційно спадає, монотонно або по чергово змінюючи знак
$ARMA(0, 2)$	Має викиди на лагах 1 і 2. Для інших лагів кореляція відсутня	Експоненційно або синусоїдально спадає
$ARMA(1, 1)$	Починаючи з лага 1 експоненційно спадає, монотонно або по чергово змінюючи знак	Починаючи з лага 1 експоненційно спадає, монотонно або по чергово змінюючи знак

Рекомендації щодо вибору $ARMA$ -моделі на підставі візуального аналізу автокореляційної та частинної автокореляційної функцій є доволі умовними і стосуються типових ситуацій. Реальні часові ряди можуть виявитися складнішими, а тому для їхньої ідентифікації графічний аналіз необхідно доповнювати формальними тестами.

Оскільки на практиці використовують моделі, для яких $p + q \leq 3$ і $d \leq 2$, то в процесі вибору типу моделі можна просто перебрати всі моделі зі значенням параметрів у зазначених межах та протестувати їх на точність і адекватність.

Представлений вище метод оцінювання параметрів моделі *ARMA* (*метод моментів*) не гарантує отримання ефективних оцінок. Існують інші методи, які дають змогу підвищити точність параметрів моделей цього класу. Зокрема, у сучасних прикладних програмах, в яких реалізовано процедури прогнозування, найчастіше використовується *метод максимальної правдоподібності*, який вважають універсальним методом оцінювання невідомих параметрів моделей часових рядів. За цим методом припускають, що емпіричні дані мають певний імовірнісний розподіл, і обчислюють імовірність події, пов'язаної з невідомими параметрами моделі. На основі емпіричних даних намагаються знайти максимальне значення ймовірності цієї події. Коефіцієнти, за яких досягається максимум імовірності події, вважають оцінками параметрів моделі. Якщо знайти оцінки за допомогою аналітичних методів не вдається, то використовують рекурентні числові методи пошуку мінімуму суми квадратів відхилень значень ряду від значень, обчислених за моделлю.

Більш детально з методом максимальної правдоподібності можна ознайомитися в спеціальній літературі. Зазначимо, що наявні сучасні програмні засоби забезпечують можливість оцінювання параметрів моделей дослідникам, які не володіють ґрунтовними знаннями з математичної статистики. Дослідник повинен уміти інтерпретувати побудовані моделі, оцінювати їхню адекватність та якість, робити практичні рекомендації.

Третій етап реалізації методології Бокса-Дженкінса передбачає тестування оціненої моделі на адекватність. У цьому разі необхідно перевірити:

- статистичну значущість оцінених коефіцієнтів моделі;
- відповідність залишків (помилки) моделі процесу «білого шуму».

Якщо в результаті перевірки буде з'ясовано, що адекватними є декілька моделей, то для вибору оптимальної моделі необхідно враховувати ще дві вимоги: вищу точність моделі; меншу кількість параметрів моделі.

Зазначені вимоги враховано в інформаційних критеріях *Акайке* і *Шварца*, які ґрунтуються відповідно на *AIC*-статистиці і *SC*-статистиці:

$$AIC = \ln\left(\frac{SSE}{n}\right) + \frac{2m}{n}; \quad (4.155)$$

$$SC = \ln\left(\frac{SSE}{n}\right) + \frac{m}{n} \ln(n), \quad (4.156)$$

де *SSE* – сума квадратів залишків; $m = p + q + f$, ($f = 1$, якщо використовують модель з вільним членом, $f = 0$ – у протилежному випадку).

За цими критеріями вибирають модель з меншими значеннями статистик *AIC* та *SC*.

Після вибору оптимальної моделі приступають до *прогнозування*. Як правило, період прогнозування рекомендують вибирати не більшим ніж третина довжини початкового часового ряду.

У практиці економічного прогнозування доволі часто використовують модель *ARMA*(1, 1):

$$y_t = \alpha_1 y_{t-1} + \varepsilon_t - \beta_1 \varepsilon_{t-1}. \quad (4.157)$$

Вибір моделі (4.157) обґрунтовують тим, що складова, яка належить до авторегресії першого порядку, поглинає всі процеси плинної середньої більш високих порядків і, навпаки, складова, яка належить до плинної середньої, поглинає процеси авторегресії високих порядків. Автокореляційна і частинна автокореляційна функції для моделей *AR*(1) і *MA*(1) експоненційно згасають після першої затримки.

Прогноз на момент часу $t + 1$ на основі моделі *ARMA*(1, 1) розраховують за формулою

$$y_{t+1}^{\text{прогн}} = \alpha_1 \left(1 - \frac{\beta_1}{\alpha_1}\right) y_t. \quad (4.158)$$

Якщо потрібно знайти прогноз на τ ($\tau > 1$) кроків уперед, то користуються таким рекурентним співвідношенням:

$$y_{t+\tau}^{\text{прогн}} = \alpha_1^\tau \left[\left(1 - \frac{\beta_1}{\alpha_1}\right) \right] y_t + \frac{\beta_1}{\alpha_1} y_{t+\tau-1}^{\text{прогн}}. \quad (4.159)$$

У загальному випадку формули для прогнозування процесів, що описуються моделлю $ARMA(p, q)$, отримують об'єднанням формул (4.128) — (4.130) з формулами (4.141) — (4.143). У такому разі реальні параметри α і β замінюють їхніми оцінками $\hat{\alpha}$ і $\hat{\beta}$, а випадкові величини ε_t — залишками e_t , отриманими в процесі оцінювання моделі, або помилками попередніх прогнозів. Випадкові величини для періоду, на який складають прогноз, невідомі, і їх замінюють нульовими значеннями (вважають, що їхнє середнє значення дорівнює нулю). Якщо з'являються значення ряду «у майбутньому», то їх замінюють прогнозними значеннями, отриманими на відповідний момент часу.

Якщо прогноз розроблено на основі даних, отриманих у результаті застосування різницевих операторів, то, використовуючи формули приведення ряду до стаціонарного вигляду, необхідно повернутися до початкового часового ряду. Зокрема, у випадку, коли початковий часовий ряд є інтегрованим рядом першого порядку, використовують такі формули:

$$z_t = \Delta y_t = y_t - y_{t-1}; \quad y_t = z_t + y_{t-1}; \quad y_{t+1}^{\text{прогн}} = z_{t+1}^{\text{прогн}} + y_t. \quad (4.160)$$

Узагальненням моделі $ARMA(p, q)$ на випадок часових рядів з проявом сезонності є модель $SARMA$ (*Seasonal ARMA*). Цю модель отримують у результаті додавання до правої частини моделі $ARMA$ P авторегресійних компонент

$$\alpha_s y_{t-s} + \alpha_{2s} y_{t-2s} + \dots + \alpha_{P_s} y_{t-P_s}, \quad (4.161)$$

та Q компонент плинної середньої

$$\beta_s \varepsilon_{t-s} + \beta_{2s} \varepsilon_{t-2s} + \dots + \beta_{Q_s} \varepsilon_{t-Q_s}, \quad (4.162)$$

де s — тривалість періоду сезонності.

Модифікацією моделі $ARIMA(p, d, q)$ є модель $SARIMA(p, d, q) \cdot (P, D, Q)$ — модель $SARMA(p, q) \cdot (P, Q)$ для ряду, до якого d разів застосовували звичайне диференціювання і D разів — сезонне. Бажано, щоб виконувалася умова $d + D \leq 2$.

Визначення параметрів моделі *SARIMA*, як і моделі *ARIMA*, ґрунтується на дослідженні автокореляційної та частинної автокореляційної функцій, у яких всі типові прояви (викиди) будуть віддалені один від одного на величину періоду сезонності.

Моделі класу *ARIMA* — це потужний арсенал засобів представлення і моделювання широкого спектру характеристик емпіричних часових рядів. Разом з тим, моделі *ARIMA* мають і недоліки:

- побудова моделі передбачає наявність великої кількості спостережень (зокрема, для моделювання несезонних процесів рекомендують оперувати не менше ніж 40 спостереженнями, а для моделювання сезонних явищ необхідно зібрати дані за 6–10 років);

- не існує простого способу коригування параметрів моделі в разі надходження нових даних (можливо, виникне потреба побудувати цілком нову модель);

- процес прогнозування є доволі трудомістким, оскільки може потребувати дослідження багатьох моделей з метою вибору оптимальної, а тому його реалізація вимагає використання спеціалізованих програмних продуктів.

4.3. Багатофакторне прогнозування на основі класичної регресії та регресійних моделей на змішаних факторних множинах

Економічні явища і процеси часто визначаються великою кількістю чинників, які діють одночасно та в сукупності.

Для кількісного аналізу впливу чинників (незалежних змінних) x_1, x_2, \dots, x_m на результатний показник y застосовують моделі множинної регресії, які слугують засобами багатофакторного прогнозування.

У загальному випадку модель множинної регресії можна записати у вигляді

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

де y – залежна змінна; $x_j, j = \overline{1, m}$ – незалежні змінні; ε – випадкова складова.

Незалежні змінні називають *пояснювальними змінними*, а залежну змінну — *пояснюваною змінною*.

Між цими змінними існує стохастичний зв'язок. На початковому етапі дослідження невідомі ні кількість, ні склад пояснювальних змінних, ні аналітична форма зв'язку. Тому необхідно здійснити *специфікацію моделі*, яка охоплює виокремлення із загальної сукупності факторних ознак найбільш впливових та вибір аналітичної форми залежності.

Помилкова специфікація моделі може бути спричинена таким:

- у моделі не враховані впливові факторні ознаки;
- у моделі представлено факторні ознаки, які несуттєво впливають на значення результатного показника;
- математична форма зв'язку між змінними в моделі не відповідає такій, що існує.

