

К. Д. Семенова, К. І. Тарасова

БІЗНЕС-СТАТИСТИКА

Підручник

**Київ
2018**

УДК 334.012.42:31 (075.8)

JEL L21, M29, C44

C 30

Рекомендовано Вченою радою Одеського національного економічного університету (протокол № 10 від 26 червня 2018 року)

Рецензенти:

Янковий О. Г., завідувач кафедри економіки підприємства та організації підприємницької діяльності Одеського національного економічного університету, доктор економічних наук, професор

Лоханова Н. О., завідувач кафедри бухгалтерського обліку та аудиту Одеського національного економічного університету, доктор економічних наук, професор

Семенова К. Д., Тарасова К. І.

Бізнес-статистика : Підручник / К. Д. Семенова, К. І. Тарасова. – К : ФОРМ Гуляєва В.М. 2018. – 210 с.

ISBN 978-617-7503-19-3

Підручник підготовлено відповідно до типової навчальної програми дисципліни «Бізнес-статистика». Методологія визначення основних показників, які розглядаються у підручнику, базується на міжнародних стандартах і сучасній практиці статистики України.

У підручнику автори висвітлили у зручній для вивчення та засвоєння формі проблеми, які постали перед економістами в царині статистичного дослідження бізнесу.

Підручник призначений для студентів вищих навчальних закладів, які навчаються за економічними спеціальностями, а також для аспірантів, викладачів, економістів-практиків, підприємців, бажаючих розширити можливості використання статистичних методів для аналізу діяльності підприємств.

ISBN 978-617-7503-19-3

© Семенова К. Д., Тарасова К. І., 2018

ЗМІСТ

ПЕРЕДМОВА.....	5
РОЗДІЛ 1. Роль статистики в бізнесі.....	6
1.1. Статистика та її роль у дослідженні бізнес-процесів.....	6
1.2. Сутність бізнес-процесу.....	7
1.3. Класифікація бізнес-процесів.....	9
1.4. Основні етапи аналізу бізнес-процесів.....	14
Питання для самоконтролю.....	19
РОЗДІЛ 2. Структури даних. Класифікація, розподіл і перетворення даних.....	20
2.1. Сутність даних як ресурсу підприємства.....	20
2.2. Способи класифікації наборів даних.....	21
2.3. Джерела даних.....	27
2.4. Типи розподілу даних у сукупностях.....	32
2.5. Основні методи зберігання даних.....	34
Питання для самоконтролю.....	36
РОЗДІЛ 3. Випадкові величини і закони розподілу як основа бізнес-аналітики.....	37
3.1. Випадкові величини в бізнес-аналізі.....	37
3.2. Закони розподілу випадкових величин: сутність і види.....	43
3.3. Нормальний розподіл.....	46
3.4. Біноміальний розподіл.....	50
3.5. Апроксимація біноміального розподілу нормальним.....	55
3.6. Розподіл Пуассона.....	57
3.7. Експоненціальний розподіл.....	60
3.8. Інші види розподілу.....	61
Питання для самоконтролю.....	65
РОЗДІЛ 4. Статистичні гіпотези і статистичні висновки.....	66
4.1. Сутність і види статистичних гіпотез.....	66
4.2. Критерій як інструмент перевірки гіпотези. Потужність критерію....	69
4.3. Двостороння перевірка гіпотези. Перевірка гіпотези про рівність середньої генеральної сукупності певному заданому значенню.....	73
4.4. Одностороння перевірка гіпотези.....	79
Питання для самоконтролю.....	87
РОЗДІЛ 5. Моделювання зв'язків між кількісними змінними.....	88
5.1. Види взаємозв'язків між кількісними змінними	88
5.2. Детерміновані моделі та їх аналіз.....	90
5.3. Регресійно-кореляційний аналіз.....	94
5.4. Розрахунок резервів на основі регресійної моделі.....	104
Питання для самоконтролю.....	105
РОЗДІЛ 6. Дисперсійний аналіз у дослідженні бізнес-процесів.....	106
6.1. Види дисперсій. Правило декомпозиції дисперсій.....	106

6.2. Сутність дисперсійного аналізу.....	110
6.3. Однофакторний дисперсійний аналіз.....	112
6.4. Багатофакторний дисперсійний аналіз.....	118
Питання для самоконтролю.....	122
РОЗДІЛ 7. Аналіз закономірностей для якісних даних.....	123
7.1. Особливості аналізу якісних даних. Сутність критерію «хі-квадрат».	123
7.2. Використання критерію «хі-квадрат» щодо подібності розподілів....	126
7.3. Використання статистичних критеріїв для оцінки тісноти зв'язку між якісними змінними.....	130
Питання для самоконтролю.....	139
РОЗДІЛ 8. Статистичні методи контролю якості.....	141
8.1. Сутність статистичного контролю якості. Статистичний контроль виробничого процесу.....	141
8.2. Контрольні листки.....	143
8.3. Діаграма Парето.....	145
8.4. Метод розшарування.....	150
8.5. Гістограма.....	153
8.6. Діаграма розкиду.....	157
8.7. Діаграма Ішікави.....	160
8.8. Контрольна карта.....	163
Питання для самоконтролю.....	166
РОЗДІЛ 9. Показники ефективності діяльності підприємства.....	167
9.1. Сутність показників ефекту та ефективності.....	167
9.2. Система показників ефективності господарської діяльності підприємства.....	171
9.3. Приклад аналізу показників ефекту та ефективності.....	188
9.4. Спеціальні показники ефективності діяльності підприємства.....	191
Питання для самоконтролю.....	197
СПИСОК ВИКОРИСТАНОЇ ТА РЕКОМЕНДОВАНОЇ ЛІТЕРАТУРИ....	199
ДОДАТКИ.....	201

ПЕРЕДМОВА

Ідея будь-якого бізнесу спирається на те, що всі доступні ресурси важливо використовувати повно та ефективно, особливо коли мова йде про інформацію. Це підкреслює актуальність навчання застосовувати методи аналізу інформації з метою розробки ефективних управлінських рішень.

Робота сучасних економістів у підприємницьких структурах, а також фахівців органів державного управління, експертів у галузі бізнес-аналітики неможлива без знання економічного змісту та аналітичних можливостей системи статистичних показників, що використовуються для аналізу інформації. Однак в умовах розвитку нових технологій, постійного оновлення асортименту продукції та послуг використання розповсюджених стандартних математичних методів і моделей не задовольняє економістів, вони прагнуть здійснювати дослідження, спираючись на фактичні дані, і розробляти на їх основі специфічні моделі, які могли б урахувати особливості власного бізнесу, закономірності його розвитку.

Запропонований підручник містить системний виклад загальних категорій, принципів і методів проведення бізнес-аналізу, теоретичних основ моделювання і прогнозування із застосуванням сучасних програмних продуктів.

Перевагою підручника є те, що основні теоретичні положення демонструються на конкретних практичних ситуаціях, що є досить важливим для економіста, який прагне чітко уявляти собі умови та особливості використання певних методів аналізу.

Підручник дає відповіді на ключові питання, що виникають у фахівців, які використовують великі масиви статистичних даних, що характеризують ресурси та результати підприємницької діяльності.

Структура підручника допоможе господарюючим суб'єктам приймати управлінські рішення не наосліп, а на основі адекватної оцінки ймовірності досягнення поставленої мети.

Підручник «Бізнес-статистика» призначений, у першу чергу, для студентів, які навчаються за спеціальністю «Економіка». Однак він може бути корисним і для студентів інших спеціальностей, а також для аспірантів, науковців, економістів-практиків.

РОЗДІЛ 1. РОЛЬ СТАТИСТИКИ В БІЗНЕСІ

1.1. Статистика та її роль у дослідженні бізнес-процесів

Незалежно від рівня та стадії розвитку суспільства, характеру політичної системи, статистика завжди була і є необхідним та ефективним інструментом управління життєдіяльністю як держави у цілому, так і її окремих суб'єктів.

Роль статистики в бізнесі ще більше зростає в сучасних умовах розвитку країни, коли підприємці вимушені приймати управлінські рішення в умовах нестаціонарного середовища, іноді не маючи повної та достовірної інформації не лише стосовно зовнішнього середовища, а й щодо внутрішніх бізнес-процесів. Саме статистичні методи допомагають здійснювати аналіз доступної інформації і розробляти на її основі обґрунтовані тактичні та стратегічні рішення.

Методи статистики дозволяють провести аналіз досягнень фірми, розглянути сильні та слабкі сторони її комерційної діяльності; вони є тим допоміжним засобом, що сприяє розробленню програм розвитку бізнесу, впровадженню інновацій, проведення інвестиційної політики.

Крім того, статистика є тим інструментом, який дозволяє досліджувати інфляційні процеси в державі, аналізувати стан фінансових і товарних ринків, конкурентоспроможність підприємств і галузей економіки, продукції та послуг, правильно оцінювати результати виробничо-фінансової діяльності корпорацій і фірм, виявляти внутрішньовиробничі резерви.

Отже, можна дати наступне визначення **бізнес-статистики** – це наука, яка вивчає сукупність методів статистичного збору та обробки даних для дослідження ринкового, комерційного та господарського стану і розвитку окремого суб'єкта підприємницької діяльності.

У сучасних умовах одним із основних завдань, які можуть бути вирішені за допомогою методів бізнес-статистики, є пошук статистичних даних про кількісні та якісні показники діяльності як діючих, так знову відкритих фірм, які є потенційними конкурентами; обробку і аналіз отриманих даних для вибору правильної стратегії поведінки, яка дозволяє утримати свої позиції в даному сегменті ринку.

Таким чином, бізнес-статистика виступає з одного боку, як джерело інформаційної бази та методологічна основа аналізу

економічних процесів, а з іншого – є тим науковим фундаментом, на якому базується прийняття об'єктивно обумовлених рішень на різних гілках управління.

1.2. Сутність бізнес-процесу

Бізнес – це сукупність ділових відносин підприємців, спрямованих на задоволення потреб споживачів, з одного боку, і на отримання прибутку, з іншого. Він відіграє визначальну роль у сучасній економіці. Саме бізнесу суспільство зобов'язане практично всіма економічними благами, що визначають умови існування сучасної людини – від житла та одягу до харчування та інтелектуального дозвілля.

Суть бізнесу полягає у з'єднанні матеріальних, фінансових, трудових та інформаційних ресурсів з метою виробництва товарів і надання послуг, призначених для продажу іншим економічним агентам – державі, підприємствам, організаціям або громадянам. Якщо мова йде про довгострокове з'єднання цих ресурсів, виникає, зазвичай, організація або організаційна одиниця особливого типу під назвою «підприємство», яка виконує дії з реалізації бізнес-процесів.

Існують різні підходи до визначення сутності бізнес-процесу. Основні з них представлено на рис. 1.1

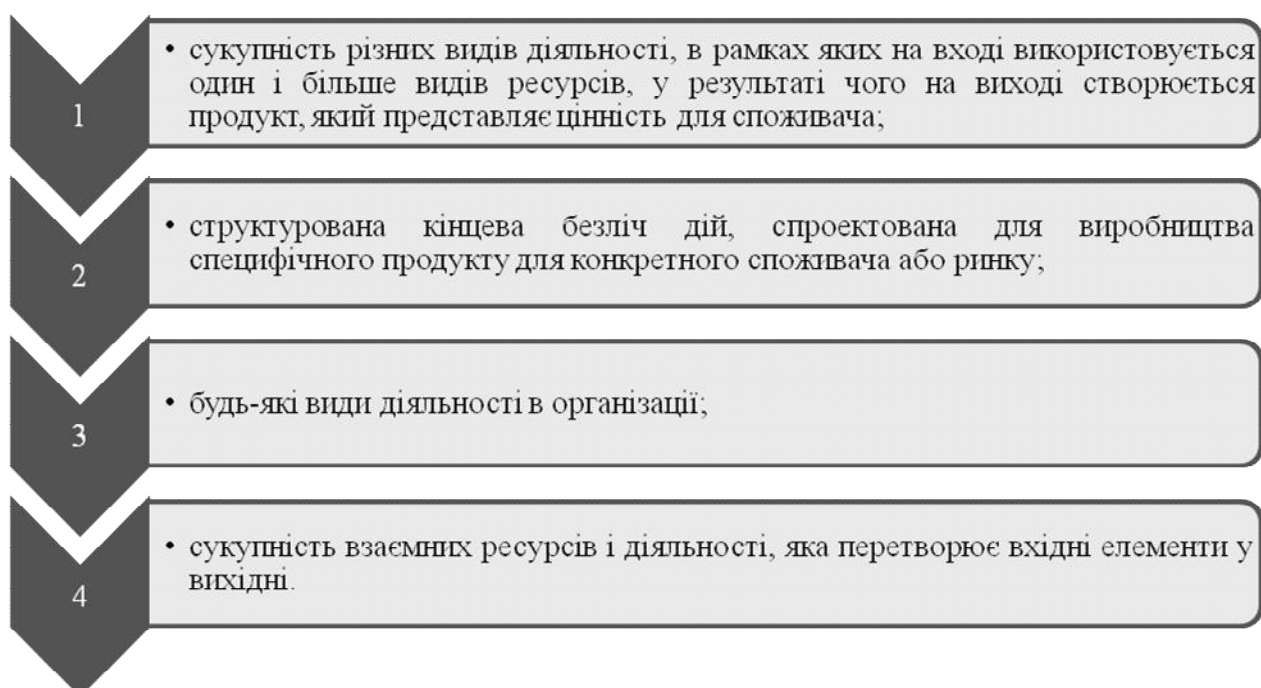


Рис. 1.1. Різні підходи до визначення бізнес-процесу

У комплексі складових – це один логічно цілий акт бізнес-діяльності, який організовується для досягнення результатів підприємства з метою задоволення потреб клієнта.

На сучасному етапі розвитку економічної думки деякі вітчизняні та закордонні вчені сприймають підприємство саме як поєднання бізнес-процесів, замість сукупності його підрозділів [1, с. 19]. Основними підставами для формування «процесної» точки зору на підприємство є такі:

- управління процесами на підприємстві (а не окремими його відділами) дозволяє створити краще підґрунтя для контролю за використанням ресурсів та часом виконання робіт;

- управління процесом, який є наскрізним для декількох підрозділів підприємства, знижує ризик субоптимізації, тобто під час призначення відповідальних за процес вдається уникнути розподілу відповідальності за фрагментами, що часто буває на спеціалізованих підприємствах;

- кожен процес має свого споживача, а зосередження на кожному процесі окремо сприяє кращому його задоволенню;

- визначення меж розглянутого процесу, а також постачальників і споживачів, дозволяє забезпечити кращу взаємодію та розуміння вимог, які слід задовольнити.

Бізнес-процес є об'єктом, який функціонує під впливом суб'єкта – керівника бізнес-процесу, що представляє собою посадову особу або колегіальний орган управління, який має у своєму розпорядженні ресурси, необхідні для виконання функції перетворення, і несе відповідальність за результати бізнес-процесу. Вплив керівника на бізнес-процес виражається в управлінні, а керівник бізнес-процесу є його невід'ємною частиною.

У свою чергу функція перетворення – це одна або декілька дій (операцій), що є здійснюваними в логічній послідовності та спрямованими на зміну вхідного об'єкту в результат з певними характеристиками за допомогою ресурсів.

Підсумовуючи викладене вище, можна дати наступне визначення: **бізнес-процес** – це сукупність видів діяльності підприємства, результатом яких є створення продукту і, як наслідок, отримання прибутку.

Під час аналізу бізнес-процесів підприємство ставить наступні цілі:

- зрозуміти загальну структуру організації діяльності підприємства та визначити динаміку її розвитку;
- виявити поточні проблеми та окреслити можливості для їх подолання;
- створити зрозумілу для всіх учасників (розробників, користувачів, замовників тощо) систему цілей і завдань;
- сформулювати вимоги до необхідного програмного забезпечення.

Для того, щоб найбільш повно розкрити сутність поняття «бізнес-процес», необхідно визначити ряд найважливіших його характеристик, які дозволяють більш конкретно уявляти і більш точно виділяти бізнес-процеси. Так, незалежно від сфери діяльності, кожен бізнес-процес може бути описаний одними й тими ж характеристиками [10]:

1. Кордон – це початок і кінець виконання процесу.
2. Власник – посадова особа, яка не тільки володіє необхідними ресурсами для роботи, а й планує, аналізує, управляє процесом і несе відповідальність за отриманий результат.
3. Вхід – інформаційне повідомлення, отримане в будь-якій прийнятій на підприємстві формі, що визначає потребу в інформації або матеріалах і знаменує початок процесу.
4. Вихід – інформація або матеріалізований продукт, який споживається зовнішнім по відношенню до виконавця клієнтом.
5. Виконавець – персонал компанії, що зайнятий у процесі.
6. Ресурси – матеріальна або інформаційна частина діяльності, яка не змінюється під час виконання операції (а лише сприяє перетворенню вхідної інформації або матеріалу в кінцевий продукт).
7. Контроль якості – галузеві або внутрішні (прийняті керівництвом компанії) показники, що дозволяють визначити ефективність роботи.

1.3. Класифікація бізнес-процесів

Існує безліч класифікацій бізнес-процесів. Моделювання роботи підприємств і, як частина цього завдання, визначення класифікації бізнес-процесів виділилося в окрему самостійну галузь досліджень, якою займаються багато вчених. Наприклад, дослідники з Плімутського університету (США) розробили ієрархію бізнес-

процесів, яка ділиться на три основні групи: «виробництво», «управління» та «підтримка».

Більш прикладний і простий підхід був запропонований у ході виконання норвезького проекту *TOPP* з порівняльного бенчмаркінгу. Виконавці цього проекту запропонували розділити всі бізнес-процеси підприємства на первинні та підтримуючі (допоміжні) відповідно до теорії Портера про ланцюги вартості. Деякі з підтримуючих процесів були далі виділені в окремий клас – процеси розвитку. Ці три групи процесів відповідно до проекту *TOPP* визначаються наступним чином:

- первинні процеси – це основні процеси підприємства, які створюють цінності. Вони пронизують підприємство цілком, починаючи від його відносин із споживачами та закінчуючи відносинами з постачальниками;

- підтримуючі (допоміжні) процеси – це процеси, які не створюють безпосередньо додану цінність, однак вони потрібні для забезпечення проходження основних процесів. Такими допоміжними процесами можуть виступати, наприклад, управління фінансами та менеджмент персоналу;

- процеси, що розвиваються – це такі процеси, які дозволять створити ланцюжок цінностей в основному та допоміжному процесам на новому рівні показників. Наприклад, розробка нової продукції та пошук нового постачальника сировини.

Результати, що отримані норвезькими фахівцями в рамках програми *TOPP*, набули подальшого розвитку в рамках наступної програми – Європейської мережі дослідження перспективних показників (*ENAPS*).

У результаті виконання робіт за цією програмою були прийняті дещо інші назви груп бізнес-процесів: первинні бізнес-процеси були названі власне бізнес-процесами і були розбиті на чотири підгрупи основних процесів. Дві інші були названі вторинними процесами, які, у свою чергу, поділились на групи процесів підтримки і процесів розвитку. Загальна схема класифікації бізнес-процесів програми *ENAPS* наведена на рис. 1.2.

Питанням класифікації бізнес процесів займаються й вітчизняні вчені. Так, А. Варзунов виділяє наступні типи бізнес-процесів відповідно до їх *основних функцій* [2]:

- 1. Основні бізнес-процеси** – це процеси, що генерують доходи підприємства. До них відносяться процеси, які орієнтовані на

виробництво товару або надання послуги, що є цільовими об'єктами створення підприємства та забезпечують отримання доходу. Саме основні бізнес-процеси формують результат і споживчі якості, за які зовнішній клієнт готовий платити кошти. Наприклад, для деревообробного заводу основним бізнес-процесом може бути виробництво меблів.



Рис. 1.2. Загальна класифікація бізнес-процесів за критеріями програми ENAPS

2. Супутні бізнес-процеси – це процеси, що орієнтовані на виробництво товару або надання послуги, які є результатами супутнього основному виду виробничої діяльності. Прикладом може слугувати процес ремонту стороннього транспорту на власній ремонтній базі підприємства з виробництва автомобілів.

3. Допоміжними бізнес-процесами є процеси, що призначені для життєзабезпечення основних і супутніх процесів підприємства, які орієнтовані на підтримку їх специфічних рис. Наприклад, для промислового підприємства допоміжним бізнес-процесом є процес ремонту виробничого обладнання.

4. Забезпечувальні процеси підтримують інфраструктуру підприємства. Вони призначені для життєзабезпечення всіх інших процесів і орієнтовані на підтримку їх універсальних рис. На підприємствах будь-якої галузі – це процес фінансового забезпечення, кадрового забезпечення, інженерно-технічного забезпечення тощо.

5. Бізнес-процеси управління – це процеси, що охоплюють увесь комплекс функцій управління на рівні кожного бізнес-процесу та підприємства в цілому. Це процеси стратегічного, оперативного та поточного планування, формування та здійснення управлінських дій.

6. Бізнес-процеси розвитку – це процеси вдосконалення виготовленого товару або послуги, процеси розвитку технологій, процеси модифікації устаткування, а також будь-які інші інноваційні процеси на підприємстві. Наприклад, це проведення науково-дослідних і дослідно-конструкторських робіт у машинобудуванні, процес технічного переозброєння в електроенергетиці тощо.

Основні характеристики розглянутих процесів наведені в табл. 1.1.

Існують і більш прості та узагальнені класифікації бізнес-процесів.

Так, за **виконанням ролі** бізнес-процеси поділяють на:

- **функціональні**, які безпосередньо формують зміст бізнес-системи та виконують її функції. Наприклад, до бізнес-процесів цього типу можна віднести вибір цільового ринку підприємства та власного кола покупців; розвиток місії та стратегії підприємства; розробку продукту та послуг; просування та продаж виготовлених товарів; виробництво продукції та надання послуг; виставлення рахунків покупцям;

- **структурні**, що є спрямованими на підтримку та розвиток інфраструктури бізнес-системи. Вони забезпечують власне її існування. Такими бізнес-процесами виступають: розвиток та управління підприємством; управління інформаційними потоками; управління фінансами та матеріальними ресурсами підприємства; програмне управління впливів на навколишнє середовище;

управління зовнішніми ресурсами, управління організаційним розвитком.

Таблиця 1.1

Характеристика бізнес-процесів залежно від їх основної функції

Назва бізнес-процесу	Характеристика бізнес-процесу
Основні бізнес-процеси	- дзеркально відображають бізнес-напрямки діяльності; - виступають джерелом генерування доходів; - визначають профіль бізнесу; - мають стратегічне значення; - можуть розвиватися або відмирати залежно від потреб ринку та стратегії розвитку підприємства.
Супутні бізнес-процеси	- самостійно від основної діяльності приносять дохід, але значно менший за обсягом.
Допоміжні бізнес-процеси	- забезпечують виробництво ресурсами, при цьому не створюють вартості та не додають до неї.
Забезпечувальні бізнес-процеси	- виходи можуть продаватися на зовнішньому ринку; - не мають стратегічного значення; - можуть перетворитися в основний бізнес-процес; - можуть відмерти у випадку наявності конкурентоспроможних альтернатив на зовнішньому ринку або передачі їх виконання на аутсорсинг.
Бізнес-процеси управління	- мають типову внутрішню структуру: планування, організація, облік, контроль, регулювання; - відмінність між процесами управління визначається специфікою об'єктів управління, якими процес управляє.
Бізнес-процеси розвитку	- на 80% – це проекти, які виконуються лише один раз; - вимагають інших технік управління, які називають технологіями управління проектами; - висувають інші вимоги до проектного менеджера, на відміну від вимог до менеджера операційного.

За ступенем впливу на додану вартість бізнес-процеси бувають:

– **основні** – це бізнес-процеси, які забезпечують випуск продукції для зовнішніх споживачів;

– **забезпечувальні** – це бізнес-процеси, що забезпечують протікання основних процесів, за допомогою надання в потрібний час необхідної кількості різних видів ресурсів;

– **процеси менеджменту** – це бізнес-процеси, які забезпечують функціональну та своєчасну реорганізацію об'єкта управління.

За *межами реакції* підприємства бізнес-процеси поділяють на:

– **зовнішні процеси** – процеси, які мають вхід або вихід поза підприємством;

– **внутрішні процеси** – процеси, входи та виходи яких цілком знаходяться в рамках підприємства.

Якщо описувати структуру бізнес-процесів зверху-донизу, тобто надати **ієрархічну класифікацію**, то бізнес-процеси можна умовно поділити на три основні підгрупи:

– **бізнес-процеси верхнього рівня** – це ключові, найбільш значущі для підприємства бізнес-процеси, які виражають цільову функцію його діяльності та визначають її структуру;

– **бізнес-процеси середнього рівня** – це процеси, що спрямовані на реалізацію тактичних цілей;

– **бізнес-процеси нижнього рівня** – це процеси, що спрямовані на реалізацію операційних цілей і виконання стандартних процедур.

1.4. Основні етапи аналізу бізнес-процесів

Аналіз бізнес-процесів – це систематичне отримання даних з метою ідентифікації, визначення, оцінки і подання процесу як основи для його організації та покращення. Приводом для проведення такого аналізу, зазвичай, є конкурентне становище підприємства на ринку. Аналіз існуючих бізнес-процесів може прояснити поточну ситуацію та підштовхнути до її поліпшення.

Аналіз бізнес-процесів можна надати у вигляді наступних взаємопов'язаних етапів:

Етап 1. Постановка мети описання бізнес-процесу.

Метою описання бізнес-процесу, зазвичай, вибирається одна або декілька цілей з переліку найбільш часто виконуваних завдань.

Етап 2. Структуризація мети проекту.

Структуризація мети проекту, описання аналізу та організації бізнес-процесів, яка складається з наступних кроків (рис. 1.3).

Етап 3. Визначення показників для вимірювання ступеня досягнення цілей.

Для кожної сформульованої мети повинен бути сформульований один або декілька ключових показників ефективності, що дозволяють установити планові значення та моніторити їх дотримання.

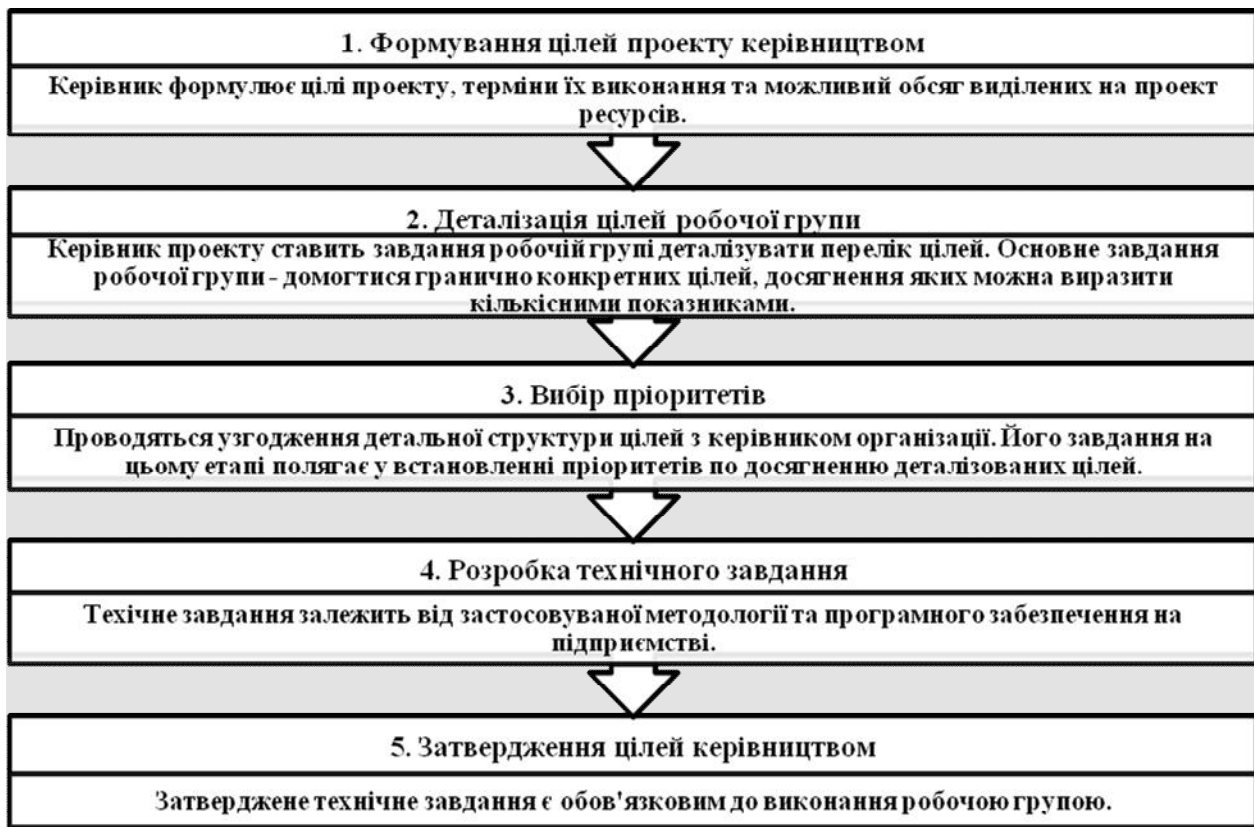


Рис. 1.3. Структуризація мети проекту

Показники повинні володіти низкою властивостей, які забезпечать власника процесу інформацією для описання повної картини, що відбувається на підлеглому йому об'єкті. До цих властивостей відносять:

- 1) повне, адекватне, об'єктивне відображення справ, що проходять у бізнес-процесі;
- 2) забезпечення періодичного збору та обробки даних;
- 3) цінність даних повинна перевищувати витрати на їх отримання та обробку;
- 4) показники повинні забезпечувати якісну та кількісну характеристику бізнес-процесу;
- 5) представлена інформація повинна бути однозначною і зрозумілою;
- 6) отримані показники повинні мати властивість порівнянності з аналогічними показниками.

Показники, які необхідно використовувати при аналізі бізнес-процесу, поділяються на чотири групи:

1. *Показники вхідних об'єктів і ресурсів.* Вони повинні відображати всі параметри і характеристики (якісні та кількісні), за

якими можливо вчасно прийняти рішення щодо усунення дій в бізнес-процесі, які призводять до помилок.

2. *Показники ефективності функції перетворення.* Вони відображають ступінь виконання запланованої діяльності та досягнення запланованих результатів, зв'язок досягнутих результатів, витрачених ресурсів і вхідних об'єктів.

3. *Показники результату функції перетворення.* Необхідність отримання цих показників визначається затребуваністю результатів споживачами (клієнтами), тобто готовністю споживача (клієнта) придбавати за певну плату результат бізнес-процесу. Одна з основних цілей реалізації бізнес-процесів, отримання результатів функцій перетворень з подальшою їх реалізацією та отримання від цього вигоди – це зростання добробуту власників бізнесу. Результати функцій перетворення характеризуються затребуваністю споживача продуктів (послуг) підприємства, яка характеризується в подальшому сумою отриманого прибутку від їх реалізації та рівнем рентабельності.

4. *Показники задоволеності замовника.* Спочатку необхідно відібрати будь-яку кількість параметрів, які підприємство вважає важливими та оцінити їх за 5-бальною шкалою при відсутності інших одиниць вимірювання. Зазвичай оцінюють ставлення покупців до продукту або послуги, сервіс, операційні стандарти, привабливість реклами і ціни на товари порівняно з конкурентами. Після визначення декількох оптимальних кількісних і якісних показників задоволеності клієнта їх необхідно об'єднати в один єдиний критерій – індекс задоволеності клієнта (customer satisfaction index, CSI). CSI є одним з видів підсумкових показників, з яким керівництво підприємства може знайомитися з певною періодичністю, наприклад, один раз на місяць. Даний індекс буде показувати наскільки добре воно задовольняє потреби своїх клієнтів.

Етап 4. Описання бізнес-процесів підприємства.

Основним завданням цього етапу є проведення ранжування бізнес-процесів з метою вибору найбільш пріоритетних із них. Відібрані бізнес-процеси необхідно детально описати і призначити відповідальних за їх виконання. У результаті буде побудована матриця розподілу відповідальності, що дозволяє визначити наявність провалів у розподілі функцій і їх дублювання.

Етап 5. Безпосередній аналіз бізнес-процесів.

Безпосередньо процес аналізу бізнес-процесів складається з двох складових: якісного та кількісного аналізу.

У свою чергу **якісний аналіз** – це сукупність декількох різних методів, які надають словесне описання бізнес-проекту:

- 1) якісний аналіз на основі суб'єктивних думок експертів (наприклад, SWOT-аналіз);
- 2) візуальний якісний аналіз графічних схем бізнес-процесу;
- 3) порівняння стану процесу з вимогами та стандартами.

Якісний аналіз носить суб'єктивний характер, оскільки будується на думці експертів і не має математичної бази.

Більш детального дослідження заслуговує кількісний аналіз бізнес-процесів, оскільки він є більш детально розробленим і надає об'єктивний результат. Більша частина методів кількісної оцінки заснована на зборі, обробці та статистичному аналізі інформації про процеси. До цих методів відносять, у першу чергу: статистичний аналіз ефективності процесу; моделювання процесів; АВС-аналіз процесу. Крім того, здійснюється аналіз задоволеності клієнтів, порівняльний аналіз процесів тощо.

Виділяють чотири основні етапи статистичного аналізу бізнес-процесів:

- 1) планування дослідження (аналізу);
- 2) попередній аналіз інформації;
- 3) оцінка (дозволяє отримати на основі даних числове значення невідомої величини);
- 4) перевірка гіпотез (дані використовуються для прийняття рішення про відповідність висунутого припущення дійсності).

Прикладами оцінки невідомих величин у бізнесі можуть бути:

- обсяг продаж у наступному місяці;
- реакція населення м. Одеси на новий продукт;
- рівень браку у виробничому процесі;
- зміна продуктивності праці при зміні стратегії розвитку підприємства.

Приклади гіпотез, які можна перевірити з використанням інформації:

- середній мешканець м. Одеси в наступному місяці планує витратити на придбання нашого продукту хоча б 500 грн.;
- завтра на виборах переможе Степаненко В.;
- засіб марки "Х" ефективно пере та відбілює;

- нове медичне обладнання безпечно та ефективно;
- помилка у фінансовому звіті менша, ніж певна величина.

Статистичний аналіз ефективності бізнес-процесів передбачає можливість їх оцінювати. Оцінка дозволяє швидко зрозуміти, які процеси вимагають пильної уваги і повинні змінюватися в першу чергу. Це дозволяє знайти слабку ланку в усій системі бізнес-процесів.

Будь-який процес можна охарактеризувати за такими **критеріями**.

Результативність процесу: досягає процес необхідних результатів чи ні. Наприклад, якщо в результаті процесу «Виготовлення сметани», вийшла сметана – процес є результативним.

Ефективність – здатність бізнес-процесу забезпечувати задану результативність при мінімальному використанні ресурсів і відсутності непродуктивних витрат, тобто ефективність показує скільки ресурсів витрачає процес на отримання результату. Якщо відомо, що на приготування 1 кг сметани має піти 5 л молока, а було витрачено 8 л, то процес не є ефективним.

Визначеність процесу означає ступінь відповідності реального процесу його документальному описанню. Якщо сметана була виготовлена згідно певного стандарту, то процес є визначеним.

Повторюваність є однією з найважливіших характеристик процесу. Вона показує чи можна в результаті даного процесу отримувати однакові результати кожного разу, чи ні. Якщо підприємство постійно видає різну (за смаковими та іншими характеристиками) сметану одного виду, то щось не так з процесом.

Здатність до адаптації – це характеристика гнучкості бізнес-процесу, що означає здібність змінюватися залежно від умов, тобто здатність відповідати майбутнім змінам вимог споживачів або задовольняти певні поточні потреби деяких із них.

Тривалість процесу – час, необхідний для його виконання, або проміжок часу між початком процесу і його завершенням.

Вартість – це сукупність усіх витрат на виконання процесу один раз. Для цього необхідно підрахувати скільки молока витрачається на виготовлення, наприклад, 1 т сметани, скільки коштує робочий час працівників, які її виготовляють, скільки коштує використання устаткування тощо.

Одне з центральних місць в аналізі бізнес-процесів займає їх моделювання – створення організаційних моделей з метою вивчення

та поліпшення діяльності підприємства. Існують різні види моделювання бізнес-процесів. Основними з них є такі.

1. Імітаційне (динамічне) моделювання – це розробка інформаційних моделей бізнес-процесів за допомогою програмних засобів для імітації виконання бізнес-процесів у часі.

2. Графічне моделювання бізнес-процесів – розробка певного набору графічних моделей бізнес-процесів, які відображають зміст бізнес-процесу, послідовність і взаємозв'язок робіт, що входять до його складу, їх виконавців, матеріальні та інформаційні потоки тощо.

3. Карта бізнес-процесів – це діаграма, що відображає укрупнену сукупність бізнес-процесів підприємства.

4. Дерево бізнес-процесів – діаграма, на якій подається ієрархічно упорядкована сукупність бізнес-процесів підприємства.

Серед методів аналізу бізнес-процесів важливе місце займає АВС-аналіз. Це метод, який дозволяє класифікувати ресурси підприємства за ступенем їх важливості. Він є одним із методів раціоналізації та може застосовуватися на будь-якому підприємстві. В основі даного методу лежить принцип Парето: 20% усіх товарів дають 80% обороту. По відношенню до АВС-аналізу правило Парето може прозвучати так: надійний контроль 20% позицій дозволяє на 80% контролювати всю систему, будь-то запаси сировини та комплектуючих, або запаси готових виробів тощо.

Питання для самоконтролю

1. Дайте визначення бізнес-статистиці.
2. Наведіть визначення бізнес-процесу та назвіть його структурні елементи.
3. Дайте перелік основних характеристик будь-якого бізнес-процесу.
4. Наведіть класифікацію бізнес-процесів за критеріями програми *TOPP*.
5. Охарактеризуйте класифікації бізнес-процесів згідно з програмою *ENAPS*.
6. Охарактеризуйте класифікацію бізнес-процесів відповідно до функцій, які вони виконують.
7. Наведіть класифікацію бізнес-процесів за методом впливу на додану вартість.
8. Дайте визначення бізнес-процесам верхнього, середнього та нижнього рівня.
9. Які бізнес-процеси називають внутрішніми та зовнішніми?
10. Назвіть основні етапи аналізу бізнес-процесів.
11. Назвіть основні кроки структуризації цілей проекту.
12. Які критерії повинні бути властиві показникам аналізу бізнес-процесу?
13. Якими критеріями може бути охарактеризований процес?
14. У чому полягає сутність АВС-аналізу?

РОЗДІЛ 2. СТРУКТУРИ ДАНИХ. КЛАСИФІКАЦІЯ, РОЗПОДІЛ І ПЕРЕТВОРЕННЯ ДАНИХ

2.1. Сутність даних як ресурсу підприємства

У більшості сфер діяльності людини, що пов'язані з функціонуванням будь-якого бізнес-процесу, прийняття управлінських рішень базується на використанні й аналізі наявних ресурсів (фінансових, трудових, матеріальних, земельних тощо). Для того, щоб використання ресурсів було оптимальним, необхідно вміти ефективно управляти ними. Управління ресурсами, в загальному сенсі, означає здатність здійснювати над ними певне коло процедур: розподілу, планування, використання, заощадження. Успішне вирішення цих завдань потребує вивчення основних властивостей цих ресурсів – отримання певних *даних*.

Таким чином, одним із основних видів ресурсів підприємства є **інформація**, тобто сукупність відомостей, які сприймаються з навколишнього середовища, передаються у навколишнє середовище або зберігаються всередині інформаційної системи. Інформація, яка необхідна для управління іншими ресурсами підприємства, витягується із зібраних, проаналізованих *даних* про властивості певного виду ресурсу. Можливість моделювання і збереження у пам'яті ЕОМ будь-яких даних (бухгалтерських та інвентарних документів, звітів, картотеки тощо) дозволяє розглядати дані в якості одного з видів ресурсу підприємства.