Першим кроком, який потрібно зробити в процесі побудови множинної регресійної моделі, є відбір факторних ознак, які впливають на результатний показник, тобто сформувані сукупність пояснювальних змінних. Початкову сукупність пояснювальних змінних визначають на підставі системного аналізу якісного змісту об'єкта дослідження з виокремленням тих змінних, для яких існує можливість прямого чи опосередкованого кількісного вираження за доступною статистичною інформацією. Кінцевий відбір пояснювальних змінних відбувається на етапах оцінювання і тестування побудованої моделі.

Якщо до моделі не увійдуть впливові змінні, то оцінки коефіцієнтів регресії і дисперсія залишків можуть бути зміщеними, а значення дисперсій оцінок коефіцієнтів регресії — заниженими.

У випадку включення до моделі пояснювальної змінної, яка суттєво не впливає на пояснювану змінну, оцінки параметрів рівняння регресії та їхніх

дисперсій будуть незміщеними, але точність оцінок, отриманих за допомогою методу МНК, може знизитися.

Після того, як сформовано сукупність пояснювальних змінних, потрібно встановити вигляд співвідношення (форму зв'язку) між результатним показником і факторними ознаками, що його обумовлюють. Вказати точний функціональний вигляд такого співвідношення доволі складно, тому мова йде про апроксимацію зв'язку між змінними відносно простими функціями. На практиці перевагу віддають лінійним моделям або таким, які можна звести до лінійного вигляду шляхом заміни і перетворення початкових змінних. Використання нелінійних моделей, як правило, пов'язано з ускладненням оцінювання параметрів моделі та зниженням точності результатів моделювання.

Позначимо:

$$Y = \begin{bmatrix} y_1 \\ y_2 \\ \dots \\ y_n \end{bmatrix}, \quad X = \begin{bmatrix} 1 & x_{11} & x_{12} & \dots & x_{1m} \\ 1 & x_{21} & x_{22} & \dots & x_{2m} \\ \dots & \dots & \dots & \dots & \dots \\ 1 & x_{n1} & x_{n2} & \dots & x_{nm} \end{bmatrix}, \quad \beta = \begin{bmatrix} \beta_0 \\ \beta_1 \\ \dots \\ \beta_m \end{bmatrix},$$

$$B = \begin{bmatrix} b_0 \\ b_1 \\ \dots \\ b_m \end{bmatrix}, \quad \varepsilon = \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \dots \\ \varepsilon_n \end{bmatrix}, \quad E = \begin{bmatrix} e_1 \\ e_2 \\ \dots \\ e_n \end{bmatrix},$$

де Y – вектор (матриця-стовпець) значень спостережень за пояснюваною змінною; X – матриця значень спостережень за пояснювальними змінними (до матриці штучно включено перший стовпець, всі елементи якого дорівнюють одиниці); β – вектор (матриця-стовпець) параметрів моделі; B – вектор (матриця-стовпець) оцінок параметрів моделі; ε – вектор (матриця-стовпець) не спостережуваних (випадкових) величин; E – вектор (матриця-стовпець) залишків (відхилень між фактичними і теоретичними (оціненими) значеннями пояснюваної змінної).

З урахуванням прийнятих позначень загальна і вибіркова моделі множинної лінійної регресії відповідно набувають такого матричного вигляду:

$$Y = X\beta + \varepsilon; \quad (4.163)$$

$$\hat{Y} = XB + E. \quad (4.164)$$

Побудова класичної регресійної моделі ґрунтується на припущеннях:

- модель специфікована правильно (зв'язок між пояснюваною змінною і пояснювальними змінними є лінійним, вибрані факторні ознаки є впливовими);

- математичне сподівання випадкової складової ε для кожного спостереження дорівнює нулю: $M(\varepsilon_i) = 0, i = \overline{1, n}$;

- дисперсія випадкової складової ε є сталою для всіх спостережень:

$$\sigma_{\varepsilon_i}^2 = const = \sigma^2, i = \overline{1, n};$$

- значення випадкових складових для i -го і j -го спостережень не корелюють між собою: $cov(\varepsilon_i, \varepsilon_j) = 0, i = \overline{1, n}; j = \overline{1, n}; i \neq j$;

- випадкові складові моделі та пояснювальні змінні розподілені нормально для одних і тих же спостережень, тобто не корелюють між собою (припущення виконується автоматично, якщо пояснювальні змінні є не випадковими величинами): $cov(x_{ij}, \varepsilon_i) = 0, i = \overline{1, n}; j = \overline{1, m}$;

- змінні x_1, x_2, \dots, x_m не залежні між собою (відсутня мультиколінеарність): $cov(x_i, x_j) = 0, i = \overline{1, n}; j = \overline{1, m}; i \neq j$.

Оцінки параметрів моделі (9.1) обчислюють за такою формулою:

$$B = (X^T \cdot X)^{-1} X^T \cdot Y. \quad (4.165)$$

Зазначимо, що оцінки $b_0, b_1, b_2, \dots, b_m$ невідомих параметрів $\beta_0, \beta_1, \beta_2, \dots, \beta_m$ моделі множинної лінійної регресії (4.163), які визначено за МНК з дотриманням основних припущень, є найкращими лінійними оцінками. Оцінений коефіцієнт регресії b_j характеризує *чистий вплив*

пояснювальної змінної x_j на пояснювану змінну y за умови, що модель містить усі фактори, які впливають на результатний показник.

Для дослідження впливу факторних ознак на результатний показник потрібно враховувати, що значення коефіцієнтів регресії не завжди можна порівнювати між собою, оскільки вони залежать від вибраних одиниць виміру. З метою встановлення порівняльної міри впливу кожної пояснювальної змінної на пояснювану змінну рівняння регресії представляють у стандартизованому (нормалізованому) вигляді. Для цього всі змінні перетворюють за допомогою співвідношень:

$$y^* = \frac{y - \bar{y}}{\sigma_y}, \quad (4.166)$$

$$x_j^* = \frac{x_j - \bar{x}_j}{\sigma_{x_j}}, \quad (4.167)$$

де y^* – стандартизовані значення пояснюваної змінної; x_j^* – стандартизовані значення j -ї пояснювальної змінної; \bar{y} , \bar{x}_j – середні значення пояснюваної і j -ї пояснювальної змінної відповідно; σ_y , σ_{x_j} – середньоквадратичні відхилення пояснюваної і пояснювальних змінних відповідно.

Вільний член у стандартизованому рівнянні множинної лінійної регресії відсутній, тобто вибіркова модель має вигляд

$$\hat{y}^* = b_1^* x_1^* + b_2^* x_2^* + \dots + b_m^* x_m^*, \quad (4.168)$$

де b_j^* – оцінені стандартизовані коефіцієнти регресії (бета-коефіцієнти), які характеризують рівень впливу пояснювальної змінної на пояснювану (вимірюються значеннями середньоквадратичних відхилень).

Що більшим за абсолютною величиною є значення бета-коефіцієнта b_j^* , тим відчутнішим є вплив j -ї пояснювальної змінної на результат.

Іншим показником порівняння впливу пояснювальних змінних на пояснювану є середній коефіцієнт еластичності (E_{x_j})

$$E_{x_j} = b_j \frac{\bar{x}_j}{\bar{y}}, \quad i = \overline{1, m}, \quad (4.169)$$

який слугує мірою впливу пояснювальної змінної на пояснювану у відсотках.

Він показує на скільки відсотків у середньому зміниться результатний показник за зміни відповідної факторної ознаки на 1 % від своєї середньої величини (за умови постійних значень усіх інших пояснювальних змінних моделі). Порівняння середніх показників еластичності забезпечує ранжування факторних ознак за силою їхнього впливу на результатний показник.

Для оцінювання статистичної якості побудованої моделі використовують такі основні статистичні характеристики: стандартну помилку моделі; множинні коефіцієнти детермінації і кореляції; критерії перевірки істотності зв'язку та впливу окремих факторів на результатний показник.

Величину

$$\hat{\sigma}_\varepsilon^2 = \frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{n-m-1} = \frac{\mathbf{Y}^T \mathbf{Y} - \mathbf{B}^T \mathbf{X}^T \mathbf{Y}}{n-m-1} \quad (4.170)$$

називають *оцінкою дисперсії відхилень (оцінкою непоясненої дисперсії)*, а $\hat{\sigma}_\varepsilon = \sqrt{\hat{\sigma}_\varepsilon^2}$ — *стандартною помилкою* множинної лінійної регресійної моделі.

Для моделі множинної лінійної вибіркової регресії має місце рівність

$$SST = SSR + SSE, \quad (4.171)$$

де SST — загальна сума квадратів відхилень пояснюваної змінної від середнього значення, що зумовлена впливом усіх можливих факторів; SSR — сума квадратів відхилень, що зумовлена впливом факторних ознак, включених до моделі; SSE — сума квадратів залишків, яка характеризує ступінь впливу неврахованих факторів.

Частку загальної варіації, яку пояснює варіація включених до моделі факторів, називають *множинним коефіцієнтом детермінації* (R^2):

$$R^2 = \frac{SSR}{SST} = 1 - \frac{SSE}{SST} = \frac{\mathbf{B}^T \mathbf{X}^T \mathbf{Y} - n\bar{y}^2}{\mathbf{Y}^T \mathbf{Y} - n\bar{y}^2}. \quad (4.172)$$

Арифметичне значення кореня квадратного з коефіцієнта детермінації $R = +\sqrt{R^2}$ називають *коефіцієнтом множинної кореляції*. Він характеризує щільність зв'язку між результатним показником і факторами, включеними до рівняння множинної лінійної регресії.

Коефіцієнт детермінації застосовують для тестування адекватності моделі множинної лінійної регресії. Для цього формулюють такі статистичні гіпотези:

- H_0 : для генеральної сукупності $\beta_1 = \beta_2 = \dots = \beta_m = 0$, тобто між пояснюваною і пояснювальними змінними моделі відсутній лінійний зв'язок;

- H_1 : для генеральної сукупності не всі коефіцієнти регресії одночасно дорівнюють нулю, тобто зв'язок між пояснюваною і пояснювальними змінними моделі статистично значущий.

Для тестування вказаних вище гіпотез використовують критерій Фішера, значення якого розраховують на підставі коефіцієнта детермінації за формулою

$$F^{\text{розр}} = \frac{V_2}{V_1} \cdot \frac{R^2}{1-R^2}, \quad (4.173)$$

і порівнюють з критичним (табличним) значенням $F^{\text{крит}}(\alpha; V_1; V_2)$ за заданого рівня значущості α і $V_1 = m$ та $V_2 = n - m - 1$ ступенях вільності (m – кількість пояснювальних змінних моделі, n – обсяг вибірки).

Якщо $F^{\text{розр}} > F^{\text{крит}}$, то гіпотезу H_0 відхиляють і приймають альтернативну гіпотезу H_1 про статистичну значущість лінійного зв'язку.

Для перевірки статистичної значущості параметрів β_j рівняння регресії формулюють такі статистичні гіпотези:

- H_0 : для генеральної сукупності $\beta_j = 0$;

- H_1 : для генеральної сукупності $\beta_j \neq 0$.

На рівні значущості α і $V = n - m - 1$ ступенів вільності нульову гіпотезу H_0 відхиляють, якщо

$$t_{b_j}^{\text{розр}} = \frac{|b_j|}{\hat{\sigma}_{b_j}} > t_{\text{кр}}(\alpha, V), \quad (4.174)$$

де $\hat{\sigma}_{b_j}$ – стандартна вибіркова помилка j -го коефіцієнта рівняння регресії (b_j).

Оцінка дисперсії j -го коефіцієнта рівняння регресії дорівнює

$$\hat{\sigma}_{b_j}^2 = c_{jj} \hat{\sigma}_\varepsilon^2, \quad j = \overline{0, m}, \quad (4.175)$$

де c_{jj} – діагональні елементи матриці $\mathbf{C} = (\mathbf{X}^T \mathbf{X})^{-1}$.

Статистична незначущість коефіцієнта регресії свідчить про те, що на основі вибірки, використаної для побудови рівняння регресії, не можна гарантувати надійність оцінки впливу відповідної пояснювальної змінної на пояснювану.

Знаючи стандартну вибірку помилку ($\hat{\sigma}_{b_j}$) коефіцієнта множинної лінійної регресії (b_j), можна розрахувати довірчий інтервал для значення параметра (β_j) в генеральній сукупності:

$$b_j - t_{\text{кр}}(\alpha, V) \hat{\sigma}_{b_j} \leq \beta_j \leq b_j + t_{\text{кр}}(\alpha, V) \hat{\sigma}_{b_j}, \quad (4.176)$$

де $t_{\text{кр}}$ – критичне значення розподілу Стьюдента для рівня значущості α і $V = n - m - 1$ ступенів вільності.

Граничною помилкою параметра β_j називають величину

$$\Delta b_j = t^{\text{крит}}(\alpha, V) \hat{\sigma}_{b_j}. \quad (4.177)$$

Якщо кількість факторних ознак дорівнює m , то можна побудувати 2^m різних регресійних моделей. Очевидно, що перед дослідником стоїть завдання вибрати оптимальну («найкращу») модель, яка відображає основні закономірності об'єкта прогнозування з достатнім рівнем статистичної надійності. З поміж існуючих методів вибору оптимальної моделі найчастіше користуються двома методами — покрокового включення або покрокового виключення змінних. За методом покрокового включення до моделі на кожному кроці долучають пояснювальну змінну, яка дає найбільший внесок у

регресію. Метод покрокового виключення діє у зворотному напрямі — з моделі покроково вилучають статистично незначущі пояснювальні змінні.

Прогнозування на основі моделі множинної лінійної регресії полягає в оцінюванні очікуваних значень пояснюваної змінної за заданих значень пояснювальних змінних.

Точковий прогноз $y_{\text{точк}}^{\text{прогн}}$ на основі множинної лінійної регресії отримують шляхом підстановки прогнозних значень пояснювальних змінних $x_1 = x_1^{\text{прогн}}$, $x_2 = x_2^{\text{прогн}}$, ..., $x_m = x_m^{\text{прогн}}$ в оцінене рівняння регресії:

$$y_{\text{точк}}^{\text{прогн}} = b_0 + b_1 x_1^{\text{прогн}} + b_2 x_2^{\text{прогн}} + \dots + b_m x_m^{\text{прогн}}. \quad (4.178)$$

Інтервальний прогноз $y_{\text{інт}}^{\text{прогн}}$ (проміжок, в який із заданою ймовірністю і за заданих значень пояснювальних змінних потрапляє точкове прогнозне значення пояснюваної змінної) дорівнює:

$$y_{\text{точк}}^{\text{прогн}} - t^{\text{крит}} \hat{\sigma}_y \leq y_{\text{інт}}^{\text{прогн}} \leq y_{\text{точк}}^{\text{прогн}} + t^{\text{крит}} \hat{\sigma}_y, \quad (4.179)$$

де $\hat{\sigma}_y$ — оцінка помилки прогнозного значення пояснюваної змінної.

Величину $\hat{\sigma}_y$ обчислюють за формулою

$$\hat{\sigma}_y = \hat{\sigma}_\varepsilon \sqrt{1 + \mathbf{X}_0^T (\mathbf{X}^T \cdot \mathbf{X})^{-1} \mathbf{X}_0}, \quad (4.180)$$

де $\mathbf{X}_0^T = (1; x_1^{\text{прогн}}; x_2^{\text{прогн}}; \dots; x_m^{\text{прогн}})$ — розширений транспонований вектор прогнозних значень пояснювальних змінних; $\hat{\sigma}_\varepsilon$ — стандартна вибіркова помилка рівняння регресії.

Зазначимо, що прогнозування на основі рівняння регресії є надійним лише тоді, коли задані прогнозні значення пояснювальних змінних суттєво не виходять за межі їхніх вибіркових значень. Причому, точність прогнозу зростає з наближенням прогнозних значень пояснювальних змінних до середніх вибіркових значень. Тому використання лінії регресії «далеко» за межами досліджуваного діапазону значень пояснювальних змінних може призвести до значних помилок.

Передбачення значень багатьох економічних показників призводить до необхідності включення до прогностичної моделі як кількісних, так і

номінальних (якісних) ознак. Увести номінальні за своєю сутністю факторні ознаки в регресійну модель неможливо, оскільки їхні рівні набувають якісних, а не кількісних значень. У той же час номінальні факторні ознаки можуть істотно впливати на щільність зв'язку між пояснюваною та пояснювальними змінними. Наприклад, на обсяг реалізації окремих видів продукції у часі істотно впливає номінальна ознака «сезонність», а на вартість об'єкта нерухомості — номінальні ознаки: «наявність комунікацій», «тип будівлі», «кон'юнктура ринку» тощо. Використання номінальних ознак у регресійному аналізі передбачає їхнє оцифрування, тобто приписування кожній градації якісної шкали відповідного числа з дотриманням таких умов: шкала градацій має бути повною; градації не повинні перетинатися.

Оцифровані та включені до регресійної моделі номінальні ознаки називають *фіктивними* або *структурними змінними*. Фіктивна змінна — це умовний код, який вказує на належність або неналежність одиниці сукупності до певної градації, тобто відображає два протилежних стани номінальної факторної ознаки. Найчастіше використовують *бінарні фіктивні змінні*, які набувають двох значень (0 і 1) в залежності від деякої умови. Якщо номінальна ознака має k градацій, то для її представлення в моделі використовують $(k - 1)$ фіктивну змінну.

Фіктивні змінні включають до регресійної моделі з метою оцінювання:

- впливу структурних (групових) відмінностей одиниць сукупності на пояснювану змінну;
- впливу номінальної факторної ознаки на значення пояснюваної змінної.

Регресійна модель може містити одну або декілька фіктивних змінних у поєднанні з кількісними пояснювальними змінними або без них, а форма такого поєднання залежить від припущень щодо характеру впливу фіктивних змінних на пояснювану змінну. Регресійні моделі, які містять лише якісні пояснювальні змінні, називають *ANOVA-моделями* (*моделями дисперсійного аналізу*). Регресійні моделі, які містять пояснювальні змінні як кількісного,

так і якісного характеру, називають *ANCOVA-моделями* (моделями коваріаційного аналізу).

Якщо пояснювана змінна залежить лише від однієї номінальної пояснювальної змінної, яка представлена k градаціями, то регресійна модель має вигляд:

$$y = \alpha_0 + \alpha_1 d_1 + \alpha_2 d_2 + \dots + \alpha_{k-1} d_{k-1} + \varepsilon, \quad (4.181)$$

де y – пояснювана змінна; d_1, d_2, \dots, d_{k-1} – бінарні фіктивні пояснювальні змінні; $\alpha_0, \alpha_1, \dots, \alpha_{k-1}$ – параметри моделі; ε – випадкова величина.

Нехай значення результатного показника y формується під впливом m факторних ознак x_1, x_2, \dots, x_m , виражених у метричній шкалі вимірювання, і p факторних ознак u_1, u_2, \dots, u_p , виражених у номінальній шкалі вимірювання. Якщо позначити через q_k ($k = \overline{1, p}$) кількість градацій k -ї номінальної ознаки, то до регресійної моделі можна включити $q_1 + q_2 + \dots + q_p = p$ змінних, а саму модель записують так:

$$y = \gamma_0 + \sum_{i=1}^m \beta_j x_j + \sum_{i=1}^p \sum_{l=1}^{q_p-1} \alpha_{il} d_{il} + \varepsilon, \quad (4.182)$$

де d_{il} – i -та фіктивна змінна, зумовлена l -ю ознакою; γ_0 – вільний член; β_j – коефіцієнти регресії, які відображають «чистий» вплив фактора x_j ; α_{il} – коефіцієнти регресії, які відображають «чистий» вплив l -ї градації i -го номінального фактора.