Використання даних як ресурсу передбачає:

- наявність уявлення про саме поняття сутності даних;
- вміння збирати та аналізувати дані про той чи інший об'єкт;
- визначення природи і властивостей даних, для чого необхідно знати як і з якою метою вони використовуються, де зберігаються і звідки беруться; таким чином, для керівництва даними необхідно зібрати найбільш вичерпну інформацію про них;
- наявність середовища для зберігання даних, які можуть бути надійно збережені тільки за умови чітких процедур планування, збирання та накопичення;
- можливість отримання доступу до ресурсу даних усюди, де потрібна інформація, яка є необхідною для управління іншими ресурсами, тобто інтегровані дані;

- наявність адміністратора даних.

Історично розглядання даних як ресурсу стало можливим із появою зовнішніх носіїв інформації та розміщенням на них наборів даних.

Дані мають цілу низку *специфічних властивостей*:

- вони є невичерпними, що забезпечує можливість їх багаторазового та багатоцільового використання;
- вони не відчужені під час продажу та обміну;
- обсяг даних постійно зростає;
- склад даних постійно змінюється, реагуючи на зміни інформаційних потреб підприємства;
- можливість оперативного пошуку даних;
- наявність засобів зберігання великого обсягу даних;
- наявність доступу до віддалених ресурсів (за допомогою веб-рішень та розвитку хмарних технологій);
- можливість фіксації на одному носії різних за характером даних.

Саме дані сьогодні стають одним із найважливіших видів ресурсів сучасного підприємництва, а згодом їх значимість буде тільки зростати. Одне із свідчень феномену полягає в тому, що дані стають *товаром*, сукупна вартість якого на ринку порівнянна з вартістю традиційних ресурсів.

Великий обсяг інформації та можливість її подання в електронному вигляді призвели до появи розвиненого ринку інформаційних ресурсів, товарів і послуг. У даний час у багатьох країнах (включаючи й Україну) сформувався для цього національний ринок; можна побачити й явні ознаки відповідного світового ринку інформації. Цей ринок багато в чому подібний до ринку інших ресурсів, оскільки має певну номенклатуру, в якості якої на ньому виступають специфічні набори даних або послуги з їх аналізу, зберігання, трансформації тощо.

2.2. Способи класифікації наборів даних

Набір даних містить одне або декілька значень для кожного з окремих об'єктів, які називаються **елементарними одиницями**. У якості таких об'єктів можуть виступати люди, домогосподарства, міста, телевізійні приймачі або що завгодно, що представляє інтерес для вивчення.

Для кожного з об'єктів реєструють одну й ту саму ознаку (або ознаки). Ознака, яку реєструють (наприклад, вартість), називається **змінною**.

Існують чотири основних способи **класифікації наборів даних**:

- за кількістю змінних (2.2.1);
- за типом вимірювання (2.2.2);
- залежно від того, чи є важливою впорядкованість набору в часовому зрізі (2.2.3);
- залежно від того, чи збиралася інформація спеціально для даного аналізу або дані збиралися раніше кимось іншим для своїх власних потреб (2.2.4. – 2.3).

2.2.1. За кількістю змінних розрізняють набори даних: **одновимірний, двовимірний і багатовимірний**.

Одновимірний набір даних (одна змінна) містить інформацію тільки про одну ознаку, що зареєстрована для кожного об'єкта. Одновимірний набір даних дозволяє визначити типове значення та характеристику мінливості даних, а також виділити специфічні особливості або проблеми в даних. Типовий приклад одновимірних даних наведений в табл. 2.1.

Таблиця 2.1

**Фінансовий результат діяльності підприємств України,
січень-вересень 2017 р.**

Галузь економіки	Фінансовий результат, млн. грн.
Оптова та роздрібна торгівля, ремонт автотранспортних засобів і мотоциклів	31890,0
Транспорт, складське господарство, поштова та кур'єрська діяльність	11178,5
Інформація та телекомунікації	11003,6
Фінансова та страхова діяльність	49330,3

Можна надати й інші приклади одновимірних наборів даних.

1. Доходи окремих людей, які визначені в рамках перепису населення. Статистичне дослідження цих даних дозволило б виявити структуру доходів, їх середній рівень, варіацію доходів залежно від стажу, рівня освіти, географічного розташування тощо.

2. Тип пакування, який обрано членами фокус-групи в ході маркетингового дослідження. Подальший аналіз даних дозволить обрати найкращий варіант пакування нової продукції.

Двовимірний набір даних (дві змінні) містить дві ознаки, значення яких реєструються для кожного об'єкта. Двовимірні дані, на додаток до інформації про кожну змінну як набір одновимірних даних, дозволяють вивчити зв'язок між двома змінними та передбачити значення однієї змінної на основі значення іншої. Прикладом може слугувати табл. 2.2, яка містить дані про фінансовий результат діяльності підприємств деяких галузей економіки України за три квартали 2017 р., а також зміну цього показника порівняно з відповідним кварталом попереднього року.

Таблиця 2.2

Динаміка фінансового результату промислових підприємств України, січень-вересень 2017 р.

Галузь економіки	Фінансовий результат, млн. грн.	Зміна фінансового результату у % до відповідного періоду 2016 р.
Оптова та роздрібна торгівля, ремонт автотранспортних засобів і мотоциклів	31890,0	+ 620,0
Транспорт, складське господарство, поштова та кур'єрська діяльність	11178,5	+ 16,3
Інформація та телекомунікації	11003,6	+ 720,4
Фінансова та страхова діяльність	49330,3	+ 18,1

Розглянемо ще декілька прикладів двовимірних наборів даних.

1. Дані за звітний період про витрати на виробництво продукції (**перша змінна**) та обсяг виготовленої продукції (**друга змінна**) великих підприємств однієї галузі, які виробляють однотипну продукцію. Статистичний аналіз таких двовимірних даних зміг би виявити взаємозв'язок між витратами підприємств та обсягом виробництва. Зокрема, в ході аналізу можливо було б виокремити постійні та змінні витрати, а також визначити ефективність функціонування підприємств галузі.

2. Дані опитування серед туристів міста, чи користувались вони послугами екскурсійної агенції (**перша змінна**; відповіді – «так»/«ні» або 1/0) та чи пам'ятають вони рекламу цієї агенції (**друга змінна**, відповідь до якої можна записати таким самим чином). Такі дані, наприклад, допоможуть визначити ефективність існуючої рекламної кампанії підприємства.

Багатовимірний набір даних (багато змінних) містить три або більше ознак, значення яких реєструються для кожного об'єкта. Багатовимірні дані дають можливість вивчити зв'язок попарно між змінними та передбачити значення однієї змінної на основі значення інших. У якості прикладу наведемо табл. 2.3, де деталізовано інформацію про функціонування галузей економіки України.

Таблиця 2.3

Статистична характеристика діяльності промислових підприємств України, січень-вересень 2017 р.

Галузь економіки	Фінансовий результат, млн. грн.	Зміна фінансового результату у % до відповідного періоду 2016 р.	Рентабельність діяльності, %	Кількість зайнятих працівників, тис. осіб
Оптова та роздрібна торгівля, ремонт автотранспортних засобів і мотоциклів	31890,0	+ 620,0	25,9	2182,3
Транспорт, складське господарство, поштова та кур'єрська діяльність	11178,5	+ 16,3	6,6	912,4
Інформація та телекомунікації	11003,6	+ 720,4	19,3	282,8
Фінансова та страхова діяльність	49330,3	+ 18,1	3,2	209,5

Існують й інші приклади наборів багатомірних даних.

1. Заробітна плата робітника (**виділена змінна**) та характеристики, які на неї впливають (**інші змінні** – рівень освіти, стаж роботи, категорія робіт і рівень продуктивності праці). Статистичне дослідження зможе виявити основні фактори, що впливають на рівень заробітної плати, наприклад, у середині підприємства.

2. Ціна на квартири однієї будівельної компанії (**виділена змінна**) та **ряд змінних**, які на неї впливають (поверх, на якому

знаходиться квартира; кількість кімнат; загальна площа; наявність душової kabіни; кількість санвузлів тощо). Аналіз даних показав би принцип, згідно з яким компанія проводить оцінювання побудованих осель, а також виявити, яка комбінація характеристик найбільш суттєво підвищує ціну квартири.

2.2.2. Друга характеристика набору даних, за типом вимірювання, розрізняє: *категорії* та *числа*.

Якщо змінна містить інформацію про те, до якої з декількох нечислових категорій належить об'єкт, то вона називається *якісною змінною* або *категорією*.

Існують два типи якісних змінних (категорій): *порядкові (ординальні)* та *номінальні*.

Якщо якісні дані можна звичайним чином і змістовно впорядкувати, виділити перший, другий, третій, найкращий, найгірший показник, то мова йде про порядкові дані, для яких під час статистичного аналізу розраховують *медіану*. У якості прикладу можна назвати посади для кожного члена управлінського персоналу (президент, віце-президент, чальник відділу, заступник начальника тощо), кваліфікації робітників (розряди), привабливість міст країни для туризму за результатами опитування та ін.

Якщо порядок у даних відсутній, то мова йде про *номінальну якісну змінну*. Наприклад, номінальними якісними змінними є стать людини, віросповідання, цех (його назва або номер). При статистичному аналізі номінальних якісних змінних зазвичай визначають *моду*.

Незважаючи на те, що часто значення якісної змінної можна записати за допомогою чисел, така змінна все ж залишається якісною, оскільки ці числа не мають будь-якої інтерпретації.

Значення змінних, які реєструються як числа, що мають змістовний сенс, називають *кількісними даними*. Вони можуть бути *дискретними* та *безперервними*.

Дискретна – це така змінна, яка може набувати значень тільки з деякого списку певних чисел. Прикладами такої змінної є кількість комп'ютерів на підприємстві; число викликів майстра підприємства для післяпродажного обслуговування; число клієнтів, які звернулися на підприємство за певний проміжок часу; число випадків відмов обладнання протягом доби тощо.

Безперервною вважають будь-яку змінну, яка не є дискретною. Вона може приймати значення з деякого проміжку. Прикладами

безперервної змінної є заробітна плата робітників підприємства, стаж їх роботи, прибуток підприємств промисловості тощо.

2.2.3. Залежно від того, чи є набір даних часовим рядом чи ні, розрізняють: *часові ряди (ряди динаміки)* і *дані одного часового зрізу*.

Якщо послідовність запису даних має змістовний сенс (хронологію), то відповідний набір даних є часовим рядом (наприклад, рис. 2.1).



Рис. 2.1. Динаміка індексу цін виробників промислової продукції України, 1996-2017 рр.

Розглянемо ще декілька прикладів часових рядів.

1. Ціна на цукор за останні 50 років з урахуванням інфляції. Якщо припустити, що зміни, які були в минулому, будуть повторюватися й у майбутньому, то можна використовувати ці тренди для довгострокового прогнозування.

2. Місячні обсяги продажу цифрової техніки за останні десять років. Статистичний аналіз цих даних допоможе виявити зміну в їх структурі під впливом розвитку нових технологій, а також наявну сезонну особливість із піком протягом грудня-січня.

Якщо послідовність запису даних не важлива, то відповідний набір містить дані про *однин часовий зріз*. Слова «про один часовий

зріз» у даному випадку означають лише те, що немає ніякого впорядкування у часі, а є лише інформація по деяких об'єктах у певний момент часу (свого роду «моментальний знімок»). Наприклад, курс долара в деяких обмінних пунктах міста на сьогоднішню дату, виміряна тривалість сну перед іспитом обраної групи студентів тощо.

Аналіз часових рядів є більш складним, ніж аналіз даних по одному часовому зрізу, оскільки вимагає ретельного обліку порядку спостережень.

2.2.4. Залежно від того, чи збиралася інформація спеціально для цього аналізу або дані збиралися раніше кимось іншим для своїх потреб, бувають дані: *первинні* та *вторинні*. Ми розглянемо їх більш детально у наступному питанні.

2.3. Джерела даних

Достатньо важливим питанням дослідження даних виступають джерела їх походження. Вибір цих джерел здійснюється підприємством залежно від їх надійності, доступності, вартості та важливості для економічної діяльності. Якщо підприємство здійснює цей збір самостійно (або доручає іншим особам за відповідним договором), то в результаті отримують *первинні дані*. Дані, що зібрані іншими підприємствами або людьми для власних цілей, називають *вторинними*.

Первинними даними для підприємства можуть слугувати:

– *безпосереднє спостереження*, яке здійснюється шляхом реєстрації даних та їх ознак на основі підрахунку, зважування, обмірювання, зняття показників приладів, реєстрації цін і обсягу проданої продукції, інвентаризації залишків товарно-матеріальних цінностей на складах підприємства тощо;

– *документальний облік* – спосіб спостереження, при якому використовують різні документи первинного обліку підприємств. Наприклад, документальне дослідження постачальника: якості його обладнання, сировини, рівня браку, динаміки цін на продукцію. Таке дослідження зазвичай проводить відділ постачання або технічного забезпечення. Безпосереднє спостереження та документальний облік дають найбільш достовірні дані;

– *опитування*, в ході якого джерелом даних є відомості, які дають опитувані особи. Наприклад, дані опитування відділу

маркетингу серед споживачів продукції підприємства з приводу розширення асортименту.

Основними *перевагами* первинних даних є їх унікальний, ексклюзивний характер, недоступність конкурентам та іншим стороннім особам, можливість отримання якісної інформації, що характеризує ставлення покупців до товарів ті їх якості, цін, самого підприємства та його бренду, можливість виявити мотиви поведінки покупців тощо.

У той же час, збір первинних даних має низку суттєвих *недоліків*, як, наприклад, великі витрати фінансових і трудових ресурсів, відносно тривалий характер отримання та обробки інформації, можлива неточність результатів, що виникає у випадку порушення правил збору первинної інформації. Тому до збору первинних даних прибігають тоді, коли необхідно отримати відомості для вирішення конкретного завдання, впевнитися у повноті, актуальності та достовірності вторинних даних, або отримати недоступну конкурентам інформацію з метою отримання конкурентних переваг.

Розглянемо детальніше *джерела вторинних даних*. Зазвичай вони включають як внутрішні, так і зовнішні публікації та документи, які дозволяють отримати різноманітну інформацію: про стан розвитку галузі, про динаміку цін, про рух робочої сили, напрямки науково-дослідної діяльності та нових технологій, нові джерела енергії та загальний стан її споживання тощо.

Усі *зовнішні джерела* вторинних даних можна розділити на чотири основні групи (табл. 2.4).

В якості *внутрішніх джерел* вторинних даних підприємства розглядають власні документи та звіти попередніх досліджень, наприклад, дані про запаси продукції, інформація про клієнтів та обсяги їх закупівель, сезонний та географічний розподіл продажів, дані про прибутки та збитки підприємства в цілому та його окремих підрозділів тощо. Ці дані є достовірними, важливими та ексклюзивними. Однак, попередні дані можуть не враховувати всі потреби маркетингових відділів підприємства; вони можуть бути дубльованими, а доступ до них може бути обмежений внаслідок внутрішніх перепон. Основним призначенням внутрішніх джерел даних є оцінка прийнятих маркетингових рішень, рівня обслуговування, якості продукції, ретроспективний аналіз ринку тощо.

Класифікація джерел даних

Тип джерела	Приклад	Переваги	Недоліки	Основне призначення
Офіційні видання та документи	<ul style="list-style-type: none"> - специфічні видання; - економічна та технічна література; - річні звіти організацій і підприємств; - документи професійних асоціацій; - засоби масової інформації. 	<ul style="list-style-type: none"> - високий ступінь достовірності та доступності; - невеликі витрати. 	<ul style="list-style-type: none"> - неспівставність даних; - неповнота інформації; - дані можуть бути використані конкурентами; - швидке старіння даних. 	<ul style="list-style-type: none"> - більш чітке визначення проблеми дослідження; - аналіз ринку; - пошук даних для більш детального розуміння первинної інформації.
Неофіційні джерела	<ul style="list-style-type: none"> - спілкування з постачальниками, клієнтами, посередниками, торговим персоналом тощо. 	<ul style="list-style-type: none"> - можливість отримання ексклюзивних даних. 	<ul style="list-style-type: none"> - неструктурованість даних; - трудомісткість організації контактів; - нерегулярність контактів; - сумнівна достовірність даних; - багато зайвих даних. 	<ul style="list-style-type: none"> - збір даних про клієнтів, партнерів, конкурентів або інших осіб, які мають суттєвий вплив на діяльність підприємства.
Специфічні джерела інформації	<ul style="list-style-type: none"> - придбання товарів конкурентів або використання їхніх послуг; - відвідування виробництва на підприємствах тієї ж галузі; - інші специфічні акції. 	<ul style="list-style-type: none"> - відносна доступність даних; - конкретна спрямованість отриманої інформації. 	<ul style="list-style-type: none"> - неповнота даних; - зайві дані; - висока трудомісткість; - високий рівень фінансових витрат. 	<ul style="list-style-type: none"> - збір даних про переваги конкурентних товарів, послуг, технологій.
Синдикативні джерела даних – дані досліджень спеціалізованих фірм	<ul style="list-style-type: none"> - дані про товарні ринки; - бази даних; - відомості про споживачів. 	<ul style="list-style-type: none"> - високий рівень якості даних; - регулярність оновлення даних; - вартість даних нижча за індивідуальні маркетингові дослідження. 	<ul style="list-style-type: none"> - дані можуть бути куплені конкурентами; - неспівставність даних різних досліджень; - неможливість впливати на склад даних дослідження. 	<ul style="list-style-type: none"> - відстеження зміни цінностей споживачів; - сегментація ринку; - дослідження ринкових тенденцій; - відстеження наявності товарів.

Усі джерела даних можна також розділити на два великих класи: *документальні* та *електронні* (рис. 2.2).



Рис. 2.2. Класифікація джерел даних залежно від їх носія

Серед цього переліку відмітимо наукові журнали: згідно даних статистичних досліджень вони складають близько 70% усіх наукових документів, а близько 80% фахівців різних рівнів вважають науковий журнал основним джерелом науково-технічної інформації.

Однак особливої уваги заслуговують електронні дані, що знаходяться в мережі Інтернет або у специфічних базах. На кінець 2017 р. в мережі знаходилося біля 1,3 млрд. сайтів, що майже в 2 рази більше, ніж 5 років тому. Протягом дня тільки пошукова система Google робить біля 6 млн. запитів, у рамках того ж дня офіційно продається більше 600 тис. комп'ютерів і 3,5 млн. смартфонів. Все це свідчить про зростаючу важливість електронних даних.

Класифікація Інтернет-ресурсів даних наведена в табл. 2.5.

Таблиця 2.5

Класифікація Інтернет-ресурсів залежно від способу наведення даних

Група	Характеристика	Види
Ресурси прямого доступу	Основним типом інформаційних ресурсів у Мережі є сторінка гіпертексту та (або) інформація у довільній формі: графічній, звуковій, відео та ін.	Офіційні сторінки організацій; персональні сторінки; різноманітні портали.
Ресурси вторинних даних	Ресурси, що містять огляди Інтернету, тематичні каталоги. Доступ до потрібної інформації здійснюється через посилання на інші сайти.	Каталоги посилань; торрент-трекери тощо.
Ресурси інтерактивних даних	Ресурси, що дають можливість запитувати та розміщувати власну інформацію.	Форуми; блоги; чати; телеконференції та ін.

Залежно від типу даних електронні ресурси даних можна також поділити на **тематичні** (що містять найбільш цікаву інформацію, присвячену певній галузі), **наукові публікації** (електронний варіант статей, збірників тощо), **рекламні** (більшість web-сайтів комерційних компаній, представлених у Мережі, які мають тією чи іншою мірою рекламний характер; даний вид інформації може також допомогти в пошуках потрібних даних), **довідкові** (різноманітні довідкові матеріали, нормативні бази) та **новини** (ця інформація легко доступна в Мережі, являє собою вид необроблених даних, які є цінними не стільки самі собою, скільки в контексті інших даних або в динаміці, і часто потребують подальшої обробки та аналізу).

Стосовно банків та баз даних звернемо увагу, що тільки в **Україні** для підприємців працює більше 140 відкритих джерел цього типу. До них відносяться:

- єдиний державний реєстр юридичних осіб та фізичних осіб-підприємців (irc.gov.ua/ua/Poshuk-v-YeDR);
- пошук виданих всіляких ліцензій (irc.gov.ua/ua/Poshuk-v-YeLR);
- інформація з фондового ринку України (smida.gov.ua/db);
- перевірка дипломів (osvita.net/checkdoc);

– реєстр зареєстрованих товарних знаків і послуг (uipv.org/ua/bases2);

– державна служба фінансового моніторингу (sdfm.gov.ua/articles) тощо.

До *міжнародних баз даних*, які можуть цікавити національні підприємства, можна віднести базу *LinkedIn*, яка обіймає велику кількість підприємств та потенційних працівників; *Enterprise Europe Network*, що допомагає знайти партнерів по всій Європі і стати на шлях інноваційного розвитку; *International Network, Local Hubs*, що спрямована на розвиток громад і відкриття нового бізнесу.

2.4. Типи розподілу даних у сукупностях

Перш ніж приступати до аналізу наявних матеріалів (вихідних даних), слід визначити їх розподіл, оскільки різні розподіли даних описуються за допомогою різних прийомів статистики. Визначення *виду розподілу* даних допоможе в майбутньому вибрати найбільш прийнятний статистичний критерій для перевірки гіпотез. Крім того, якщо розподіл далекий від нормального, то не можна автоматично робити узагальнюючі висновки про сукупність у цілому.

У більшості досліджень у першу чергу проводиться перевірка на «*нормальність розподілу*».

Нормальний розподіл (*розподіл Гауса*) являє собою положу криву у формі дзвону без випадкових відхилень (рис. 2.3).

Така крива є ідеальним набором даних, у якому більшість чисел сконцентровані в середній частині діапазону значень, а решта значень із згасанням симетрично розташовані по обидва боки від вершини дзвона.

Перевірка нормальності розподілу може проводитися *трьома способами*:

– за допомогою описової статистики (на комп'ютері: середня (\bar{x}), медіана (Me), дисперсія (σ^2), стандартне відхилення (σ), мінімальне і максимальне значення, коефіцієнти асиметрії (A) та ексцесу) (E));

– графічно;

– з використанням статистичних критеріїв.



Рис. 2.3. Крива нормального розподілу (розподіл Гауса)

Розподіл може бути *одновершинним*, *двовершинним* (*бімодальним*) і більше.

Серед одновершинних розподілів є *симетричні* та *асиметричні* (скошені), *гостро- і плосковершинні*.

У симетричному розподілі рівновіддалені від центру значення змінної мають однакові частоти, в асиметричному – вершина розподілу зміщена. Якщо вершина зміщена вліво, маємо правосторонню асиметрію, і навпаки (рис. 2.4).

У симетричному розподілі характеристики центру – середня, мода, медіана – мають однакові значення, в асиметричному між ними існують певні відмінності. У випадку правосторонньої асиметрії $\bar{x} > Me > Mo$, а у випадку лівосторонньої, навпаки, $\bar{x} < Me < Mo$. Чим більшою є асиметрія, тим більшим є відхилення $\bar{x} - Mo$.

Для оцінки симетричності розподілу розраховують коефіцієнт асиметрії:

$$A = \frac{\bar{x} - Mo}{\sigma} \quad (2.1)$$

Цей коефіцієнт характеризує напрямок і міру скошеності (зміщення) в середині розподілу: при правосторонній асиметрії $A > 0$, при лівосторонній – $A < 0$. Існує й більш детальна градація:

- якщо $|A| < 0,25$, то асиметрія вважається незначною;
- якщо $0,25 < |A| < 0,50$, то асиметрія вважається помірною;

– якщо $|A| > 0,50$, то асиметрія вважається суттєвою.

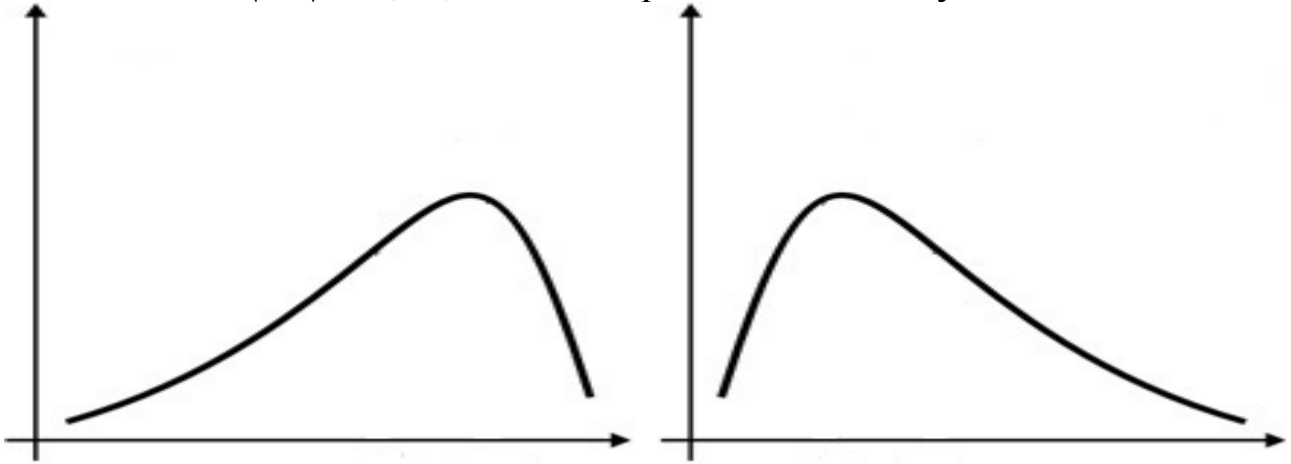


Рис. 5.4. Правосторонній та лівосторонній асиметричний розподіл

Одновершинні розподіли можуть бути *гостро-* і *плосковершинними*. Для їх оцінки розраховують коефіцієнт ексцесу, який ґрунтується на стандартизованому моменті 4-го порядку:

$$E = \mu_4 : \sigma^4 \quad (2.2)$$

де μ^4 – момент четвертого порядку, який розраховується наступним чином:

$$\mu_4 = \frac{\sum (x - \bar{x})^4 \cdot f}{\sum f} \quad (2.3)$$

У симетричному, близькому до нормального, розподілі $E = 3$; при гостровершинному $E < 3$; при плосковершинному $E > 3$.

2.5. Основні методи зберігання даних

Дані є похідними від інформації, вони являють собою запис у відповідному кодї спостереження, факт, об'єкт, що є придатним для інтерпретації, передачі, обробки та отримання інформації. Отже, поняття інформації передбачає існування матеріального носія та форми, способу, яким вона пов'язана зі своїм носієм. Таким чином, **дані** – це інформація, яка представлена у вигляді, що дозволяє автоматизувати її збір, зберігання та подальшу обробку людиною або інформаційним засобом [17, с. 3].

Відомі два способи організації масивів даних: *файлова організація* та організація у формі *баз даних*. З точки зору прикладної

програми **файл** – це іменована область зовнішньої пам'яті, на яку можна записувати дані або зчитувати з неї. Система управління файлами бере на себе розподіл зовнішньої пам'яті, відображення назв файлів у відповідні адреса зовнішньої пам'яті та забезпечення доступу до даних.

Перш за все файли застосовуються для зберігання текстових даних: фінансових документів, звітів тощо. Такі файли зазвичай утворюються та модифікуються за допомогою різноманітних текстових та графічних редакторів (*MS Word, MS Excel, Adobe Photoshop*).

Однак із розвитком цифрових технологій традиційна файлова система зберігання перестає задовольняти наростаючий потік даних та вимоги щодо їх обробки, що призводить до появи іншої форми організації даних – баз даних (*БД*).

У широкому сенсі **база даних** – це набір постійно збережених даних, що використовуються прикладними програмними системами підприємства.

Перша відмінність БД від файлової організації даних полягає в тому, що БД зберігає дані в структурованому і спеціально організованому вигляді, що виключає спільне зберігання різнорідної інформації або її дублювання. Наприклад, у текстових файлах на порядок розміщення даних не накладається скільки-небудь серйозних обмежень, і дані можуть бути розташовані довільно. Деякі дані можуть неодноразово повторюватися. У електронних таблицях дані по рядках і стовпчиках розташовуються вже впорядковано, але все ще досить довільно. Особа сама вирішує в момент створення таблиці, як краще та наочніше розмістити дані. І лише в БД структура даних суворо фіксована та визначається стандартом використовуваної моделі даних.

Друга відмінність полягає в тому, що БД зберігає не тільки дані, але й описання їх структури (мета-дані). *Мета-дані* зберігаються окремо від самих даних у так званому словнику (системному каталозі) даних. Таким чином, будь-яка БД може працювати з різними наборами даних, оскільки структура їх зберігання доступна при читанні даних. У файлової системи спосіб зберігання даних – справа кожної програми, що здійснює зберігання та обробку даних. Структура даних вбудована в програму доступу і не може бути прочитана іншими програмами.

Системне програмне забезпечення керує всіма компонентами обладнання підприємства і забезпечує доступ до нього всіх інших додатків, що працюють на комп'ютері. Прикладами системного програмного забезпечення БД можуть виступати *Windows, Linux, UNIX, MVS, Mac OS, Open Solaris* тощо.

Саме створення баз даних стало першим кроком **інформатизації підприємств** – комплексу робіт із створення єдиного інформаційного середовища підприємства, яке включає сукупність технічних, програмних, телекомунікаційних засобів, що дозволяють застосовувати в процесах управління підприємством нові інформаційні технології, здійснювати збір, зберігання та обробку даних.

Питання для самоконтролю

1. Що дало підставу вважати дані ресурсом підприємства?
2. Назвіть специфічні властивості даних.
3. Назвіть чотири основні класифікації наборів даних.
4. Що таке одновимірні, двовимірні та багатовимірні дані?
5. Що можна робити із багатовимірним набором даних?
6. Охарактеризуйте класифікацію наборів даних залежно від критерію часу.
7. Що є складнішим для аналізу: ряди динаміки чи ряди одного часового зрізу?
8. Дайте визначення первинних даних.
9. Як можна класифікувати первинні дані?
10. Дайте визначення вторинних даних.
11. Яка відмінність між первинними та вторинними даними?
12. Які існують способи класифікації вторинних даних?
13. Наведіть класифікацію Інтернет-ресурсів.
14. Наведіть приклади національних і міжнародних баз даних.
15. Який розподіл даних можна назвати нормальним?
16. Назвіть способи перевірки розподілу на нормальність.
17. Наведіть класифікації розподілу даних.
18. Що характеризує коефіцієнт асиметрії і які значення він приймає?
19. Якими можуть бути одновершинні розподіли?
20. Що характеризує коефіцієнт ексцесу і які значення він має?
21. Що таке база даних?
22. Для чого використовують файлову систему збереження даних?
23. Що таке інформатизація підприємств?

РОЗДІЛ 3. ВИПАДКОВІ ВЕЛИЧИНИ І ЗАКОНИ РОЗПОДІЛУ ЯК ОСНОВА БІЗНЕС-АНАЛІТИКИ

3.1. Випадкові величини в бізнес-аналізі

Випадковою величиною називають величину, яка внаслідок проведення дослідження може прийняти те чи інше значення (лише одне), причому до проведення дослідження невідомо яким саме воно буде. Випадкові величини позначають літерами X, Y, \dots , а їх окремі значення відповідно x, y, \dots .

У бізнес-аналітиці доволі часто виникають ситуації, які потребують роботи з випадковими величинами, наприклад, під час визначення ефективності портфеля інвестицій або в рамках проведення маркетингового дослідження з метою виявлення ступеня задоволення продукцією підприємства. Кожного разу, коли частиною випадкового експерименту є одне єдине число, ми маємо справу з випадковими величинами або так званими **випадковими змінними**.

Економістам важливо вміти розраховувати та аналізувати узагальнюючі характеристики об'єктів, які є випадковими величинами, а також виявляти ймовірність подій, які залежать від результатів спостережень, наприклад, імовірність того, що 90% клієнтів залишаться задоволеними послугами або продукцією підприємства.

Випадкові величини також можна розглядати у вигляді джерел даних. Багато з тих наборів даних, які ми обговорювали у **Розділі 2**, були отримані в результаті спостережень та фіксації значень деяких випадкових величин. У такому випадку самі випадкові величини є генеральною сукупністю, у той час як отримані значення випадкових величин є результатом вибірки.

Наведемо декілька *прикладів* випадкових величин. Варто зауважити, що кожна з цих величин є випадковою лише до того моменту, поки не буде зафіксований конкретний результат спостереження.

Випадок перший. Підприємство доставило в магазин на продаж товарів загальною вартістю 1 млн. грн. Таким чином, обсяг продажу цих товарів покупцям протягом періоду буде охарактеризовано деяким числом в інтервалі від нуля до мільйону.

Випадок другий. За 30 днів підприємством було виготовлено 2500 одиниць продукції типу А. Кількість бракованих одиниць продукції буде знаходитися в межах від 0 до 2500.

Випадок третій. Кількість кваліфікованих фахівців, які претендують на певну посаду. Якщо загальна кількість осіб, які подали документи HR-менеджеру, буде дорівнювати 20 особам, то випадкова величина може мати значення від 0 до 20.

Випадок четвертий. Кількість викликів, які поступають на телефонну лінію технічного обслуговування підприємства, може варіювати від нуля до певного цілого значення числа.

Випадок п'ятий. Заробітна плата працівника підприємства, який буде учасником соціологічного опитування. Якщо рівень заробітної плати на підприємстві варіює від 4 тис. грн. до 18 тис. грн., то і випадкова величина заробітної плати даного працівника буде знаходитися в цьому діапазоні.

Випадкову величину також можна визначити як описання численного результату певного експерименту. Саме значення називають **значенням, що спостерігають**. Наприклад, розмір прибутку підприємства в наступному кварталі – це випадкова величина, яка детермінує й описує число, що буде отримане в результаті експерименту, який полягає в тому, щоб дочекатися кінця кварталу та розрахувати отриманий розмір прибутку. Реальне значення, яке буде досягнуто в майбутньому, наприклад 30 млн. грн., є значенням випадкової величини, що спостерігають.

Варто підкреслити, що існує відмінність між випадковою величиною (яка відноситься до будь-якого випадкового процесу) і значенням, що спостерігають (яке є фіксованим числом, що реєструється в ході спостереження).

Правило визначення ймовірностей значень випадкових величин називають **правилом розподілу ймовірностей**.

Серед випадкових величин можна виділити дискретні та неперервні.

Дискретною випадковою величиною називається така величина, кількістю можливих значень якої є або скінчена, або нескінченно злічена множина.

Неперервною випадковою величиною називається така величина, можливі значення якої заповнюють певний інтервал (скінчений або нескінченний) числової осі.

У деяких випадках знайти відмінність між дискретною та неперервною випадковими величинами є досить складно. Скажімо, величина прибутку підприємства в наступному кварталі може скласти 320457 грн. або 345004,89 грн., або будь-яке інше число, яке буде меншим установленої межі, наприклад 350000 грн. У суворому розумінні, ця величина прибутку є дискретною величиною, оскільки всі ці значення можна перелічити, однак з практичної точки зору, оскільки інтервал між сусідніми значеннями є дуже малим (1 копійка), то з цією випадковою величиною можна працювати як з неперервною.

Якщо для дискретної випадкової величини є відомим перелік усіх можливих значень із ймовірностями їх отримання, то про відповідний бізнес-процес відомо абсолютно все. Із використанням такого переліку можна розрахувати будь-яку характеристику, що є цікавою для дослідника (наприклад, імовірність утрат у результаті певної виробничої операції та їх можливе значення в гривнях). Розглянемо декілька прикладів дискретних випадкових величин.

1. Обсяг виробництва деталі А на підприємстві протягом наступного тижня. Можливими значеннями будуть 0,1,2,3

2. Кількість працівників, які подали заявку на проведення відпустки у будинку відпочинку підприємства. Лист можливостей у цьому випадку має такі самі значення: 0,1,2,3

3. Розмір бюджету грантової пропозиції на проведення екологічного фестивалю за чотирьох можливих варіантів: 25 тис. дол., 50 тис. дол., 75 тис. дол. та 100 тис. дол. США. Отже, перелік можливих значень буде складатися саме з цих чисел: 25000, 50000, 75000, 100000.

Такий перелік значень разом із ймовірністю їх виникнення є розподілом імовірності для дискретної випадкової величини.

Розглянемо дискретну випадкову величину X , значення якої x_1, x_2, \dots, x_n нам відомі. Знання про можливі значення випадкової величини ще не дають нам повного описання самої випадкової величини, тому що ми не можемо сказати, як часто можуть з'явитися ті чи інші її значення. Для цього необхідно знати закон розподілу ймовірностей випадкової величини.

Внаслідок проведення дослідів випадкова величина X набуде одного із своїх можливих значень, тобто відбудеться одна подія із повної групи несумісних подій: $X = x_1, X = x_2, \dots, X = x_n$.

Позначимо ймовірності цих подій літерами p з відповідними індексами: $P(X=x_1) = p_1, P(X=x_2) = p_2, \dots, P(X=x_n) = p_n$.

Ймовірності повинні приймати значення тільки додатних чисел або нуля, а сума ймовірностей усіх можливих значень випадкової величини X дорівнює одиниці:

$$\sum_{i=1}^n P(X = x_i) = \sum_{i=1}^n p_i = 1 \quad (3.1)$$

Ця сумарна ймовірність якимось чином розподілена між окремими значеннями випадкової величини. Дискретна випадкова величина буде повністю описаною з точки зору ймовірності, якщо буде вказано, яку ймовірність має кожна з подій. Так ми встановимо закон розподілу випадкової величини.

Заданий таким чином розподіл дозволяє знайти середнє значення, середньоквадратичне відхилення та ймовірність будь-якої події, що відповідає цій випадковій величині.

Розглянемо це за допомогою прикладів. Припустимо, підприємство вирішило розширити асортимент виробництва продукції і готує в наступному кварталі випустити на ринок новий інноваційний товар, аналогів якому на даний час майже не має. Залежно від того, як буде сприйнятий товар потенційними покупцями, маркетологи підприємства розробили декілька сценаріїв фінансового результату наприкінці періоду (табл. 3.1).

Таблиця 3.1

Можливі сценарії фінансового результату підприємства

Сценарій	Фінансовий результат, млн. грн.	Ймовірність
Дуже гарний	20,0	0,200
Добрий	10,0	0,400
Нормальний	5,0	0,160
Поганий	-1,0	0,140
Дуже поганий	-5,0	0,100

Цей розподіл імовірностей дуже легко використовувати для розрахунку ймовірностей усіх подій, пов'язаних з отриманням фінансового результату. Наприклад, ймовірність отримання фінансового результату в розмірі 20,0 млн. грн. складає 0,200 або 20,0%. Ймовірність отримання фінансового результату в розмірі

10,0 млн. грн. і більше можна визначити за допомогою підсумовування ймовірностей: $0,200+0,400 = 0,600$.

Середнє або очікуване значення (математичне сподівання) дискретної випадкової величини – це певне число, що буде характеризувати типове значення цієї величини, подібно тому, як певний набір даних характеризується середнім значенням. Середнє значення для випадкової величини позначається як μ або $E(X)$ і розраховується за наступною формулою:

$$\mu = E(X) = \sum XP(x) \quad (3.2)$$

Таким чином, середнє або очікуване значення випадкової величини визначається як сума значень помножених на їх імовірності.

Якщо б імовірності отримання різних варіантів фінансового результату були однаковими, ми б отримали просте осереднення існуючих значень. Однак, оскільки ймовірності різні, ми повинні використовувати зважування всіх значень, у якому в якості вагів виступають відповідні ймовірності.