Регресійні моделі з фіктивними змінними слугують ефективним засобом прогнозування сезонних явищ і процесів.

Якщо має місце адитивна сезонність і відсутня тенденція, то модель для квартальних даних можна записати у такому вигляді:

$$y_t = \alpha_0 + \alpha_1 d_1 + \alpha_2 d_2 + \alpha_3 d_3 + \varepsilon_t, \quad (4.183)$$

де y_t – рівень часового ряду для періоду $t = \overline{1, n}$; $\alpha_0, \alpha_1, \alpha_2, \alpha_3$ – параметри моделі; d_1, d_2, d_3 – фіктивні змінні для I, II і III кварталів відповідно.

У цьому разі четвертий квартал розглядають як базу порівняння, а фіктивні змінні набувають таких значень:

- $d_1 = 1$, якщо спостереження належить до I кварталу, і $d_1 = 0$ в усіх інших випадках;

- $d_2 = 1$, якщо спостереження належить до II кварталу, і $d_2 = 0$ в усіх інших випадках;

- $d_3 = 1$, якщо спостереження належить до III кварталу, і $d_3 = 0$ в усіх інших випадках.

У такому випадку за базу порівняння вибрано IV квартал.

Вільний член α_0 у моделі (4.183) інтерпретують як середній рівень для IV кварталу, а коефіцієнт регресії α_j — як різницю між середнім рівнем j -го кварталу і середнім рівнем IV кварталу (для IV кварталу $d_1 = d_2 = d_3 = 0$). Використовуючи параметри моделі (4.183), можна оцінити середні значення для кожного кварталу ($\bar{y}_j = \alpha_0 + \alpha_j$) та абсолютну величину сезонних коливань ($S_j = \bar{y}_j - \bar{y}$).

Якщо параметри моделі (4.183) статистично значущі, то в результаті підстановки значень фіктивних змінних можна отримати прогнознi значення.

У випадку наявності тенденції адитивну модель сезонності з фіктивними змінними записують так:

$$y_t = \tilde{y}_t + \alpha_1 d_1 + \alpha_2 d_2 + \alpha_3 d_3 + \varepsilon_t, \quad (4.184)$$

де \tilde{y}_t — деякий тренд або багатofакторна регресійна модель.

Для лінійного характеру тренда модель (4.184) набуває вигляду:

$$y_t = \alpha_0 + \beta_1 t + \alpha_1 d_1 + \alpha_2 d_2 + \alpha_3 d_3 + \varepsilon_t, \quad (4.185)$$

де t - фактор часу, який набуває значень порядкових номерів кварталів.

Параметри моделі (4.185) мають таку інтерпретацію:

- α_0 — рівень IV кварталу для року $t = 0$ (попереднього до періоду, який досліджують);

- β_1 — вплив тенденції у разі елімінування сезонності;

• $\alpha_1, \alpha_2, \alpha_3$ — зміна рівня для відповідного кварталу в порівнянні з IV кварталом.

У випадку мультиплікативної сезонності модель (4.185) трансформують на таку модель:

$$y_t = ab^t c_1^{d_1} c_2^{d_2} c_3^{d_3} \cdot \varepsilon_t. \quad (4.186)$$

Злогарифмувавши праву і ліву частину рівняння (4.186), отримаємо лінійну модель

$$\ln y_t = \ln a + t \ln b + d_1 \ln c_1 + d_2 \ln c_2 + d_3 \ln c_3 + \ln \varepsilon_t, \quad (4.187)$$

параметри якої можна оцінити за МНК. Після цього шляхом потенціювання переходимо до моделі (4.186).

У моделі (4.186) параметр b — це середній коефіцієнт зростання рівня часового ряду незалежно від впливу сезонності, а параметр c_j характеризує вплив сезонності j -го кварталу відносно базового кварталу незалежно від тенденції часового ряду.

До переваг методу моделювання сезонності на основі регресійних рівнянь з фіктивними змінними відносять:

- метод простий у реалізації (параметри оцінюють за МНК);
- не проводять декомпозицію часового ряду, а тому не потрібно згладжувати окремі компоненти;
- у процесі моделювання не втрачають спостереження;
- метод дає змогу врахувати сезонність у регресійній моделі довільної форми.

4.4. Прогнозування часових рядів з використанням нейронних мереж

Лінійна апроксимація за умови існування нелінійних зв'язків між даними не слугує надійним засобом отримання достовірних прогнозів. Для прогнозування складних явищ і процесів з нелінійними моделями даних використовують апарат нейронних мереж.

Штучна нейронна мережа (нейронна мережа, НМ) — це математична модель об'єкта або процесу разом з її програмним та апаратним забезпеченням, побудована на основі принципів організації функціонування біологічних нейронних мереж.

Спроможність моделювати нелінійні залежності, узагальнювати і виділяти приховані залежності між вхідними і вихідними даними, визначати неінформативні для аналізу дані та відсіювати їх, адаптуватися до мінливих зовнішніх впливів, працювати з неповними даними тощо є головною перевагою НМ порівняно з традиційними статистичними методами, що використовують для розв'язання великого переліку задач і зокрема прогнозування.

Нейронна мережа — це сукупність нейронів і зв'язків між ними (рис. 4.9).

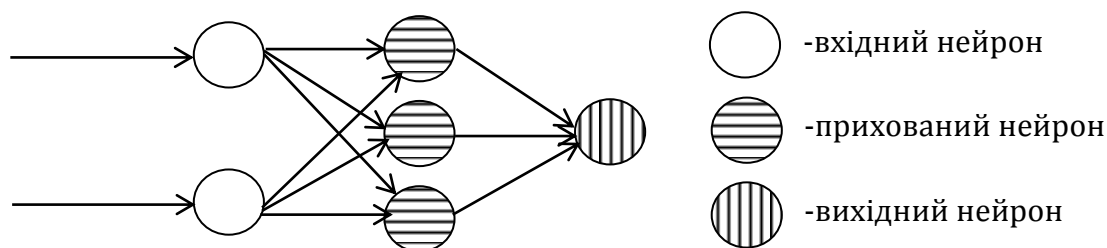


Рис. 4.9. Схема простої нейронної мережі

Одиницею обробки інформації в НМ є *штучний нейрон* (далі — нейрон) — аналог біологічного нейрона (рис. 4.10). Він має декілька каналів вводу інформації (*дендритів*) та один канал виводу інформації (*аксон*). Аксон нейрона з'єднаний з іншими нейронами зв'язками, які називають *синапсами*. Синапс характеризується одним параметром — ваговим коефіцієнтом (w), значення якого впливає на зміни інформації, коли вона передається від одного нейрона до іншого.

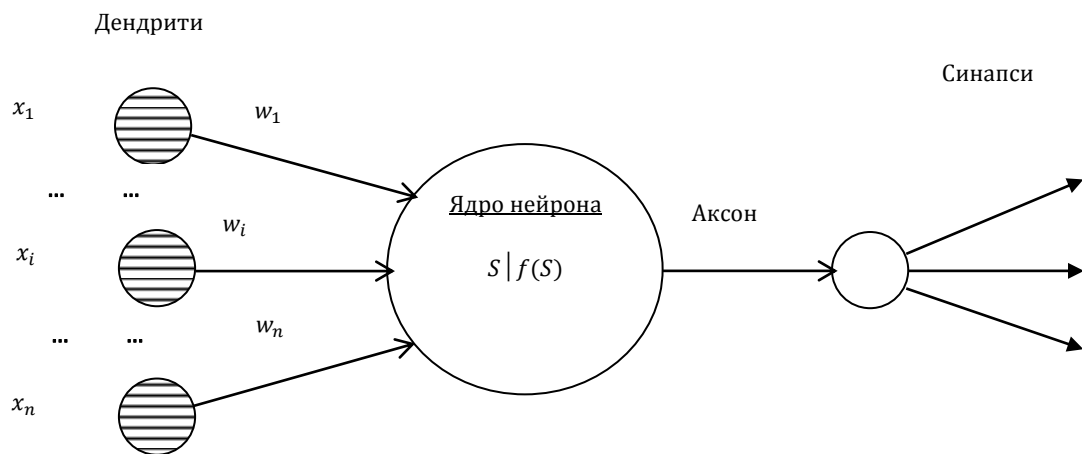


Рис. 4.10. Схема штучного нейрона

Функцію, аргументом якої є зважена сума входів нейрона, а значенням — вихід нейрона, в теорії НМ називають *активаційною функцією* (функцією активації):

$$y = f(S),$$

$$S = \sum_{i=1}^n w_i \cdot x_i,$$

де n — кількість входів нейрона; x_i ($i = \overline{1, n}$) — значення i -го входу нейрона;

w_i ($i = \overline{1, n}$) — вага i -го входу нейрона; S — зважена сума входів нейрона;

$f(S)$ — активаційна функція; y — значення вихідного сигналу нейрона (активаційної функції).

Функція активації як засіб перетворення зваженого входу нейрона у вихідний сигнал застосовується на останньому етапі обробки сигналу нейроном з метою обмеження амплітуди вихідного сигналу. Від вибраної функції активації та її параметрів залежить якість навчання НМ на конкретних прикладах. Найчастіше вживані активаційні функції представлено в табл. 4.9.

Активаційні функції

Назви функцій	Функціональні залежності	Області значень
Ступінчаста порогова	$y = \begin{cases} 0, \text{ якщо } S < S^* \\ 1, \text{ якщо } S \geq S^* \end{cases}$	0; 1
Тотожна (лінійна)	$y = S$	$(-\infty; \infty)$
Напівлінійна	$y = \begin{cases} 0, \text{ якщо } S < 0 \\ S, \text{ якщо } S \geq 0 \end{cases}$	$[0; \infty)$
Напівлінійна з посиленням	$y = \begin{cases} 0, \text{ якщо } S \leq 0 \\ S, \text{ якщо } 0 < S < 1 \\ 1, \text{ якщо } S \geq 1 \end{cases}$	$[0; 1]$
Лінійна з посиленням	$y = \begin{cases} -1, \text{ якщо } S \leq -1 \\ S, \text{ якщо } -1 < S < 1 \\ 1, \text{ якщо } S \geq 1 \end{cases}$	$[-1; 1]$
Сигмоїдна (логістична)	$\frac{1}{1+e^{-aS}}$ де a – параметр, що визначає нахил функції	$(0; 1)$
Гіперболічний тангенс	$\frac{e^S - e^{-S}}{e^S + e^{-S}}$	$(-1; 1)$
Радіальна базисна (Гаусса)	$y = e^{-S^2/2}$	$(0; 1)$

Метою використання функції активації є задавання нелінійних зв'язків між нейронами.

За характером структури зв'язків НМ поділяють на *повнозв'язані* і *неповнозв'язані*.

У повнозв'язаних НМ кожний нейрон передає свій вихідний сигнал іншим нейронам і собі також.

Неповнозв'язані НМ поділяють на *одношарові* та *багатошарові*. Шар — це один або декілька нейронів, на входи яких подається один і той же загальний сигнал.

В одношарових НМ сигнали з вхідного шару відразу передаються вихідному шару, тобто нейрони вхідного шару вирішують лише задачу розподілення сигналів і не виконують обчислювальних операцій. Необхідні обчислення і видачу результатів забезпечує вихідний шар нейронів.

Багатошарові НМ, крім вхідного і вихідного шарів, містять один або декілька *прихованих шарів*, які розташовані між ними і призначені для обробки інформації. Нейрони вхідного шару приймають виміряні значення характеристик досліджуваних об'єктів (процесів) і передають їх на перший прихований шар без перетворення. Приховані шари і вихідний шар відображають специфіку знань та перетворюють вхідні дані. Вихідний шар остаточно генерує розв'язок задачі у вигляді скалярних або векторних значень. Багатошарова НМ спроможна моделювати ситуації, яким притаманний високий рівень складності і невизначеності. Встановлення кількості прихованих шарів та кількості елементів у них є важливим проблемним питанням побудови НМ.

У той час, коли кількість нейронів у вхідному та вихідному шарах визначається досить легко на основі постановки задачі, то визначення кількості нейронів у прихованих шарах є неоднозначною задачею. Нині немає строгих правил для визначення кількості нейронів у прихованих шарах. Тому на практиці часто використовують доволі трудомісткий процес — починають з одного нейрона і додають по одному нейрону в прихований шар до тих пір, поки вихідний результат не буде достатньо точним.

З поміж багатошарових НМ виділяють:

- *мережі прямого поширення* (без зворотних зв'язків);
- *рекурентні мережі* (зі зворотними зв'язками);
- *мережі радіальних базисних функцій* (RBF-мережі).

У мережах *прямого поширення* сигнали передаються строго в напрямку від вхідного шару до прихованих шарів і від прихованих шарів до вихідного шару. У зворотному напрямку сигнали не передаються, і нейрони одного шару не пов'язані між собою. Такі мережі успішно використовують для вирішення задач прогнозування.

НМ, у яких виходи нейронів наступних шарів мають семантичні зв'язки з нейронами попередніх шарів, називають *рекурентними*. Виділяють декілька різновидів рекурентних НМ. Прикладом рекурентної НМ є мережа Елмана (рис. 4.11).

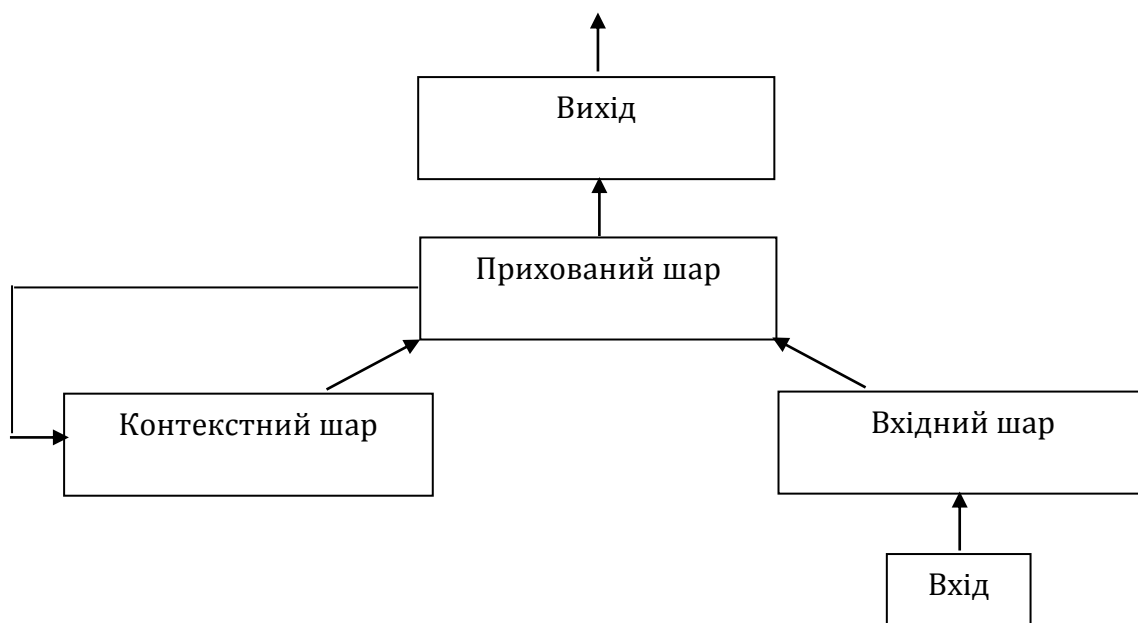


Рис. 4.11. Рекурентна мережа Елмана

Ця мережа побудована таким чином, що вихідні сигнали прихованого шару передаються не тільки нейронам вихідного шару, а й додатковим нейронам, які називають *контекстними*. Вони виконують функцію зберігання інформації про попередній стан нейронів прихованого шару. Така архітектура НМ є ефективною у випадку відтворення часових рядів.

RBF-мережі — це двошарові нейронні мережі прямого поширення, які будують на основі використання радіальних базисних функцій. Активаційна

функція кожного нейрона прихованого шару – це радіальна базисна функція. Оскільки радіальна базисна функція є нелінійною, то для моделювання довільної функції достатньо використовувати лише один шар прихованих нейронів.

Першими були створені одношарові НМ, які називають перцептронами (від лат. *perception* — сприйняття). Їхнє призначення — класифікація лінійно розділених об'єктів. До складу такої мережі входять елементи трьох типів: А-елементи, S-елементи та R-елементи. А-елементи (асоціативні елементи) — це шар нейронів, які з'єднані з множиною входів (S-елементів, рецепторів, сенсорів). Кожен S-елемент може перебувати в одному з двох станів — спокою або збудження. В останньому випадку відбувається передача одиничного сигналу асоціативним елементам. Асоціативні елементи активізуються, якщо кількість сигналів від S-елементів на їхньому вході перевищує деяку величину. Сигнали від збуджених А-елементів передаються суматору (R-елементам). Причому, сигнал від кожного асоціативного елемента має свою вагу. Реагуючі елементи на основі порогових активаційних функцій формують вихідний сигнал.

Надалі були розроблені багатошарові перцептрони, які успішно застосовували для вирішення великого переліку складних задач, зокрема прогнозування нелінійних тенденцій.

Наявність нелінійної функції активації, одного або декількох прихованих шарів нейронів, високий рівень семантичних з'єднань разом зі спроможністю до навчання забезпечують обчислювальну потужність багатошарового перцептрона.

Однією з головних переваг НМ перед традиційними алгоритмами є спроможність навчатися, що з технічної точки зору зводиться до пошуку адекватних коефіцієнтів зв'язку між нейронами. У результаті навчання виявляються і узагальнюються закономірності між вхідними і вихідними даними. За умови успішного навчання НМ буде формувати правильні

відповіді на підставі нових даних (навіть у випадку неповних і частково спотворених даних).

Теоретичною основою для побудови НМ слугує твердження: для довільної множини пар вхідного і вихідного векторів існує двошарова однорідна НМ з послідовними зв'язками, кінцевою кількістю нейронів і сигмоїдальними функціями активації.

Набір нейронів і структура семантичних зв'язків, як правило, формуються один раз на початку створення НМ і не змінюються в процесі її експлуатації. Процес налаштування семантичних ваг називають *навчанням НМ*. Залежно від способів навчання розрізняють:

- *навчання з учителем;*
- *навчання без учителя;*
- *навчання з підкріпленням.*

Навчання з учителем передбачає наявність навчальних прикладів (навчальної множини, даних спостереження), кожен з яких є парою: <вхід — бажаний вихід>.

У процесі навчання кожен приклад подається на вхід НМ і підлягає обробці, результатом якої є вихідний сигнал. Його значення порівнюють з відповідним значенням бажаного виходу.

З математичної точки зору, навчання з учителем є аналогом задачі інтерпретації або апроксимації деякої таблично заданої функції.

Репрезентативність є основною вимогою до навчальної та тестової вибірок, тобто вони мають правильно відображати досліджуваний об'єкт (процес). Опрацьовуючи під час навчання вхідні та відповідні їм вихідні значення даних, НМ знаходить деякі залежності між ними. В ідеальному випадку після навчання вона має видавати бажаний вихідний сигнал у разі надходження на вхід довільного сигналу з навчального набору даних.

Спосіб навчання НМ, за якого не вказують правильні відповіді на вхідні сигнали, називають навчанням без учителя. Алгоритм навчання без учителя

застосовують у випадках, коли відомі лише вхідні сигнали. Переважно параметри НМ налаштовують так, щоб вона видавала однакові результати для достатньо близьких значень вхідних сигналів.

Навчання НМ з підкріпленням є частковим випадком навчання з учителем. У процесі такого способу навчання мережа навчається, взаємодіючи з деяким середовищем. Такий спосіб навчання використовують у тому випадку, коли необхідно вибрати кращий варіант з доступної множини рішень.