Для розглянутого вище прикладу середній очікуваний фінансовий результат буде розрахований наступним чином:

Фінансовий результат = $20,0*0,200 + 10,0*0,400 + 5,0*0,160 + (-1,0)*0,140 + (-5,0)*0,100 = 8,160$ млн. грн.

Отже, очікуваний фінансовий результат складе 8,16 млн. грн. Це значення характеризує всі можливі результати (20,0; 10,0; 5,0; -1,0; -5,0) одним числом, яке враховує ймовірність виникнення кожного з них.

Середньоквадратичне відхилення дискретної випадкової величини приблизно вказує на скільки реальні значення цієї випадкової величини можуть відрізнятись від середньої. У багатьох випадках господарської діяльності підприємств цей показник характеризує ризик, що вказує якою невизначеною є ситуація. Формула розрахунку середньоквадратичного відхилення наведена нижче:

$$\sigma = \sqrt{\sum (X - \mu)^2 P(x)} \quad (3.3)$$

Розрахуємо середньоквадратичне відхилення для нашого прикладу (табл. 3.2).

У цьому випадку середньоквадратичне відхилення буде дорівнювати наступній величині: $\sigma = \sqrt{60,054} = 7,749$ млн. грн., що

свідчить про значний рівень ризику випуску нового виду продукції. Очікуваний фінансовий результат може виявитися вищим або нижчим за 8,160 млн. грн. на 7,749 млн. грн.

Таблиця 3.2

Можливі сценарії фінансового результату підприємства

Сценарій	Фінансовий результат, млн. грн. X	Імовірність $P(x)$	$X - \mu$	$(X - \mu)^2$	$(X - \mu)^2 P(x)$
Дуже гарний	20,0	0,200	11,840	140,186	28,037
Добрий	10,0	0,400	1,840	3,386	1,354
Нормальний	5,0	0,160	-3,160	9,986	1,598
Поганий	-1,0	0,140	-9,160	83,906	11,747
Дуже поганий	-5,0	0,100	-13,160	173,186	17,319
Усього	x	1,000	x	x	60,054

Аналіз випадкових величин допоможе і в інших ситуаціях, як то в інвестиційному аналізі. Розглянемо це на наступному прикладі.

Допустимо, підприємство має у своєму розпорядженні 100,0 тис. у.о, які воно бажає інвестувати у певний проект.

Менеджери підприємства повинні обрати одну з трьох проектних пропозицій, яка буде видаватися найбільш привабливою як з точки зору суми повернених коштів, так і з точки зору ймовірності їх отримання (ризик). Вихідні дані проектів наведені в табл. 3.3.

Таблиця 3.3

Інвестиційні проекти підприємства

Проект	Сума повернених коштів, млн. грн.	Імовірність
A	110,0	1,000
B	120,0	0,500
	130,0	0,500
C	250000,0	0,001
	0,0	0,990

Знайдемо середні значення суми повернених коштів для кожного з проектів за допомогою формули 3.2. Так, для проекту A це буде 110,0 тис. у.о, для проекту B – 125,0 (120,0*0,500 + 130,0*0,500) тис. у.о, для проекту C – 250,0(250000,0*0,001 + 0,0*0,990) тис. у.о.

Отже:

$$E(X) = \mu_A = 110,0 \text{ тис. у. о.}$$

$$E(X) = \mu_B = 125,0 \text{ тис. у. о.}$$

$$E(X) = \mu_C = 250,0 \text{ тис. у. о.}$$

Якщо аналізувати лише розміри середніх величин, то можна зробити висновок, що найбільш привабливим для підприємства є *проект С*, однак він є найбільш ризикованим, адже ймовірність отримання коштів складає лише 1 %. Продовжуючи наш аналіз розрахуємо середньоквадратичне відхилення для аналізу ризиків кожного з проектів:

$$\sigma_A = \sqrt{(110,0 - 110,0)^2 * 1,000} = 0,0 \text{ тис. у. о.}$$

$$\sigma_B = \sqrt{(120,0 - 125,0)^2 * 0,500 + (130,0 - 125,0)^2 * 0,500} = 5,0 \text{ тис. у. о.}$$

$$\sigma_C = \sqrt{(250000,0 - 250,0)^2 * 0,001 + (0,0 - 250,0)^2 * 0,990} = 7897,8 \text{ тис. у. о.}$$

Розрахунки середньоквадратичного відхилення підтверджують, що *проект С* є найбільш ризикованим серед інших проектів, адже у випадку його реалізації ризик дорівнює 7997,8 тис. у.о. У той же час неможливо однозначно відповісти, який проект буде більш привабливим для підприємства, оскільки ставлення до ризику у підприємців буває різним. Деякі з них оберуть можливість отримання меншого прибутку з більшою ймовірністю, інші – більші прибутки за меншої ймовірності їх отримання.

3.2. Закони розподілу випадкових величин: сутність і види

Законом розподілу випадкової величини називають будь-яке співвідношення, що встановлює зв'язок між можливими значеннями випадкової величини та відповідними ймовірностями.

Найпростішою формою задання закону розподілу випадкової величини X є таблиця, в якій наведені можливі значення випадкової величини з відповідними ймовірностями (табл. 3.4).

Така таблиця також носить назву *ряду розподілу випадкової величини X* . Її можна представити й у графічному вигляді. У цьому випадку можливі значення випадкової величини відкладають по осі абсцис, а по осі ординат – відповідні ймовірності. Вершини з'єднують відрізками прямих, а таку фігуру називають

багатокутником розподілу. Сума ординат багатокутника розподілу завжди дорівнює одиниці.

Таблиця 3.4

Форма задання закону розподілу

x_i	x_1	x_2	...	x_n
p_i	p_1	p_2	...	p_n

Розглянутий ряд розподілу є дуже зручною формою представлення закону розподілу для дискретної випадкової величини із скінченим числом можливих значень. Для неперервної випадкової величини ряд розподілу взагалі не можна побудувати. Тобто необхідно мати такий розподіл імовірності, щоб її можна було застосовувати для найрізноманітніших випадкових величин.

Найзагальнішою формою закону розподілу випадкової величини X є так звана **функція розподілу**. Кожен закон розподілу – це деяка функція, що повністю описує випадкову величину з імовірнісної точки зору.

Функцією розподілу, або **інтегральним законом розподілу** випадкової величини X називається задання ймовірності виконання нерівності $X < x$, що розглядається як функція від аргументу x :

$$F(x) = P(X < x) \tag{3.4}$$

Функція розподілу повністю характеризує випадкову величину з точки зору ймовірності, тому вона є однією з форм закону розподілу.

Для дискретної випадкової величини функція розподілу матиме вигляд:

$$F(x) = \sum_{x_i < x} P(X = x_i) \tag{3.5}$$

З цього виразу випливає, що функція розподілу дискретної випадкової величини стрибкоподібно зростає.

Функція розподілу неперервної випадкової величини має один недолік – за нею досить важко робити висновки про характер розподілу випадкової величини у невеликому колі тієї чи іншої точки числової осі. Для цього використовують функцію, що носить назву **щільності розподілу ймовірності** або **диференціального закону розподілу випадкової величини**.

Крива, що зображує щільність розподілу $f(x)$ випадкової величини, називається **кривою розподілу**.

Випадкова величина X називається **неперервною**, якщо її функція розподілу $F(x)$ неперервна на всій осі Ox , а щільність розподілу $f(x)$ існує скрізь, окрім, можливо, скінченої кількості точок.

На практиці висновки щодо розподілу ймовірностей випадкової величини X часто доводиться робити лише за результатами випробувань. Повторюючи випробування, ми повинні кожного разу реєструвати, чи стала нам цікавою подія A , чи ні.

Відносною частотою (або **часткою**) випадкової події A називається відношення числа n_A появ цієї події до загального числа n проведених випробувань. При цьому ми приймаємо, що значення відносних частот випадкових подій близькі до їх можливостей. Це є тим більш вірним, чим більшим є число проведених досліджень. При цьому частоти, як і ймовірності, необхідно відносити не до окремих значень випадкової величини, а до інтервалів значень. Це означає, що весь діапазон можливих значень випадкової величини X потрібно розбити на інтервали. Проводячи серії випробувань, які дають емпіричні значення величини X , потрібно фіксувати числа n_x влучень результатів у кожен інтервал. При великій кількості випробувань n співвідношення n_x/n (частоти потрапляння в інтервали) повинні бути близькими до можливості попадання в ці інтервали. Залежність частот n_x/n від інтервалів визначає емпіричний розподіл імовірностей випадкової величини X , графічне представлення якої називається **гістограмою**.

Закон розподілу повністю характеризує випадкову величину з точки зору ймовірності. Однак, при вирішенні багатьох практичних задач немає необхідності характеризувати випадкову величину повністю, а достатньо мати тільки деяке загальне уявлення про неї. У теорії ймовірностей для характеристики випадкової величини використовують величини, які носять назву **числових характеристик випадкової величини**. Їх призначення – у стислій формі показати найбільш суттєві особливості того чи іншого розподілу.

Про кожен випадкову величину необхідно знати, перш за все, її середнє значення, біля якого групуються можливі значення випадкової величини, а також число, що характеризує ступінь розкиду цих значень навколо середнього. Крім цих характеристик використовують й інші, як то: *дисперсія, коефіцієнт асиметрії, коефіцієнт ексцесу, мода, медіана та моменти розподілу*.

Основними законами розподілу випадкових величин є:

1. Нормальний розподіл.
2. Біноміальний розподіл.
3. Розподіл Пуасона.
4. Експоненціальний розподіл.
5. Рівномірний розподіл.
6. Геометричний розподіл.
7. Логнормальний розподіл.
8. Розподіл Вейбула тощо.

Розглянемо ці види розподілу більш детально.

3.3. Нормальний розподіл

Нормальний закон розподілу або закон Гауса, який ми частково розглянули в другому розділі, грає виключно важливу роль у бізнес-статистиці та займає серед інших законів розподілу особливе місце. Головна характеристика, що виділяє нормальний закон серед інших, полягає в тому, що він є граничним законом, до якого наближаються інші закони розподілу при типових умовах.

Нормальний розподіл – це неперервний розподіл, який має графічне зображення у вигляді симетричної дзвіноподібної кривої, наведеної на рис. 3.1.



Рис. 3.1. Крива Гауса. Нормальний розподіл випадкової величини

Нормальний закон розподілу характеризується щільністю ймовірності у вигляді:

$$P(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}, \quad (3.6)$$

де π - співвідношення довжини кола та його діаметра, що дорівнює приблизно 3,142;

e – основа натурального логарифму, що дорівнює приблизно 2,718;

P – імовірність;

μ – математичне сподівання;

σ^2 – дисперсія;

x – випадкова величина.

Отже, конкретна форма нормального розподілу залежить від 2-х параметрів: математичного сподівання та дисперсії. Параметр μ визначає центр розподілу, якому відповідає максимальна висота графіка. Дисперсія σ^2 характеризує розмах варіації, тобто «розмазаність» даних. Таким чином, математичне сподівання зміщує центр розподілу вправо або вліво, не впливаючи на саму форму кривої щільності; дисперсія визначає загостреність кривої. Коли дані мають малий розкид, то вся їх маса сконцентрована в центрі. Якщо ж у даних великий розкид, то вони будуть розкидані по широкому діапазону.

Дзвіноподібна крива дозволяє більш наочно представити ймовірність для випадку нормального розподілу. Більш вірогідними видаються спостереження значень, які розташовані поблизу центру кривої, біля її самої високої точки (рис. 3.2, зелена область). Поблизу країв кривої, де вона йде нижче, спостереження відповідних значень стають менш правдоподібними (рис. 3.2, синя область). Імовірність того, що значення попаде в деякий інтервал на числовій прямій, дорівнює площині відповідної області під кривою, як це зображено на рис. 3.2.

Стандартний нормальний розподіл – це нормальний розподіл із значенням математичного сподівання $\mu = 0$ та середньоквадратичним відхиленням $\sigma = 1$. Для визначення випадкової величини вчені зазвичай використовують літеру Z .

Один із методів розрахунку ймовірності для нормального розподілу полягає у використанні таблиць ймовірності стандартного розподілу (*Додаток А*), оскільки взагалі неможливо розробити таблиці для кожної комбінації середньоквадратичного відхилення та

математичного сподівання. За допомогою стандартного нормального розподілу можна подати будь-який нормальний розподіл у тому випадку, коли аналізовані величини виступають не в реальних одиницях вимірювання (наприклад гривнях, метрах, кілограмах тощо), а в кількості середньоквадратичних відхилень у більший або менший бік від середнього значення.

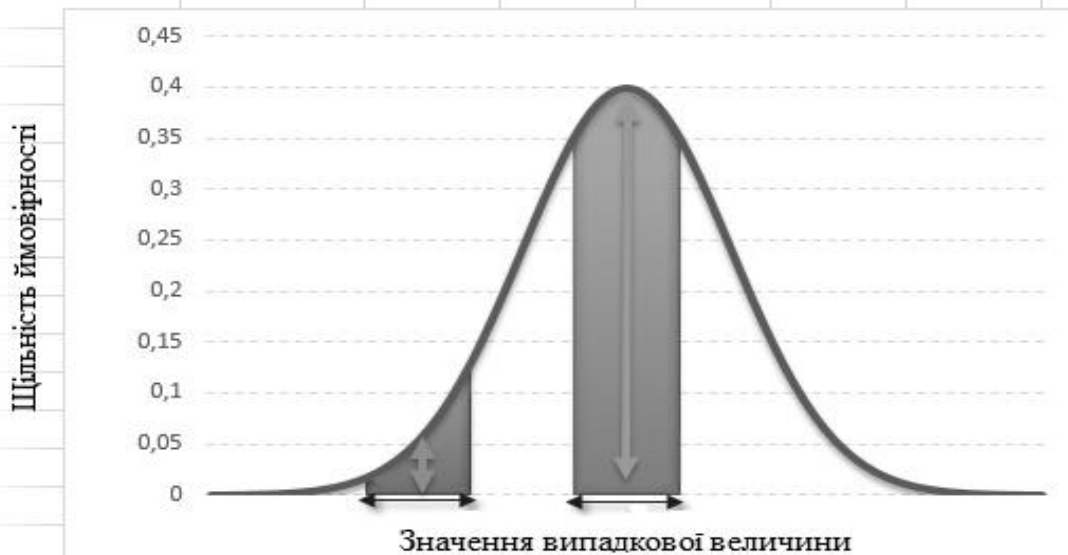


Рис. 3.2. Імовірність того, що випадкова величина приймає значення, які лежать у певному інтервалі

Додаток А ймовірностей для стандартного нормального розподілу містить у собі ймовірність того, що випадкова величина Z , яка має стандартне нормальне відхилення, приймає значення менше деякого заданого числа z . Так, ймовірність того, що величина Z менша 1,51, складає 0,9345.

Розглянемо приклад вирішення задач на визначення ймовірності за умови нормального розподілу. Припустимо, фінансовий відділ підприємства спрогнозував прибуток у третьому кварталі поточного року в розмірі 21 000 000 грн. Фактичні значення були дещо більшими і наприкінці періоду підприємство отримало прибуток у розмірі 23 300 000 грн. Прогнозований розмір прибутку на четвертий квартал складає 22 000 000 грн., середньоквадратичне відхилення, визначене з минулого досвіду, дорівнює 2 000 000 грн. З урахуванням того, що прибуток підприємства відповідає закону нормального розподілу, необхідно визначити ймовірність того, що наступний

квартал буде охарактеризований спадом ділової активності, а прибуток підприємства буде менший за 17 000 000 грн.

Проаналізувавши задану інформацію, ми можемо зробити висновок, що прибуток у розмірі 23 300 000 грн. і прогнозоване значення на третій квартал 21 000 000 грн. описують минулі події, а отже не мають ніякого впливу на вирішення поточної задачі.

У той же час варто приділити увагу наступним фактам:

- у розглянутій задачі присутній нормальний розподіл, отже його математичне сподівання згідно умови складає $\mu = 22\,000\,000$ грн.;
- стандартне відхилення складає $\sigma = 2\,000\,000$ грн.

Для того, щоб вирішити дану задачу нам необхідно провести нормування цих значень (крім математичного сподівання та середньоквадратичного відхилення), що дозволить використовувати для пошуку відповіді таблицю ймовірностей нормального розподілу із додатку А. Нормальне значення (Z) являє собою кількість середньоквадратичних відхилень від середнього значення в більший чи менший бік. Процес нормування значень можна виконати за допомогою спеціальної функції програми STATISTICS або з використанням наступної формули:

$$z = \text{Нормоване значення} = \frac{\text{значення} - \mu}{\sigma} \quad (3.7)$$

У нашому прикладі розрахунок буде мати наступний вигляд:

$$z = \frac{17\,000\,000 - 22\,000\,000}{2\,000\,000} = -2,50$$

Це означає, що в даному випадку величина 17 000 000 грн. відповідає величині $z = -2,50$. Таким чином, ми трансформували вихідну задачу в наступну задачу визначення ймовірності для стандартного нормального розподілу. Згідно з даними додатку А, ймовірність того, що квартал буде охарактеризований спадом ділової активності, складає 0,0062 або 0,62%. Тобто негативний результат для діяльності компанії є майже неймовірним, що обіцяє гарні перспективи для подальшої її діяльності.

Крім наведеного прикладу, існують ще три інших варіанти вирішення задач на пошук імовірностей. Тобто, в цілому ці задачі можна розділити на чотири типи, кожному з яких буде властивий певний спосіб вирішення (табл. 3.5). Значення z , z_1 , z_2 – нормовані значення з постановок задач.

Варіанти вирішення задач на знаходження ймовірностей

Умови завдання	Спосіб вирішення
Необхідно знайти ймовірність того, що величина Z :	знайти в таблиці ймовірностей ймовірність для z
є меншою за z	
є більшою за z	відняти від 1 попередній результат
знаходиться між z_1 та z_2	знайти в таблиці ймовірності для z_1 та z_2 , а потім відняти від більшої меншу
знаходиться за межами інтервалу між z_1 та z_2	відняти від 1 попередній результат

3.4. Біноміальний розподіл

Біноміальний закон розподілу описує ймовірність настання події A m раз в n незалежних випробуваннях, за умови, що ймовірність p настання події A в кожному випробуванні постійна.

Наприклад, у бізнес-статистиці достатньо часто використовують при розрахунках відсотки. Якщо кількість настання події виражається як відсоток до загальної кількості можливостей, то кількість разів настання події повинно мати біноміальний розподіл. У якості прикладу величин, які мають біноміальний розподіл, можна навести наступні:

1. Кількість бракованих одиниць продукції з партії товару в 100 одиниць.
2. Чисельність чоловіків у колективі відділу маркетингу з 20 осіб.
3. Кількість голосів «за» (або «проти») під час захисту дисертаційної роботи.
4. Чисельність осіб серед опитаних 200, які виказали бажання поїхати на екскурсію до музею.
5. Кількість осіб із тест-групи, які виказали бажання придбати продукцію, що вони спробували.

Графічно розподіл випадкових величин, що підкоряється біноміальному закону, можна подати в наступному вигляді (рис. 3.3).

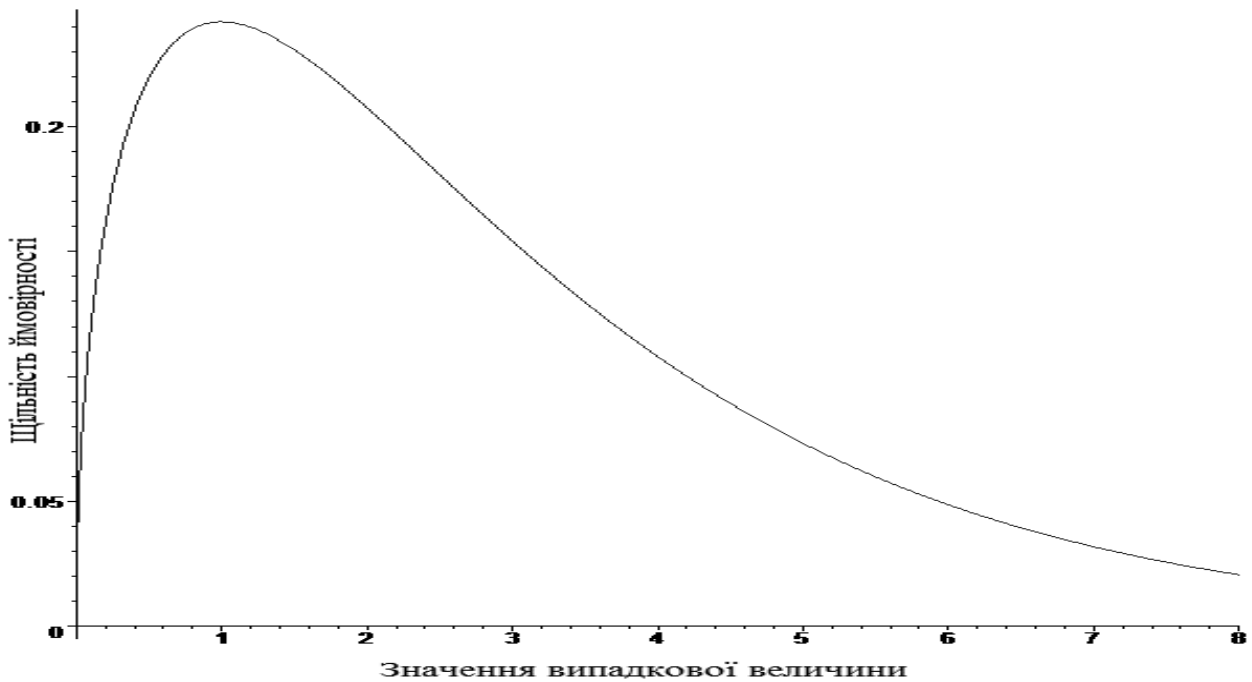


Рис. 3.3. Схематичне зображені біноміального розподілу

Розглянемо деяку конкретну подію, яка під час експерименту трапляється або не трапляється. Наявність тільки двох варіантів результату пояснює приставку «бі» у назві розглянутого закону. Випадкова величина X , яка являє собою число настання певної події в результаті n випробувань, має біноміальний розподіл тільки в наступних випадках:

- якщо всі спроби під час експерименту є незалежними одна від одної; ця вимога означає неможливість побачити результат заздалегідь. Наприклад, під час голосування людина голосує «за», оскільки бачить, що більшість інших осіб голосує так само, а вона не хоче виділятися. Для того, щоб розподіл був біноміальним, вибір повинен бути незалежним від будь-яких факторів;

- якщо в кожній з n спроб імовірність настання події Ω одна й та ж сама.

Біноміальна пропорція P – це подання випадкової змінної X , що має біноміальний розподіл, у вигляді частки кількості спроб n :

$$P = \frac{X}{n} = \frac{\text{Кількість випадків настання події}}{\text{Кількість спроб}} \quad (3.8)$$

Необхідно звернути увагу на те, що Ω – це фіксована величина, яка визначає ймовірність настання події, а P – випадкова величина, що залежить від даних, які спостерігають.

Розглянемо біноміальний закон на конкретному прикладі. Припустимо, маркетинговий відділ підприємства почергово

презентує керівництву три нових типи продукції. З попереднього досвіду відомо, що близько 70,0 % (тобто $\Omega = 0,700$) пропозицій відділу приймаються дирекцією для виконання. Необхідно визначити ймовірність того, що всі три пропозиції будуть прийняті на підприємстві, а також імовірність того, що не буде прийнята жодна з них. Для того, щоб вирішити цю задачу існує два способи, перший з яких являє собою древо ймовірностей. Древо ймовірностей (рис. 3.4) дає повне описання ситуації та ілюструє результат для кожного з трьох проектів відділу.

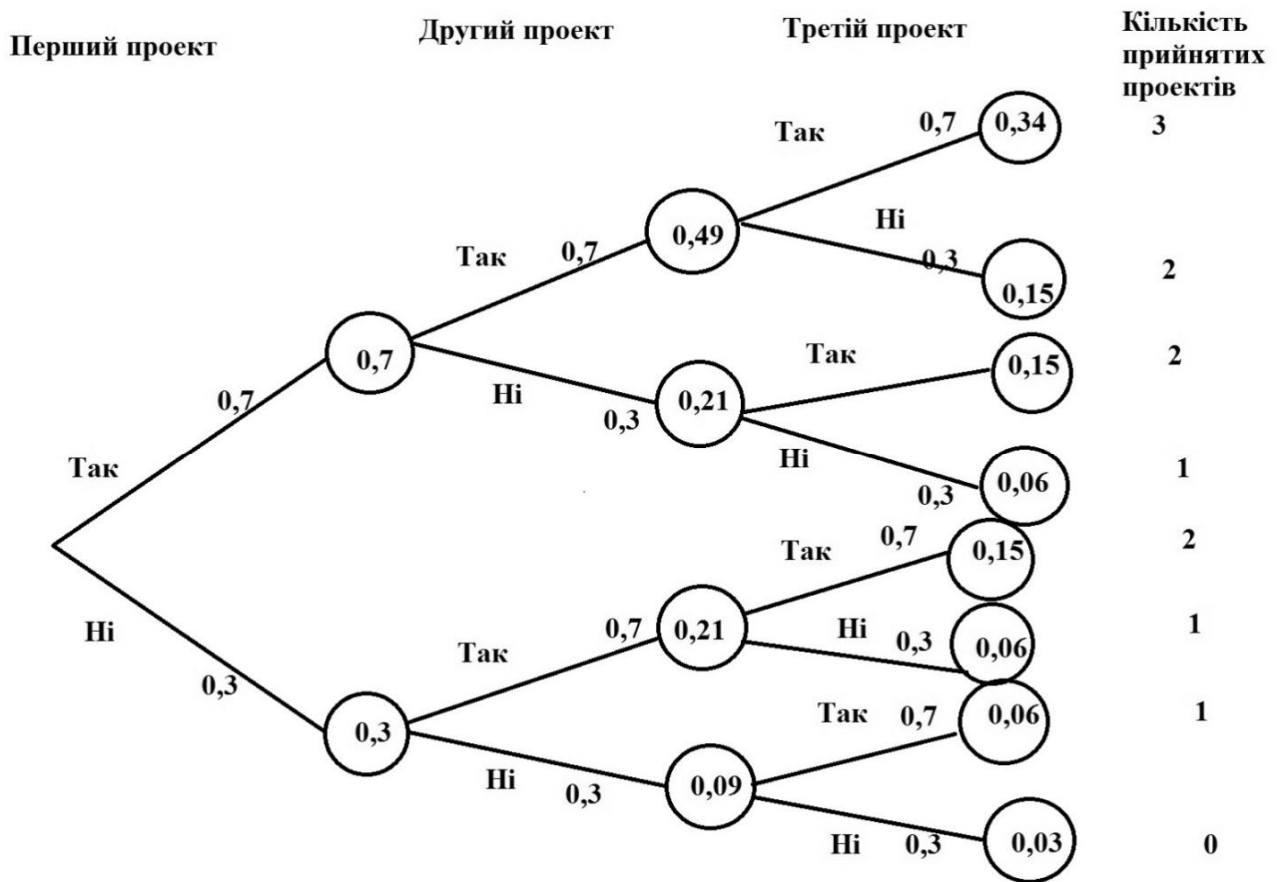


Рис. 3.4. Древо ймовірностей для послідовності трьох проектів

Таким чином, наведений рисунок свідчить, що ймовірність того, що керівництво підприємства схвалить усі три види нової продукції, складає 34%, а ймовірність того, що не буде прийнятий жоден із проектів, дорівнює 3%.

Варто звернути увагу на те, що показані впродовж гілок ймовірності завжди дорівнюють 0,7 та 0,3, оскільки проекти розглядаються незалежно один від одного. Також відмітимо, що існує

три способи ухвалення двох видів нової продукції і, відповідно, три гілки дерева, які це показують.

Для того, щоб отримати розподіл імовірності кількості схвалених проектів, потрібно підсумувати всі ймовірності різних способів реалізації даного проекту (табл. 3.6).

Таблиця 3.6

Ймовірність отримання певної кількості замовлень

Кількість схвалених нових видів товару	Відсоток схвалених нових видів товару	Імовірність
0	0,0	0,03
1	33,3	0,18 (0,06+0,06+0,06)
2	66,7	0,45 (0,15+0,15+0,15)
3	100,0	0,34

Виходячи із розподілу ймовірностей можна знайти будь-яку необхідну ймовірність за допомогою складання відповідних значень. Так, ймовірність схвалення не менш ніж 2 проектів складає $0,45 + 0,34 = 0,79$ або 79 %.

У даному випадку ми мали лише три проекти, що зробило наше дерево ймовірностей не дуже громіздким. Однак, бувають ситуації, коли величин у рази більше, тоді використання дерева ймовірностей є недоцільним, що підводить нас до другого варіанту розрахунків, пов'язаних із використанням математичного сподівання і стандартного відхилення.

Середня кількість випадків настання події (математичне сподівання) у випадку біноміального розподілу виражається за допомогою формули $E(X) = n\Omega$, тобто воно розраховується як добуток числа ймовірностей реалізації події на ймовірність її реалізації. Математичне сподівання для частки дорівнює ймовірності настання події:

$$E\left(\frac{X}{n}\right) = E(p) = \Omega \quad (3.9)$$

Існують також формули для визначення середньоквадратичного відхилення для величини, якій властивий біноміальний розподіл:

$$\sigma = \sqrt{n\Omega(1-\Omega)} \quad (3.10)$$

$$\sigma = \sqrt{\frac{\Omega(1-\Omega)}{n}} \quad (3.11)$$

Формула 3.10 використовується для розрахунку середньоквадратичного відхилення для кількості випадків настання події X , а формула 3.11 – для частки або відсотку $P = \frac{X}{n}$.

У розглянутому раніше прикладі з новою продукцією відділу маркетингу $n = 3$, а $\Omega = 0,7$. Якщо ми використаємо наведені формули, то математичне сподівання та середньоквадратичне відхилення можна буде знайти наступним чином (табл. 3.7).

Таблиця 3.7

Розрахунок показників математичного сподівання та середньоквадратичного відхилення для кількості нових товарів відділу маркетингу

Показник	Кількість випадків настання події X	Частка або відсоток, $P = \frac{X}{n}$
Математичне сподівання	$E(X) = 3 * 0,7 = 2,1$ нових товарів	$E(p) = \Omega = 0,7$ або 70 %
Середньоквадратичне відхилення	$\sigma = \sqrt{3 * 0,7(1 - 0,7)} = 0,794$ нових товарів	$\sigma = \sqrt{\frac{0,7 * (0,7 - 1)}{3}} = 0,265$ або 26,5 %

Отже, варто очікувати, що 2,1 одиниць загальної кількості нових товарів будуть схвалені керівництвом; у деяких випадках їх буде менше (3) або більше (2). Величина цієї невизначеності розраховується за допомогою середньоквадратичного відхилення від середньої, в даному випадку це 0,794 нових запропонованих товарів. Останнє число, 26,5 %, що характеризує стандартне відхилення для відсотка, трактується як відсоток від одиниці замість відсотка від деякої кількості. Це означає, що в ситуації, коли ми очікуємо 70 % імовірності настання події (ймовірності схвалення нового товару), то в реальності ми можемо отримати на 26,5 % одиниць більше цього результату ($70,0 + 26,5 = 96,5$ %) або на 26,5 % менше цього результату ($70,0 - 26,5 = 43,5$ %).

Розглянемо інший приклад, коли величини n та Ω є відомими, а нам необхідно знайти ймовірність того, що X буде дорівнювати певному значенню x . Для цього існує спеціальна формула, яку використовують у біноміальному розподілі:

$$P(X = x) = \frac{n!}{x!(n-x)!} \Omega^x (1-\Omega)^{n-x} = \frac{1 * 2 * 3 * \dots * n}{(1 * 2 * 3 * \dots * x) * (1 * 2 * 3 * \dots * (n-x))} \Omega^x (1-\Omega)^{n-x} \quad (3.12)$$

Якщо застосувати цю формулу для кожного зі значень x від 0 до n , можна повністю розрахувати розподіл імовірностей.

Припустимо, маркетинговий відділ презентував п'ять нових проектів ($n = 5$) для розгляду. Імовірність успіху кожного з них дорівнює 60 % ($\Omega = 0,6$). Розрахуємо ймовірність того, що керівництвом буде схвалено 2 проекти, тобто $x = 2$. Розрахунок буде мати наступний вигляд:

$$P(X = 2) = \frac{5!}{2!(5-2)!} 0,6^2 * 0,4^3 = \frac{1*2*3*4*5}{1*2*1*2*3} 0,36*0,0384 = 0,138.$$

Таким чином, ми можемо зробити висновок, що з імовірністю 13,8 % керівництво схвалить одночасно два нових проекти. Якщо існує необхідність визначити ймовірність успіху для 2 та більше разів, потрібно використати формулу ще три рази (для $x = 3$, $x = 4$, $x = 5$). Необхідна ймовірність буде розрахована як сума отриманих значень.

Ці розрахунки також можна проводити в редакторі *Excel* за допомогою функції *БИНОМРАСП*.

3.5. Апроксимація біноміального розподілу нормальним

Між біноміальним і нормальним розподілами випадкових величин існують дві великі відмінності: по-перше, в результаті будь-якого нормального розподілу можна отримати дробові числа, наприклад, 7,8; 5,563 тощо, у той час як результатом біноміального розподілу завжди виступає тільки ціле число; по-друге, нормальний розподіл завжди є симетричним, тоді, коли біноміальний – асиметричний (окрім випадків, коли $\Omega = 0,5$).

Однак для зручності розрахунків біноміальний розподіл можна апроксимувати за допомогою нормального розподілу якщо n є достатньо великим, а ймовірність Ω не дуже близька до 1 чи 0. Краще за все, якщо n більше 5, а Ω знаходиться в районі 0,5 [23, с. 61]. Апроксимація допоможе визначити ймовірності для біноміального розподілу складних і громіздких розрахунків на більш прості.

Використовуючи нормальний розподіл у якості заміни біноміальному, ми робимо припущення, що значення дискретної випадкової величини 2 знаходиться в проміжку між 1,5 і 2,5, що називається **поправкою на неперервність**.

Розглянемо можливість апроксимації на конкретному прикладі. Так, підприємство кожен день виробляє велику кількість шоколадних

батончиків, 40 % із яких є бракованими. Для перевірки якості працівник обирає 10 батончиків, виготовлених протягом дня. Яка ймовірність того, що 7 чи більше з них є бракованими?

Для вирішення цієї задачі, перш за все, використаємо більш громіздкий метод і скористаємось формулою 3.12. Нам буде необхідно використати формулу 4 рази для $x = 7$, $X = 8$, $x = 9$ та $x = 10$.

У результаті розрахунків отримуємо:

$$P(x = 7) = \frac{10!}{7!(10-7)!} 0,4^7 (1-0,4)^3 = \frac{1*2*3*4*5*6*7*8*9*10}{1*2*3*4*5*6*7*1*2*3} 0,4^7 * 0,6^3 = 0,042,$$

$$P(x = 8) = \frac{10!}{8!(10-8)!} 0,4^8 (1-0,4)^2 = \frac{1*2*3*4*5*6*7*8*9*10}{1*2*3*4*5*6*7*8*1*2} 0,4^8 * 0,6^2 = 0,011,$$

$$P(x = 9) = \frac{10!}{9!(10-9)!} 0,4^9 (1-0,4)^1 = \frac{1*2*3*4*5*6*7*8*9*10}{1*2*3*4*5*6*7*8*9*1} 0,4^9 * 0,6^1 = 0,002,$$

$$P(x = 10) = \frac{10!}{10!(10-10)!} 0,4^{10} (1-0,4)^0 = \frac{1*2*3*4*5*6*7*8*9*10}{1*2*3*4*5*6*7*8*9*10*0} 0,4^{10} * 0,6^0 = 0,000.$$

Підсумувавши всі отримані ймовірності ($0,042 + 0,011 + 0,002 + 0,000 = 0,055$), ми можемо зробити висновок, що з імовірністю 5,5 % у вибірці з 10 шоколадних батончиків ми знайдемо 7 і більше бракованих одиниць.

Розрахунки із заміною біноміального розподілу нормальним достатньо прості. Дискретне значення, що дорівнює 7, необхідно представити неперервною випадковою величиною в межах від 6,5 до 7,5. Замість того, щоб знаходити ймовірність дискретної величини 7 та більшої кількості бракованих одиниць, знайдемо ймовірність значення неперервної випадкової величини більшої ніж 6,5 бракованих одиниць.

За формулою 3.2 ми можемо визначити, що математичне сподівання дорівнює 4 бракованим одиницям ($\mu = 10 * 0,4 = 4$). Середньоквадратичне відхилення знаходиться за наступною формулою:

$$\sigma = \sqrt{n\Omega(1-\Omega)} \quad (3.13)$$

Воно складає 1,549 одиниць ($\sigma = \sqrt{10 * 0,4(1-0,4)} = 1,549$).

За формулою 3.7 розрахуємо значення z для 6,5:

$$z = \frac{6,5 - 4}{1,549} = 1,613.$$

У таблиці ймовірностей нормального розподілу знаходимо, що ймовірність того, що серед 10 обраних шоколадних батончиків буде знаходитися 7 або більше бракованих одиниць, дорівнює $1 - 0,9463 = 0,0537$ або 5,4%. Отже, отриманий результат майже не відрізняється від результатів розрахунків при біноміальному розподілі.

Розглянемо ще один приклад заміни біноміального розподілу нормальним для ситуації з відсотковим співвідношенням. З минулих спостережень орнітологів відомо, що із 1000 лебедів, що прилітають зимувати на озеро, 40 є чорними. Яка ймовірність того, що цієї зими на озеро прилетить більше 50 чорних лебедів?

Математичне сподівання, тобто середня величина, буде дорівнювати $\mu = \frac{40}{1000} = 0,040$. Стандартне відхилення в цьому випадку буде розраховуватися за формулою:

$$\sigma = \sqrt{\frac{\Omega(1-\Omega)}{n}} \quad (3.14)$$

та буде дорівнювати $\sigma = \sqrt{\frac{0,040(1-0,040)}{1000}} = 0,006$. Для подальшого вирішення здійснимо нормування величини z за вже відомою формулою 3.7:

$$z = \frac{0,050 - 0,040}{0,006} = 1,667.$$

Таким чином, згідно з додатком А, ймовірність того, що на озеро прилетить більше 50 чорних лебедів, складає $1 - 0,9525 = 0,0475$ або 4,8 %.

3.6. Розподіл Пуассона

Розподіл Пуассона є подібним до біноміального розподілу і пов'язаний із підрахунком кількості разів настання певної події. Відмінність є лише в тому, що у випадку розподілу Пуассона немає заданої кількості можливих спроб n . Якщо деяка подія трапляється випадково та незалежно від кожної з спроб, а середня кількість настання події із зростанням числа спроб не змінюється, то кількість випадків настання певної події у фіксованій кількості буде підкорятися розподілу Пуассона.

Таким чином, **розподіл Пуассона** – це розподіл дискретної величини, який залежить тільки від очікуваної середньої кількості можливостей настання події.

Наведемо декілька прикладів розподілу випадкових величин, які можуть бути охарактеризовані розподілом Пуассона:

1. Кількість покупців, які придбають продукцію підприємства наступного тижня.

2. Кількість дзвінків у службу підтримки протягом наступної зміни.

3. Кількість дефектів у виготовленій продукції.

4. Кількість туристів, що відвідають місто в наступному сезоні.

5. Біноміально розподілена величина X за великої кількості n та малого значення Ω .

Графічно розподіл Пуассона наведено на рис. 3.5.