Властивість НМ навчатися робить їх більш привабливими порівняно з системами, які функціонують за наперед заданими правилами. Виділяють два класи методів навчання НМ:

- *детерміновані* — коригування параметрів мережі проводять на основі їхніх поточних значень, величин входів, фактичних і бажаних виходів;

- *стохастичні* — параметри змінюються випадковим чином, причому зберігаються лише ті зміни, які зумовили покращання якості моделі.

Різницю між бажаним (очікуваним, цільовим) значенням виходу моделі (y) і фактичним значенням виходу моделі (y^*) називають *помилкою навчання* (E). Якщо НМ має лише один вихідний нейрон, то $E = y - y^*$. У випадку n вихідних нейронів помилку навчання визначають як середній квадрат помилок на кожному виході, тобто:

$$E = \frac{1}{n} \sum_{i=1}^n (y_i - y_i^*)^2.$$

Помилка навчання є показником точності налаштування НМ на навчальній множині і може бути використана як умова зупинки навчання. Проте вона не може слугувати оцінкою точності функціонування моделі з даними, які не приймали участі у навчанні, тобто узагальненою оцінкою спроможності мережі. З цією метою використовують помилку навчання на тестовій множині (*помилку узагальнення*).

Навчання НМ ґрунтується на прямій і зворотній передачі інформації про помилки навчання (рис. 4.12).

Вхідні сигнали послідовно надходять у НМ, яка їх обробляє і формує відповіді. За результатами сформованої відповіді обчислюється помилка навчання. Якщо помилка навчання задовільна, то мережа переходить до опрацювання наступної групи вхідних сигналів, інакше — проводиться процедура мінімізації помилки навчання.

Процедури 2-6 повторюються для кожної пари навчальної множини, допоки помилка навчання на всій навчальній множині не досягає прийняттого рівня.

Після того, як сформовано навчальну вибірку та обрано спосіб обчислення функції помилки, навчання НМ набуває характеру задачі багатоваріантної оптимізації. Для різного виду функцій помилок використовують окремі класи методів оптимізації.

Одним із методів пошуку мінімального значення функції багатьох змінних є метод *градієнтного спуску*. На першому етапі цього методу вибирають деяку початкову точку M_0 та обчислюють у ній градієнт (вектор, який задає напрямок найшвидшого зростання функції). Далі роблять крок в антиградієнтному напрямку. У результаті отримують нову точку M_1 , значення функції в якій зазвичай не більше за значення функції в точці M_0 . Якщо значення функції збільшилося, то потрібно зменшити крок. Після цього в новій точці M_2 обчислюють градієнт і рухаються в антиградієнтному напрямку. Момент завершення ітераційного процесу настає тоді, коли рух з отриманої точки за малого кроку призводить до збільшення значення функції або градієнт у цій точці рівний нулеві (точка локального мінімуму).

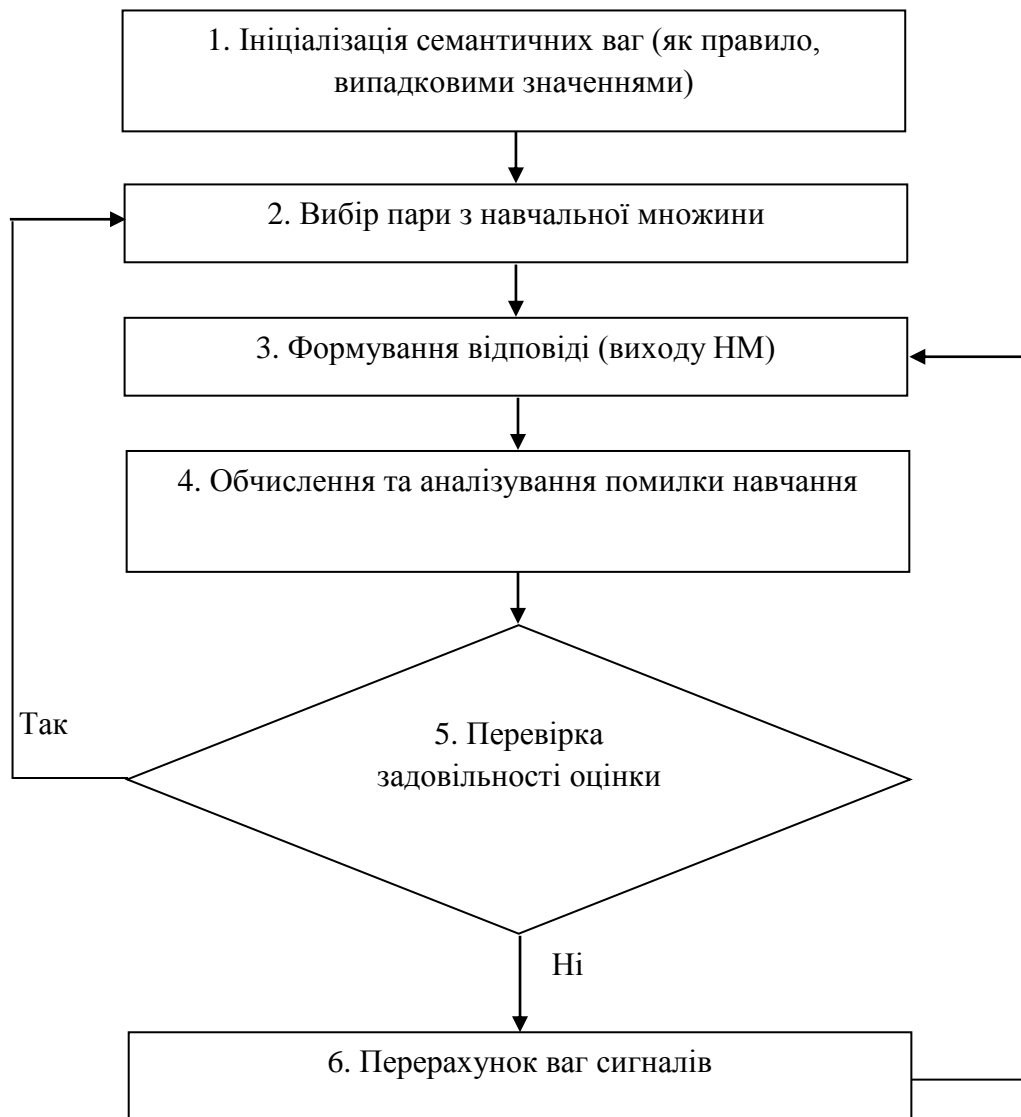


Рис. 4.12. Схема процесу навчання НМ

Процес навчання НМ є ітеративним, а його кроки називають *епохами* або *циклами*. Одна епоха — це один перегляд всіх прикладів навчальної множини з одночасним коригуванням семантичних ваг.

На вхід НМ подаються вектори початкових даних, а на вихідний вузол — повідомлення про бажане (цільове) значення результату. Дію, яка відбувається, можна вважати контрольованим навчанням. Контрольоване навчання НМ розглядають як процес розв’язання оптимізаційної задачі — задачі пошуку семантичних ваг, які мінімізують розбіжність між отриманим і бажаним результатами.

Для обчислення семантичних ваг користуються такою залежністю:

$$W(t + 1) = W(t) - \eta V,$$

де $W(t)$ – вектор семантичних ваг для епохи t ; $V = \frac{\partial E(t)}{\partial W(t)}$ – градієнт функції помилки навчання для епохи t ; η – коефіцієнт швидкості навчання.

Коефіцієнт швидкості навчання — це параметр, який керує величиною коригування ваг на кожній ітерації і змінюється в діапазоні $[0; 1]$. Вибір значення η є проблемною задачею:

- великі значення η відповідають великому значенню кроку корекції; у такому разі для знаходження мінімуму функції помилки буде потрібно менше ітерацій, але потенційно збільшується помилка навчання;

- малі значення η відповідають меншому кроку корекції семантичних ваг, а тому збільшується точність налаштування алгоритму мінімізації функції помилки, але також збільшується кількість епох навчання.

Існує декілька модифікацій методу градієнтного спуску (метод спряжених градієнтів, метод стохастичного градієнтного спуску, метод Левенберга-Маркара тощо), які реалізовано в програмних пакетах нейромережових технологій обробки даних.

У процесі навчання НМ можливе явище *перенавчання*, коли навчена модель добре розпізнає приклади з навчальної множини («зазубрила» пред'явлені їй спостереження), однак не розпізнає або погано розпізнає інші приклади, які не були використані у процесі навчання. Перенавчання є результатом надмірного підлаштування параметрів моделі до залежностей, які є в навчальній множині. Перенавчання є результатом довготривалого навчання надто великої кількості параметрів (ваг).

Одним із способів запобігання перенавчанню НМ є паралельне спостереження за зміною помилки мережі на навчальній і тестовій множинах. За деяку кількість кроків навчання помилки зменшуються на обох множинах. Проте на якомусь кроці помилка на тестовій множині

починає збільшуватися, а на навчальній множині продовжує зменшуватися (рис. 4.13).

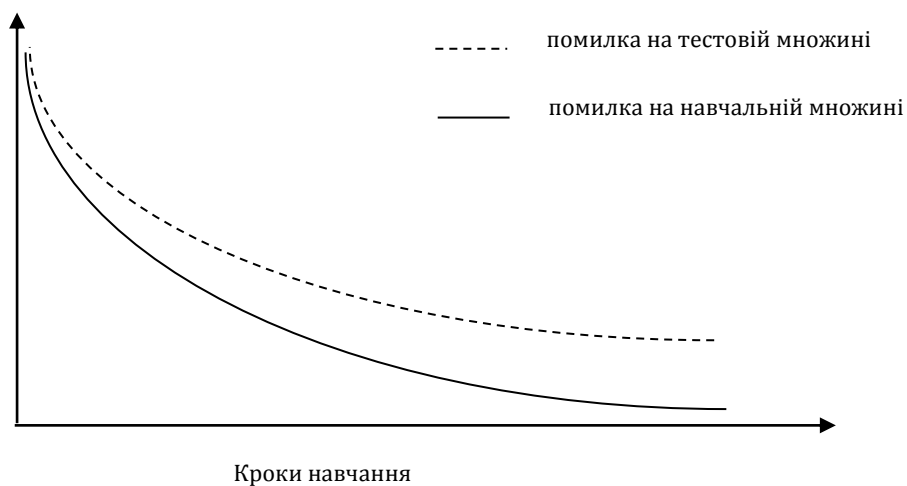


Рис. 4.13. Процес зміни помилки навчання

Уся інформація, яку НМ використовує для розв'язання поставленої задачі, зосереджена в навчальній множині (в навчальних прикладах). Тому якість навчання НМ насамперед залежить від кількості прикладів у навчальній множині, а також від того, наскільки добре ці приклади характеризують задачу.

Для зупинки процесу навчання використовують такі критерії:

- виконання заданої кількості ітерацій;
- досягнення функцією якості навчання встановленого мінімального значення;
- стабілізація функції якості навчання;
- функція якості навчання на тестовій вибірці починає зростати, а на навчальній — спадати.

Нейромережевий підхід до задач прогнозування придатний як для лінійних, так і для нелінійних залежностей, але є особливо ефективним для розвідувального аналізу даних. Якщо між вхідними і вихідними даними

існує деяка залежність, яку не вдається виявити традиційними статистичними методами, то НМ спроможна автоматично налаштуватися на неї з заданим рівнем точності, що допомагає в процесі розроблення прогнозів.

НМ достатньо проста у використанні. Користувачеві необхідно мати лише мінімальний рівень знань про те, як вибрати архітектуру мережі, формувати тестову і навчальну множини, інтерпретувати результати процесу моделювання.

Серйозним обмеженням практичного застосування НМ у сфері соціально-економічного прогнозування є дві проблеми:

- необхідність використання великої кількості спостережень для побудови прийнятної моделі;
- неспроможність будувати інтервальні прогнози.

Життєвий цикл НМ, призначеної для вирішення задач прогнозування, охоплює такі основні етапи:

Етап 1. Постановка задачі та вибір архітектури НМ:

- формування задачі прогнозування;
- обґрунтування застосування нейромережевого підходу;
- підбір архітектури НМ.

Етап 2. Формування інформаційної бази:

- визначення вхідних і вихідних змінних моделі НМ;
- формування вибіркової множини даних;
- попередня обробка вибіркової множини даних;
- розділення вибіркової множини даних на навчальну і тестову.

Етап 3. Визначення структури НМ:

- задавання кількості прихованих шарів;
- задавання кількості нейронів у кожному шарі.

Етап 4. Налаштування параметрів НМ та алгоритму навчання:

- вибір активаційних функцій;

- вибір значення коефіцієнта швидкості навчання;
- задавання умов завершення навчання.

Етап 5. Додаткове навчання НМ (вхідні і вихідні дані, отримані в процесі експлуатації, можуть бути використані для подальшого перерахунку семантичних ваг).

Нині розроблено багато програмних пакетів, які реалізують нейромережеві технології. Частина з них до певної міри є універсальними, інші — вузькоспеціалізовані. Одним з найпопулярніших і найефективніших програмних продуктів для прогнозування з використанням НМ є *STATISTICA Automated Neural Networks (SANN)*. SANN пропонує широкий спектр переваг та унікальних можливостей:

- набір типових архітектур НМ та алгоритмів навчання;
- наявність засобів автоматизації роботи з мережею, які потенційно збільшують чисельність користувачів і супроводжують дослідника на всіх етапах побудови моделі;
- підтримка завантаження й аналізу декількох моделей;
- контроль параметрів, які впливають на якість мережі, зокрема функцій активації та помилок;
- зберігання найкращих мереж.

У SANN реалізовано всі типи НМ, які нині використовують для вирішення практичних задач, а також новітні алгоритми швидкого навчання.

Основні положення та вказані моделі використовувалися при побудові:

- економіко-математичної моделі з дифузією, що зростає степеневно для знаходження банківськими установами величини ринкового портфеля акцій та визначення внутрішньої волатильності на ринку в будь-який момент часу, а також дослідження динаміки ринку і здійснення моніторингу фінансових потоків, за допомогою розкладу за системою функцій Бесселя першого роду

за умови, яка враховує лінійну комбінацію фінансового потоку та швидкості його зміни за різними чинниками.

Благун І. С., Буртняк І. В. Моделювання ефективності функціонування банків за допомогою фінансових потоків. Вісник Хмельницького національного університету. Серія: Економічні науки. 2020. № 4. С. 51–57.

- гравітаційної моделі просторового розподілу грошових доходів населення, теоретична та практична значущість якої полягає у визначенні нових місць отримання грошових доходів населення та оцінюванні впливу таких потенційних центрів на існуючу соціально-економічну систему. Розв'язання проблеми розміщення нового центру, а також оцінка міри впливу окремих центрів з урахуванням нового центру, дозволили визначити просторову структуру джерел отримання грошових доходів та розробити прогноз модифікацій такої структури на основі аналізу її імовірнісних характеристик.

Благун І. С., Дмитришин Л. І. Гравітаційна модель просторового розподілу грошових доходів населення. Економічна кібернетика : міжнародний науковий журнал. 2012. № 4-6. С. 76–78.

- прогнозних моделей попиту на туристичні послуги з урахуванням сезонної хвилі, що дозволяють кількісно оцінити розвиток туризму і дати його прогноз на найближчу перспективу. Така модель є ефективним інструментом побудови план-матриці розвитку підприємства на основі аналізу динаміки як кількісних, так і якісних показників його фінансового стану.

Благун І. С., Кейван О. І. Прогнозування попиту на туристичні послуги. Бізнес Інформ. 2012. № 8. С. 7–11.

- інтерактивної моделі розвитку партнерства на основі поведінкових детермінант взаємодії. Це дозволило проаналізувати чинники і змінні, що впливають на формування ринкових взаємовідносин та їх результативність. Визначено, що основою стійких конкурентних переваг є партнерських маркетинг на підприємствах.

Благу́н І. С., Кукурудз Р. Інтерактивна модель розвитку партнерства на основі поведінкових детермінованих взаємовідносин. Modeling the development of the economic systems. 2022. № 4. С. 56–62.

- моделі дослідження попиту на туристичні послуги на основі економетричної моделі Торнквіста. Це дозволило зробити висновок про те, що моделювання попиту та витрат на туристичні послуги вимагає аналізу середньої величини доходів та видатків населення. Визначені показники еластичності ефективного попиту на туризм дозволили зробити висновок, що туристичні послуги для вітчизняних домогосподарств відносяться до групи предметів та послуг розкоші. Аналіз значень показників еластичності доходу у кожній групі показав, що збільшення доходів домогосподарств сприяє зростанню попиту на туристичні послуги й зростанню частки витрат на такі цілі в бюджетах домогосподарств.

Благу́н І. С., Ле́щук Г. В., Ки́фор М. В. Економічне моделювання попиту на туристичні послуги в регіонах. Регіональна економіка. 2019. № 4. С. 87–93.

- трендової моделі туристичних потоків на прикладі Івано-Франківської області. Засобами експоненціального згладжування розроблені прогнози з урахуванням вибору кінцевої форми моделей на основі критерію найменшого середнього або згаслого прогнозу.

Благу́н І. С., Ле́щук Г. В., Ки́фор М. В. Прогностична модель оцінки туристичних потоків з урахуванням фактора адитивної сезонності на

прикладі Івано-Франківської області. Проблеми економіки. 2019. № 4 (42). С. 250–256.

- гравітаційних моделей двосторонньої торгівлі сільськогосподарськими продуктами між Україною та ЄС. Це дозволило визначити силу впливу чинників, пов'язаних із процесом глобалізації, на обсяги торгових потоків. Виокремлено детермінанти які можуть розглядатися, як альтернатива фізичній відстані між країнами в рівняннях гравітації.

Благун І. С., Надвірнянський Ю. Р. Застосування гравітаційних моделей для аналізу торгівлі між Україною та ЄС. Бізнес Інформ. 2022. № 12. С. 140–145.

- моделі сталого розвитку підприємництва, в основі якої є підвищення рівня якості життя населення. Вона відображає безперервний процес балансування економічних, соціальних та екологічних цілей господарюючих суб'єктів, що передбачає гнучку зміну розподілу ресурсів у міру зміни середовища та відходить від використання класичного лінійного програмування в економіці.

Благун І. С., Романюк М. Д., Судук Н. В., Мендела Є. М. Забезпечення сталого розвитку України та країн світу на основі коригування моделей лінійної економіки. Бізнес Інформ. 2022. № 1. С. 95–101.

- моделі управління виробництвом багаторівневих інтегрованих структур лісопромислового комплексу, реалізація якої дозволяє забезпечити узгодження агрегованих показників роботи підприємств із загальним критерієм ефективності виробничо економічного характеру. В її основі лежить група виробничих задач вибору технологій, пов'язаних системою транспортних задач перевезення сировини, готової продукції.

Запропонований автором метод розв'язання лінійного варіанта багатоетапної транспортно виробничої задачі, заснований на блоковій структурі зв'язків між обмеженнями з використанням схем подвійної декомпозиції, при якій отримані часткові задачі зберігають свою вихідну структуру і зміст, оскільки блочні задачі повністю відповідають задачам виробничого планування підсистеми, а “центральна”, хоча і втрачає властивості транспортної, але залишається близькою до неї.

Благул І. С., Судук Н. В. Моделювання управління виробництвом багаторівневих інтегрованих структур лісопромислового комплексу. Галицький економічний вісник. 2012. № 5 (38). С. 11–21.