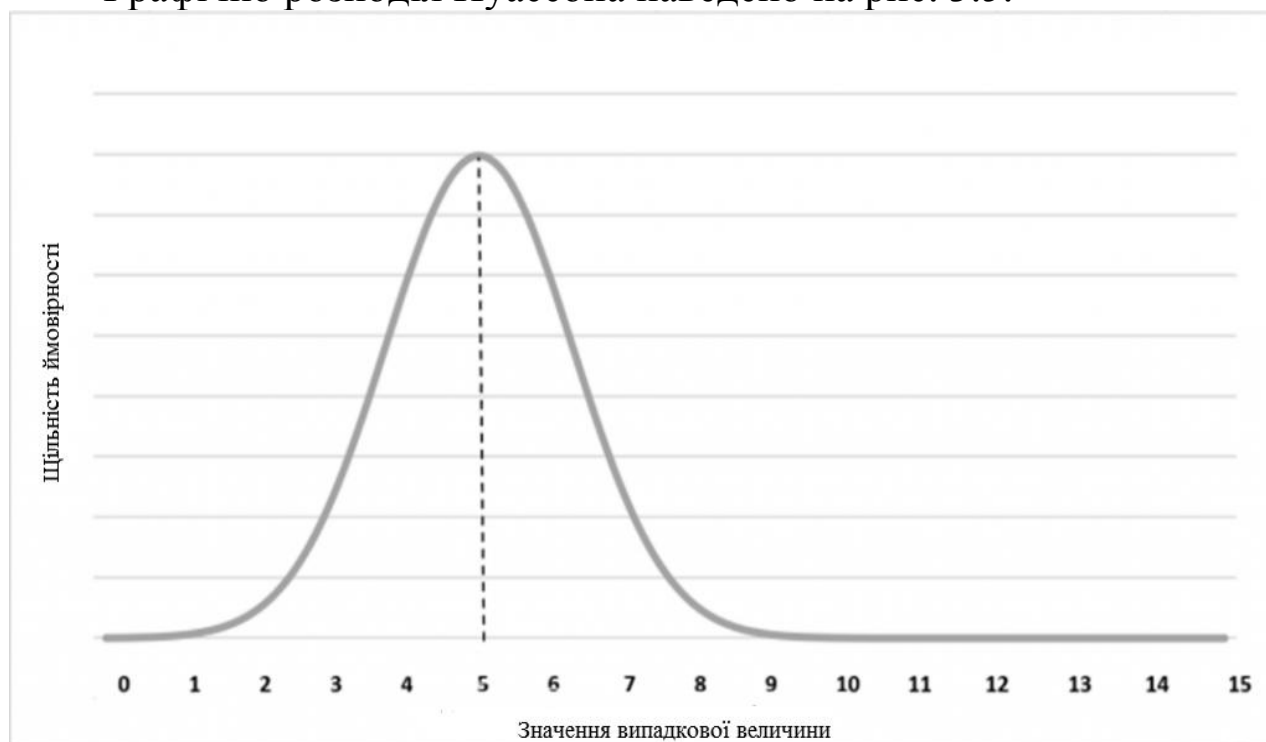


Рис. 3.5. Графічне зображення розподілу Пуассона

Варто звернути увагу на те, що схематичне зображення за своєю формою нагадує нормальний розподіл. Це свідчить про те, що в ситуації очікування великої кількості настання подій, розподіл Пуассона наближається до нормального.

Основними особливостями розподілу Пуассона є:

1. Середньоквадратичне відхилення завжди дорівнює квадратному кореню із середнього значення (математичного сподівання):

$$\sigma = \sqrt{\mu} \quad (3.15)$$

2. Імовірність того, що випадкова величина X , яка характеризується розподілом Пуассона, дорівнює деякій величині, розраховується за допомогою наступної формули:

$$P(X = a) = e^{-\mu} \left(\frac{\mu^a}{a!} \right) \quad (3.16)$$

Розглянемо цей варіант розподілу за допомогою прикладів.

Оскільки інструкції до використання товару є вичерпними та дуже детальними, середня кількість дзвінків, що надходять до служби підтримки підприємства за день, дорівнює 1,7 дзвінка. Необхідно визначити ймовірність того, що завтра до служби підтримки не буде жодного дзвінка, буде 1 дзвінок, буде 2 дзвінки.

Для розрахунків використаємо формулу 3.16:

$$P(x = 0) = e^{-1,7} \left(\frac{1,7^0}{0!} \right) = 0,183 * \frac{1}{1} = 0,183,$$

$$P(x = 1) = e^{-1,7} \left(\frac{1,7^1}{1!} \right) = 0,183 * \frac{1,7}{1} = 0,311,$$

$$P(x = 2) = e^{-1,7} \left(\frac{1,7^2}{2!} \right) = 0,183 * \frac{2,890}{2} = 0,264.$$

Отже, з імовірністю 18,3 % можна стверджувати, що завтра до служби підтримки не буде жодного дзвінка; з імовірністю 31,1 % – що буде один дзвінок; з імовірністю 26,4 % – що буде два дзвінки.

Для визначення ймовірності за допомогою редактора *Excel* необхідно скористатись функцією *ПУАССОН*.

Розподіл Пуассона, як і біноміальний розподіл, можна апроксимувати нормальним за умови великого середнього значення. Розглянемо це на прикладі.

Бригада робітників підприємства протягом зміни виготовляє в середньому 450 одиниць продукції А. Необхідно знайти ймовірність того, що протягом наступної зміни бригада виготовить 500 або більше одиниць продукції. Таким чином, згідно з методикою, описаною в параграфі 3.5, нам необхідно представити дискретну кількість деталей 500 у вигляді нормального розподілу в проміжку між 499,5 та 500,5.

Середньоквадратичне відхилення за формулою 3.15 буде дорівнювати $\sigma = \sqrt{450} = 21,213$. Нормоване значення кількості

виготовлених одиниць продукції розрахуємо за допомогою формули 3.7: $z = \frac{499,5 - 450}{21,213} = 2,333$.

Скориставшись таблицею з *Додатку А*, розрахуємо необхідну ймовірність як $1 - 0,9901 = 0,0099$ (0,99 %). Таким чином, з ймовірністю 1 % можна стверджувати, що протягом наступної зміни бригада робітників виготовить 500 або більше продукції А.

3.7. Експоненціальний розподіл

Експоненціальний розподіл – це неперервний розподіл із сильною асиметрією (рис. 3.6). У лівій частині графіку, при наближенні до 0, крива наближається до осі Ox , а в правій частині вона поступово знижується.

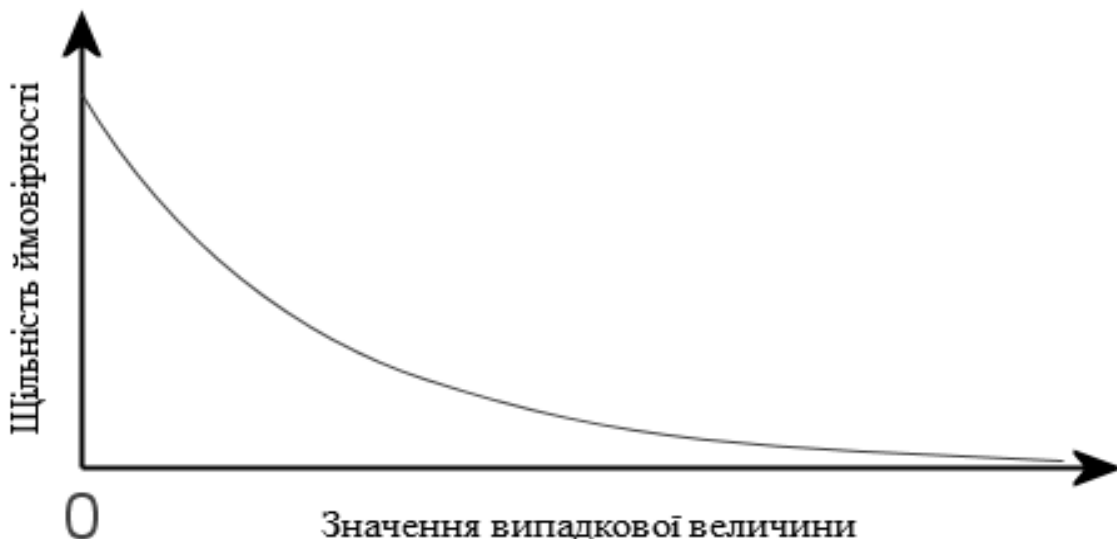


Рис. 3.6. Схематичне зображення експоненціального розподілу

Якщо події трапляються випадково, незалежно одна від одної і з постійною частотою, час очікування між двома послідовно наступаючими подіями має експоненціальний розподіл. При цьому загальна кількість подій підкорюється розподілу Пуассона.

Наведемо декілька прикладів випадкових величин, яким характерний експоненціальний розподіл:

1. Проміжки часу між появою покупців у магазині.
2. Тривалість типової розмови за допомогою Skype.

3. Витрати часу на технічне обслуговування одного клієнта в автомайстерні.

4. Час роботи процесору комп'ютера до його відмови.

5. Тривалість відповіді одного студента біля дошки.

Експоненціальний розподіл можна охарактеризувати наступною властивістю: імовірність того, що випадкова величина X , що підкоряється експоненціальному розподілу, буде менше величини a , розраховується за формулою:

$$P(X < a) = 1 - e^{-\frac{a}{\mu}} \quad (3.17)$$

Якщо події трапляються незалежно одна від одної і з постійною частотою, то між експоненціальним розподілом та розподілом Пуассона існує певний взаємозв'язок. Кількість подій для будь-якого проміжку часу має розподіл Пуассона, а час очікування між подіями – експоненціальний розподіл.

Розглянемо приклад. До маркетингового відділу підприємства незалежно один від одного протягом години дзвонять 30 осіб для того, щоб зробити замовлення. Необхідно знайти ймовірність того, що час очікування дзвінка нового клієнта, який підкорюється експоненціальному розподілу, буде менше п'яти хвилин.

Оскільки протягом кожної години в середньому дзвонять 30 клієнтів, то середнє значення випадкової експоненціально розподіленої величини буде складати $\frac{1}{30} = 0,033$ години або $0,33 * 60 = 1,98$ хвилини. Таким чином, необхідна ймовірність буде розрахована як $P(X < 5) = 1 - e^{-\frac{5}{1,98}} = 1 - 0,080 = 0,920$. Отже, з імовірністю 92 % можна очікувати, що в маркетинговий відділ підприємства протягом п'яти хвилин подзвонить новий клієнт для того, щоб зробити замовлення.

Для того, щоб розрахувати цю ймовірність в *Excel*, необхідно скористатися функцією *EXP*.

3.8. Інші види розподілу

Розглянуті в попередніх параграфах види розподілу випадкових величин є найбільш розповсюдженими їх видами в реальному житті. Однак, ми також хочемо звернути вашу увагу на існування й інших їх видів.

Рівномірний розподіл – це розподіл, який характеризується тим, що ймовірність будь-якого інтервалу залежить тільки від його довжини. Припустимо, що автобус певного маршруту курсує з інтервалом один раз в 10 хвилин, і ви у випадковий момент часу підійшли до зупинки. Імовірність того, що наступний автобус підійде через хвилину, дорівнює $\frac{1}{10}$; імовірність того, що автобус буде через 3 хвилини, також $\frac{1}{10}$; через 7 хвилин – $\frac{1}{10}$.

Графічно зобразити рівномірний розподіл можна наступним чином (рис. 3.7).

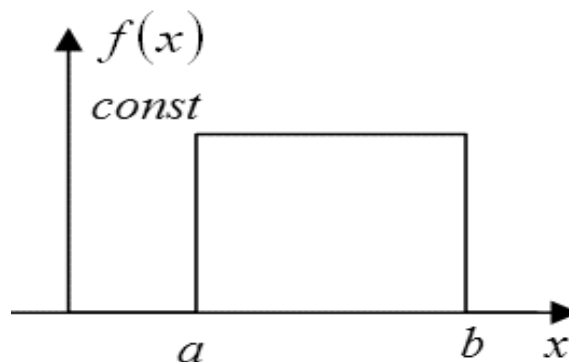


Рис. 3.7. Схематичне зображення рівномірного розподілу

Геометричний розподіл – це розподіл дискретної випадкової величини, яка дорівнює кількості випробувань випадкового експерименту, до спостереження першого «успіху». Нехай проводиться серія випробувань, у кожному з яких випадкова подія A може з'явитися з імовірністю p ; причому, випробування закінчуються при першій же появі даної події. Тоді випадкова величина, що характеризує кількість здійснених спроб, як раз і має геометричний розподіл. Графічно розподіл представлений на рис. 3.8.

Випадкова величина X називається **логарифмічно-нормально (логнормально) розподіленою**, якщо її логарифм підкоряється нормальному закону розподілу. Це означає, що значення логнормальної випадкової величини формується під впливом дуже великої кількості незалежних факторів, причому вплив кожного з цих факторів є рівномірно незначним і має рівну ймовірність щодо знаку. При цьому, на відміну від нормального розподілу, послідовний характер впливу випадкових факторів є таким, що випадковий приріст, викликаний дією кожного наступного фактору, є пропорційним уже досягнутому до цього моменту значенню

досліджуваної величини (в цьому випадку говорять про *мультиплікативний вплив факторів*).

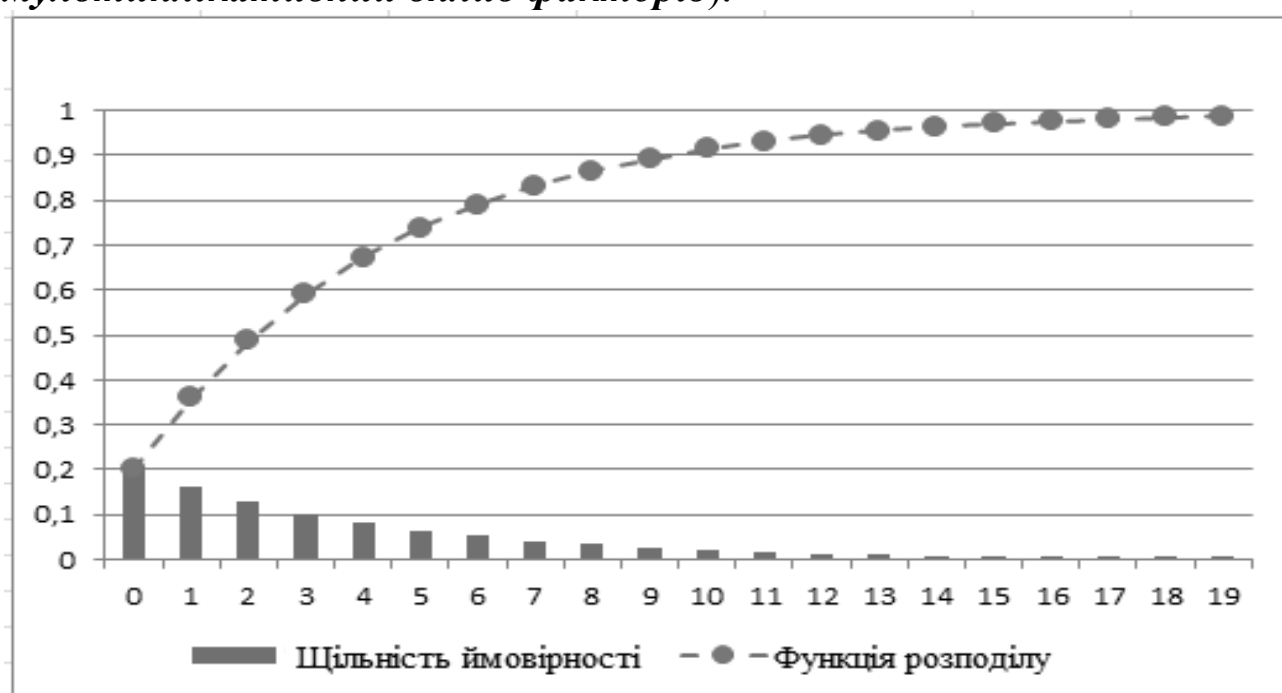


Рис. 3.8. Схематичне зображення геометричного розподілу

Схематичне зображення логнормального розподілу наведено на рис. 3.9.

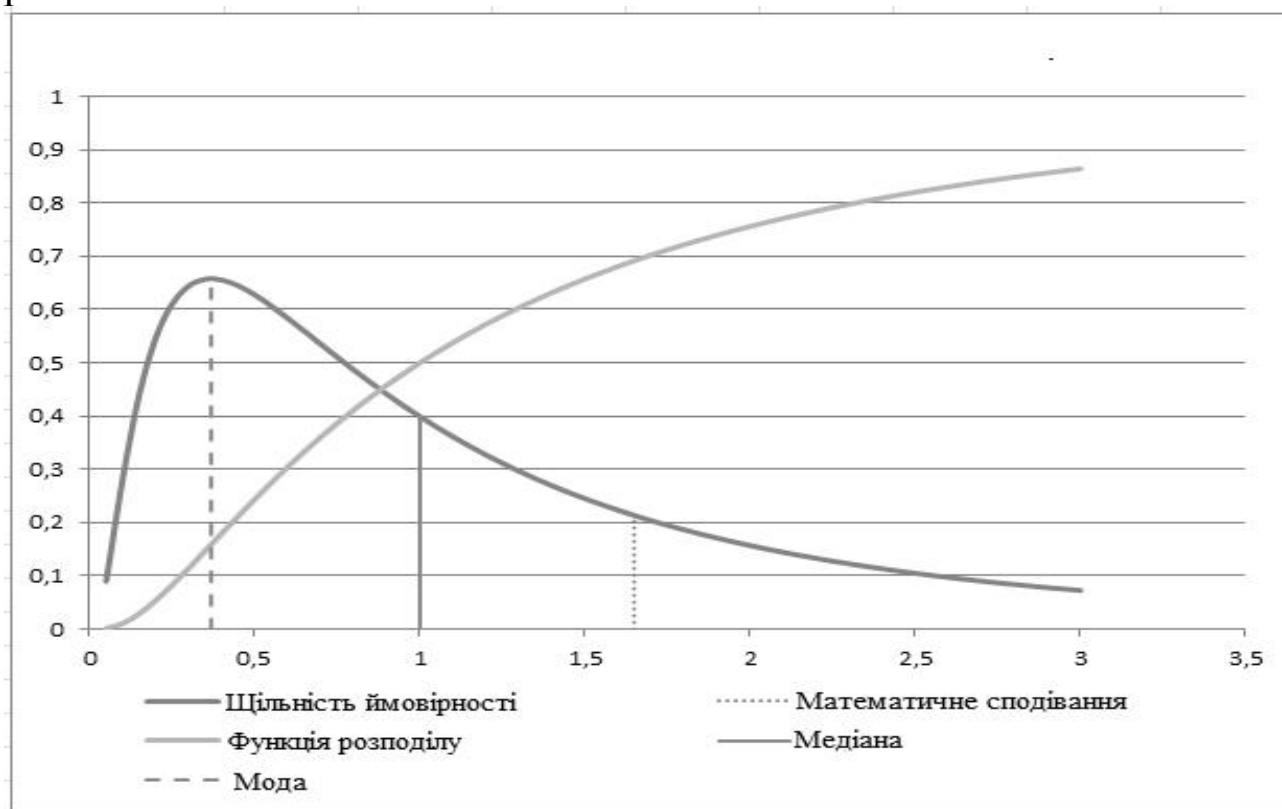


Рис. 3.9. Схематичне зображення логнормального розподілу

Розподіл Вейбула являє собою двопараметричний розподіл. Описуваний цим розподілом закон є універсальним, оскільки при відповідних значеннях параметрів перетворюється в нормальний, експоненціальний або інші види розподілів. Розподіл Вейбула залежить від двох параметрів, перший з яких (альфа) визначає форму розподілу, а другий (бета) – визначає його масштаб.

Розподіл Вейбула є адекватною моделлю для описання часу безперервної роботи безлічі технічних засобів:

- часу відмови внаслідок зносу (відмова повинна траплятися через несправність найменш надійної комплектуючої);
- часу відмови матеріалу через його псування (відмова повинна траплятися через наявність будь-якого внутрішнього дефекту).

Якщо параметр альфа дорівнює одиниці, то ми отримуємо експоненціальний розподіл, а сама причина буде зовнішньою.

Схематичне зображення розподілу Вейбула для заданих величин альфа та бета наведено на рис. 3.10.

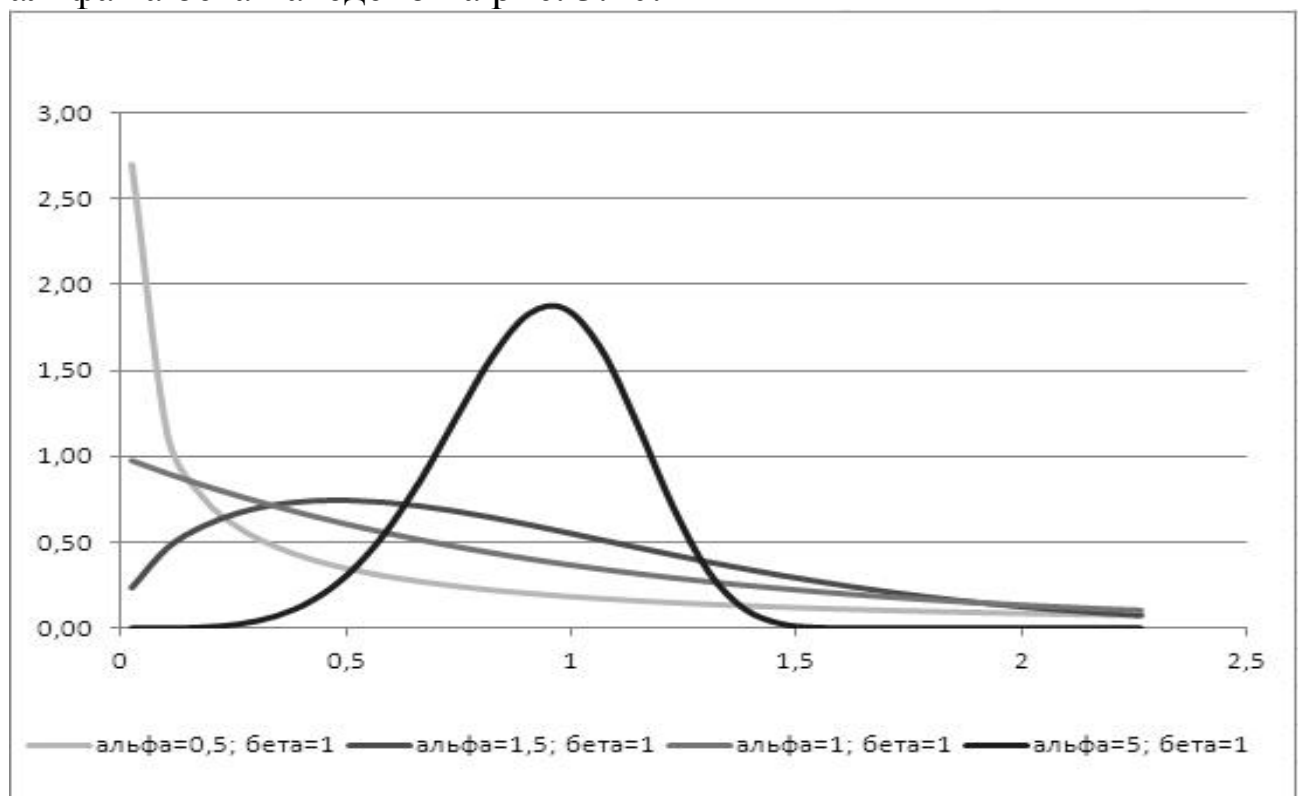


Рис. 3.10. Приклад розподілів Вейбула

Відмітимо, що існує й багато інших законів розподілу випадкових величин, з якими можна зіштовхнутися на практиці. До них відносять розподіл Бернуллі, гама-розподіл, розподіл Ерланга тощо.

Питання для самоконтролю

1. Що таке випадкова величина?
2. Яка відмінність між випадковою величиною та числом?
3. Наведіть приклади випадкових величин.
4. Що таке дискретна випадкова величина?
5. Що таке неперервна випадкова величина?
6. Що таке правило розподілу ймовірностей?
7. Дайте визначення закону розподілу випадкових величин.
8. Назвіть закони розподілу, які ви знаєте.
9. Що таке функція розподілу?
10. Охарактеризуйте нормальний закон розподілу.
11. Охарактеризуйте біноміальний закон розподілу.
12. Назвіть основний недолік використання дерева ймовірностей при розрахунках біноміального закону розподілу.
13. У чому сутність апроксимації біноміального закону нормальним?
14. Дайте визначення розподілу Пуассона.
15. Наведіть приклади біноміального розподілу та розподілу Пуассона.
16. В яких випадках розподіл дискретної випадкової величини можна представити у вигляді розподілу неперервної?
17. Назвіть основні особливості розподілу Пуассона.
18. Що таке експоненціальний розподіл? Наведіть приклади.
19. Охарактеризуйте рівномірний розподіл. Наведіть приклади.
20. Який розподіл називається геометричним?
21. Дайте визначення логнормального розподілу.
22. Що таке розподіл Вейбула? Наведіть приклади використання цього розподілу в реальному житті.

РОЗДІЛ 4. СТАТИСТИЧНІ ГІПОТЕЗИ І СТАТИСТИЧНІ ВИСНОВКИ

4.1. Сутність і види статистичних гіпотез

Статистичною гіпотезою називається будь-яке припущення відносно виду або параметрів невідомого закону розподілу. В конкретній ситуації *статистичну гіпотезу* формулюють як припущення, з певним рівнем значущості, про властивості *генеральної сукупності* на основі оцінок вибіркової сукупності.

Статистичну гіпотезу прийнято позначати латинською літерою *H* (від лат. *Hypothesis*). Так, може бути висунута статистична гіпотеза відносно того, що середнє значення в генеральній сукупності дорівнює деякій величині *a* або є меншим за деяку величину *b*. У вигляді умовних позначень це може бути записано наступним чином: *H: $\mu = a$* або *H: $\mu < b$* . Це гіпотетичне твердження є або справедливим (істинним), або помилковим (хибним), що потребує його перевірки.

Розрізняють *прості* та *складні* статистичні гіпотези. Статистична гіпотеза вважається **простою**, якщо вона повністю визначає теоретичну функцію розподілу випадкової величини. Наприклад, наведена раніше гіпотеза *H: $\mu = a$* є саме простою, оскільки значення середньої може бути описано лише певним значенням *a*.

Складна гіпотеза містить скінченну чи безкінечну кількість простих гіпотез, причому певна область значень параметру повинна бути вказаною. Наприклад, гіпотеза *H: $\mu < b$* складається з безлічі простих гіпотез *H: $\mu = c$* , в яких *c* – будь-яке значення, що є меншим, ніж *b*.

Якщо гіпотеза описує параметри генеральної сукупності, то вона називається **параметричною**; якщо гіпотеза описує закон розподілу – **непараметричною**.

Також статистичні гіпотези поділяються на нульові та альтернативні. Гіпотезу про те, що дві сукупності, які порівнюють між собою за однією чи декількома ознаками, є однаковими між собою, називають **нульовою гіпотезою** або **нуль-гіпотезою**. Ця гіпотеза позначається як *H₀* і є таким твердженням, що приймається, коли немає переконливих аргументів для його спростовування.

Наприклад, гіпотеза « $H_0 : f_{i1} - f_{i2} = 0$ » читається так: «висунута нульова гіпотеза про відсутність значущої відмінності між значеннями f_{i1} і f_{i2} ». Зазвичай нульова гіпотеза – це те, що ми хочемо спростувати, якщо перед нами стоїть завдання довести значущість відмінностей.

Нульова гіпотеза спростовується тоді, коли за вибіркою отримується результат, який при істинності нульової гіпотези є малоімовірним. Межею малоімовірного чи неймовірного вважають $\alpha = 0,05$ тобто 5% (інколи 0,01 або 0,001). Якщо орієнтуватися на правило «трьох сигм», то ймовірність помилки α повинна дорівнювати 0,0027. Однак, для цього рівня ймовірності критерії рідко табулюються: зазвичай значення критеріїв у статистико-математичних таблицях розраховані для ймовірності помилки 0,05; 0,01; 0,001 [6, с. 272].

Альтернативна або дослідницька гіпотеза є логічним запереченням нульової гіпотези та позначається як H_1 . Природно, що це гіпотеза про існування відмінностей. Наприклад, гіпотеза « $H_1 : f_{i1} - f_{i2} \neq 0$ » читається так: «висунута альтернативна гіпотеза про наявність значущої відмінності між значеннями f_{i1} і f_{i2} ». Зазвичай альтернативна гіпотеза – це те, що ми хочемо довести. Проте існують завдання, коли бажано підтвердити нульову гіпотезу і переконатися, наприклад, що вибірки не відрізняються між собою за певними показниками.

Наведемо приклади нульових та альтернативних статистичних гіпотез, які були сформульовані відносно генеральної сукупності. Необхідно звернути увагу на те, що в кожному випадку обидві гіпотези не можуть бути істинними одночасно і для того, щоб вибрати одну з них, необхідно використовувати належні дані.

1. *Ситуація*. Випадково обрана група із 100 людей взяла участь у дегустації молочних продуктів підприємства у ході рекламної акції. Після цього реєструється кількість людей із групи, які купували дегустовані продукти протягом наступного тижня.

Нульова гіпотеза. Рекламна акція підприємства не мала жодного ефекту. Інакше кажучи, відсоток покупців продукції в генеральній сукупності серед тих, хто брав участь у дегустації, дорівнює відсотку покупців продукції в генеральній сукупності, які не брали участь у дегустації. Рівень продажів залишається на минулому рівні, який, з досвіду, дорівнює 21,5 %.

Альтернативна гіпотеза. Проведена рекламна акція мала значний ефект. Тобто відсоток покупців продукції підприємства в генеральній сукупності, що приймали участь у дегустації, відрізняється від звичайного рівня продажів (21,5 %), який відповідає тим покупцям генеральної сукупності, що не брали участь у дегустації.

Обговорення. Зверніть увагу на те, що обидві гіпотези представляють собою ствердження відносно генеральної сукупності покупців, а не вибірки із 100 осіб, що приймали участь в дегустації. Вибіркові дані, що зібрані в результаті спостереження поведінки 100 випадково відібраних осіб, допоможе вирішити, яку з гіпотез необхідно прийняти. Оскільки нульова гіпотеза містить у собі точне значення відсотка (21,5 %), вона є більш визначеною, ніж альтернативна гіпотеза, яка говорить про більш широкий діапазон значень (про будь-яке значення, окрім 21,5 %). Якщо фахівці підприємства приймуть рішення, що рекламна акція з дегустацією була ефективною, то вони зроблять більш суворе ствердження, яке необхідно довести.

2. *Ситуація.* Підприємство протягом тижня оцінює нові верстати, які теоретично повинні прискорити виготовлення деталі *A*, що суттєво підвищить обсяги її виробництва.

Нульова гіпотеза. Нові верстати в довгостроковій перспективі ніяк не впливають на швидкість виготовлення та, відповідно, денні обсяги виробництва, які з минулого досвіду складають 121 деталь.

Альтернативна гіпотеза. У довгостроковій перспективі нове обладнання якимось чином впливає на швидкість виготовлення і обсяги виробництва, тобто значення випадкової величини може бути більшим чи меншим ніж 121 одиниця деталі *A* за день.

Обговорення. Нульова гіпотеза є більш певною. Обидві гіпотези сформульовані відносно генеральної сукупності (кількість виготовлених деталей у довгостроковій перспективі) замість оцінювання результатів роботи за минулий тиждень (вбірка). Таким чином, постачальнику нового устаткування необхідно довести, що воно не є неефективним. Це ні в якому разі не є завданням самого підприємства, що тестувало обладнання.

3. *Ситуація.* Власнику підприємства, який навчався у Європі та полюбляє все європейське, пред'явлено позов у несправедливому фаворитизмі співробітників, що закінчили іноземні вищі заклади

освіти, адже їх заробітна плата завжди є вищою, ніж у тих, хто навчався в національному закладі.

Нульова гіпотеза. Розміри заробітної плати працівників, які навчалися за кордоном, і тих, які не навчалися там, є рівними, без урахування випадкових відхилень у значеннях. Інакше кажучи, заробітна плата робітників не дуже б змінилася, якби власник знову сформував зарплатний фонд і розподілив саму зарплати між працівниками у випадковому порядку.

Альтернативна гіпотеза. Відмінність між заробітною платою тих, хто навчався за кордоном і тих, хто навчався в національних вищих навчальних закладах, є надто великою, щоб бути випадковою.

Обговорення. Перш за все, зверніть увагу на те, що розглянута сукупність є ідеалізованою, оскільки співробітники підприємства не можуть бути розглянуті як випадкова вибірка. Гіпотези відносяться до деякої ідеалізованої сукупності, яка представляє собою співробітників рівних з точки зору заробітної плати, а самі можливі відмінності у рівнях заробітної плати можуть бути пояснені випадковістю її розподілу між ними. Якщо нульова гіпотеза буде спростована, то власник підприємства буде мати значні проблеми.

Однак, варто пам'ятати, що статистичні методи зазвичай описують тільки самі числа, але не пояснюють, чому саме числа є такими, як вони є. Відмінність у рівнях заробітної плати може бути пояснена дискримінаційним фактором залежності міста навчання, але може бути обґрунтована й іншими факторами, як то особистими навичками та вміннями, якістю отриманої освіти, рівнем володіння іноземними мовами, отриманим досвідом тощо. Перевірка статистичної гіпотези, що розглядає лише розмір заробітної плати та місце попереднього навчання, не може показати, які саме фактори вплинули на ці відмінності.

4.2. Критерій як інструмент перевірки гіпотези.

Потужність критерію

Статистичні висновки робляться на підставі прийняття однієї гіпотези та відхилення іншої, а саме рішення приймається з певною достовірністю. Перевірка гіпотез здійснюється на основі так званих *статистичних критеріїв*.

Статистичним критерієм, або просто критерієм, називають випадкову величину K , що слугує для перевірки нульової гіпотези.

Для різних гіпотез статистичний критерій є неоднаковим, він дозволяє визначити, чи суперечить нульова гіпотеза фактичним даним, чи ні.

У цілому перевірку статистичної гіпотези на основі статистичного критерію можна представити послідовністю наступних етапів:

1. Формування завдань дослідження у вигляді статистичної гіпотези.
2. Вибір статистичної характеристики гіпотези.
3. Вибір нульової та альтернативної гіпотез на основі аналізу помилкових рішень та їх наслідків.
4. Визначення області допустимих значень, критичної області, а також значення статистичного критерію на основі відповідної таблиці.
5. Обчислення фактичного значення статистичного критерію.
6. Перевірка висунутої гіпотези на основі порівняння фактичних значень і критичного значення критерію.
7. Прийняття або відхилення гіпотези залежно від результатів перевірки.

Під час перевірки гіпотез згідно одного з критеріїв можливими є прийняття 2-х помилкових рішень:

- неправильне відхилення нульової гіпотези – помилка I роду. Імовірність появи цієї помилки зазвичай складає близько 5 %, а її саму можна контролювати, оскільки нульова гіпотеза є повністю визначеною, що передбачає існування точного значення ймовірності;
- неправильне прийняття нульової гіпотези – помилка II роду.

Коли нульова гіпотеза є вірною (I) або невірною (II), можуть бути прийняті два помилкових рішення: нульова гіпотеза відхиляється і приймається альтернативна; нульова гіпотеза приймається.

Можливі рішення цієї ситуації наведені в табл. 4.1.

Таблиця 4.1

Можливі рішення під час перевірки гіпотез

Рішення згідно критерію	Фактично	
	H_0 вірна	H_0 невірна
H_0 не відхиляється	помилка I роду	правильне рішення
H_0 відхиляється	правильне рішення	помилка II роду

Наприклад, якщо встановлено, що нова харчова добавка підвищує імунітет краще за інші, хоча насправді це не так – то допущена помилка I роду. Якщо було вирішено, що нова і стара харчові добавки є однаковими, а насправді одна є кращою (гіршою) за іншу, то була допущена помилка II роду.

Імовірності, що відповідають невірним рішенням, називаються **ризиками** (I та II). Ризик I дорівнює ймовірності помилки α (рівень значущості), ризик II дорівнює ймовірності помилки β . Оскільки значення α завжди є більшим нуля, то ймовірність виникнення помилки β є завжди присутньою. За умови заданого значення α та вибірки розміру n значення β тим більше, чим менше значення α . Якщо значення n є великим, то відповідні значення α і β можуть бути скільки завгодно малими, тобто рішення будуть більш обґрунтованими [6, с. 273].

Зазвичай на практиці значення α є заданими, а значення β намагаються зробити якомога меншими. Імовірність $1 - \beta$ називається **потужністю критерію**, чим більшим є її значення, тим меншою є ймовірність появи помилки II роду.

Область, в яку потрапляння значення статистичного критерію приводить до відхилення H_0 , називається **критичною областю**. Імовірність попадання в цю область дорівнює прийнятому рівню значущості. Доповнює критичну область ще одна – **область допустимих значень**. Якщо значення критерію потрапляє в область допустимих значень, то нульова гіпотеза не суперечить фактичним даним.

Точки, що розділяють область допустимих значень і критичну область, називають **межею критичної області** або **граничними точками**. Залежно від формулювання альтернативної гіпотези критична область може бути **двосторонньою** (рис. 4.1) або **односторонньою** (рис. 4.2) – **лівостороння** або **правостороння**. Якщо обчислене значення критерію потрапляє в критичну область, нульова гіпотеза відхиляється, оскільки вона суперечить фактичним даним.

Для перевірки істинності чи хибності гіпотез зазвичай використовують t -критерій Стьюдента, критерій Колмагорова-Смірнова та критерій хі-квадрат тощо.

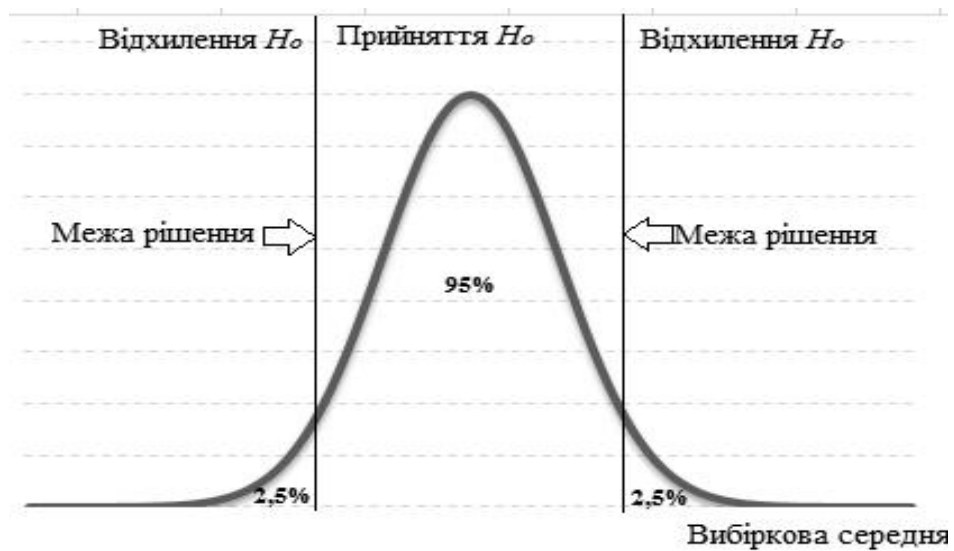


Рис. 4.1. 5% двостороння критична область

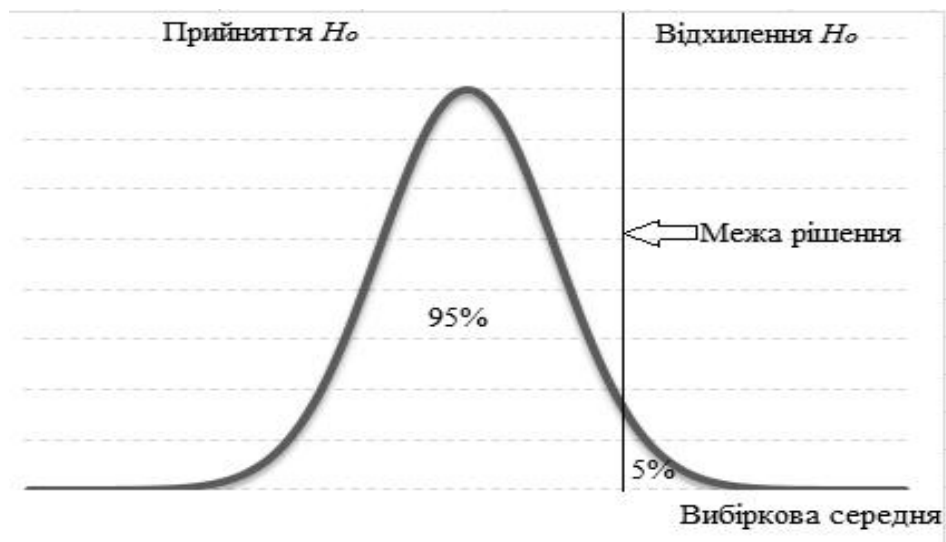


Рис. 4.2. 5% одностороння критична область

Зокрема, для визначення достовірності відмінностей між середніми двох вибірок застосовують метод Стьюдента, а для того щоб робити висновки щодо відмінності між трьома або більшою кількістю вибірок – F -тест або дисперсійний аналіз (ANOVA). Якщо дослідник має справу з даними, що отримані в неметричних (номінальних або порядкових) шкалах, або вибірки надто малі для впевненості в тому, що генеральні сукупності, з яких вони взяті, підкоряються нормальному розподілу, використовують непараметричні методи – критерій χ^2 -квадрат, критерій Манна-Уїтні, критерій Уїлкоксона тощо.

4.3. Двостороння перевірка гіпотези. Перевірка гіпотези про рівність середньої генеральної сукупності певному заданому значенню

Одним із найпростіших випадків перевірки гіпотези є перевірка рівності між певним числом (μ_0) і середнім значенням генеральної сукупності (μ). Так, нульова гіпотеза буде стверджувати, що значення є рівними між собою ($H_0: \mu = \mu_0$), а альтернативна гіпотеза буде свідчити навпаки, що значення між собою є нерівними ($H_1: \mu \neq \mu_0$).

Означена перевірка гіпотези буде двосторонньою, оскільки альтернативна гіпотеза передбачає будь-яке значення середньої в генеральній сукупності: воно може бути як більшим за обране значення μ_0 , так і меншим.

Звертаємо вашу увагу на те, що в цій теоретичній задачі будуть присутніми цілих три числа: невідоме середнє значення генеральної сукупності (μ), задане значення (μ_0), з яким порівнюють середнє по сукупності, та \bar{x} – відоме вибіркове середнє, яке буде використовуватися для прийняття рішення щодо істинності чи хибності гіпотези. Зі всіх трьох перелічених чисел лише значення \bar{x} є випадковою величиною, оскільки воно буде розраховано відповідно до даних вибірки з генеральної сукупності. Ця величина є оціночним значенням, яке представляє μ .

Таким чином, сутність перевірки гіпотези буде полягати у порівнянні значень μ_0 та \bar{x} , тобто у порівнянні значень уже відомих величин. Якщо ці значення відрізняються надто сильно, що не можна пояснити простою випадковістю, то нульову гіпотезу відхиляють. Якщо ж відмінність між значеннями є не надто великою, то нульову гіпотезу приймають. Критерієм «незначущості» відмінностей між значеннями виступає стандартна похибка $S_{\bar{x}}$, яка вимірює ступінь випадковості \bar{x} . Таким чином, коли значення \bar{x} та μ_0 знаходяться один від одного на відстані достатньої кількості $S_{\bar{x}}$, то це є достатнім доказом того, що μ та μ_0 є нерівними. Розглянемо два методи перевірки істинності гіпотези й отримання результату: використання довірчого інтервалу та використання методу t -статистики (критерію Стьюдента).

Довірчим інтервалом називають такий обчислений на існуючих даних інтервал, який з відомою ймовірністю може містити необхідний досліднику невідомий параметр генеральної сукупності.

Ця ймовірність визначається з урахуванням випадкового експерименту, який починається з формування випадкової вибірки. Також існує можливість вибирати ймовірність твердження, яку називають **довірчим рівнем**, **коефіцієнтом довіри** або **довірчою ймовірністю**. Традиційно цей показник установлюють рівним 95 %, але часто використовують значення 90, 99 і навіть 99,9 %. Межі довірчого інтервалу мають наступний вигляд:

$$\bar{x} - t_{табл} S_x \text{ та } \bar{x} + t_{табл} S_x, \quad (4.1)$$

де значення t беруть із таблиць про двосторонній довірчий рівень, наприклад, 95%, який ми будемо використовувати у подальших розрахунках. Таблиця значень t наведена в *Додатку Б*.

Таким чином, для того, щоб прийняти чи спростувати гіпотези $H_0: \mu = \mu_0$ та $H_1: \mu \neq \mu_0$ необхідно побудувати 95 % довірчий інтервал за допомогою наведених формул і виходячи зі значень \bar{x} і S_x . Після цього потрібно впевнитись, що значення μ_0 знаходиться в межах довірчого інтервалу. Якщо це дійсно так, то μ_0 може бути розглянуто як допустиме значення середнього значення генеральної сукупності, отже необхідно прийняти нульову гіпотезу як істинну. У протилежному випадку приймають альтернативну гіпотезу. Сутність даного підходу проілюстрована на рис. 4.3.



Рис. 4.3. Схематичне зображення перевірки гіпотези за допомогою довірчого інтервалу

Характеристика можливих результатів дослідження за допомогою довірчого інтервалу наведена в табл. 4.2.

Таблиця 4.2

**Методика прийняття рішення відносно перевірки гіпотези
стосовно середньої в генеральній сукупності**

Значення μ_0 знаходиться в інтервалі від $\bar{x} - t_{табл} S_x$ до $\bar{x} + t_{табл} S_x$	Значення μ_0 не знаходиться в інтервалі від $\bar{x} - t_{табл} S_x$ до $\bar{x} + t_{табл} S_x$
Нульова гіпотеза H_0 приймається як допустима можливість	Альтернативна гіпотеза H_1 приймається як допустима можливість
Альтернативна гіпотеза H_1 спростовується	Нульова гіпотеза H_0 спростовується
Відмінності між значеннями \bar{x} та μ_0 є випадковими	Відмінності між значеннями \bar{x} та μ_0 не можуть бути обумовлені лише випадковістю
Результат перевірки не є статистично значущим	Результат перевірки є статистично значущим

Розглянемо використання довірчого інтервалу на прикладі, який ми згадували на початку розділу. Нехай існує підприємство, яке протягом тижня тестувало нове обладнання на своєму виробництві. Теоретично нове обладнання працює швидше, що означає збільшення обсягів виробництва деталей типу *A* протягом дня. Вихідні дані прикладу наведені в табл. 4.3.

Таблиця 4.3

Вихідні дані прикладу

Показник	Умовне позначення	Значення	Одиниця вимірювання
Середній денний обсяг виготовленої продукції протягом минулого тижня (з використанням нового обладнання)	\bar{x}	135	штуки
Стандартна похибка	S_x	15	штуки
Розмір вибірки	n	7	дні
Середній обсяг виготовленої продукції протягом довгострокового періоду (без використання нового обладнання)	μ_0	121	штуки

Наведені дані містять у собі 7 днів спостереження за обсягами виробництва деталі *A* за допомогою нового обладнання. Генеральна сукупність складається з усіх можливих денних обсягів виробництва

деталі з використанням нової техніки; зокрема середнє генеральної сукупності μ є середнім значенням обсягу виробництва, отримане за тривалий період використання нового обладнання (в таблиці це значення відсутнє, оскільки є невідомим). Вибіркове значення \bar{x} є найкращою оцінкою μ .

Інформація, яка наведена в табл. 4.3, свідчить, що використання нового обладнання можливо є ефективним. Середньоденний обсяг виробництва деталі A , досягнутий з використанням нової техніки ($\bar{x} = 135$ шт.), є на 16 одиниць більшим, ніж середній очікуваний обсяг виробництва, розрахований за минулий довгостроковий період ($\mu_0 = 121$). Постає питання, чи є ця відмінність випадковою, чи вона дійсно викликана використанням нового обладнання.

Висунемо гіпотези. Таким чином, нульова гіпотеза буде стверджувати, що під час використання нового обладнання невідоме середнє значення обсягу продукції за тривалий період часу μ **дорівнює** заданому значенню $\mu_0 = 121$ одиниць деталі A , що аналогічно обсягу виробництва продукції без використання нової техніки ($H_0: \mu_0 = 121$).

Альтернативна гіпотеза, навпаки, стверджує, що під час використання нового обладнання, невідоме середнє значення обсягу продукції за тривалий період часу μ **не дорівнює** заданому значенню $\mu_0 = 121$ одиниць деталі A – обсягу виробництва без використання нової техніки ($H_1: \mu_0 \neq 121$).

Наступним кроком нашого дослідження є розрахунок довірчого інтервалу за формулою 4.1 та значення t із *Додатку Б*. З кількістю ступенів свободи 6 ($n-1=7-1=6$) табличне значення $t = 2,447$. Отже, з імовірністю 95 % можна стверджувати, що під час використання нового обладнання протягом довгострокового періоду середнє значення обсягу виробництва деталі A буде знаходитись в інтервалі від 98 до 172 одиниць.

Для того, щоб перевірити правдивість гіпотези, необхідно визначити, чи входить значення $\mu_0 = 121$ до довірчого інтервалу. Як ми бачимо – входить, адже 121 знаходиться між значеннями 98 та 172. Це свідчить, що ми можемо прийняти нульову гіпотезу $H_0: \mu_0 = 121$ як допустиму можливість і спростувати альтернативну гіпотезу $H_1: \mu_0 \neq 121$. Відмінність між середнім вибірковим значенням обсягу виробництва продукції $\bar{x} = 135$ і заданим значенням $\mu_0 = 121$ може бути обумовлена лише випадковістю. Результат перевірки не є статистично значущим.

Варто відмітити, що перевірка гіпотези не довела ефективність використання нового обладнання для виготовлення деталі А. Воно може бути як ефективним, так і неефективним – на даний час не має доказів ні для того, ні для іншого. Варіантом вирішення такого завдання на практиці може бути збільшення вибірки, тобто використання нового обладнання ще протягом деякого періоду часу, наприклад ще два тижні або місяць (за умови вигідних пропозицій з боку постачальника цієї техніки).

Наведена формула 4.1, яку ми використовували для вирішення цього завдання, є дійсною для нормального розподілу випадкової величини. Для інших варіантів розподілу, разом з іншими формулами, вона може мати інший вигляд. Наведемо формули і позначення для нормального та біноміального законів розподілу в табл. 4.4.

Таблиця 4.4

Формули розрахунку показників та їх умовні позначення

Показник	Нормальний розподіл	Біноміальний розподіл
Середнє значення в сукупності	μ	Ω
Задане значення	μ_0	Ω_0
Нульова гіпотеза	$H_0: \mu = \mu_0$	$H_0: \Omega = \Omega_0$
Альтернативна гіпотеза	$H_1: \mu \neq \mu_0$	$H_1: \Omega \neq \Omega_0$
Дані	x_1, x_2, \dots, x_n	x настання подій в n випробуваннях
Оцінка	\bar{x}	$p = \frac{x}{n}$, де n – обсяг вибіркової сукупності
Стандартна похибка	$S_x = S\sqrt{n}$	$S_p = \sqrt{\frac{p(1-p)}{n}}$
Довірчий інтервал	від $\bar{x} - t_{табл} S_x$ до $\bar{x} + t_{табл} S_x$	від $\bar{x} - t_{табл} S_p$ до $\bar{x} + t_{табл} S_p$
t-статистика	$t_{статистика} = \frac{\bar{x} - \mu_0}{S_x}$	$t_{статистика} = \frac{p - \Omega_0}{S_p}$

Перейдемо до другого методу перевірки гіпотез із тих, що ми згадували – методу t-статистики. Сутність цього методу полягає в тому, щоб розрахувати значення t за формулою $t_{статистика} = \frac{\bar{x} - \mu_0}{S_x}$, щоб після цього з використанням t-таблиці (Додаток Б) вирішити, яку

саме гіпотезу необхідно прийняти. За умови правильності розрахунків, результат повинен бути той самий, що й при використанні методу довірчого інтервалу. Тому немає жодного значення, який із методів використовувати.

Відповідно до другого методу, перевірку статистичної гіпотези починають з того, що для визначення відмінності між двома гіпотезами на основі найкращих даних розраховують так звану тест-статистику. Отриману величину тест-статистики (як то, t -статистику) далі порівнюють з відповідним критичним значенням із таблиці критичних значень (t -таблиці); на основі цього порівняння й визнають, яку гіпотезу необхідно прийняти. Тест-статистика показує, скільки стандартних помилок знаходиться між значеннями \bar{x} та μ_0 .

Для нормального закону розподілу t -статистика розраховується за наступною формулою:

$$t_{\text{статистика}} = \frac{\bar{x} - \mu_0}{S_{\bar{x}}} \quad (4.2)$$

Для біноміального закону вона прийме наступний вид:

$$t_{\text{статистика}} = \frac{p - \Omega_0}{S_p} \quad (4.3)$$

Залежно від того, чи перевищує розрахована t -статистика табличне значення, можливі наступні варіанти розвитку подій (табл. 4.5).

Таблиця 4.5

Використання t -статистики для перевірки гіпотез

t -статистика за абсолютною величиною (за модулем) є меншою, ніж t -значення із t -таблиці	t -статистика за абсолютною величиною (за модулем) є більшою, ніж t -значення із t -таблиці
Нульова гіпотеза H_0 приймається як допустима можливість	Альтернативна гіпотеза H_1 приймається як допустима можливість
Альтернативна гіпотеза H_1 спростовується	Нульова гіпотеза H_0 спростовується
Відмінність між значеннями \bar{x} та μ_0 є випадковою	Відмінність між значеннями \bar{x} та μ_0 не може бути обумовлена лише випадковістю
Результат перевірки не є статистично значущим	Результат перевірки є статистично значущим

Звертаємо вашу увагу на те, що таблиці бракує ще одного варіанту порівняння – коли фактичне значення t -статистики *співпадає*

з t-значенням із таблиці. У цьому випадку μ_0 точно співпадає з межею довірчого інтервалу. Така ситуація буває вкрай рідко, тим не менше, все ж трапляється. Тоді необхідно провести розрахунок з більшим ступенем точності та залишити більше знаків після коми. Або, другий варіант, можна прийняти нульову гіпотезу і зробити зауваження, що значення є граничним.

Корисно також запам'ятати наступне правило: якщо значення t-статистики за абсолютною величиною є меншим 2, то нульову гіпотезу спростовують, і навпаки. Однак, це правило можна використовувати тільки в тих випадках, коли вибіркова сукупність є більшою, ніж 40 одиниць, тоді використовують число 2 як апроксимацію t-значення 1,96.

Розрахуємо значення t-статистики для попереднього прикладу з використанням нового обладнання на підприємстві для виробництва деталі А. Згідно з умовою, нам було відомо, що $n = 7$, $\bar{x} = 135$, $S_x = 15$, а задане значення $\mu_0 = 121$. Тоді, відповідно до формули 4.2, отримуємо:

$$t_{\text{статистика}} = \frac{135 - 121}{15} = 0,933.$$

Оскільки отримане значення t-статистики за модулем є меншим, ніж табличне значення 2,447 (Додаток Б), то ми можемо зробити висновок, що необхідно прийняти нульову гіпотезу і спростувати альтернативну. Отже, за допомогою методу t-статистики ми отримали той самий результат, що й під час використання методу довірчого інтервалу.

4.4. Одностороння перевірка гіпотези

Варто повторити, що перевірка гіпотези, про яку йшла мова до цього моменту, була двосторонньою, оскільки розглядалася нульова гіпотеза $H_0: \mu = \mu_0$ проти альтернативної $H_1: \mu \neq \mu_0$. Ця альтернативна гіпотеза є двосторонньою, адже середнє в сукупності може виявитися як більшим, так і меншим, ніж задане опорне значення.

Однак, в аналізі економічних процесів часто трапляються більш конкретні ситуації, коли дослідника цікавить точніший результат, наприклад, чи є середнє в сукупності більшим заданого значення (чи меншим, залежно від ситуації). Скажімо, підприємство закупить нове обладнання тільки в тому випадку, якщо відсоток бракованих виробів

виявиться меншим деякого конкретного числа, або покупці придбають нову харчову добавку тільки в тому випадку, коли кількість корисних речовин в ній виявиться більшою, ніж певне значення.

У той же час відмітимо, що використання односторонньої перевірки не є обов'язковим. Для подібних випадків можна скористатися і вже відомою двосторонньою перевіркою. Якщо результат двосторонньої перевірки свідчить, що нульова гіпотеза спростовується, а приймається альтернативна, то з урахуванням того, меншим чи більшим є середнє значення вибірки \bar{x} , ніж задане опорне значення, можна зробити наступні висновки:

- в ситуаціях, коли $\bar{x} > \mu_0$, середнє значення вибірки \bar{x} є значно більшим заданого значення μ_0 ;
- в ситуаціях, коли $\bar{x} < \mu_0$, середнє значення вибірки \bar{x} є значно меншим заданого значення μ_0 .

Однак, бувають ситуації, коли необхідно використовувати саме односторонню перевірку, наприклад, коли її результати свідчать про те, що необхідно прийняти нульову гіпотезу, а результати двосторонньої – альтернативну. Одностороння перевірка приділяє увагу тільки одній стороні, тому може краще знайти відмінність між середнім і певним заданим значенням. У той же час, існує недолік, адже ця перевірка нічого не знає про іншу сторону значень від заданого числа.

Під час процедури односторонньої перевірки гіпотези нульова гіпотеза стверджує, що значення μ знаходиться по одну сторону від μ_0 , а альтернативна гіпотеза – по іншу. Випадок, коли $\mu = \mu_0$, завжди включають до нульової гіпотези, яку необхідно спростувати, оскільки це гарантує, що прийняття альтернативної гіпотези дозволяє зробити більш точний висновок «набагато більше, ніж ...» або «набагато менше, ніж...».

Одностороння гіпотеза про те, що μ є меншим, ніж μ_0 , формулюються наступним чином:

$$H_0: \mu \geq \mu_0, \quad (4.4)$$

$$H_1: \mu < \mu_0 \quad (4.5)$$

Отже, нульова гіпотеза стверджує, що невідоме середнє значення генеральної сукупності є щонайменше таким же великим як і значення μ_0 . Альтернативна гіпотеза говорить, що невідоме середнє значення генеральної сукупності є меншим за значення μ_0 .

Одностороння гіпотеза про те, що μ є більшим, ніж μ_0 , формулюється так:

$$H_0: \mu \leq \mu_0, \quad (4.6)$$

$$H_1: \mu > \mu_0 \quad (4.7)$$

Таким чином, нульова гіпотеза стверджує, що невідоме середнє значення генеральної сукупності є не більшим, ніж значення μ_0 . Альтернативна гіпотеза говорить, що невідоме середнє значення генеральної сукупності є більшим за задане опорне значення μ_0 .

Саме використання односторонньої перевірки йде з деяким обмеженням: дослідник повинен бути впевнений, що незалежно від того, як поведуть себе дані, ця перевірка буде продовжуватись використовуватися лише на цьому самому боці. Якщо внаслідок зміни характеру даних виникає необхідність або сумнів щодо можливості використання перевірки іншого боку, то необхідно залишити односторонню перевірку і перейти до двосторонньої.

Також відмітимо, що одностороння перевірка залишає деякі сумніви, адже містить суб'єктивну думку дослідника про те, чи є значення тільки більшим, чи тільки меншим за певне число. У ситуаціях, коли необхідно переконати людей із протилежною точкою зору, краще використовувати двосторонню перевірку. У той же час, якщо дослідник представляє результати своїх досліджень колу одностороннім, які підтримують його точку зору, можна використовувати односторонню перевірку.

Як і для двосторонньої перевірки, для односторонньої існують два способи перевірки гіпотези: метод довірчого інтервалу та метод використання t-статистики. Розглянемо це за допомогою прикладу.

Відділ управління персоналом промислового підприємства збирається організувати виїзний тренінг співробітників з метою зміцнення командного духу та розробки інноваційних ідей. Виходячи з проведеного теоретичного аналізу, тренінг буде успішним і принесе плідні результати тільки в тому випадку, коли більше 30 % співробітників вирішать його відвідати. Ці 30,0 % є значенням μ_0 , яке було отримано без урахування вибірки. Для того, щоб вирішити питання щодо необхідності організації такої поїздки, HR-відділ провів опитування серед випадкової кількості співробітників підприємства ($n = 43$) і розрахував односторонній довірчий інтервал. Виходячи з проведеного опитування 45 % співробітників виказали бажання прийняти участь у тренінгу ($\bar{x} = 45$). Керівництву відділу на 95 % впевнено, що щонайменше 37,8% працівників поїдуть на

тренінг. Оскільки задане значення $\mu_0 = 30,0\%$, що характеризує ефективність тренінгу, знаходиться поза довірчого інтервалу (отже, неприйнятно вважати, що середнє значення дорівнює $30,0\%$), існує переконливий доказ того, що середнє в генеральній сукупності є більшим за $30,0\%$.

Ця ситуація коротко наведена в табл. 4.6.

Таблиця 4.6

Тест для відсотка працівників підприємства, що бажають прийняти участь у тренінгу (з використанням довірчого інтервалу)

Показник	Загальний вираз	Конкретні значення
Нульова гіпотеза	$H_0: \mu \leq \mu_0$	$H_0: \mu \leq 30\%$
Альтернативна гіпотеза	$H_1: \mu > \mu_0$	$H_1: \mu > 30\%$
Середня	\bar{x}	45%
Стандартна похибка (розрахована за формулою з таблиці 4.4)	S_x	3,324 %
Розмір вибірки	n	43
Задане значення	μ_0	30 %
Довірчий інтервал	$\bar{x} - t_{табл} S_x$ до $\bar{x} + t_{табл} S_x$	Керівництво відділу на 95 % впевнено, що середнє значення генеральної сукупності принаймні дорівнює 37,8 %
Рішення	<i>Прийняти альтернативну гіпотезу H_1</i>	Відділ сподівається, що більше ніж 30 % співробітників вирішать відвідати тренінг

Таким чином, рішення задачі полягає в тому, щоб прийняти альтернативну гіпотезу H_1 , оскільки задане значення знаходиться за межами довірчого інтервалу (тобто $30,0\%$ не означає щонайменше $37,8\%$).

Наведемо загальний принцип односторонньої перевірки гіпотези для обох її типів (перевірок того, що \bar{x} є значимо більшим чи значимо меншим за μ_0) з використанням як методу довірчого інтервалу, так і t -статистики (табл. 4.7).

Схема вирішення поставленого завдання

Одностороння перевірка того, що μ є більшим за μ_0	Одностороння перевірка того, що μ є меншим за μ_0
Перевіряємо нульову гіпотезу $H_0: \mu \leq \mu_0$ проти альтернативної $H_1: \mu > \mu_0$	Перевіряємо нульову гіпотезу $H_0: \mu \geq \mu_0$ проти альтернативної $H_1: \mu < \mu_0$
Твердження про довірчий інтервал: керівництво відділу на 95% впевнено, що середнє генеральної сукупності принаймні не менше за $\bar{x} - t_{табл} S_p$	Твердження про довірчий інтервал: керівництво відділу на 95% впевнено, що середнє генеральної сукупності не більше за $\bar{x} + t_{табл} S_x$
t-статистика визначається за формулою 4.3. Варто пам'ятати, що в односторонній перевірці абсолютне (модальне) значення не використовується	t-статистика визначається за формулою 4.3. Варто пам'ятати, що в односторонній перевірці абсолютне (модальне) значення не використовується
Чи дійсно $\bar{x} - t_{табл} S_p \leq \mu_0$? У рамках методу довірчого інтервалу формулюється питання: чи знаходиться значення μ_0 всередині нього. У рамках методу t-статистики формулюємо питання, чи є розраховане значення t меншим, чи дорівнює t-табличному	Чи дійсно $\bar{x} + t_{табл} S_x \geq \mu_0$? У рамках методу довірчого інтервалу формулюється питання: чи знаходиться значення μ_0 всередині нього. У рамках методу t-статистики формулюємо питання, чи є розраховане значення t більшим, чи дорівнює t-табличному, взятому з мінусом
Якщо відповідь на це питання позитивна, то:	Якщо відповідь на це питання позитивна, то:
Нульова гіпотеза H_0 приймається як допустима можливість	Нульова гіпотеза H_0 приймається як допустима можливість
Альтернативна гіпотеза H_1 спростовується	Альтернативна гіпотеза H_1 спростовується
Середнє значення \bar{x} не є значимо більшим за μ_0	Середнє значення \bar{x} не набагато менше за μ_0
Якщо \bar{x} є більшим за μ_0 , то це можна пояснити випадковістю	Якщо \bar{x} є меншим за μ_0 , то це можна пояснити випадковістю
Результат перевірки не є статистично значущим	Результат перевірки не є статистично значущим
Чи дійсно $\bar{x} - t_{табл} S_x > \mu_0$? У рамках методу довірчого інтервалу формулюється питання: чи знаходиться значення μ_0 поза нього. У рамках методу t-статистики формулюємо питання, чи є розраховане значення t більше t-табличного.	Чи дійсно $\bar{x} + t_{табл} S_x < \mu_0$? У рамках методу довірчого інтервалу формулюється питання: чи знаходиться значення μ_0 поза нього. У рамках методу t-статистики формулюємо питання, чи є розраховане значення t меншим t-табличного, взятого з мінусом.

Продовження таблиці 4.7

Одностороння перевірка того, що μ є більшим за μ_0	Одностороння перевірка того, що μ є меншим за μ_0
Якщо відповідь на це питання позитивна, то:	Якщо відповідь на це питання позитивна, то:
Альтернативна гіпотеза H_1 приймається як допустима можливість	Альтернативна гіпотеза H_1 приймається як допустима можливість
Нульова гіпотеза H_0 спростовується	Нульова гіпотеза H_0 спростовується
Середнє значення \bar{x} є значно більшим за μ_0	Середнє значення \bar{x} є значно меншим за μ_0
Відмінність між значеннями неможливо пояснити простою випадковістю	Відмінність між значеннями неможливо пояснити простою випадковістю
Результат перевірки є статистично значущим	Результат перевірки є статистично значущим

Отже, під час використання методу довірчого інтервалу варто пам'ятати, що існують два різних односторонніх довірчих інтервали. Необхідно обрати той, який відноситься до сторони, стосовно якої сформульована альтернативна гіпотеза. Наприклад, якщо досліджувана альтернативна гіпотеза має вигляд $H_1: \mu > \mu_0$, то необхідний односторонній довірчий інтервал буде складатися з усіх значень μ , які принаймні не є меншими за відповідне число, розраховане за формулою $\bar{x} - t_{\text{табл}} S_{\bar{x}}$ з використанням значення t , взятого із таблиці.

На рис. 4.4 показано, що для прийняття рішення про те, що \bar{x} значно більше за μ_0 , відстань між ними повинна бути достатньо великою, для того, щоб цю відмінність неможливо було пояснити лише випадковістю [16, с.472]. Односторонній довірчий інтервал використовує той самий напрям, що й альтернативна гіпотеза, тобто ті значення μ , які не є меншими за граничне значення довірчого інтервалу. Прийняти рішення про те, що середнє вибірки є значимо більшим можна тільки в тому випадку, коли задане значення μ_0 знаходиться набагато нижче від середнього вибірки.

На рис. 4.5. наведено схематичне зображення односторонньої перевірки у протилежному напрямку. Односторонній довірчий інтервал використовує той самий напрям, що й альтернативна гіпотеза, тобто ті значення μ , які є меншими або дорівнюють



Рис. 4.4. Схема використання односторонньої перевірки для прийняття рішення про те, чи дійсно значення μ є більшим за задане значення μ_0 .

граничному значенню довірчого інтервалу. Прийняти рішення про те, що середнє значення вибірки є значимо меншим, можна тільки в тому випадку, коли задане значення μ_0 знаходиться набагато вище від середнього з вибірки.



Рис. 4.5. Схема використання односторонньої перевірки для прийняття рішення про те, чи дійсно значення μ є меншим за задане значення μ_0 .

Перевірка односторонньої гіпотези за допомогою t -статистики, в свою чергу, буде проводитись наступним чином: дослідник повинен порівняти розраховану t -статистику із табличним значенням t або з табличним значенням t взятим із мінусом (залежно від того, з якого боку робиться перевірка). Більше того, табличне значення t -статистики є однаковим, як для односторонньої, так і для двосторонньої перевірки; відмінність лише в тому, яким чином її використовують.

Нами був розглянутий приклад відвідування співробітниками підприємства виїзного тренінгу, де стверджувалось, що ця подія буде ефективною тільки в тому випадку, коли її відвідають не менше 30 % працівників. Відповідні дані наведені в таблиці 4.6. Виконаємо таку саму перевірку прийняття гіпотези, тільки за допомогою методу t -статистики (табл. 4.8).

Таблиця 4.8

Тест для відсотка працівників підприємства, що бажають прийняти участь у тренінгу (з використанням методу t -статистики)

Показник	Загальний вираз	Конкретні значення
Нульова гіпотеза	$H_0: \mu \leq \mu_0$	$H_0: \mu \leq 30 \%$
Альтернативна гіпотеза	$H_1: \mu > \mu_0$	$H_1: \mu > 30 \%$
Середнє	\bar{x}	45 %
Стандартна похибка (розрахована за формулою з таблиці 4.4)	S_x	3,324 %
Задане значення	μ_0	30 %
t -статистика	$t_{статистика} = \frac{\bar{x} - \mu_0}{S_x}$	$t_{статистика} = \frac{45,0 - 30}{3,324} = 4,513$
Критичне значення	$t_{табл}$	1,625
Рішення	Прийняти альтернативну гіпотезу H_1	Відділ сподівається, що більше ніж 30 % співробітників вирішать відвідати тренінг

Таким чином, виходячи із даних табл. 4.8, ми приймаємо альтернативну гіпотезу $H_1: \mu > \mu_0$ та спростовуємо нульову гіпотезу $H_0: \mu \leq \mu_0$, оскільки $t_{статистика}$ є більшою за $t_{табл}$ (тобто 4,513 є більшим за 1,625). Отже, як і в прикладі із двосторонньою перевіркою, обидва методи дали один і той самий результат.

Підводячи висновки вищевикладеного, відмітимо, що перевірку статистичних гіпотез використовують для обґрунтування рішення при наявності альтернативного варіанту розвитку подій. Досить часто процедуру перевірки використовують для виявлення закономірностей і спростування випадковостей.

Питання для самоконтролю

1. Надайте визначення статистичній гіпотезі.
2. Яку статистичну гіпотезу називають простою, а яку – складною? Наведіть приклади.
3. Чим відрізняються одна від одної параметричні та непараметричні гіпотези?
4. Дайте визначення нульовій гіпотезі.
5. Що таке альтернативна гіпотеза?
6. Наведіть ситуації з прикладами нульової та альтернативної гіпотези.
7. У чому сутність статистичного критерію? Які статистичні критерії ви знаєте?
8. Перелічіть етапи перевірки статистичної гіпотези за допомогою статистичного критерію.
9. У чому сутність помилки I роду під час перевірки гіпотези?
10. Охарактеризуйте помилки II роду під час перевірки гіпотези.
11. Надайте поняття критичної області та області допустимих значень.
12. Які існують види критичної області залежно від формулювання альтернативної гіпотези?
13. Які основні методи використовують для перевірки гіпотези?
14. Що таке довірчий інтервал?
15. Коротко опишіть етапи перевірки двосторонньої гіпотези за допомогою довірчого інтервалу.
16. Коротко опишіть етапи перевірки двосторонньої гіпотези за допомогою методу t-статистики.
17. Що таке задане значення? Звідки воно береться? Чи можна взяти його із таблиці?
18. Дайте характеристику односторонній перевірці.
19. В яких ситуаціях краще використовувати односторонню перевірку?
20. Чи можна використовувати двосторонню перевірку гіпотези замість односторонньої?
21. Сформулюйте гіпотези для односторонньої перевірки.
22. Як виконують односторонню перевірку гіпотези на основі довірчого інтервалу?
23. Як виконують односторонню перевірку гіпотези на основі t-статистики?

РОЗДІЛ 5. МОДЕЛЮВАННЯ ВЗАЄМОЗВ'ЯЗКІВ МІЖ КІЛЬКІСНИМИ ЗМІННИМИ

5.1. Види взаємозв'язків між кількісними змінними

Світ, що нас оточує, сповнений всіляких взаємозв'язків: між політикою держави та станом економіки, обсягом випуску продукції та попитом на неї на ринку, доходами та витратами підприємства тощо.

У тому випадку, коли відомі двовимірні або багатовимірні дані (наприклад, продуктивність праці робітників, їх оснащеність основними засобами, заробітна плата та ін.), завжди є можливість вивчати не лише кожний вимір окремо (як частина одновимірної сукупності даних), а й здійснювати їх спільний аналіз, у тому числі шляхом виявлення взаємозв'язків між ними, що дозволяє отримати значно більший ефект від аналізу.

Вивчаючи взаємозв'язки між дво- і багатовимірними даними, слід завжди пам'ятати про наступні три *основні цілі* [16, с. 518-519].

Перша. Описання і розуміння взаємозв'язку. При вивченні складної системи дуже важливо знати, які чинники найбільш тісно взаємодіють один з одним, а які взагалі не впливають один на одного. Знання цієї інформації може надати значну допомогу в довгостроковому плануванні і прийнятті інших стратегічних рішень.

Друга. Прогнозування і передбачення. Розуміння певного взаємозв'язку може дозволити використовувати інформацію про один із вимірників для більш якісного передбачення іншого.

Третя. Регулювання й управління процесом. Коли ми втручаємося в будь-який процес, необхідно визначити обсяг цього втручання. Якщо існує безпосередній взаємозв'язок між втручанням і результатом і ми цей взаємозв'язок розуміємо, то таке знання може допомогти нам здійснити оптимальне регулювання.

Між досліджуваними змінними (економічними показниками) можуть бути різні види і форми зв'язків.

За характером залежності явищ розрізняють зв'язки функціональні та стохастичні.

Функціональні зв'язки характеризуються повною відповідністю між причиною (факторна ознака) і наслідком (результативна ознака). Внаслідок цього функціональна залежність виражається точною математичною формулою, яка може бути

застосована до будь-яких явищ, що розглядаються. При функціональному зв'язку кожному значенню однієї ознаки (або декількох) відповідає цілком певне значення іншої, результативної ознаки. Функціональний зв'язок виражається у вигляді того або іншого рівняння функції і зберігає свою силу (виявляється) у кожному окремому випадку для кожної окремо взятої одиниці даної сукупності.

Проте, в економічних дослідженнях частіше зустрічаються зв'язки іншого роду, коли одній і тій же величині ознаки-чинника відповідають різні значення результативного показника, створюючи ряд розподілу, тобто спостерігається залежність розподілу значень результативної ознаки від значень ознаки-чинника. Такого роду зв'язки називаються **стохастичними**. При стохастичному зв'язку зі зміною факторної ознаки змінюється розподіл одиниць за результативною ознакою.

Окремим видом стохастичних зв'язків є **кореляційні** зв'язки. Термін кореляція (correlation) означає співвідношення, відповідність. Кореляційний зв'язок між результативною ознакою та однією з певної кількості факторних ознак може проявитися тільки в загальному, в середньому, за інших однакових умов, коли вплив факторів, що не є об'єктом дослідження, усувається шляхом їх осереднення. Відповідно до закону великих чисел таке осереднення на основі взаємопогашення відхилень певних одиниць сукупності у той або інший бік від середньої досягається при достатньо великій їх кількості. І чим більшою є статистична сукупність, тим точніше встановлене співвідношення виражає закономірність кореляційних зв'язків.

Як функціональні, так і кореляційні зв'язки *за їх напрямком* можуть бути **прямими** та **оберненими**. Якщо із зростанням факторної ознаки результативна ознака також зростає, то це буде **прямий** зв'язок. Якщо ж із збільшенням факторної ознаки результативна зменшується або, навпаки, із зменшенням факторної ознаки результативна зростає, то це – **обернений** зв'язок.

За *аналітичним вираженням (формою)* зв'язки можуть бути **прямолінійними** (лінійними) і **криволінійними** (нелінійними). Якщо залежність результативної ознаки від чинника може бути виражена рівнянням прямої лінії, то зв'язок називається **прямолінійним** (лінійним), якщо ж залежність виражається рівнянням будь-якої кривої (гіперболи, параболи тощо), то зв'язок називається **криволінійним**.

Якщо досліджується залежність результативної ознаки тільки від одного чинника, то зв'язок називається **однофакторним**, при вивченні взаємозв'язку між декількома ознаками говорять про зв'язок **багатофакторний**.

Залежно від тісноти зв'язку між змінними виділяють такі їх види: практично відсутній, слабкий, помірний, сильний.

5.2. Детерміновані моделі та їх аналіз

Для вивчення взаємозв'язків між ознаками використовуються різноманітні моделі, які за характером залежностей між явищами можна розділити на два класи: 1) детерміновані; 2) імовірнісні.

Залежно від того, враховує економіко-математична модель елемент випадковості чи ні, вона може бути віднесена до класу стохастичних або детермінованих. У детермінованих моделях ні цільова функція, ні рівняння зв'язку не містять випадкових факторів. А це означає, що для даної множини вхідних значень на виході може бути отриманий лише один-єдиний результат. Для стохастичних економіко-математичних моделей характерна наявність факторів, що мають імовірнісний характер, і вони представлені певними законами розподілу. Значення вихідних параметрів таких моделей можуть бути передбачені лише з певною ймовірністю.

Слід відмітити певну обмеженість можливостей детермінованого факторного моделювання та факторного аналізу. Це обумовлено декількома причинами. Зокрема, повнота вивчення економічних явищ і процесів та показників, що їх відображають, залежить від правильного і повного відображення зв'язку між показниками-факторами. Крім того, вплив окремих факторів на зміну результативного показника вивчається ізольовано, прямим рахунком, сукупний вплив факторів отримують простим сумуванням, яке відображає цей ізольований вплив. При цьому не враховуються можливості існування між результативним показником і факторами, а також між самими факторами складних стохастичних залежностей, через які вплив одних факторів може спотворюватися впливом інших. У детермінованому моделюванні не враховується те, що дія багатьох факторів на результат відбувається одночасно, а характер їх зв'язку в багатьох випадках є нелінійним.

В економічній літературі представлені різні моделі, що використовуються в детермінованому економічному аналізі. Їх

використання залежить від напрямків аналізу, економічної сутності взаємопов'язаних факторних показників. У цьому випадку взаємозв'язок між факторними показниками може бути різний і відображатися за допомогою арифметичних дій – додавання, віднімання, ділення та множення.

У табл. 5.1 представлено види детермінованих моделей, що використовуються в економічному аналізі кількісних змінних.

Таблиця 5.1

Види (типи) детермінованих моделей зв'язку

Вид (тип) моделі	Математичне вираження	Умови використання
Адитивна	$Y = \Sigma X = X_1 + X_2 + \dots + X_n$	Результативний показник є алгебраїчною сумою факторних показників
Мультиплікативна	$Y = \Pi X = X_1 \cdot X_2 \cdot \dots \cdot X_n$	Результативний показник є добутком декількох факторів
Кратна	$Y = \frac{X}{Z}$	Результативний показник є часткою від ділення одного фактору на інший
Змішана (комбінована)	$Y = \frac{X + K}{Z}; \quad Y = (X + K) \cdot Z$	Результативний показник поєднує в собі декілька форм зв'язку

На практиці найбільш поширеними, серед наведених у табл. 5.1 моделей, є адитивні та мультиплікативні моделі.