СПИСОК ВИКОРИСТАНИХ ДЖЕРЕЛ

1. Abdey J. Business Analytics. Applied Modelling and Prediction. SAGE, 2023. 704 p.
2. Ariel B., Bland M., Sutherland A. Experimental Designs. SAGE, 2022. 240 p.
3. Arkes J. Regression Analysis. A Practical Introduction. 2nd ed. New York : Routledge, 2023. 412 p.
4. Bayesian Methods in Statistics. From Concepts to Practice. SAGE, 2022. 272 p.
5. Blahun I. S., Leshuk H. V. Probabilistic and statistical methods of risk assessment of investment projects of a region. Austrian Journal of Humanities and Social Sciences. 2017. No. 5-6. P. 53–59.
6. Blahun I. S., Leshuk H. V. Sensitivity analysis of indicators of efficiency of investment projects in the economic space of a region. The European Journal of Economics and Management. 2017. Vol. 3. P. 39–45.
7. Blahun I. S., Savchyn I. Z. Detection method of interdependences between dynamics of economic development indicators at regional and national levels. European Journal of Economics and Management. 2017. No. 3. P. 47–56.
8. Chandola T., Booker C. Archival and Secondary Data. SAGE, 2022. 184 p.
9. Crookes D. J. Mathematical Models and Environmental Change. Case Studies in Long Term Management. New-York : Routledge, 2023. 96 p.
10. Eichhorn J. Survey Research and Sampling. SAGE, 2022. 136 p.
11. Ferguson B., Lim G. Dynamic Economic Models in Discrete Time Theory and Empirical Applications. London : Routledge, 2016. 176 p.
12. Gopal R., Philips D., Weyde T. Foundations of Programming, Statistics, and Machine Learning for Business Analytics. SAGE, 2023. 512 p.

13. Gordon M. E. *Business Analytics. Combining data, analysis and judgement to inform decisions.* SAGE, 2023. 240 p.
14. Greene W. H. *Econometric Analysis.* 8th ed. Pearson, 2018. 1168 p.
15. Jones J. S., Goldring J. *Exploratory and Descriptive Statistics.* SAGE, 2022. 256 p.
16. Kacapyr E. *Essential Econometric Techniques A Guide to Concepts and Applications.* New York : Routledge, 2022. 228 p.
17. MacInnes J. *Statistical Inference and Probability.* SAGE, 2022. 224 p.
18. Martin P. *Linear Regression. An Introduction to Statistical Models.* SAGE, 2022. 200 p.
19. Martin P. *Regression Models for Categorical and Count Data.* SAGE, 2022. 272 p.
20. Maziarz M. *The Philosophy of Causality in Economics. Causal Inferences and Policy Proposals.* New York : Routledge, 2021. 222 p.
21. McBee M. *Statistical Approaches to Causal Analysis.* SAGE, 2022. 264 p.
22. McCoach D. B., Cintron D. *Introduction to Modern Modelling Methods.* SAGE, 2022. 304 p.
23. Privitera G. J. *Statistics for the Behavioral Sciences.* 4th ed. SAGE, 2023. 960 p.
24. Stock J., Watson M. *Introduction to Econometrics (Pearson Series in Economics).* 4th ed. Pearson, 2019. 800 p.
25. Sydsaeter K., Hammond P., Strom A. *Essential Mathematics for Economic Analysis.* 6th ed. Pearson, 2022. 976 p.
26. Timming A. R. *Applied Statistics. Business and Management Research.* SAGE, 2022. 456 p.
27. Yang K. *Analysing Intersectionality. A Toolbox of Methods.* SAGE, 2023. 200 p.

28. Бандоріна Л. М., Лозовська Л. І., Савчук Л. М. Моделювання економіки. Дніпро : УДУНТ, 2022. 154 с.
29. Барковський В., Барковська Н., Лопатін О. Теорія ймовірностей та математична статистика. Київ : Центр навч. літ., 2019. 424 с.
30. Благун І. С. Оцінка ефективності комерційних банків за допомогою методу DEA. Бізнес Інформ. 2020. № 11. С. 192–197.
31. Благун І. С., Боднарук І.Л. Теоретичні основи рейтингового управління в регіональних економічних системах. Інноваційна економіка. 2015. № 5 (60). С. 124–130.
32. Благун І. С., Буртняк І. В. Моделі аналізу та оцінки діяльності банків в умовах нестабільного економічного середовища. *Вісник Хмельницького національного університету. Серія: Економічні науки.* 2020. № 4. С. 41–45.
33. Благун І. С., Буртняк І. В. Моделювання ефективності функціонування банків за допомогою фінансових потоків. *Вісник Хмельницького національного університету. Серія: Економічні науки.* 2020. № 4. С. 51–57.
34. Благун І. С., Дмитришин Л. І. Гравітаційна модель просторового розподілу грошових доходів населення. *Економічна кібернетика : міжнародний науковий журнал.* 2012. № 4-6. С. 76–78.
35. Благун І. С., Дмитришин Л. І. Моделювання взаємозв'язку «нерівність доходів–економічне зростання» за допомогою теорії каскадних переваг. *Проблеми економіки.* 2012. № 4. С. 216–221.
36. Благун І. С., Дмитришин Л. І. Просторово-структурний аналіз доходів населення. *Регіональна економіка.* 2013. № 1 (67). С. 99–106.
37. Благун І. С., Ільчук П. Г. Методи оцінювання рівня інтернаціоналізації підприємств. *Економіка і прогнозування.* 2014. № 4. С. 97–109.

38. Благун І. С., Ільчук П. Г. Моделі взаємозв'язку рівня інтернаціоналізації та фінансових результатів діяльності українських підприємств. *Проблеми економіки*. 2015. № 1. С. 259–264.
39. Благун І. С., Кацедан А. В. Моделювання взаємозв'язків чинників регіональної конкурентоспроможності з її потенціалом розвитку. *Бізнес Інформ*. 2014. № 11. С. 103–115.
40. Благун І. С., Квасній З. В. Моделювання зв'язків сектора МСП та іноземних інвестицій із працевлаштуванням та економічним зростанням у регіонах. *Бізнес Інформ*. 2014. № 7 (438). С. 71–76.
41. Благун І. С., Кейван О. І. Прогнозування попиту на туристичні послуги. *Бізнес Інформ*. 2012. № 8. С. 7–11.
42. Благун І. С., Кічор В. П., Селюченко Н. Є. Економічне прогнозування: теоретичні та прикладні аспекти. Львів : Растр-7, 2020. 292 с.
43. Благун І. С., Кічор В. П., Скворцов Д. І. Статистичний аналіз і моделювання соціально-економічних об'єктів та процесів. Львів : Растр-7, 2022. 400 с.
44. Благун І. С., Кічор В. П., Фещур Р. В., Воробець С. Й. Математичні методи в економіці. Тернопіль : Богдан, 2011. 264 с.
45. Благун І. С., Кукурудз Р. Інтерактивна модель розвитку партнерства на основі поведінкових детермінованих взаємовідносин. *Modeling the development of the economic systems*. 2022. № 4. С. 56–62.
46. Благун І. С., Лещук Г. В., Кифор М. В. Економічне моделювання попиту на туристичні послуги в регіонах. *Регіональна економіка*. 2019. № 4. С. 87–93.
47. Благун І. С., Лещук Г. В., Кифор М. В. Прогностична модель оцінки туристичних потоків з урахуванням фактора адитивної сезонності на прикладі Івано-Франківської області. *Проблеми економіки*. 2019. № 4 (42). С. 250–256.

48. Благун І. С., Надвірнянський Ю. Р. Застосування гравітаційних моделей для аналізу торгівлі між Україною та ЄС. Бізнес Інформ. 2022. № 12. С. 140–145.
49. Благун І. С., Романюк М. Д., Судук Н. В., Мендела Є. М. Забезпечення сталого розвитку України та країн світу на основі коригування моделей лінійної економіки. Бізнес Інформ. 2022. № 1. С. 95–101.
50. Благун І. С., Судук Н. В. Моделювання управління виробництвом багаторівневих інтегрованих структур лісопромислового комплексу. Галицький економічний вісник. 2012. № 5 (38). С. 11–21.
51. Вовк Л. Математичний інструментарій моделювання економічних процесів. Київ : Ліра-К, 2017. 252 с.
52. Касьяненко В. О., Старченко Л. В. Моделювання та прогнозування економічних процесів. Київ : Унів. кн., 2023. 185 с.
53. Козак Ю. Г. Математичне моделювання для економістів. Київ : Центр учб. літ., 2020. 252 с.
54. Мармоза А. Теорія статистики. Київ : Центр навч. літ., 2019. 592 с.
55. Хома І., Бондаренко Л., Чубка О. Моделювання фінансових рішень в умовах невизначеності. Київ : Кондор, 2023. 284 с.
56. Шпігельхальтер Д. Мистецтво статистики. Прийняття аргументованих рішень на основі даних. Київ : КМ-БУКС, 2023. 384 с.

Наукове видання

БЛАГУН Іван Семенович

**МОДЕЛЮВАННЯ ТА ПРОГНОЗУВАННЯ
СОЦІАЛЬНО-ЕКОНОМІЧНИХ СИСТЕМ**

Монографія

Видання в авторській редакції

Головний редактор – *Василь Головчак*

Підп. до друку 21.03.2024. Формат 60x84/16. Гарнітура “Times New Roman”.
Тираж 300 прим. Ум. друк. арк. 16,6. Зам. №56 .

Видавець

Прикарпатський національний університет
імені Василя Стефаника

76018, м. Івано-Франківськ, вул. С. Бандери, 1,
тел. 75-13-08, e-mail: vdvcit@pnu.edu.ua

Свідоцтво суб'єкта видавничої справи ДК № 7616 від 26.05.2022

Виготовлювач

ФОП Голіней О. В.

76008, Україна, м. Івано-Франківськ, вул. Галицька, 128
тел.: 0664816601, e-mail: gsm1502@ukr.net

ISBN 978-966-640-556-5