Будь-яка із наведених моделей може бути деталізована одним із наступних шляхів:

- подовження (формальне розкладання);
- розширення;
- скорочення.

В адитивних факторних моделях доцільно застосовувати подовження розкладання (деталізацію) результативного показника на частини, тобто здійснювати так зване дезагрегування факторів.

Наприклад, повна собівартість реалізованої продукції підприємства може бути представлена у вигляді такої адитивної моделі:

$$Y = X_1 + X_2 + X_3, \tag{5.1}$$

де X_1 – виробнича собівартість продукції;

X_2 – адміністративні витрати;

X_3 – витрати на збут.

Ця трьохфакторна модель (5.1) може бути дезагредована у п'ятифакторну:

$$Y = X_1 + X_2 + X_3 + X_4 + X_5, \quad (5.2)$$

де X_1 – матеріальні витрати (витрати сировини, матеріалів, електроенергії, комплектуючих тощо);

X_2 – витрати на оплату праці;

X_3 – амортизація ОС;

X_4 – адміністративні витрати;

X_5 – витрати на збут.

За потреби можна ще більше розчленувати витрати підприємства за їх видами.

Очевидно, що із деталізацією аналітичні та інформаційні можливості дослідження значно збільшуються, оскільки з'являється можливість визначити вплив значно більшої кількості факторів на результат. При цьому рівень деталізації результативного показника залежить від мети та завдань аналізу.

У мультиплікативних факторних моделях використовується прийом їх розширення за рахунок розділення результативного показника (добутку) на окремі складові-співмножники. Так, виробнича собівартість продукції може бути представлена як добуток двох факторів-співмножників:

$$Y = X_1 \cdot X_2, \quad (5.3)$$

де X_1 – обсяг виготовленої продукції;

X_2 – собівартість одиниці продукції.

Дана мультиплікативна модель (5.3) може бути перетворена шляхом дезагрегування у трьохфакторну:

$$Y = X_1 \cdot X_2 \cdot X_3, \quad (5.4)$$

де X_1 – середньооблікова кількість робітників;

X_2 – продуктивність праці одного робітника (виробіток);

X_3 – собівартість одиниці продукції.

Отримана трьохфакторна модель може бути дезагредована у чотирьох-, п'ятифакторну та ін.

Розподіл абсолютного приросту результативного показника в адитивних моделях є достатньо простим: він є сумою приростів кожного фактора, що був включений у модель.

У мультиплікативних моделях розподіл абсолютного приросту результативного показника може здійснюватися одним із двох методів: ланцюгових підстановок і виявлення ізольованого впливу факторів. Згідно правил методу ланцюгових підстановок зміна

кожного фактора в моделі здійснюється за принципом ланцюга: спочатку змінюється перший фактор, а всі інші фіксуються на рівні базисного періоду, далі – другий фактор (при цьому перший уже фіксується на рівні звітного періоду, а всі інші – на рівні базисного і т. д.). Очевидно, що в такому випадку важливе значення має правильне розташування факторів у моделі. Першим у мультиплікативній моделі розташовується об'ємний показник, далі – структурний, а потім – якісні показники. При цьому перший показник-фактор у моделі є первинним, другий – вторинним до першого, але первинним до третього і т. д. Важливо враховувати, що кожен фактор у мультиплікативній моделі повинен мати реальний економічний зміст і добуток кожних факторів послідовно, починаючи з першого фактора, також повинен бути реальним економічним показником.

Основною перевагою методу ланцюгових підстановок є простота розрахунків і можливість надати чітку інтерпретацію їх результатів. У той же час, даний метод має суттєвий недолік – неможливо виділити чистий вплив кожного фактора на результативний показник.

Вирішити це питання дозволяє метод виявлення ізольованого впливу факторів. Згідно правил даного методу кожен фактор змінюється ізольовано, тобто, усі інші фактори фіксуються на рівні базисного періоду, незалежно від їх розташування в моделі. Однак, слід враховувати, що у випадку використання даного методу при визначенні приросту результативного показника виникає додатковий приріст, який характеризує спільний вплив факторів. Саме цей приріст у деяких випадках досить складно пояснити, наприклад тоді, коли прирости результативного показника за рахунок кожного фактора були позитивними (мали знак «плюс»), а даний показник мав знак «мінус». Тому ця обставина, в першу чергу, привела до того, що більш поширеним на практиці став метод ланцюгових підстановок.

Перехід від абсолютних приростів результативного показника до відносних здійснюється шляхом ділення кожного абсолютного приросту (і загального, і факторних) на базисний рівень результативного показника (незалежно від того, який метод розподілу приросту за факторами використовується).

5.3. Регресійно-кореляційний аналіз

Детерміновані моделі, як було відмічено в параграфі 5.2, мають певні обмеження щодо відображення зв'язків, які існують між показниками у реальному житті. Більш поширеними з цієї точки зору є ймовірнісні моделі.

В основі ймовірнісних моделей лежить передумова про стохастичний (кореляційний або регресійний) зв'язок між залежною змінною та чинниками, що її обумовлюють. Іншими словами, поряд з основними вирішальними факторами розглядаються також другорядні та випадкові чинники, що визначають варіацію результативної ознаки. Це досить адекватна постановка завдання, оскільки в реальній дійсності існує саме такий механізм формування рівня більшості економічних показників підприємства [24, с. 83].

Одним із найбільш поширених методів, які застосовуються для аналізу стохастичних зв'язків, є регресійно-кореляційний метод аналізу.

Регресійно-кореляційний аналіз (РКА) полягає в побудові й аналізі економіко-математичної моделі у вигляді рівняння регресії, що виражає залежність результативного показника від визначальних його факторів:

$$\hat{Y}_x = f(x_1, x_2, \dots, x_n) \quad (5.5)$$

РКА складається з наступних етапів (стадій), які тісно пов'язані між собою:

1. Попередній (апріорний) аналіз.
2. Збір інформації та її первинна обробка.
3. Побудова моделі (рівняння регресії).
4. Оцінка й аналіз моделі.

Під **попереднім (апріорним) аналізом** розуміється процес дослідження даного явища до збору початкової інформації. На цьому етапі формуються основні напрямки всього регресійно-кореляційного аналізу, а саме:

- у загальному вигляді формулюються завдання дослідження;
- здійснюється вибір і визначається методика вимірювання результативного показника;
- здійснюється попередній відбір факторів, які будуть включені в модель, і вирішується питання щодо способів їх вимірювання;

– обґрунтовуються напрямки причинно-наслідкових зв'язків між результативною та факторними ознаками;

– обґрунтовується форма зв'язку між результативною та факторними ознаками.

При формуванні попереднього переліку факторних ознак необхідно враховувати такі обставини:

– факторні ознаки повинні бути кількісними, краще за все – безперервними;

– факторні ознаки повинні розраховуватися по відношенню до однієї бази;

– не рекомендується включати в модель показники, які взаємно дублюють один одного, тобто відображають одну й ту ж сторону досліджуваного явища.

Результати регресійно-кореляційного аналізу, як і будь-якого статистичного дослідження, значною мірою залежать від якості вихідної інформації. Тому необхідно, щоб інформаційна база відповідала певним вимогам:

– статистична сукупність (кількість одиниць або спостережень) повинна бути достатньою за обсягом, для того щоб через дію закону великих чисел визначені в ході РКА статистичні характеристики були типовими і надійними;

– вихідні дані повинні бути якісно та кількісно однорідними (якісна однорідність припускає близькість умов формування результативної і факторних ознак; кількісна однорідність полягає у відсутності одиниць спостереження, які за своїми числовими характеристиками істотно відрізняються від основної маси даних);

– факторні ознаки, що включаються в модель, не повинні знаходитися між собою в тісному взаємному зв'язку;

– кількість спостережень повинна щонайменше в 6 разів перевищувати кількість факторів, які будуть включені в регресійну модель.

Побудова регресійної моделі починається із встановлення типу аналітичної функції, що характеризує механізм взаємозв'язку між результативною ознакою і одним або декількома ознаками-чинниками. При цьому рівняння регресії може включати один (однофакторна модель) або декілька (багатофакторна модель) факторів.

У випадку побудови однофакторної моделі вибір типу аналітичної функції здійснюється наступним чином:

– якщо зі зміною фактора результативний показник змінюється більш-менш рівномірно, такий зв'язок описується лінійною функцією;

– якщо при рівномірному збільшенні значення факторного показника результативний збільшується або зменшується прискорено (або спочатку збільшується, а потім зменшується), то таку залежність можна представити у вигляді параболи (полінома) другого порядку;

– обернену залежність між двома ознаками можна виразити або рівнянням прямої з від'ємним коефіцієнтом регресії, або рівнянням гіперболи;

– якщо відносна зміна результативного показника повинна бути пропорційною відносній зміні ознаки-фактора, то в якості форми зв'язку можна прийняти ступеневу функцію.

Якщо будується багатофакторна регресійна модель, то вибір типу аналітичної функції потребує більш детального вивчення. Якщо ж за тими чи іншими причинами неможливо встановити форму зв'язку між змінними, то зазвичай використовують лінійну форму зв'язку.

Побудова моделі, тобто розрахунок коефіцієнтів регресії здійснюється шляхом реалізації методу найменших квадратів, для чого необхідно вирішити систему нормальних рівнянь. У сучасних умовах це здійснюється шляхом використання програмних продуктів.

Найбільш простим є використання редактора *Excel*.

Для цього необхідно активізувати «Пакет аналізу» і далі здійснити наступні кроки:

1. Обрати опцію «Регресія».
2. У діалоговому вікні з'являються адреси результативної і факторних ознак, їх необхідно вказати (із попередньо складеного в *Excel* масиву вихідних даних).
3. Встановити рівень надійності коефіцієнтів регресії (за замовчуванням – 95 %).
4. За необхідності активізувати опцію «Остатки».
5. Командою ОК вивести результати на новий робочий аркуш.

Після цього на робочому аркуші з'являються результати РКА у формі розрахункової таблиці.

Знайдені параметри моделі (рівняння регресії) знаходяться в 3-му блоці таблиці у стовпчику «Кoeffициенты». Вільний член рівняння знаходиться в рядку з позначкою «Y-пересечение».

Після побудови моделі, як відмічалось раніше, здійснюється її аналіз.

По-перше, здійснюється *оцінка точності побудованої моделі*. Для цього зазвичай використовують множинний індекс детермінації, який у випадку лінійної форми зв'язку називають множинним коефіцієнтом детермінації:

$$R^2 = \frac{\sigma_{\phi}^2}{\sigma^2}, \quad (5.6)$$

де σ_{ϕ}^2 – факторна дисперсія результативної ознаки, яка

розраховується за формулою: $\sigma_{\phi}^2 = \frac{\sum_{i=1}^n (\hat{Y}_i - \bar{Y})^2}{n}$;

σ^2 – загальна дисперсія результативної ознаки, яка визначається за

формулою: $\sigma^2 = \frac{\sum_{i=1}^n (Y_i - \bar{Y})^2}{n}$.

Коефіцієнт детермінації може набувати значень від 0 до 1. При цьому чим ближчим є його значення до 1, тим більш точно отримане рівняння регресії описує залежність результативного показника від факторних.

Множинний коефіцієнт детермінації показує, скільки відсотків загальної варіації результативної ознаки пояснюється варіацією факторів, включених у регресійну модель.

Наступним етапом аналізу є *оцінка тісноти зв'язку* між результативною ознакою і включеними в модель факторами. Для цього використовується множинний коефіцієнт кореляції, який визначається як корінь квадратний із коефіцієнта детермінації:

$$R = \sqrt{R^2} \quad (5.7)$$

Коефіцієнт кореляції може приймати значення від 0 до 1. При $R = 0$ зв'язок між факторними та результативною ознаками відсутній, при $R = 1$ зв'язок є функціональним.

Для оцінки щільності (тісноти) зв'язку можна використати шкалу Чеддока:

Значення коефіцієнта кореляції	0,1 – 0,3	0,3 – 0,5	0,5 – 0,7	0,7 – 0,9	0,9 – 0,99
Характеристика тісноти зв'язку	слабка	помірна	помітна	висока	дуже висока

Оцінка адекватності регресійної моделі здійснюється за допомогою F -критерію Фішера, який розраховується за такою формулою:

$$F = \frac{R^2}{1-R^2} \cdot \frac{n-m}{m-1}, \quad (5.8)$$

де n – кількість одиниць досліджуваної сукупності;

m – кількість параметрів рівняння регресії.

Розрахункове значення F порівнюється з критичним (табличним) для прийнятого рівня значущості α і кількості ступенів вільності $k_1 = m - 1$ і $k_2 = n - m$. Таблиця критичних значень F -критерію Фішера наведена в *Додатку В*.

У випадку використання редактора *Excel* розрахункове значення F -критерію, а також його статистична значущість знаходяться автоматично: вони наводяться у 2-му блоці розрахункової таблиці, про яку мова йшла вище. Даний блок називається «Дисперсионный анализ», у ньому містяться стовпчики « F » і «значимість F ».

Побудована модель вважається надійною з імовірністю 0,95 (0,90; 0,99) у тому випадку, коли заданий рівень значущості F -критерію ($\alpha = 0,05$; $\alpha = 0,10$; $\alpha = 0,01$) більший, ніж отриманий розрахунковий рівень значущості F .

Після перевірки якості моделі в цілому здійснюється **перевірка та аналіз коефіцієнтів регресії**.

Для цього, в першу чергу, необхідно перевірити відповідність знаків коефіцієнтів регресії теоретичним і логічним уявленням про напрямок впливу факторів на результативний показник. Якщо знак коефіцієнта регресії не відповідає логічним уявленням про напрямок зв'язку між факторною і результативною ознаками, то може бути прийнято одне з двох наступних рішень:

1) встановлення причини такої невідповідності шляхом додаткового аналізу об'єкта дослідження, результативного і факторного показників, якості зібраної інформації;

2) виключення даного фактора із моделі.

Наступний крок аналізу поєднує в собі дві процедури, які необхідно виконувати практично паралельно:

– оцінка значущості коефіцієнтів регресії за t -критерієм Стьюдента;

– вимірювання тісноти зв'язку між факторами за допомогою парних коефіцієнтів кореляції.

У випадку побудови однофакторної лінійної моделі t -критерій Стьюдента розраховується для кожного i -го коефіцієнта регресії за формулою:

$$t_i = a_i \frac{\sqrt{n-2}}{\sigma_\varepsilon} \sigma_x, \quad (5.9)$$

де σ_x – середнє квадратичне відхилення i -ї факторної ознаки;

σ_ε – залишкове середнє квадратичне відхилення результативної ознаки, яке розраховується за формулою: $\sigma_\varepsilon = \sqrt{\frac{\sum (Y - \hat{Y})^2}{n}}$.

Розрахункові значення t -критерію Стьюдента порівнюють з критичним (табличним) для відповідної кількості числа ступенів вільності $k=n-m$ (n – кількість одиниць сукупності; m – кількість параметрів) і прийнятого рівня ймовірності.

У випадку використання редактора *Excel* розрахункове значення t -критерію, а також його статистична значущість знаходяться автоматично: вони наводяться у 3-му блоці розрахункової таблиці, про яку мова йшла вище. У цьому блоці таблиці містяться стовпчики з розрахунковими значеннями t -статистики та їх P -значення. Перевірка надійності коефіцієнтів регресії здійснюється аналогічно перевірці надійності моделі в цілому за F -критерієм Фішера (про що мова йшла раніше).

Якщо фактор у моделі виявився ненадійним, то його необхідно виключити. Якщо таких факторів виявилось декілька, то їх виключення із моделі здійснюється поступово:

– на першому етапі відсіюється фактор з найнижчим рівнем значущості;

– на другому етапі знову будується модель уже без цього фактора і здійснюється її повна оцінка, а також оцінка кожного фактора;

– далі процедура повторюється до того часу, поки в моделі не залишаться тільки суттєві фактори.

Вимірювання тісноти зв'язку між результативною і факторними ознаками, а також попарно між самими факторними ознаками, можна здійснювати на основі розрахунку парних коефіцієнтів кореляції. Їх

розрахунок здійснюється в редакторі *Excel* шляхом використання опції «Корреляция». Якщо два фактори тісно пов'язані між собою і один із них є незначущим за t -критерієм Стьюдента, або рівень його значущості знаходиться на межі, то саме цей фактор потрібно виключити із моделі.

Після того як у моделі залишилися лише суттєві фактори, можна приступити до економічної інтерпретації коефіцієнтів регресії. У випадку лінійної форми зв'язку **коефіцієнт регресії** показує на скільки одиниць у середньому змінюється результативний показник при зміні даного факторного показника на одну свою одиницю (при фіксованому значенні всіх інших факторів, що включені в модель).

При аналізі та економічній інтерпретації рівняння множинної регресії поряд з коефіцієнтами регресії з метою порівняння сили впливу кожного фактора на результативний показник використовують й інші характеристики: коефіцієнти еластичності та β -коефіцієнти.

Коефіцієнт еластичності (E) показує, на скільки відсотків змінюється у середньому результативна ознака при зміні факторної ознаки на 1 %. Він розраховується за формулою:

$$E_i = a_i \frac{\bar{x}_i}{\bar{Y}}, \quad (5.10)$$

де a_i – коефіцієнт регресії при i -му факторі;

\bar{x}_i – середнє значення i -го фактора;

\bar{Y} – середнє значення результативної ознаки.

Очевидно, що чим більшим є значення коефіцієнта еластичності, тим більшим є вплив даного фактора на результативний показник.

β -коефіцієнт показує, на скільки середніх квадратичних відхилень зміниться в середньому результативний показник, якщо відповідний факторний показник зміниться на одне своє середнє квадратичне відхилення. Він розраховується за формулою:

$$\beta_i = a_i \frac{\sigma_{x_i}}{\sigma_y}, \quad (5.11)$$

де σ_{x_i} – середнє квадратичне відхилення i -го фактора;

σ_y – середнє квадратичне відхилення результативної ознаки.

Чим більшим є значення β -коефіцієнта, тим більшою є сила впливу даного фактора на результативний показник.

Розглянемо реалізацію регресійно-кореляційного методу на прикладі аналізу показника середньомісячної номінальної заробітної плати працівників в Україні.

Показник середньомісячної заробітної плати обрано нами тому, що заробітна плата відіграє важливу роль у суспільстві, оскільки вона є одним з основних видів доходів населення не лише України, а й багатьох інших країн світу. Крім того, заробітна плата є не лише доходом, а й найважливішим стимулюючим фактором інноваційної діяльності, що є дуже важливим для економіки кожної країни.

На розмір заробітної плати впливає цілий комплекс соціально-відтворювальних факторів. Вони визначають не лише рівень безпосередньо заробітної плати, а й вартість інших життєвих засобів, необхідних для нормального відтворення робочої сили. При цьому таких коштів має вистачити не лише для відновлення працездатності працівника, а й для утримання на певному рівні непрацездатних членів його сім'ї.

Побудову багатофакторної лінійної моделі проведемо за допомогою вбудованого блоку Excel «Регресія» (покроково алгоритм такої побудови наведено вище). При цьому, як було відмічено раніше, в якості результативної ознаки виступає розмір середньомісячної номінальної заробітної плати в розрахунку на одного штатного працівника (в гривнях).

У якості факторів обрано такі економічні показники:

x_1 – валовий внутрішній продукт на душу населення, грн.;

x_2 – валовий регіональний продукт на душу населення, грн.;

x_3 – прожитковий мінімум, грн.;

x_4 – рівень безробіття населення, %;

x_5 – рівень зайнятості населення, %;

x_6 – обсяг інвестицій в основний капітал у розрахунку на одну особу, тис. грн.

Для побудови регресійно-кореляційної моделі були використані офіційні дані по Україні за 2000-2016 рр., що містяться на сайті Державної служби статистики України та в Статистичних щорічниках України. Вихідні дані наведено у *Додатку Г*.

На першому етапі робимо теоретичні припущення щодо напрямку зв'язку між факторними та результативною ознаками: всі

фактори, крім рівня безробіття, повинні мати з результативним показником прямий зв'язок.

На підставі даних *Додатку Г* будуюмо багатофакторне рівняння регресії, яке описує зв'язок між факторами $x_1 - x_6$ і результативною ознакою. Побудова здійснювалася шляхом використання вбудованого блоку Excel «Регресія». Результати («Вывод итогов») наведено в *Додатку Д*.

Як видно із *Додатку Д*, такі фактори, як валовий регіональний продукт на душу населення, рівень безробіття, рівень зайнятості населення та обсяг інвестицій у основний капітал у розрахунку на одну особу виявилися незначущими за t -критерієм Стьюдента, тому їх необхідно виключити із моделі.

Після цього здійснюється побудова моделі без виключених факторів, причому процедура РКА проводиться із самого початку.

У результаті отримали таку регресійну модель (її характеристики наведені в *Додатку Ж*):

$$Y = -285,372 + 0,073x_1 + 0,970x_3$$

Отже, після виключення з моделі незначущих факторів залишилися такі: валовий внутрішній продукт на душу населення і прожитковий мінімум.

Критерій Фішера $F = 10029,4$ та його рівень значущості $\alpha = 8,02849E-23$ свідчать про те, що побудована модель у цілому є адекватною.

Множинний коефіцієнт детермінації досить високий, його значення складає 99,93%. Це означає, що 99,93% загальної варіації показника середньої заробітної плати обумовлено варіацією обраних факторів. Решта 0,07% варіації результативної ознаки пояснюється впливом інших, не врахованих у моделі, факторів. Тобто, модель у цілому є достатньо точною.

Оскільки значення коефіцієнта кореляції наближається до 1, то дана модель відображає тісний зв'язок між рівнем середньої заробітної плати і факторами, що до неї ввійшли.

Як ми відмічали раніше, у регресійній моделі коефіцієнти регресії характеризують середню величину зміни результативного показника при зміні кожного фактора на одиницю за умови, що інші показники залишаються без зміни, тобто закріплені на середньому рівні.

У даному випадку знаки коефіцієнтів регресії в моделі відповідають теоретичним уявленням про напрямок впливу факторів на результативну ознаку. Отже, можна навести їх економічну інтерпретацію: при збільшенні валового внутрішнього продукту на душу населення на 1 грн. розмір заробітної плати збільшується в середньому на 0,073 грн., а збільшення прожиткового мінімуму на 1 грн. призводить до збільшення розміру заробітної плати в середньому на 0,97 грн.

Для перевірки мультиколінеарності між факторами використано вбудований блок *Excel* «Корреляция». Результати перевірки наведені в таблиці 5.5.

Таблиця 5.5

Матриця коефіцієнтів кореляції

	Столбец 1	Столбец 2	Столбец 3
Столбец 1	1		
Столбец 2	0,997651225	1	
Столбец 3	0,981347543	0,867682491	1

Матриця коефіцієнтів кореляції характеризує наявність або відсутність кореляції між факторами, що ввійшли до моделі. Отже, результати перевірки показали, що мультиколінеарність між факторами є, тобто зв'язок досить високий, але ці фактори здійснюють важливий вплив на рівень результативного показника, тому доцільно залишити їх в моделі, тим більше, що коефіцієнти Стюдента цих факторів показали, що вони мають досить високий рівень вагомості в порівнянні з іншими факторами, які були включені в модель.

Визначимо коефіцієнти еластичності за формулою (5.10):

$$E_1 = 0,073 \cdot 20902,0 / 1935,4 = 0,788 (\%)$$

$$E_3 = 0,970 \cdot 726,9 / 1935,4 = 0,364 (\%)$$

Розраховані значення коефіцієнтів еластичності свідчать про те, що значний вплив на рівень заробітної плати в Україні має валовий внутрішній продукт на душу населення: з його підвищенням на один відсоток результативний показник зростає майже на 0,8 %. Цей факт є очевидним, оскільки держава впливає на рівень заробітної плати працівників через наявні ресурси (у першу чергу – валовий внутрішній продукт).

5.4. Розрахунок резервів на основі регресійної моделі

Побудоване рівняння регресії може бути використане для розрахунку резервів. У кожному конкретному випадку (залежно від сутності результативного показника) здійснюється розрахунок резервів або його зростання, або зниження.

Якщо результативний показник є *ознакою-стимулятором*, то можна визначити резерви *зростання* цього показника. Наприклад, якщо результативним показником є рентабельність підприємства, то логічним є розрахунок резервів зростання даного показника, тобто даний показник є *ознакою-стимулятором*.

Якщо ж результативний показник є *ознакою-дестимулятором*, то є сенс розраховувати *резерви зниження* цього показника. Якщо, наприклад, результативним показником є витрати на гривню реалізованої продукції, то, очевидно, що потрібно прагнути до їх зменшення, тобто даний показник є *ознакою-дестимулятором*, тому мова може йти про розрахунок резервів його зниження.

Визначення резервів на основі регресійної моделі може здійснюватися в тому випадку, коли досліджувана сукупність є просторовою, тобто вихідна інформація – це *дані по сукупності об'єктів* (підприємств, робітників тощо). Якщо ж вихідні дані характеризують динаміку результативного і факторних показників по окремому об'єкту, то резерви не розраховуються.

Процес розрахунку резервів на основі регресійної моделі складається з наступних етапів.

На першому етапі здійснюється розрахунок середнього рівня результативної ознаки.

На другому етапі вся сукупність об'єктів ділиться на дві групи: 1) група лідерів – це передові об'єкти, в яких значення результативної ознаки більші, ніж середнє значення; 2) група аутсайдерів – це відстаючі об'єкти, для яких значення результативної ознаки менші, ніж середнє значення.

На третьому етапі в кожній виділеній групі розраховуються середні значення результативної і факторних ознак, що увійшли до рівняння регресії, тобто визначаються групові середні.

На четвертому етапі для кожної з факторних ознак і результативної ознаки знаходяться різниці середніх значень передових і відстаючих об'єктів.

На п'ятому етапі отримані різниці середніх значень по кожній факторній змінній множаться на відповідні коефіцієнти регресії. Це і є резерви зростання (зменшення) результативного показника за рахунок зміни кожного фактора.

Питання для самоконтролю

1. Які існують види зв'язків?
2. Що є характерним для функціонального, стохастичного та кореляційного зв'язку?
3. Коли для моделювання використовується адитивна модель?
4. Коли для моделювання використовується мультиплікативна модель?
5. У чому є принципова відмінність між методами ланцюгових підстановок і виявлення ізольованого впливу факторів?
6. У чому сутність регресійно-кореляційного аналізу?
7. З яких етапів складається регресійно-кореляційний аналіз?
8. Які вимоги висуваються до факторних ознак при включенні їх в регресійну модель?
9. Які вимоги пред'являє регресійно-кореляційний аналіз до якості вихідної інформації?
10. Який показник використовують для оцінки точності регресійної моделі і як його розраховують?
11. Який показник використовують для вимірювання тісноти зв'язку в регресійній моделі і як його розраховують?
12. За допомогою якого показника здійснюється оцінка адекватності регресійної моделі і як його розраховують?
13. За допомогою якого показника оцінюють істотність (значущість) коефіцієнтів регресії?
14. Для чого використовують шкалу Чеддока?
15. Що характеризує матриця коефіцієнтів кореляції?
16. Який зміст мають коефіцієнти регресії?
17. Як розраховуються та який зміст мають коефіцієнти еластичності?
18. Як розраховуються та який зміст мають β -коефіцієнти?
19. Що таке «ознака-стимулятор» та «ознака-дестимулятор»?
20. Назвіть та охарактеризуйте етапи визначення резервів на основі регресійної моделі.

РОЗДІЛ 6. ДИСПЕРСІЙНИЙ АНАЛІЗ У ДОСЛІДЖЕННІ БІЗНЕС-ПРОЦЕСІВ

6.1. Види дисперсій. Правило декомпозиції дисперсій

Основою будь-якого статистичного дослідження просторових даних про соціально-економічні явища є формування якісно однорідної сукупності. Тому серед статистичних методів особливе місце займає дисперсійний аналіз, одним з основних завдань якого є вимірювання однорідності сукупності.

При цьому якщо досліджувана сукупність розбита на групи за певною ознакою, то можна визначити дисперсію як у цілому по сукупності, так і в кожній групі, крім того можна визначити міжгрупову дисперсію.

Зведеною характеристикою розсіювання (варіації) значень кількісної ознаки в генеральній сукупності навколо свого середнього значення є *загальна (генеральна) дисперсія*, яка розраховується за формулою:

$$\sigma^2 = \frac{\sum (X - \bar{X})^2}{n}, \quad (6.1)$$

де X – значення ознаки в кожній одиниці загальної сукупності;

\bar{X} – загальна середня;

n – кількість одиниць сукупності.

Розглядаючи кожну групу як самостійну сукупність можна визначити *дисперсію в групі*, яка характеризує розсіювання значень кількісної ознаки навколо групової середньої величини. Розрахунок дисперсії в групі здійснюється за формулою, аналогічною загальній дисперсії:

$$\sigma_j^2 = \frac{\sum (X_{ij} - \bar{X}_j)^2}{n_j}, \quad (6.2)$$

де X_{ij} – значення ознаки в кожній одиниці сукупності в j -й групі;

\bar{X}_j – середня величина ознаки в j -й групі;

n_j – кількість одиниць в j -й групі.

Внутрішньогрупова дисперсія характеризує випадкову варіацію ознаки, яка відбувається під впливом інших факторів (крім групувального), і не залежить від ознаки, що покладена в основу групування.

Узагальнюючою мірою внутрішньогрупової варіації є *середня з групових дисперсій*, яка розраховується за такою формулою:

$$\bar{\sigma}^2 = \frac{\sum \sigma_j^2 f_j}{\sum f_j}, \quad (6.3)$$

де f_j – частота або частка j -ї групи.

Мірою варіації групових середніх навколо загальної середньої є *міжгрупова дисперсія*. Вона характеризує відмінності між варіантами досліджуваної ознаки, які виникають під дією лише однієї ознаки – фактора, який покладено в основу групування. Тобто, міжгрупова дисперсія показує, на скільки відрізняються між собою групові середні. Вона розраховується за формулою:

$$\delta^2 = \frac{\sum (\bar{X}_j - \bar{X})^2 f_j}{\sum f_j}, \quad (6.4)$$

Міжгрупова дисперсія може дорівнювати нулю, коли середні величини в кожній групі однакові (рівні) і вона є тим більшою, чим більше відрізняються середні між собою.

Загальна дисперсія складається з двох частин. Перша характеризує внутрішньогрупову, друга – міжгрупову варіацію. Тому загальна дисперсія дорівнює сумі середньої з внутрішньогрупових і міжгрупової дисперсій:

$$\sigma^2 = \bar{\sigma}^2 + \delta^2 \quad (6.5)$$

Взаємозв'язок дисперсій називають правилом декомпозиції (розкладання) дисперсій. На практиці цей взаємозв'язок досить часто використовується для визначення однієї з дисперсій на основі значень двох інших дисперсій.

Відношення міжгрупової дисперсії до загальної характеризує частку варіації результативної ознаки, яка обумовлена варіацією (впливом) групувальної ознаки. Це відношення називається *емпіричним коефіцієнтом детермінації*:

$$\eta^2 = \frac{\delta^2}{\sigma^2} \quad (6.6)$$

Для оцінки тісноти зв'язку між групувальною і результативною ознаками розраховується *емпіричне кореляційне відношення*:

$$\eta = \sqrt{\eta^2} \quad (6.7)$$

Емпіричне кореляційне відношення може набувати значень від 0 до 1. Чим більше його значення наближається до 1, тим більш тісним є зв'язок між групувальною і результативною ознаками.

Розглянемо розрахунок емпіричного кореляційного відношення на наступному прикладі.

24 студенти групи розподілені за кількістю пропущених занять із статистики на 2 групи: 1-а – студенти, які пропустили менше, ніж 10 % загальної кількості годин, 2-а – 10 % і більше. До 1-ї групи потрапили 18 студентів, середній бал рейтингу яких склав 78, причому бали (рейтинг) окремих студентів групи відхиляються від середнього рейтингу в середньому на 11 балів. У 2-ій групі середній бал склав 46, а коливання навколо середнього рейтингу – 16.

Потрібно визначити, чи впливає наявність або відсутність пропусків занять на варіацію балів рейтингу студентів групи, тобто, чи існує залежність між пропусками занять студентів та рівнем їх успішності зі статистики.

Для зручності складемо таблицю, в яку занесемо вихідні дані (табл. 6.1).

Таблиця 6.1

Вихідні дані щодо розподілу студентів

Групи студентів за кількістю пропусків занять, %	Кількість студентів, чол. (f_i)	Середній бал (\bar{X}_i)	Варіація балів (σ_i)
До 10	18	78	11
10 і більше	6	46	16
Усього	24	-	-

Для того, щоб відповісти на поставлене питання необхідно визначити емпіричне кореляційне відношення. Для цього нам необхідно мати міжгрупову та загальну дисперсії.

Визначимо середній бал рейтингу студентів групи в цілому на основі групових середніх за формулою середньої арифметичної зваженої:

$$\bar{X} = \frac{\sum \bar{X}_i f_i}{\sum f_i}$$

Проміжні розрахунки наведено в табл. 6.2.

$$\bar{X} = \frac{1680}{24} = 70 \text{ балів}$$

Визначаємо міжгрупову дисперсію за формулою 6.4 (проміжні розрахунки наведено в табл. 6.2).

$$\delta^2 = \frac{4608}{24} = 192$$

Таблиця 6.2

До розрахунку емпіричного кореляційного відношення

Групи студентів за кількістю пропусків занять, %	(f_i)	(\bar{X}_i)	$\bar{X}_i f_i$	$(\bar{X}_i - \bar{X})^2 f_i$	σ_i^2	$\sigma_i^2 f_i$
До 10	18	78	11	1152	121	2178
10 і більше	6	46	16	3456	256	1536
Усього	24	-	-	4608	-	3714

Оскільки немає даних про рейтинг кожного студента, то не можна визначити загальну дисперсію за вихідною формулою 6.1, тому її можна визначити згідно правила декомпозиції дисперсій (формула 6.5). Для цього спочатку визначимо середню з внутрішньогрупових дисперсій за формулою 6.3 (проміжні розрахунки наведено в табл. 6.2):

$$\bar{\sigma}^2 = \frac{3714}{24} = 154,75$$

Загальна дисперсія складає:

$$\sigma^2 = 154,75 + 192 = 346,75$$

Визначаємо емпіричний коефіцієнт детермінації за формулою 6.6:

$$\eta^2 = \frac{192}{346,75} = 0,554 \quad \text{або} \quad 55,4 \%$$

Отже, 55,4 % загальної варіації балів рейтингу студентів пояснюється варіацією пропусків занять.

Емпіричне кореляційне відношення складає:

$$\eta = \sqrt{0,554} = 0,744$$

Оскільки значення емпіричного кореляційного відношення наближається до 1, то між кількістю пропусків занять із статистики та рівнем успішності студентів (балами їх рейтингу) з дисципліни існує достатньо тісний зв'язок. Причому, за результатами аналітичного групування, яке міститься в табл. 6.1, можна побачити, що зв'язок цей обернений.

Отримані результати даного прикладу свідчать про те, що зв'язок між досліджуваними ознаками є, але якщо значення емпіричного кореляційного відношення складає близько 50 %, то дати відповідь відносно існування зв'язку складно. Такі питання (і не лише) допомагає вирішувати дисперсійний аналіз.

6.2. Сутність дисперсійного аналізу

В основі дисперсійного аналізу лежить правило декомпозиції дисперсій. Однак відмінність полягає в тому, що в дисперсійному аналізі здійснюється не простото розчленування загальної варіації досліджуваної ознаки на складові, які обумовлені впливом певних факторів, а проводиться перевірка гіпотези щодо суттєвості їх впливу (або значущості відмінностей між груповими середніми).

Дисперсійний аналіз широко застосовується в багатьох сферах діяльності: при дослідженні рівня життя населення, в медицині, хімії, фізиці тощо.

Зокрема, дисперсійний аналіз використовується для перевірки наявності відмінностей доходів чи витрат різних груп населення: наприклад він дозволяє встановити, чи є суттєвими регіональні відмінності в середніх доходах, чи ні. У цьому випадку загальна дисперсія варіації доходів населення аналізується як у зв'язку з відмінностями в самих регіонах, так і у зв'язку з відмінностями середніх значень доходів між регіонами. Очевидно, що чим більша частка загальної дисперсії припадає на відмінності між середніми значеннями доходів в окремих регіонах, тим вище ймовірність того, що ці групи (регіони) дійсно різні; разом з тим, чим більшою є частка дисперсії, яка припадає на внутрішньогрупові відмінності, тим вище ймовірність того, що відмінності між середніми значеннями доходів у різних груп мають суто випадковий характер.

Слід відмітити, що в економічних дослідженнях застосування дисперсійного аналізу є не таким поширеним, при тому, що важливість його використання в певних випадках є навіть більш значущою.

Розглянемо це на прикладі. З метою підвищення якості йогурту, що виготовляє підприємство, планується використовувати одну з чотирьох харчових добавок. Підприємство двадцять разів запускала виробництво йогурту з кожною добавкою і двадцять разів без добавки. Результат представляє собою набір із п'яти рядів даних

виробництва продукції (з кожною добавкою і без добавок). При цьому ми маємо один фактор – харчову добавку, який має декілька значень. Унаслідок того, що будь-який виробничий процес пов'язаний із багатьма факторами виробництва, то складно пояснити чим обумовлена зміна якості продукції: тією чи іншою добавкою, або суто випадковими обставинами, пов'язаними лише з мінливістю виробничого процесу. Саме для цього проводиться однофакторний дисперсійний аналіз, який дає можливість встановити, чи є суттєвими відмінності між харчовими добавками та їх впливом на рівень якості йогурту. Якщо ці відмінності суттєві, то можна далі більш детально вивчати доцільність здійснення витрат для впровадження у виробництво певної харчової добавки. У протилежному випадку можна зробити висновок, що суттєвих відмінностей між добавками немає, тому їх впровадження у виробничий процес скоріш за все є недоцільним.

Отже, в основі дисперсійного аналізу знаходиться припущення про те, що результати експерименту можна представити у вигляді певної суми складових. Наприклад, якщо досліджується вплив одного фактора, то модель, яка відображає результат експерименту, має вигляд [20, с. 207]:

$$x_{ij} = \bar{X} + \alpha_i + \varepsilon_{ij}, \quad (6.8)$$

де x_{ij} – значення ознаки, що отримані на i -му рівні фактора;

\bar{X} – загальна середня;

α_i – ефект фактора на i -му рівні;

ε_{ij} – випадковий компонент, який викликаний впливом інших факторів.

Однофакторний дисперсійний аналіз є найбільш поширеним в економічних дослідженнях.

У тому випадку, коли визначається вплив двох факторів (А і В) та їх взаємодія, то модель, яка відображає результат експерименту, має вигляд:

$$x_{ijk} = \bar{X} + \alpha_i + \beta_j + \gamma_{ij} + \varepsilon_{ijk}, \quad (6.9)$$

де x_{ijk} – результат експерименту в k -му спостереженні на i -му рівні фактора А і j -му рівні фактора В;

\bar{X} – загальна середня;

α_i – ефект фактора А;

β_j – ефект фактора В;

γ_{ij} – ефект, який викликаний спільним впливом обох факторів;

ε_{ij} – випадковий компонент, який викликаний впливом інших (крім А і В) факторів.

Для проведення дисперсійного аналізу вся досліджувана сукупність розбивається на однорідні групи, які відрізняються рівнями факторів. Показником, що характеризує рівень однорідності, може бути коефіцієнт варіації.

Якщо групувальною ознакою є якісна, то кількість груп та їх найменування визначаються самою ознакою. Якщо ж групувальною ознакою є кількісна, то виділення однорідних груп здійснюється поступово. На першому етапі визначається середня величина показника, за яким сукупність поділяється на найменшу необхідну кількість груп. На другому етапі в кожній групі визначаються основні статистичні характеристики (середня величина, дисперсія, коефіцієнт варіації). Якщо отримані групи є дуже неоднорідними, то сукупність поділяють на більшу кількість груп і т. д.

6.3. Однофакторний дисперсійний аналіз

Як ми відмічали раніше, за допомогою однофакторного дисперсійного аналізу досліджується вплив одного фактора, тобто перевіряється значущість відмінностей між собою значень середніх величин декількох вибірових сукупностей.

В однофакторному дисперсійному аналізі набір даних складається з k незалежних одновимірних вибірок із k генеральних сукупностей (або з однієї генеральної сукупності), елементи яких вимірюються в однакових одиницях.

Дані для однофакторного дисперсійного аналізу зручно подавати в такій формі (табл. 6.3).

У рядку «Усього» наводиться загальна кількість одиниць сукупності (загальний обсяг вибірки) n , загальна середня (по сукупності в цілому) \bar{X} і сума квадратів відхилень від середньої величини в цілому по сукупності S^2 .

Таблицю можна також транспонувати, зробивши рядки графами.

Для перевірки гіпотези щодо суттєвості відмінностей між середніми величинами у вибірках, що сформовані за факторною ознакою, використовується *F-критерій Фішера*.

В однофакторному дисперсійному аналізі статистичні гіпотези формулюються так:

- нульова гіпотеза (H_0): усі середні рівні між собою;
- альтернативна (дослідницька) гіпотеза (H_1): не всі середні рівні між собою, принаймні існує хоча б одна пара вибірових сукупностей, середні величини в яких відрізняються між собою.

Таблиця 6.3

Форма подання даних для однофакторного дисперсійного аналізу

Номер вибірки	Значення досліджуваної ознаки	Обсяг вибірки	Середня	Сума квадратів відхилень від середніх
1	$X_{11}; X_{21}; \dots; X_{i1}; \dots; X_{n1}$	n_1	\bar{X}_1	S_1^2
2	$X_{12}; X_{22}; \dots; X_{i2}; \dots; X_{n2}$	n_2	\bar{X}_2	S_2^2
...
j	$X_{1j}; X_{2j}; \dots; X_{ij}; \dots; X_{nj}$	n_j	\bar{X}_j	S_j^2
...
k	$X_{1k}; X_{2k}; \dots; X_{ik}; \dots; X_{nk}$	n_k	\bar{X}_k	S_k^2
Усього		n	\bar{X}	S^2

Для коректного використання однофакторного дисперсійного аналізу повинні виконуватися такі *вимоги*:

1. Кількісні дані повинні бути неперервними, дискретні дані є менш бажаними.
2. Вибіркові сукупності повинні бути незалежними.
3. Незалежними повинні бути спостереження в кожній із вибірок.
4. Вибіркові сукупності повинні мати нормальний розподіл.
5. Дисперсії генеральних сукупностей, із яких здійснювалися вибірки, повинні бути однаковими ($\sigma_1^2 = \sigma_2^2 = \dots = \sigma_k^2$). Це означає, що всі вибірки здійснюються або з однієї генеральної сукупності, або середні в генеральних сукупностях, з яких здійснюються вибірки, рівні між собою, що дозволяє використовувати для перевірки гіпотез стандартні статистичні таблиці.

F-критерій Фішера для однофакторного дисперсійного аналізу визначається як співвідношення незміщених оцінок міжгрупової та середньої з внутрішньогрупових дисперсій:

$$F = \frac{\delta^2}{\bar{\sigma}^2} \quad (6.10)$$

Тобто, *F-критерій* показує, у скільки разів середні величини кожної з вибірок відрізняються між собою, тобто наскільки більше вони відрізняються між собою, ніж можна було цього очікувати, якби їх розбіжності були суто випадковими. Очевидно, що його значення можуть бути або додатними, або дорівнювати нулю. Дорівнювати нулю *F-критерій* може в тому випадку, коли міжгрупова варіація ознаки відсутня.

Для того, щоб перевірити істотність розходження дисперсій необхідно отримане значення *F-критерію* порівняти з табличним. Таблиці розподілу *F-критерію* для 5 % рівня значущості у випадку однофакторного дисперсійного аналізу містяться в *Додатку В*, також їх легко можна знайти в Інтернеті.

Слід звернути увагу на те, що у формулі *F-критерію* Фішера зіставляються не просто дисперсії, а їх незміщені оцінки. Відмінність полягає в тому, що при визначенні незміщених оцінок дисперсій суми квадратів відхилень діляться не на кількість одиниць сукупності, а на кількість ступенів свободи.

Поняття ***кількості ступенів свободи*** вів у вжиток англійський статистик Рональд Фішер (1890-1962), і воно означає кількість незалежних змінних у сумі.

Кількість ступенів свободи визначається по-різному для кожної з дисперсій. І тому критерій Фішера має два значення ступенів свободи, які відповідають двом порівнюваним дисперсіям: міжгрупової і середньої з внутрішньогрупових.

Для міжгрупової дисперсії кількість ступенів свободи складає $k - 1$ (k – кількість груп, на які розбита вся сукупність, тобто кількість одновимірних незалежних вибірок). Отже, ***незміщена оцінка міжгрупової дисперсії*** (6.4) має вигляд:

$$\delta^2 = \frac{\sum (\bar{X}_j - \bar{X})^2 f_j}{k - 1} \quad (6.11)$$

Незміщена оцінка внутрішньогрупової дисперсії (6.2) розраховується за формулою:

$$\sigma_j^2 = \frac{\sum (X_{ij} - \bar{X}_j)^2}{n_j - 1} \quad (6.12)$$

Для середньої із внутрішньогрупових дисперсій кількість ступенів свободи визначається як $n - k$. А отже і **незміщена оцінка середньої із внутрішньогрупових дисперсій** (6.3) розраховується за формулою:

$$\bar{\sigma}^2 = \frac{\sum \sigma_j^2 (f_j - 1)}{n - k} \quad (6.13)$$

Незміщена оцінка середньої з внутрішньогрупових дисперсій може бути визначена і в інший спосіб:

$$\bar{\sigma}^2 = \frac{\sum \sum (\bar{X}_j - \bar{X})^2}{n - k} \quad (6.14)$$

У чисельнику даної формули фігурує загальна сума квадратів відхилень варіантів (значень ознаки) від середніх величин в кожній групі.

Розрахунок оцінок дисперсій зручно розміщувати в спеціальній таблиці, яку називають таблицею дисперсійного аналізу [20, с. 210]. Для однофакторного дисперсійного аналізу вона має такий вигляд (табл. 6.4).

Таблиця 6.4

Форма таблиці однофакторного дисперсійного аналізу

Характеристика варіації	Сума квадратів відхилень	Кількість ступенів свободи	Оцінка дисперсії
Міжгрупова (систематична)	$S_1^2 = \sum (\bar{X}_j - \bar{X})^2 n_j$	$k - 1$	$\delta^2 = \frac{S_1^2}{k - 1}$
Внутрішньогрупова (залишкова)	$S_2^2 = \sum \sum (X_{ij} - \bar{X}_j)^2$	$n - k$	$\bar{\sigma}^2 = \frac{S_2^2}{n - k}$
Усього	$S^2 = S_1 + S_2 = \sum \sum (X_{ij} - \bar{X})^2$	$n - 1$	-

Дані табл. 6.4 дозволяють визначити співвідношення дисперсій, тобто F -критерій Фішера за формулою 6.10.

Після розрахунку F -критерію Фішера його значення порівнюється з критичним (табличним) значенням. Якщо фактичне значення F -критерію менше, ніж табличне, то слід прийняти нульову гіпотезу як прийнятну, і це означає, що розходження між

вибірковими середніми можна пояснити лише випадковістю, тобто результат не є статистично значущим.

Якщо фактичне значення F -критерію перевищує його табличне значення, то слід відхилити нульову гіпотезу і прийняти альтернативну, а це означає, що групові вибіркові середні суттєво відрізняються між собою, причому ці відмінності не можна пояснити лише випадковістю, тобто результат є статистично значущим.

Зазвичай для перевірки гіпотези використовують F -критерій із 5% рівнем значущості. При цьому F -розподіл Фішера, як ми вже відмічали вище, залежить від кількості ступенів свободи $\nu_1 = k - 1$ і $\nu_2 = n - k$.

Розглянемо реалізацію однофакторного дисперсійного аналізу на прикладі.

Є такі дані про рівень годинної заробітної плати трьох категорій робітників промислового підприємства: жінки та чоловіки із стажем роботи і молодь без стажу роботи (табл. 6.5). Необхідно визначити, чи впливає категорія, до якої відносяться робітники, на рівень їх заробітної плати.

Таблиця 6.5

Годинна заробітна плата робітників підприємства

Жінки			Чоловіки			Молодь		
номер робітника	годинна заробітна плата, грн.	$(X - \bar{X})^2$	номер робітника	годинна заробітна плата, грн.	$(X - \bar{X})^2$	номер робітника	годинна заробітна плата, грн.	$(X - \bar{X})^2$
1	48,30	15,68	11	47,28	1,49	22	46,97	7,02
2	42,56	4,75	12	49,36	10,89	23	42,16	5,66
3	44,61	0,02	13	45,12	0,88	24	45,12	0,34
4	43,12	2,62	14	43,96	4,41	25	40,78	14,14
5	41,32	11,70	15	47,54	2,19	26	42,14	5,76
6	42,20	6,45	16	44,80	1,59	27	43,19	1,82
7	43,22	2,31	17	42,31	14,06	28	45,87	1,77
8	44,64	0,01	18	44,20	3,46	29	44,56	0,00
9	46,32	2,50	19	45,12	0,88	30	48,11	12,74
10	47,12	5,66	20	48,44	5,66			
			21	48,51	6,00			
Усього	443,41	51,70	Усього	506,64	51,52	Усього	398,90	49,26

Визначимо середньогодинну заробітну плату робітників кожної категорії та по підприємству в цілому за формулою середньої арифметичної простої на основі даних, наведених у табл. 6.5:

- перша група (жінки): $\bar{X}_1 = \frac{443,41}{10} = 44,34 \text{ грн.};$
- друга група (чоловіки): $\bar{X}_2 = \frac{506,64}{11} = 46,06 \text{ грн.};$
- третя група (молодь): $\bar{X}_3 = \frac{398,90}{9} = 44,32 \text{ грн.};$
- підприємство в цілому: $\bar{X} = \frac{443,41 + 506,64 + 398,90}{30} = 44,97 \text{ грн.}$

За даними табл. 6.5 визначимо оцінки внутрішньогрупових дисперсій згідно формули 6.12:

- перша група (жінки): $\sigma_1^2 = \frac{51,70}{10-1} = 5,7444 ;$
- друга група (чоловіки): $\sigma_2^2 = \frac{51,52}{11-1} = 5,1520 ;$
- третя група (молодь): $\sigma_3^2 = \frac{49,26}{9-1} = 6,1575 .$

Оцінку середньої із внутрішньогрупових дисперсій визначимо за формулою 6.13:

$$\bar{\sigma}^2 = \frac{5,7444 \cdot 9 + 5,1520 \cdot 10 + 6,1575 \cdot 8}{30 - 3} = 5,6474 .$$

Визначимо також оцінку середньої з внутрішньогрупових дисперсій за іншою формулою – 6.14:

$$\bar{\sigma}^2 = \frac{152,48}{327} = 5,6474 .$$

На основі попередніх розрахунків середніх у кожній групі та в цілому по сукупності визначимо оцінку міжгрупової дисперсії за формулою 6.11:

$$\delta^2 = \frac{(44,34 - 44,97)^2 \cdot 10 + (46,06 - 44,97)^2 \cdot 11 + (44,32 - 44,97)^2 \cdot 9}{3 - 1} = 10,4203$$

Для зручності наведемо всі результати розрахунків у таблиці дисперсійного аналізу (табл. 6.6).

На основі даних табл. 6.6 визначаємо F -критерій Фішера згідно формули 6.10:

$$F = \frac{10,4203}{5,6474} = 1,845$$

Таблиця 6.6

Таблиця однофакторного дисперсійного аналізу

Характеристика варіації	Сума квадратів відхилень	Кількість ступенів свободи	Оцінка дисперсії
Міжгрупова (систематична)	$S_1^2 = 20,8406$	$3-1=2$	$\delta^2 = 10,4203$
Внутрішньогрупова (залишкова)	$S_2^2 = 52,48$	$30-3=27$	$\bar{\sigma}^2 = 5,6474$
Усього	$S^2 = 20,8406 + 52,48 = 73,3206$	$30-1=29$	-

Нагадаємо, що F -критерій Фішера показує на скільки вибіркові середні відрізняються між собою (чисельник формули) з урахуванням рівня варіації у вибірках (знаменник).

У даному випадку міжгрупова варіація заробітної плати (яка обумовлена тим, до якої категорії належать робітники) у 1,845 разів більша, ніж можна було очікувати виходячи лише з варіації заробітної плати окремих робітників. Але чи можна стверджувати, що ці відмінності є значущими? Для відповіді на це питання необхідно порівняти отримане значення F -критерію з критичним (табличним) його значенням.

Критичне значення F -критерію при рівні значущості 5% і кількості ступенів свободи ($3-1=2$) і ($30-3=27$) складає 3,35 (згідно таблиць розподілу, наведених у Додатку В).

Оскільки фактичне значення F -критерію Фішера (1,845) менше, ніж критичне (3,35), то результат перевірки є незначущим. Тобто, слід прийняти нульову гіпотезу щодо рівності середніх. А це означає, що рівень середньої заробітної плати робітників різних категорій не суттєво відрізняється між собою, тобто практично не існує залежності між тим, до якої з категорій відноситься робітник, і тим, яку заробітну плату він отримує.

6.4. Багатофакторний дисперсійний аналіз

Набір аналізованих даних може бути не лише одновимірним, про що йшла мова у попередніх параграфах, а й представляти собою

більш складну структуру. У цьому випадку дисперсійний аналіз може дати відповіді на більш складні питання.

При цьому дані можуть бути представлені як набір вибірок із однієї або декількох генеральних сукупностей, які виражені в однакових одиницях (так само як у випадку однофакторного дисперсійного аналізу). Можливий також інший спосіб представлення даних: у вигляді багатовимірного набору даних, у якому обов'язково одна змінна є кількісною (основна вимірювана величина), а всі інші змінні є якісними, і ці якісні змінні в сукупності визначають, яким чином кількісні спостереження групуються у вибірки.

Багатофакторний дисперсійний аналіз дає можливість оцінити ефект кожного фактора, тобто вплив кожного фактора на результативний показник, а також ефект взаємодії факторів (показати як фактори пов'язані між собою). При чому чим більше факторів, тим більше видів взаємодії між ними. Допустимо, при трьохфакторному дисперсійному аналізі крім чистого впливу кожного фактора на результативний показник визначається взаємодія між кожними двома факторами (першим і другим, другим і третім, першим і третім), а також між трьома факторами разом.

Таким чином, багатофакторний дисперсійний аналіз застосовується тоді, коли в одному дисперсійному аналізі необхідно одночасно досліджувати не лише вплив факторів, а також можливу коваріацію змінних. При цьому змінні не є незалежними одна від іншої, а навпаки, корелюють між собою.

Для коректного застосування багатофакторного дисперсійного аналізу повинні виконуватися такі ж вимоги, як і для однофакторного.

Розглянемо більш детально двохфакторний дисперсійний аналіз.

У цьому випадку вибірки утворюють таблицю, в якій одному фактору відповідають рядки, а іншому – стовпчики, тому виникає три основних питання: "Чи впливає перший фактор на появу будь-яких відмінностей результативної ознаки?" "Чи впливає другий фактор на появу будь-яких відмінностей?" "Чи залежить вплив першого фактора від другого, або обидва фактори діють незалежно один від одного?" Перші два питання мають відношення до головних ефектів кожного з факторів, а третє питання має відношення до взаємодії факторів [16, с. 828].

Фактори А і В вважаються **взаємодіючими**, якщо ефект фактора А залежить від рівня фактора В.

Інформацію для двохфакторного дисперсійного аналізу зручно наводити у вигляді такої таблиці (табл. 6.7) [20, с. 216].

В табл. 6.7: p – кількість рівнів фактора А, q – кількість рівнів фактора В, k – кількість вибірок.

За даними табл. 6.7 визначаються такі середні:

- загальна середня:
$$\bar{X} = \frac{\sum X_{ijk}}{pqn};$$
- середні за рядками:
$$\bar{X}_j = \frac{\sum X_{ijk}}{pn};$$
- середні за стовпчиками:
$$\bar{X}_i = \frac{\sum X_{ijk}}{qn};$$
- середні для кожного варіанта взаємодії факторів (для кожної окремої клітинки таблиці):
$$\bar{X}_{ij} = \frac{\sum X_{ijk}}{n}.$$

Таблиця 6.7

Форма подання даних для двохфакторного дисперсійного аналізу

Рівень фактора В	Рівень фактора А				Сума
	A_1	A_2	...	A_p	
B_1	$X_{111}; X_{112}; \dots;$ X_{11n}	$X_{211}; X_{212}; \dots;$ X_{21n}	...	$X_{p11}; X_{p12}; \dots;$ X_{p1n}	$\sum X_{i1k}$
B_2	$X_{121}; X_{122}; \dots;$ X_{12n}	$X_{221}; X_{222}; \dots;$ X_{22n}	...	$X_{p21}; X_{p22}; \dots;$ X_{p2n}	$\sum X_{i2k}$
...
B_q	$X_{1q1}; X_{1q2}; \dots;$ X_{1qn}	$X_{2q1}; X_{2q2}; \dots;$ X_{2qn}	...	$X_{pq1}; X_{pq2}; \dots;$ X_{pqn}	$\sum X_{iqk}$
Сума	$\sum X_{1jk}$	$\sum X_{2jk}$...	$\sum X_{pjk}$	$\sum X_{ijk}$

Суми квадратів відхилень від загальної середньої розкладається на складові. Схема двохфакторного дисперсійного аналізу наведена в табл. 6.8 [20, с. 217].

Оскільки двохфакторний дисперсійний аналіз дозволяє визначити вплив кожного фактора окремо, а також їх взаємну дію, то відповідно визначають три значення F -критерію Фішера:

- для фактора А $F_A = \frac{\sigma_4^2}{\sigma_1^2}$ (кількість ступенів свободи складає $p-1; N-pq$);
- для фактора В $F_B = \frac{\sigma_3^2}{\sigma_1^2}$ (кількість ступенів свободи складає $q-1; N-pq$);
- для взаємодії факторів А і В $F_{AB} = \frac{\sigma_2^2}{\sigma_1^2}$ (кількість ступенів свободи складає $(p-1)(q-1); N-pq$).

Таблиця 6.8

Форма таблиці двохфакторного дисперсійного аналізу

Джерело варіації	Сума квадратів відхилень	Кількість ступенів свободи	Оцінка дисперсії
Фактор А	$S_4^2 = nq \sum (\bar{X}_i - \bar{X})^2$	$p-1$	$\sigma_4^2 = \frac{S_4^2}{p-1}$
Фактор В	$S_3^2 = np \sum (\bar{X}_j - \bar{X})^2$	$q-1$	$\sigma_3^2 = \frac{S_3^2}{q-1}$
Взаємодія факторів А і В	$S_2^2 = n \sum (\bar{X}_{ij} - \bar{X}_i - \bar{X}_j - \bar{X})^2$	$(p-1)(q-1)$	$\sigma_2^2 = \frac{S_2^2}{(p-1)(q-1)}$
Залишкова варіація	$S_1^2 = \sum (X_{ijk} - \bar{X}_{ij})^2$	$N-pq$	$\sigma_1^2 = \frac{S_1^2}{N-pq}$
Усього	$S_0^2 = \sum (X_{ijk} - \bar{X})^2$	$N-1$	-

Це означає, що в двохфакторному дисперсійному аналізі використовуються три різних критерії:

- 1) для перевірки гіпотези про відсутність ефекту фактора А;
- 2) для перевірки гіпотези про відсутність ефекту фактора В;
- 3) для перевірки гіпотези про відсутність ефекту взаємодії факторів А і В.

Після визначення F -критеріїв Фішера перевірка гіпотез здійснюється за таким самим алгоритмом, як і в однофакторному дисперсійному аналізі.

Двохфакторний дисперсійний аналіз може мати два різновиди:

- 1) без повторень;
- 2) з повтореннями.

У першому випадку кожному рівню факторів відповідає тільки одна вибірка даних, у другому – певним рівням факторів може відповідати більше однієї вибірки даних.

Унаслідок складності розрахунків, особливо при великій кількості рівнів кожного фактора, для двохфакторного аналізу слід застосовувати або *Excel*, або спеціалізоване програмне забезпечення.

Для здійснення розрахунків в *Excel* необхідно активізувати «Пакет аналізу» і далі здійснити наступні кроки:

1. Обрати опцію «Двохфакторний дисперсійний аналіз без повторень (або з повтореннями)».

2. У діалоговому вікні з'являється адреси ознак, їх необхідно вказати (із попередньо складеного в *Excel* масиву вихідних даних).

3. Вказати кількість рядків для вибірки (враховуючи кількість ступенів свободи).

4. Встановити рівень надійності коефіцієнтів регресії (за замовчуванням – 95 %).

5. Командою ОК вивести результати на новий робочий аркуш.

Питання для самоконтролю

1. Як визначається та що характеризує загальна дисперсія?
2. Як визначається та що характеризує внутрішньогрупова дисперсія?
3. Як визначається та що характеризує міжгрупова дисперсія?
4. Як визначається та що характеризує емпіричне кореляційне відношення?
5. Які завдання вирішує однофакторний дисперсійний аналіз?
6. Які завдання вирішує двохфакторний дисперсійний аналіз?
7. Що таке нульова і альтернативна гіпотези?
8. Як здійснюється перевірка гіпотез в однофакторному дисперсійному аналізі?
9. Скільки нульових гіпотез перевіряється у двохфакторному дисперсійному аналізі?
10. Які фактори вважаються взаємодіючими?
11. Що таке «кількість ступенів свободи»?
12. Як визначається F -критерій Фішера і що він характеризує?
13. Які умови повинні виконуватися для того, щоб застосування F -критерію Фішера було коректним?
14. Чим відрізняється двохфакторний дисперсійний аналіз «без повторень» від його аналізу «з повтореннями»?

РОЗДІЛ 7. АНАЛІЗ ЗАКОНОМІРНОСТЕЙ ДЛЯ ЯКІСНИХ ДАНИХ

7.1. Особливості аналізу якісних даних. Сутність критерію «хі-квадрат»

Статистичні методи, які розглядалися у попередніх розділах даного підручника, призначені для аналізу кількісних ознак. Однак, багато ознак неможливо виміряти числом. Наприклад, можна бути чоловіком або жінкою, студентом або ні. Можна бути лікарем, юристом, геологом та ін. Тут ми маємо справу з якісними (атрибутивними) ознаками. Ці ознаки не пов'язані між собою жодними арифметичними співвідношеннями, впорядкувати їх не можна (будь-які арифметичні дії з якісними ознаками виконувати не можна, навіть якщо вони проранжовані). Єдиний спосіб описання якісних ознак полягає в тому, щоб підрахувати кількість об'єктів, що мають одне і те ж значення. Крім того, можна підрахувати, яка частка загальної кількості об'єктів має те чи інше значення.

Якісні (нечислові) змінні традиційно мають широке застосування в соціологічних опитуваннях, у ході яких збирають інформацію шляхом опитувань, проведення інтерв'ю тощо.

Останнім часом такі змінні набувають усе більшого розповсюдження і в економічних дослідженнях. Причиною тому, перш за все, є прагнення врахувати людський фактор у бізнесі. У всьому світі проводяться бізнес-опитування керівників підприємств, менеджерів щодо ситуації як на самому підприємстві, так і в навколишньому ринковому середовищі. Це можуть бути опитування, що здійснюються з метою порівняння нинішнього і попередніх (наприклад, минулих років) рівнів основних фінансово-виробничих показників, або їх можливих змін у майбутньому. Можуть проводитися також опитування щодо змін умов для здійснення бізнесу в регіоні, які безпосередньо впливають на обсяг продаж продукції підприємства. При цьому варіанти відповідей в анкетах зазвичай бувають такими: «ситуація не змінилася», «відбулося покращення ситуації», «відбулося погіршення ситуації», «складно відповісти».

Такий підхід до формулювання варіантів відповідей є особливо актуальним у сучасних умовах, коли суттєвий вплив на отримання інформації має так звана «комерційна таємниця». Так, під час

здійснення опитування менеджерів щодо обсягів розрахунків з певними постачальниками або покупцями, респонденти зазвичай не мають бажання озвучувати конкретні цифрові дані, а віддають перевагу відповідям, які лише окреслюють ті чи інші зміни ситуації: покращилася, погіршилася тощо.

Здійснюються бізнес-опитування й серед населення. Так, опитування щодо вибору переважного продукту із списку відомих дозволяє приймати управлінські рішення щодо випуску нових для підприємства видів продукції. Опитування населення щодо переваг подорожування (в межах країни, до країн Європи, Азії тощо) є підставою для прийняття управлінських рішень у фірмах, що займаються наданням туристичних послуг.

Крім того, постійний моніторинг змін, що відбуваються останнім часом на регіональних або національному ринках, призвів до необхідності розробки певних методик здійснення рейтингових оцінок різних суб'єктів господарювання: банківських установ, фірм, навчальних закладів тощо. Для цього бізнес-структурам присвоюються певні ранги. Очевидно, що ранги, в основу яких може бути покладена низка показників, не можна вважати звичайними кількісними змінними, оскільки сказати, наприклад, що другий ранг є величиною в два рази більшою (кращою), ніж перший ранг, не можна.

У деяких випадках разом із новими анкетами респонденти отримують результати їх попередніх опитувань. Така практика є особливо корисною в наступних випадках:

по-перше, для того, щоб з'ясувати чи змінилися погляди опитуваних щодо поставлених проблем;

по-друге, для підвищення ступеня відповідальності (уникнення формального підходу) до оцінки досліджуваної проблеми;

по-третє, для більш оперативного проведення дослідження, оскільки в сучасних умовах інформація зазвичай дуже швидко «старіє»;

по-четверте, для порівняння змін варіантів відповідей останнього опитування з попередніми, що відбулися внаслідок тих чи інших об'єктивних причин.

Оскільки якісні дані не мають числового вираження, то основними питаннями, які виникають при аналізі таких даних, є:

1) підрахунок частот або часток, тобто визначення кількості одиниць в кожній групі або визначення їх кількості у відсотках (частках) до загальної кількості одиниць досліджуваної сукупності;

2) знаходження способу (методу) оцінити рівень точності, з якою частоти, які визначені за даними вибіркової сукупності, відповідають частотам у всій генеральній сукупності;

3) перевірка необхідної гіпотези (чи то рівності відсотків, чи то визначення наявності або відсутності взаємозв'язку між досліджуваними ознаками) за допомогою тих чи інших статистичних критеріїв.

Для аналізу якісних даних, представлених не числами, а категоріями, використовують систему статистичних критеріїв, серед яких особливе місце займає критерій «хі-квадрат» (χ^2).

Критерій «хі-квадрат» був запропонований Карлом Пірсоном ще в 1900 році. За допомогою даного критерію здійснюється оцінка значущості відмінностей між фактичною кількістю випадків або якісних характеристик (отриманих у результаті вибіркового спостереження), що потрапили в кожну категорію, і теоретичною їх кількістю, яку можна очікувати в отриманих групах.

Іншими словами, **критерій (тест) «хі-квадрат»** заснований на частотах, які являють собою кількість одиниць вибіркової сукупності, що потрапили в кожну з категорій якісної ознаки. Завданням критерію є встановлення суттєвості (або не суттєвості) відмінностей між фактичними частотами і очікуваними, опорними частотами (у випадку справедливості нульової гіпотези) [16, с. 879].

Розраховується **критерій «хі-квадрат»** за наступною формулою:

$$\chi^2 = \sum \frac{(m - m')^2}{m'}, \quad (7.1)$$

де m – фактичні частоти;

m' – очікувані (опорні) частоти.

Необхідно звернути увагу на те, що під час розрахунку критерію підсумовування здійснюється по всіх категоріях або комбінаціях категорій.

Методика визначення очікуваних (опорних) частот залежить від того, яка саме гіпотеза в даному випадку перевіряється.

Наприклад, потрібно порівняти думку покупців про продукцію двох меблевих фірм і оцінити, чи дійсно існують відмінності між оцінками, які надаються продукції першої фірми, та оцінками, які відносяться до продукції другої фірми. Самі оцінки меблів повинні бути представлені у вигляді якісних характеристик, тобто повинні

виражатися позитивними і негативними епітетами: комфортна, красива, міцна чи незручна, груба, ненадійна та ін.

Для того, щоб зробити висновок про те, що продукція першої фірми більше подобається покупцям, ніж продукція другої фірми, необхідно довести, що позитивні оцінки зустрічаються статистично значно частіше в першій групі відгуків (про продукцію першої фірми), ніж у другій групі відгуків (про продукцію другої фірми). Здійснити порівняння частот тих чи інших якісних оцінок дозволяє виконати саме критерій «хі-квадрат».

Цей критерій цілком підходить і для порівняння того, наскільки отримана частота появи об'єктів збігається з очікуваною частотою їх появи. Таке порівняння необхідно здійснювати, наприклад, для того, щоб відповісти на питання про те, наскільки виправданою є якість прогнозів, зокрема, прогнозів погодних явищ: дощу, снігу, граду, урагану тощо. Але якщо групу очікуваних подій (прогнозів) прийняти в якості результатів спостережень у якихось уже реалізованих умовах, припустимо, в іншій країні, то це дозволить порівнювати частоти природних явищ двох країн, а не оцінювати точність прогнозів – у цьому полягає гнучкість використання критерію «хі-квадрат».

7.2. Використання критерію «хі-квадрат» щодо подібності розподілів

Критерій «хі-квадрат» щодо подібності розподілів (рівності відсотків) використовують для того, щоб відповісти на питання щодо того, чи з однаковою частотою зустрічаються значення ознаки в емпіричному і теоретичному розподілах або в двох і більше емпіричних розподілах (порівняння розподілів двох якісних ознак). Тобто, за допомогою критерію оцінюють, наскільки подібними є розподіли емпіричних і теоретичних частот. Якщо відмінності між ними незначні, то можна вважати, що відхилення емпіричних частот від теоретичних обумовлені лише випадковістю. Якщо ж, навпаки, ці розподіли виявляться достатньо різними, то можна вважати, що відмінності між ними значущі.

Таким чином, критерій «хі-квадрат» використовують для перевірки гіпотези про те, що не існують відмінності між фактичними відсотками, які побудовані за даними, взятими із деякої генеральної сукупності, і очікуваними (опорними) величинами.

Нагадаємо, що сутність і види статистичних гіпотез детально розглядаються у Розділі 4 підручника.

У даному випадку *гіпотези* формулюються так:

– нульова гіпотеза (H_0): значення відсотків у досліджуваній сукупності відповідають (рівні) набору відомих, заданих опорних величин;

– альтернативна (дослідницька) гіпотеза (H_1): значення відсотків у досліджуваній сукупності не рівні заданому набору опорних величин, щонайменше для однієї з категорій існує відмінність.

Критерій «хі-квадрат» можна застосовувати лише у випадку виконання двох умов:

1. Отримані дані є випадковою вибіркою з досліджуваної генеральної сукупності.

2. Загальна кількість спостережень (обсяг вибірки) повинна бути не меншою 50, а частоти в кожній групі повинні бути не меншими п'яти.

У формулі (7.1), яка використовується для розрахунку критерію «хі-квадрат», фігурують так звані «очікувані (опорні) частоти». Ми вже відмічали, що спосіб визначення цих частот залежить від сутності гіпотези, яка перевіряється. У даному випадку *очікуване значення частоти* для кожної категорії розраховується так:

$$m' = z \cdot n, \quad (7.2)$$

де z – задані, опорні значення *часток* в генеральній сукупності;
 n – обсяг (кількість) вибірки.

Після розрахунку критерію його значення порівнюють із критичним (табличним) значенням. При цьому зазвичай перевірка здійснюється з 95%-м рівнем довірчої ймовірності, тобто рівень надійності (значущості) складає 5 % (100 – 95).

Оскільки під час визначення критерію «хі-квадрат» здійснюється підсумовування співвідношень (формула 7.1), то величина критерію залежить від кількості доданків: за інших рівних умов чим більшою є їх кількість, тим більшим буде значення самого критерію, оскільки кожен із доданків вносить свій вклад у загальну суму. Із цього випливає, що для кожної кількості незалежних змінних критерій «хі-квадрат» буде мати власний розподіл.

Слід визначитися з тим, що таке кількість незалежних змінних. Вона завжди на одиницю менша, ніж кількість категорій номінальної ознаки. Чому це так? Тому що коли ми маємо певну вибірку

сукупність, для якої підраховано суму частот, то одну із частот ми можемо завжди визначити шляхом віднімання від загальної їх кількості решти інших частот. Це інакше називають кількістю ступенів свободи.

Отже, в даному випадку *кількість ступенів свободи* визначається наступним чином:

$$k = l - 1, \quad (7.3)$$

де l – кількість категорій (варіантів) досліджуваної ознаки.

Якщо значення «хі-квадрат» більше критичного значення з таблиці розподілу «хі-квадрат» для відповідної кількості ступенів свободи, то це є доказом того, що отримані частоти суттєво відрізняються від очікуваних частот, виходячи з опорних значень відсотків. У такому випадку слід відхилити нульову гіпотезу і прийняти досліджувану (альтернативну) гіпотезу, зробивши висновок про те, що отримані вибіркові відсотки значно (суттєво) відрізняються від заданих опорних значень.

Якщо значення критерію «хі-квадрат» менше критичного значення з таблиці розподілу «хі-квадрат», то отримані частоти не значно відрізняються від частот, які можна було б очікувати виходячи з опорних значень відсотків. У такому випадку слід прийняти нульову гіпотезу як допустиму і зробити висновок, що отримані вибіркові відсотки не мають значних відмінностей від заданих опорних значень.

Розглянемо розрахунок та перевірку критерію «хі-квадрат» на такому прикладі.

У відділі чоловічого взуття магазину протягом тижня проводилося наступне спостереження. Для кожного потенційного клієнта, що заходив до відділу магазину, фіксувався бажаний ним тип «літнього» взуття. У табл. 7.1 наведені дані за минулий тиждень, а також відповідні відсотки в цей же час у минулому році.

Необхідно:

1) оцінити зміну попиту на окремі види взуття шляхом порівняння відсотків;

2) розглядаючи відсотки за минулий рік як точні, визначити і оцінити критерій «хі-квадрат»;

3) надати адміністрації магазину поради щодо необхідності прийняття управлінських рішень.

Таблиця 7.1

**Результати опитування клієнтів за останній тиждень
і дані за минулий рік**

Тип взуття	Кількість клієнтів	
	частоти за останній тиждень, одиниць	дані за минулий рік, %
Кросівки	124	16,2
Черевики типу «босоніжки»	256	26,4
Шкіряні черевики	302	38,5
Черевики із заміників шкіри	138	18,9
Усього	820	100,0

Для відповіді на перше запитання потрібно визначити частки за останній тиждень (результати наведено в табл. 7.2).

Таблиця 7.2

Частки за останній тиждень та минулий рік

Тип взуття	Кількість клієнтів, у % до загальної кількості	
	дані за останній тиждень	дані за минулий рік
Кросівки	15,1	16,2
Черевики типу «босоніжки»	31,6	26,4
Шкіряні черевики	36,5	38,5
Черевики із заміників шкіри	16,8	18,9
Усього	100,0	100,0

Як видно із таблиці 7.2, відбулися певні зміни в структурі попиту на взуття: зросла частка тих, хто надає перевагу черевикам типу «босоніжки», у той же час зменшилася частка тих, хто бажає купувати шкіряні черевики, інші зміни були незначними. Але чи є суттєвими всі ці зміни, для того щоб пропонувати адміністрації змінювати структуру закупівель? На це питання краще відповісти шляхом перевірки тесту «хі-квадрат».

Для цього спочатку необхідно знайти очікувані частоти (формула 7.2). Результати розрахунків наведено в табл. 7.3.

Слід звернути увагу на те, що сума очікуваних частот повинна дорівнювати сумі фактичних частот (у даному випадку – 820).

Маючи очікувані частоти можемо визначити критерій «хі-квадрат» за формулою 7.1:

$$\chi^2 = \frac{(124 - 132,8)^2}{132,8} + \frac{(256 - 216,5)^2}{216,5} + \frac{(302 - 315,7)^2}{315,7} + \frac{(138 - 155,0)^2}{155,0} = 10,2$$

Таблиця 7.3

Розрахунок очікуваних частот за минулий рік

Тип взуття	Очікувані частоти
Кросівки	820 * 0,151 = 132,8
Черевики типу «босоніжки»	820 * 0,316 = 216,5
Шкіряні черевики	820 * 0,365 = 315,7
Черевики із заміників шкіри	820 * 0,168 = 155,0
Усього	820,0

Для того, щоб порівняти отримане значення критерію «хі-квадрат» із табличним, необхідно знайти кількість ступенів свободи за формулою 7.3:

$$k = 4 - 1 = 3$$

Використовуючи таблиці розподілу «хі-квадрат» (які можна легко знайти в Інтернеті) визначаємо, що для 3-х ступенів свободи і рівня значущості 5 % його значення складає 7,8.

Оскільки фактичне значення статистики "хі-квадрат" значно більше табличного, то нульова гіпотеза відхиляється (отримані частки істотно відрізняються від опорних значень відсотків). Тому можна пропонувати адміністрації магазину внести зміни в структуру продажів.

7.3. Використання статистичних критеріїв для оцінки тісноти зв'язку між якісними змінними

При наявності двовимірного набору якісних даних (коли є інформація по 2-х і більше змінних для кожного об'єкта) можна перевірити наявність зв'язку між змінними. Вважають, що дві якісні змінні є **незалежними**, якщо знання значення однієї з них не допомагає передбачити значення іншої. Іншими словами, ймовірності однієї змінної такі ж, як умовні ймовірності при заданих опорних значеннях іншої змінної.

Наприклад, уявімо, що в генеральній сукупності відсоток браку при фрезеруванні деталей у бригаді механічного цеху складає 2 %. Однак, коли працює бригада Іванова, відсоток браку становить 7 %. У цьому випадку знання значення однієї змінної (прізвище конкретного

бригадира) допомагає спрогнозувати значення іншої змінної (обсяг браку), оскільки 2 % і 7 % суттєво відрізняються між собою. Поява браку є більш імовірною під час роботи бригади Іванова і менш імовірною, коли працює будь-яка інша бригада. Отже, ці дві змінні не є незалежними.

Вимірювання тісноти зв'язку між якісними змінними засновано на побудові так званих частотних таблиць, які інакше називають таблицями спряженості.

Таблиця спряженості дещо відрізняється від звичайної таблиці розподілу, яка містить інформацію про категорії (значення) ознаки і частоти (або частки) їх виникнення. Таблиця спряженості – це таблиця спільного розподілу частот для двох категоріальних змінних, вона містить вибіркві оцінки ймовірнісних розподілів двох випадкових величин.

Слід відмітити, що таблиця спряженості є найбільш універсальним засобом вивчення статистичних зв'язків, оскільки в ній можуть бути представлені змінні з будь-яким способом вимірювання. Сам термін «таблиця спряженості» зобов'язаний своїм походженням тій обставині, що на основі аналізу подібних таблиць можна робити висновки щодо пов'язаності (спільної появи) значень одних категоріальних змінних разом з деякими значеннями інших змінних. Зв'язок між номінальними змінними, власне кажучи, і виражається у вигляді подібних сполучень.

При побудові таблиці спряженості рядкам відповідають категорії однієї змінної, а стовпчикам – іншої змінної. При цьому бажано, щоб групи явищ, які виділяються як причинний фактор, розміщувалися у лівій частині таблиці (рядки); а явища, які трактуються як наслідок дії фактору, – у верхній частині таблиці (стовпчики).

Зовні таблиця спряженості має такий вигляд (табл. 7.4).

Таблиця 7.4

Таблиця спряженості

Градації за першою ознакою	Градації за другою ознакою				Σ
	1	2	...	m	
1	a_{11}	a_{12}	...	a_{1m}	
2	a_{21}	a_{22}		a_{2m}	
...	
n	a_{n1}	a_{n2}	...	a_{nm}	
Σ					

Останнім часом у економічній літературі все частіше використовується розширене розуміння таблиці спряженості. Пропонується в якості її елементів наводити не лише частоти, але й інші показники: наприклад, у клітинках статево-вікової таблиці розподілу населення можна наводити середній рівень доходів людей, які мають дану стать і вік, що суттєво розширює можливості проведення аналізу.

За таблицями спряженості зазвичай обчислюються коефіцієнти взаємної спряженості, засновані на тестовій статистиці «хі-квадрат». У тому випадку, коли не можна виконати умови застосування статистики хі-квадрат, рекомендується користуватися так званими теоретико-інформаційними (експертними) способами вимірювання зв'язків.

Кореляцію між порядковими змінними можна вимірювати за допомогою коефіцієнтів рангової кореляції, широко використовуваними серед яких є коефіцієнти рангової кореляції Спірмена (ρ) і Кендела (τ).

Ці коефіцієнти можуть бути використані для визначення щільності зв'язку як між кількісними, так і між якісними ознаками за умови, якщо їх значення будуть певним чином упорядковані або ранжовані.

Коефіцієнт кореляції рангів (коефіцієнт Спірмена) розраховується за такою формулою (у випадку відсутності пов'язних рангів):

$$\rho = 1 - \frac{6 \sum d_i^2}{n(n^2 - 1)}, \quad (7.4)$$

де d_i^2 - квадрат різниці рангів;

n – кількість рангів.

Коефіцієнт Спірмена може приймати будь-які значення в інтервалі $[-1; 1]$. Зв'язок між ознаками можна визнати статистично істотним, якщо значення коефіцієнту перевищує 0,5.

Розглянемо методику розрахунку коефіцієнта рангової кореляції Спірмена на наступному прикладі. Серед покупців молочних продуктів і фахівців з їх якості було здійснено опитування щодо оцінки рівня якості продукції шести підприємств виробників вершкового масла.

Результати опитування наведено в табл. 7.5.

До розрахунку коефіцієнта рангової кореляції Спірмена

Підприємства-виробники	Ранг продукції підприємства, який виставлено		Квадрат різниці рангів d^2
	покупцями	фахівцями	
А	2	3	1
Б	3	2	1
В	1	1	0
Г	5	6	1
Д	6	4	4
Ж	4	5	1
Усього	-	-	8

Визначаємо Критерій Спірмена:

$$\rho = 1 - \frac{6 \cdot 8}{6(6^2 - 1)} = 0,8$$

Результати розрахунків свідчать про те, що між оцінками, які виставили фахівці та покупці продукції підприємств, є достатньо тісний зв'язок.

Коефіцієнта рангової кореляції Кендала розраховується за наступною формулою:

$$\tau = \frac{2S}{n(n-1)}, \quad (7.5)$$

де n – кількість рангів;

S – сума балів, яка знаходиться як різниця двох складових:

$$S = \sum S_1 - \sum S_2,$$

де S_1 – кількість рангів, які перевищують номер рангу, що відповідає результативній ознаці;

S_2 – кількість рангів, які менші даного рангу результативної ознаки.

Аналогічно коефіцієнту Спірмена, зв'язок між ознаками можна визнати істотним, якщо значення коефіцієнту рангової кореляції Кендала перевищує 0,5.

Визначимо коефіцієнт рангової кореляції Кендала за даними попереднього прикладу.

Для цього спочатку упорядкуємо підприємства за рангами, які їм виставили покупці (табл. 7.6).

До розрахунку коефіцієнта рангової кореляції Кендала

Підприємства-виробники	Ранг продукції підприємства, який виставлено		S_1	S_2
	покупцями	фахівцями		
В	1	1	5	0
А	2	3	3	1
Б	3	2	3	0
Ж	4	5	1	1
Г	5	6	0	1
Д	6	4	0	0
Усього	-	-	12	3

Визначаємо суму різниці рангів (при цьому S розраховується за рангами, які виставлені фахівцями, оскільки ми впорядкували підприємства за рангами покупців):

$$S = 12 - 3 = 9$$

Розраховуємо коефіцієнт Кендала:

$$\tau = \frac{2 \cdot 9}{6(6-1)} = 0,6$$

Як бачимо, коефіцієнт Кендала дав більш сувору оцінку тісноти зв'язку між рівнями якості продукції підприємств, які їм виставили покупці та фахівці (коефіцієнт Спірмена склав 0,8).

Для визначення щільності зв'язку між довільною кількістю ранжованих ознак використовується множинний коефіцієнт рангової кореляції – *коефіцієнт конкордації* (W), який визначається за формулою:

$$W = \frac{12S}{m^2 \cdot (n^3 - n)}, \quad (7.6)$$

де m – кількість факторів;

n – кількість спостережень;

S – відхилення суми квадратів рангів від середньої квадратів рангів.

За наявності пов'язаних рангів формула коефіцієнта конкордації набуває наступного вигляду:

$$W = \frac{12S}{m^2 \cdot (n^3 - n) - m \cdot \sum_{j=1}^m T_j}, \quad (7.7)$$

де $T_j = 1 / 12 \sum_{j=1}^m (t_j^3 - t_j)$;

t – кількість пов'язаних рангів за окремими ознаками.

Коефіцієнт конкордації може набувати будь-яких значень в інтервалі $[-1; 1]$.

Для оцінювання тісноти зв'язку між номінальними ознаками (які неможливо певним чином ранжувати) використовують **критерій «хі-квадрат»**. Він дає можливість відповісти на питання, чи є дві змінні незалежними, чи ні, виходячи з таблиці спостережуваних частот, складеної для двовимірного набору якісних даних (таблиці спряженості).

Іншими словами, даний критерій перевіряє нульову гіпотезу щодо того, існує чи ні зв'язок між ознаками, які складають таблицю спряженості.

При цьому нульова і альтернативна **гіпотези** формулюються наступним чином:

– нульова гіпотеза (H_0): дві досліджувані змінні не залежать одна від одної;

– альтернативна (дослідницька) гіпотеза (H_1): дві змінні пов'язані; іншими словами, вони не є незалежними одна від іншої.

Очікувані частоти в даному випадку визначають так: для кожної комбінації категорій частоту однієї категорії множать на частоту іншої категорії і отриманий результат ділять на загальну кількість спостережень (обсяг вибірки). Це можна записати у вигляді формули:

$$m' = \frac{\left(\begin{array}{c} \text{Частота категорії} \\ \text{однієї змінної} \end{array} \right) \cdot \left(\begin{array}{c} \text{Частота категорії} \\ \text{другої змінної} \end{array} \right)}{n} \quad (7.8)$$

де n – кількість спостережень, тобто загальний обсяг вибірки.

Чому саме так обчислюють очікувані частоти? Згадаймо, що відповідно до теорії ймовірності для двох незалежних подій ймовірність того, що обидві вони настануть, дорівнює добутку ймовірностей цих подій.

При цьому повинні виконуватися наступні умови:

1. Набір даних повинен бути випадковою вибіркою, що отримана з досліджуваної генеральної сукупності.

2. У кожній клітинці таблиці спряженості очікувана частота повинна бути не меншою п'яти.

Кількість ступенів свободи в даному випадку визначають як добуток частот кожної з категорій за мінусом одиниці:

$$k = (l_1 - 1) \cdot (l_2 - 1), \quad (7.9)$$

де l_1 і l_2 – відповідно частота першої та другої категорій.

Оскільки використання **критерію «хі-квадрат»** для перевірки незалежності передбачає сумування для кожної з двох категорій, то формула 7.1 набуває вигляду:

$$\chi^2 = \sum \sum \frac{(m - m')^2}{m'} \quad (7.10)$$

Якщо отримане значення критерію «хі-квадрат» перевищує критичне його значення з таблиці розподілу «хі-квадрат», то це означає, що спостережувані частоти значно відрізняються від тих, що мали б місце, якби досліджувані змінні були незалежними. У такому випадку слід відхилити нульову гіпотезу про незалежність змінних і прийняти досліджувану (альтернативну) гіпотезу, зробивши висновок, що між змінними є істотний зв'язок.

Якщо значення критерію «хі-квадрат» менше табличного (критичного) значення, то це є доказом того, що спостережувані дані не значно відрізняються від тих, яких можна було б очікувати, якби змінні в генеральній сукупності були незалежними. У цьому випадку слід прийняти нульову гіпотезу про незалежність змінних (як прийнятну можливість) і зробити висновок, що між змінними відсутній суттєвий зв'язок [16, с. 885].

Розглянемо розрахунок критерію «хі-квадрат» та перевірку гіпотези щодо незалежності змінних на наступному прикладі.

Фірма розглядає питання щодо розширення своєї діяльності в регіоні. Для цього здійснено опитування працівників фірми. Під час опитування працівникам було запропоноване таке запитання: "Як ви вважаєте, умови для бізнесу у вашому регіоні стали кращими, погіршилися або залишилися незмінними?" Усього було опитано 470 працівників. Результати опитування наведено в таблиці 7.5.

Необхідно з'ясувати:

1) чи залежать відповіді від категорії працівників, тобто, чи є залежність між категорією опитуваних і варіантом відповіді, яку вони надавали?

2) чи є підґрунтя для розширення бізнесу в регіоні?

Таблиця 7.5

Результати опитування щодо перспектив розширення бізнесу

Варіанти відповіді	Кількість опитаних, осіб	
	менеджери	інші службовці
Умови покращилися	23	92
Умови залишилися незмінними	37	168
Умови погіршилися	11	81
Не знаю	15	43
Усього	86	384

Для відповіді на поставлені запитання необхідно визначити критерій «хі-квадрат» і здійснити перевірку гіпотези щодо незалежності змінних.

На першому етапі визначаємо очікувані частоти за формулою 7.8. Результати розрахунків наведено в табл. 7.6.

Таблиця 7.6

Розрахунок очікуваних частот

Варіанти відповіді	Очікувані частоти		Разом
	менеджери	інші службовці	
Умови покращилися	$86 \cdot 115 / 470 = 21,0$	$384 \cdot 115 / 470 = 94,0$	115
Умови залишилися незмінними	$86 \cdot 205 / 470 = 37,5$	$384 \cdot 205 / 470 = 167,5$	205
Умови погіршилися	$86 \cdot 92 / 470 = 16,8$	$384 \cdot 92 / 470 = 75,2$	92
Не знаю	$86 \cdot 58 / 470 = 10,6$	$384 \cdot 58 / 470 = 47,4$	58
Усього	85,9	384,1	470

Необхідно звернути увагу на те, що сума очікуваних частот повинна дорівнювати сумі фактичних частот (у даному випадку – 470).

Для порівняння критерію «хі-квадрат» з його критичним (табличним) значенням необхідно визначити кількість ступенів свободи.

Для цього скористаємося формулою 7.9:

$$k = (4 - 1) \cdot (2 - 1) = 3$$

Отже, кількість ступенів свободи складає 3, оскільки у нас чотири категорії відповідей і дві категорії працівників.

Визначимо критерій «хі-квадрат» за формулою 7.10:

$$\chi^2 = \frac{(23 - 21)^2}{21} + \frac{(37 - 37,5)^2}{37,5} + \frac{(11 - 16,8)^2}{16,8} + \frac{(15 - 10,6)^2}{10,6} + \frac{(92 - 94)^2}{94} +$$

$$+ \frac{(168 - 167,5)^2}{167,5} + \frac{(81 - 75,2)^2}{75,2} + \frac{(43 - 47,4)^2}{47,4} = 0,19 + 0,01 + 2,00 + 1,83 +$$

$$+ 0,04 + 0,001 + 0,45 = 4,52$$

За таблицями розподілу «хі-квадрат» для 3-х ступенів свободи і рівня значущості 5 % знаходимо критичне значення, яке дорівнює 7,815.

Отримане значення «хі-квадрат» (4,52) виявилось меншим, ніж його критичне значення (7,815), тому можна зробити висновок, що між двома якісними змінними (варіант відповіді та категорія працівника) відсутній значущий зв'язок.

Оскільки обидві категорії працівників відмітили або покращення умов для бізнесу, або їх незмінний рівень, то можна розглядати питання щодо розширення бізнесу в регіоні.

Як ми відмічали раніше, однією з головних умов коректного застосування критерію «хі-квадрат» є те, що в кожній клітинці таблиці спряженості очікувана частота повинна бути не меншою п'яти, інакше у результаті розрахунку критерію можна отримати хибні висновки.

Крім того, значення критерію «хі-квадрат», що отримують у результаті розрахунків, ґрунтуються на частотах, які є цілими числами. У той же час табличні значення для його розподілу складені для безперервної шкали. Тому англійський статистик Френк Йейтс (1902-1994) запропонував поправку до даного критерію, яку зараз називають «поправкою Йейтса».

Сутність поправки полягає у відніманні числа 0,5 з різниці між фактичною і очікуваною кількістю спостережень. Тобто, формула **критерію «хі-квадрат» з «поправкою Йейтса»** набуває такого вигляду:

$$\chi^2 = \sum \sum \frac{(|m - m'| - 0,5)^2}{m'} \quad (7.11)$$

Дану формулу доцільно застосовувати в тому випадку, коли кількість частот в окремих клітинках таблиці спряженості менша п'яти.

Слід відмітити, що «поправка Йейтса» викликає багато суперечок між фахівцями, оскільки розрахований таким чином критерій «хі-квадрат» може давати дещо занижені результати. А це означає, що збільшується помилка ймовірності не знайти розходжень там, де вони є.

Іншим виходом із ситуації є визначення коефіцієнтів спряженості, які засновані на критерії «хі-квадрат». Серед них найбільш поширеними є коефіцієнти взаємної спряженості К. Пірсона (C) та А. Чупрова (K). Вважається, що ці критерії дають більш точні результати для невеликої кількості частот.

Формула **критерію Пірсона** має вигляд:

$$C = \sqrt{\frac{\chi^2}{\chi^2 + n}}, \quad (7.12)$$

де n – обсяг вибірки.

Даний критерій не враховує кількість категорій для кожної з ознак. Більш точним є **коефіцієнт спряженості Чупрова**, який розраховується за формулою:

$$K = \sqrt{\frac{\chi^2}{\sqrt{(r-1)(c-1)}}}, \quad (7.13)$$

де n – обсяг вибірки

r – кількість рядків у таблиці спряженості,

c – кількість стовпчиків у таблиці спряженості.

Модифікацією цього коефіцієнта є **коефіцієнт взаємної спряженості Г. Крамера**, в знаменнику формули розрахунку якого береться мінімальне з двох значень: кількість рядків за мінусом одиниці, або кількість стовпчиків за мінусом одиниці:

$$K = \sqrt{\frac{\chi^2 \div n}{\min\{r-1, c-1\}}} \quad (7.14)$$

Усі коефіцієнти спряженості можуть набувати значень від 0 до 1. Якщо значення коефіцієнту наближається до 0, то це свідчить про відсутність зв'язку між досліджуваними ознаками, а якщо до 1, то між ними існує тісний взаємозв'язок.

Питання для самоконтролю

1. У чому полягає відмінність між якісними та кількісними ознаками?
2. Що таке номінальна ознака?
3. У чому сутність критерію «хі-квадрат» і як він розраховується?

4. Що розуміють під опорними частотами?
5. Як визначаються опорні частоти під час перевірки гіпотези на подібність розподілів?
6. Що таке нульова і альтернативна гіпотези?
7. Які умови повинні виконуватися для того, щоб застосування критерію «хі-квадрат» було коректним?
8. Що таке кількість ступенів свободи?
9. Як розраховуються опорні частоти під час перевірки гіпотези на незалежність змінних?
10. У чому відмінність між таблицею спряженості та звичайної таблиці?
11. Як розраховується коефіцієнт кореляції рангів Спірмена?
12. Як розраховується коефіцієнт кореляції рангів Кендала?
13. При яких значеннях коефіцієнтів Спірмена та Кендала зв'язок між ознаками можна визнати істотним?
14. Як розраховується коефіцієнт конкордації?
15. Яких значень може набувати коефіцієнт конкордації?
16. Як визначаються опорні частоти під час перевірки гіпотези на незалежність змінних?
17. Що таке «поправка Йейтса»? Коли вона необхідна?
18. Як розраховується критерій Пірсона?
19. Які недоліки має критерій Пірсона?
20. Як розраховується критерій Чупрова?
21. Як розраховується критерій Крамера?
22. Яких значень можуть набувати критерії спряженості?

РОЗДІЛ 8. СТАТИСТИЧНІ МЕТОДИ КОНТРОЛЮ ЯКОСТІ

8.1. Сутність статистичного контролю якості. Статистичний контроль виробничого процесу

Сьогодні висока якість продукції розглядається як одна з важливих умов розвитку економіки, від якої залежать темпи промислового зростання країни, ефективність використання трудових ресурсів, успіхи зовнішньої торгівлі та її національний престиж. Необхідною умовою досягнення успіху підприємства щодо виробництва якісної, а відтак конкурентоспроможної продукції, є створення та налагодження виробничої системи на всіх його рівнях.

Матеріали, сировина, комплектуючі вироби (деталі) стають частиною продукції, що виробляється. Якість технологічного обладнання, приладів, засобів контролю тощо також безпосередньо впливає на якість виробленої продукції. Важливим завданням підприємства є забезпечення виробництва надходженням якісних матеріалів із зовнішніх джерел (від субпідрядників). Проте найважливішим завданням будь-якого підприємства є налагодження системи виробничих процесів, адже саме щодо цього сформовано цілу науку про найбільш економічні способи перероблення сировинних матеріалів та отримання якісної готової продукції.

Якщо **процес** – це сукупність взаємозв'язаних ресурсів і діяльності, який перетворює вхідні елементи на вихідні, то **виробничий процес** – систематична та цілеспрямована зміна в часі і просторі кількісних і якісних характеристик засобів виробництва і робочої сили для отримання готової продукції з вихідної сировини згідно із заданою програмою.

Ефективне управління якістю, виробничими процесами неможливе без використання статистичних методів, здатних своєчасно, оперативно та об'єктивно відображати зміни в процесі. Згідно даних оцінки фахівців, статистичні методи використовують як основний аналітичний інструмент у 70 % виробничих процесів з контролю якості та на всіх рівнях виробництва.

Метод статистичного контролю якості з'явився у 30-х роках минулого століття, що було пов'язано із розповсюдженням промислового застосування контрольних карт. Батьківщиною цього статистичного контролю вважають Великобританію, про що свідчить

прийняття у 1935 р. стандартів серії 600, які засновані на статистичному аналізі Е. Пірсона.

На той час у промисловості статистичні методи контролю якості застосовувалися вже не перший десяток років. Однак не можна сказати, що ці методи були невід'ємною, органічною частиною робіт із поліпшення якості продукції або повсякденним інструментом виявлення причин помилок при виготовленні продукції. Причиною цього був надто складний апарат контролю, який був орієнтований тільки на інженерно-технічних співробітників. Мова статистичних методів була тільки математичною та важкозрозумілою для простих людей, що створило уявлення про статистичні методи як про дещо дуже складне та навіть застаріле.

Кардинальні зміни наступили в 1950-х роках, коли японські вчені зрозуміли, що їм неефективно використовувати існуючі американські, британські чи інші поширені методи, їм необхідно розробити та використовувати власні. Відомий японський фахівець із дослідження якості К. Ішікава вважав, що вчені пропонували людям надто складні методи контролю тоді, коли б вистачило і простих [5, 21]. Саме тому японці адаптували існуючі прості методи статистичного контролю до реальних умов функціонування підприємств.

Так, у 1950-х роках в Японії були засновані дві премії Демінга: для підприємства та окремого його співробітника. Ці премії присуджувалися одній особі або групі осіб, які допомагали розповсюджувати статистичні методи контролю якості та розвивали їх. Ці премії є однією з найвищих нагород у країні, що відносяться до статистичних методів управління якістю.

Японська практика свідчить, що широке застосування автоматизованих виробничих процесів і використання робототехніки майже неможливе без застосування статистичних методів контролю. Ці методи є актуальними і нині, вони є вкрай необхідними на виробництві сьогодення.

Результат впровадження статистичних методів контролю якості був не миттєвим, але дуже помітним і важливим. Вже наприкінці 1970-х років Японія стала світовим лідером із масового виробництва таких товарів як кіно-, фото-, копії-техніка, автомобілі та телевізори. І сьогодні їх товари впевнено існують на американських та західноєвропейських ринках.

На даний час за ступенем складності статистичні методи контролю якості можна поділити на такі три категорії:

I. Елементарні статистичні методи (сім японських інструментів контролю якості).

II. Проміжні статистичні методи.

III. Передові статистичні методи.

Детально розглянемо першу категорію, оскільки саме 7 японських методів контролю якості є найбільш поширеними та зручними, а без їх знання використання складніших методів виявляється неможливим.

До цих методів належать:

1) контрольні листки;

2) діаграма Парето (*Pareto Diagram*);

3) метод розшарування (*Stratification*);

4) гістограма (*Histogram*);

5) діаграма розкиду (*Scatter Diagram*);

6) причинно-наслідкова діаграма Ісікави (*Cause and Effect Diagram*);

7) контрольна карта (*Control Chart*).

У своїй сукупності ці методи утворюють ефективну систему методів контролю та аналізу якості. Вони можуть застосовуватися у будь-якій послідовності, на будь-якому циклі виробництва, в будь-якому поєднанні, в різних аналітичних ситуаціях, їх можна розглядати і як цілісну систему, і як окремі інструменти аналізу. За їх допомогою аналізують кількісні критерії якості, використовуючи прийоми математичної статистики, які доступні всім учасникам виробничого процесу.

8.2. Контрольні листки

Контрольні листки – інструмент для збирання даних та їх автоматичного впорядкування для полегшення подальшого використання зібраної інформації. Тільки після збору вихідних даних можливе застосування будь-якого з інших шести інструментів контролю якості.

Контрольний листок являє собою паперовий бланк, на якому заздалегідь друкують параметри виробів, відповідно до яких можна заносити контрольовані дані за допомогою простих позначок чи символів. Під час складання контрольних листків необхідно звернути

увагу на те, щоб було зазначено, хто, на якому етапі процесу та протягом якого часу збирав дані, а також щоб форма листка була простою та зрозумілою без додаткових інструкцій.

У табл. 8.1 наведено приклад контрольного листка якості для умовних деталей підприємства «А» і «Б».

Таблиця 8.1

Листок контролю якості для деталей

Назва документа		Контрольний листок за видами дефектів
Підприємство _____	Виріб _____	Кількість оглянутих деталей _____
Цех _____	Операція _____	
Ділянка _____	Контролер _____	
Відмічати I II III IIII		
Деталь «А»		
Тип дефекту	Дані контролю	Частота
Деформації	III IIIII II	14
Подряпини	IIIIIIIIIIIIIIIIII	22
Тріщини	III	3
П'ятна	III III III	11
Розрив	III II	6
Інше	II	2
Деталь «Б»		
Тип дефекту	Дані контролю	Усього
Деформації	III III III	11
Подряпини	III IIIII III	15
Тріщини	III III	7
П'ятна	III IIIII	12
Розрив	III	3
Інше	III III	8

На базі зібраних за допомогою контрольного листка даних можна скласти зведену таблицю дефектів виробництва (табл. 8.2).

Дані таблиці свідчать, що частіше усього з усіх дефектів зустрічаються подряпини.

Сумарна кількість дефектів деталей

Тип дефектів всіх деталей	Кількість дефектів	
	одиниць	% до загальної кількості
Деформації	25	21,9
Подряпини	37	32,5
Тріщини	10	8,8
П'ятна	23	20,2
Розрив	9	7,8
Інше	10	8,8
Усього	114	100,0

На виробництві застосовують такі *види контрольних листків*:

- контрольний листок для реєстрації розподілу вимірюваного параметра в ході виробничого процесу;
- контрольний листок для реєстрації видів невідповідностей;
- контрольний листок для оцінки відтворюваності та працездатності технологічного процесу.

Контрольні листки можна застосовувати під час дослідження як атрибутивних, так і кількісних ознак якості.

8.3. Діаграма Парето

Наступним методом оцінки якості продукції на підприємстві, який є найбільш розповсюдженим, є **діаграма Парето**. Цей метод контролю має довгу історію, яка розпочалась ще в 1897 р., коли економіст Ст. Парето запропонував формулу описання нерівномірності розподілу благ у суспільстві. Цю ідею підтримав і М. Лоренц, який відобразив цю ситуацію на графіку (**крива Лоренца**). Обидва вчених довели, що більшу частку доходів має найменша частка людей.

Дещо пізніше відомий американський фахівець з управління якістю Дж. Джуран запропонував застосовувати підхід Ст. Парето у галузі контролю якості. Це дозволило розділити фактори, які впливають на якість продукції, на нечисленні істотно важливі та численні несуттєві. Дослідження Дж. Джурана показали, що, зазвичай, переважна кількість дефектів і пов'язаних з ними втрат для підприємства виникає через відносно невелику кількість причин. Цей підхід аналізу якості був названий діаграмою Парето.

Діаграма Парето являє собою різновид стовпчикової діаграми, яка застосовується для наочного відображення розглянутих факторів у порядку зменшення їх значущості.

Можна виділити два види діаграми Парето (табл. 8.3).

Побудову діаграми Парето починають із класифікації виникаючих проблем за окремими факторами. Після цього статистичний матеріал по кожному фактору збирають і аналізують для того, щоб з'ясувати, які з цих факторів превалюють при вирішенні проблем [9].

У прямокутній системі координат на осі абсцис відкладають рівні відрізки, які відповідають обраним факторам, а на осі ординат – величину їх внеску в дану проблему. При цьому порядок розташування факторів має бути таким, щоб вплив кожного наступного з них, розташованого на осі абсцис, зменшувався в порівнянні з попереднім. У результаті побудови отримують діаграму, стовпчики якої відповідають окремим факторам, які є причинами виникнення проблеми з якістю.

У цілому побудова діаграми Парето складається з дванадцяти взаємопов'язаних етапів. Розглянемо це більш детально.

Етап 1. Вибір проблем виробництва, які необхідно досліджувати. Наприклад, це можуть бути проблеми безпосередньо з якістю, аварійні ситуації на виробництві, додаткові витрати підприємства.

Етап 2. Вибір типу даних та їх класифікаційних ознак. Наприклад, можна класифікувати дані згідно з типом дефектів, місцем виникнення, технологічними причинами, видом обладнання тощо.

Етап 3. Визначення методу збору даних. Це може бути безпосереднє спостереження, документальний спосіб, опитування.

Етап 4. Розробка контрольного листка для збирання даних (із переліком видів даних, які призначені для збору). Розробка форми реєстрації даних.

Етап 5. Заповнення реєстраційної форми і підведення підсумків.

Етап 6. Розробка таблиці, яка дозволить перевіряти дані по кожному фактору окремо (з графами для підсумків). Розрахунок накопиченої суми числа дефектів, відсотків до загального підсумку і накопичених відсотків.

Характеристика видів діаграми Парето

Діаграма Парето за результатами діяльності		Діаграма Парето за причинами виникнення проблем з якістю	
Характеристика	Призначена для виявлення головної проблеми. Відображає небажані результати діяльності.	Характеристика	Відображає причини проблем, що виникають у ході виробництва. Використовується для виявлення головної з них.
Відображає негативні результати діяльності підприємства, пов'язані з:		Виявляє основні проблеми на виробництві, головною причиною яких є:	
<i>елемент</i>	<i>деталізація</i>	<i>елемент</i>	<i>деталізація</i>
якість	дефекти, відмови, помилки, рекламації, ремонти, повернення продукції	виконавець роботи	зміна, бригада, вік робітників, їх досвід роботи, кваліфікація, індивідуальні характеристики
собівартість	витрати, обсяг втрат, собівартість одиниці продукції	обладнання	верстати, агрегати, інструменти, оснащення, організація використання, моделі, штампи, знос обладнання
терміни постачання	нестача запасів, помилки в складанні рахунків, зрив термінів постачань	сировина	виробник, вид сировини, підприємство-постачальник, партія, якість сировини
безпека	нешасні випадки, трагічні помилки, аварії на виробництві.	методи праці	умови виробництва, замовлення-наряди, прийоми роботи, послідовність операцій
		виміри	точність, вірність і повторюваність (вміння дати однакову вказівку в наступних вимірах одного і того ж значення), стабільність (повторюваність протягом тривалого періоду), спільна точність, тип вимірювального приладу (аналоговий або цифровий).

Дані повинні розташовуватися в порядку їх вагомості (табл. 8.4).

Таблиця 8.4

Результати реєстрації даних за типами дефектів для побудови діаграми Парето

Тип дефекту	Кількість дефектів		Накопичена сума дефектів	
	одиниць	% до загальної кількості	одиниць	% до загальної кількості
Деформація	310	31,0	310	31,0
Подряпини	270	27,0	580	58,0
Тріщини	240	24,0	820	82,0
П'ятна	100	10,0	920	92,0
Розриви	40	4,0	960	96,0
Дефекти комплектації	25	2,5	985	98,5
Інше	15	1,5	1000	100,0
Усього	1000	100,0	x	x

Групу «інше» варто розміщувати в останньому рядку незалежно від її числових значень, оскільки вона складається із сукупності ознак, числовий результат по кожному з яких буде меншим, ніж найменше значення, отримане для ознаки, виділеної в окремий рядок таблиці.

Етап 7. Побудова горизонтальної та вертикальної осей. Вертикальна вісь повинна містити відсотки, а горизонтальна – інтервали відповідно до числа контрольованих ознак. Горизонтальну вісь необхідно розбити на інтервали відповідно до кількості контрольованих ознак.

Етап 8. Побудова стовпчикової діаграми (рис 8.1).

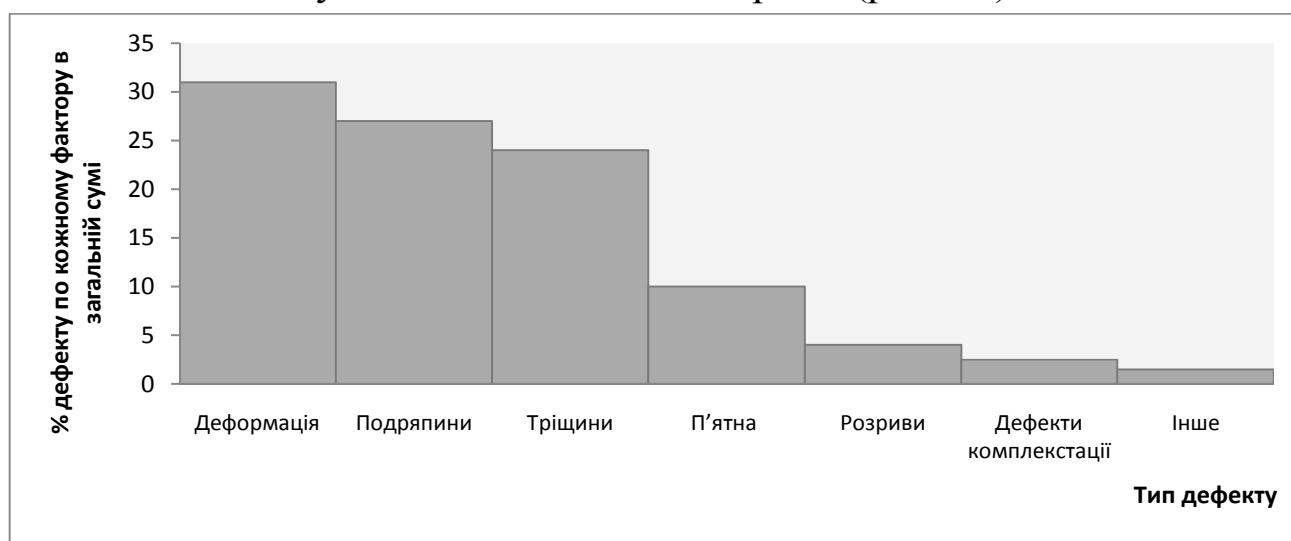


Рис. 8.1. Діаграма Парето

Етап 9. Проведення на діаграмі кумулятивної кривої (кривої Парето) (рис. 8.2).

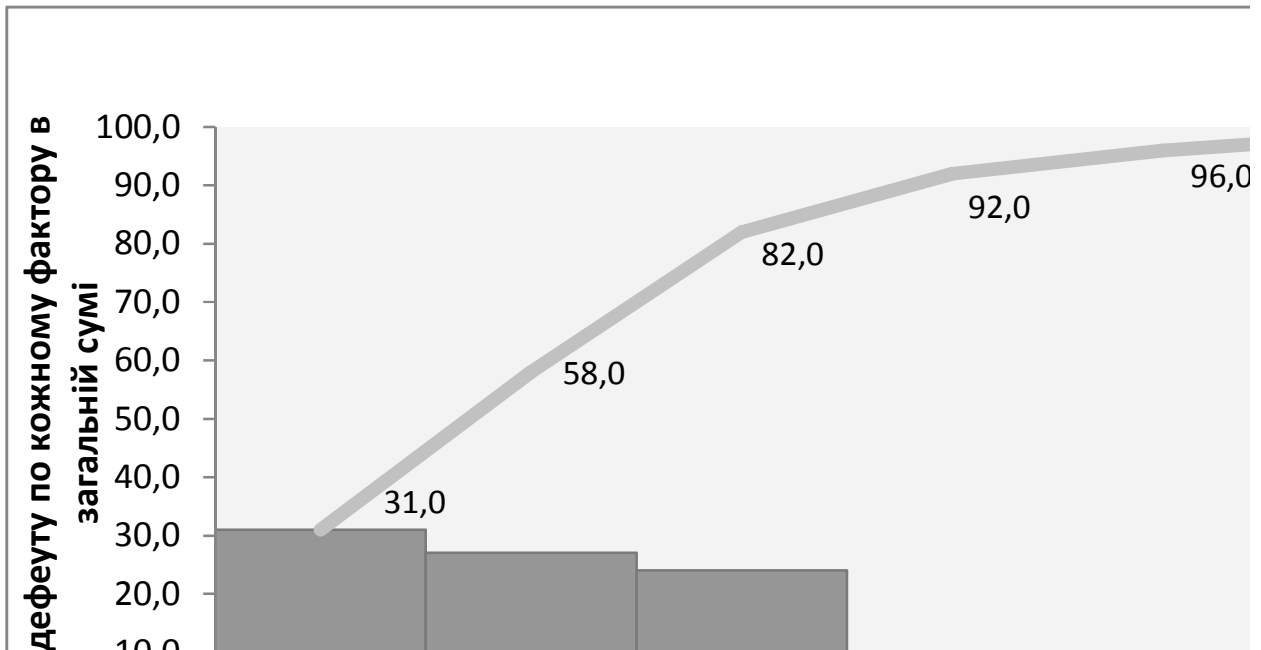


Рис. 8.2. Кумулятивна крива на діаграмі Парето

Аналіз побудованої діаграми, представленої на рис. 8.2, показує, що деформація, подряпини та тріщини в деталях складають 82,0 % всіх невідповідностей. Отже, з усунення саме цих дефектів слід розпочинати роботу щодо забезпечення якості продукції.

Діаграма Парето часто виявляє закономірність, яка отримала назву «**Правило 80/20**», засноване на принципі Парето, згідно якого велика частина наслідків викликається відносно нечисленними причинами: зазвичай 80 % виявлених невідповідностей пов'язана лише з 20 % всіх можливих причин.

Крім виявлення та ранжування факторів за їх значущістю діаграма Парето з успіхом застосовується для наочної демонстрації ефективності тих чи інших заходів у галузі забезпечення якості: достатньо побудувати та порівняти дві діаграми Парето – до і після реалізації будь-яких заходів.

Відносно побудови та застосування діаграми Парето ми також рекомендуємо наступне:

– оптимальним є використання різних класифікацій і складання великої кількості діаграм Парето. Суть проблеми легше піймати, спостерігаючи явище з різних точок зору, тому важливо випробувати різні шляхи класифікації даних, поки не будуть визначені нечисленні істотно важливі фактори, що, власне, і є метою аналізу Парето;

– група факторів «інше» не повинна складати великий відсоток від загальної кількості. Великий відсоток цієї групи вказує на те, що об'єкти спостереження класифіковані невірно і занадто багато об'єктів потрапило в одну групу. Це вказує на необхідність використання іншого принципу класифікації;

– якщо дані можна представити в грошовому вираженні, найкраще показати це на вертикальній осі діаграми Парето. Якщо існуючу проблему не можна оцінити в грошовому вираженні, саме дослідження може виявитися мало ефективним, оскільки витрати – важливий критерій вимірювань в управлінні якістю продукції;

– якщо небажаний фактор можна усунути за допомогою простого рішення, це потрібно зробити негайно, яким би незначним він не був. Оскільки діаграма Парето розцінюється як ефективний засіб вирішення проблем з якістю продукції, краще розглядати тільки нечисленні істотно важливі причини. Однак усунення неважливої причини простим шляхом може послужити прикладом ефективного вирішення проблеми, а набутий досвід, інформація та моральне задоволення – мати позитивну дію на подальшу процедуру вирішення проблем;

– одним із найбільш важливих варіантів діаграми Парето є класифікація факторів за причинами поганої якості продукції.

Окрім сфери виробництва діаграму Парето можна використовувати у **фінансовій сфері** (аналіз отриманого прибутку залежно від виду виготовленої продукції; аналіз рентабельності окремих видів продукції), **сфері збуту** (прогнозування кількості покупців окремо по кожному виду товару, аналіз кількості повернених виробів окремо по кожному виду), **сфері матеріально-технічного постачання** (аналіз кількості днів затримки постачання по окремих видах сировини, аналіз видів утрат залежно від затримки сировини на складах по окремих видах матеріалів чи виготовленої продукції), **сфері діловиробництва** (аналіз кількості пропозицій по окремих співробітниках, відсотка виконання плану по підрозділах).

8.4. Метод розшарування

Метод розшарування (стратифікації) є одним з найбільш простих методів контролю якості, відповідно до якого інформацію групують залежно від умов її отримання. Кожна група даних обробляється та аналізується окремо від інших груп.

Наприклад, розшарування (стратифікацію) даних можна провести за наступними *ознаками*:

- залежно від виконавця робіт (належність до певного підрозділу, стаж роботи, кваліфікація);
- залежно від типу обладнання (нове або старе обладнання, марка обладнання, країна-виробник тощо);
- залежно від сировини (за місцем її виробництва, відповідно до фірми-виробника, з урахуванням якості сировини тощо);
- залежно від способу виробництва (температура обробки, кількість стадій обробки, інші особливості процесу виробництва).

Під час розшарування зібраних даних необхідно прагнути до того, щоб відмінності усередині групи були якомога меншими, а відмінності між групами – якомога більшими.

Використання методу розшарування для аналізу якості продукції дозволяє отримати уявлення про приховані причини дефектів, а також допомагає виявити причину появи дефекту. При проведенні розшарування за фактором «виконавець» (при значному розходженні в даних) можна визначити вплив того чи іншого виконавця на якість виробу; якщо розшарування проведено за фактором «обладнання» – вплив використаного обладнання тощо.

Розглянемо метод розшарування на прикладі умовної деталі «А», під час виробництва якої було виявлено декілька різноманітних дефектів якості. Розіб'ємо причини цих дефектів (100 %) на 4 різні групи: постачальники сировини, робочі зміни на виробництві, тип обладнання та особистості 6 різних операторів. Представимо результати групування даних у вигляді графіків (рис. 8.3 - 8.6).

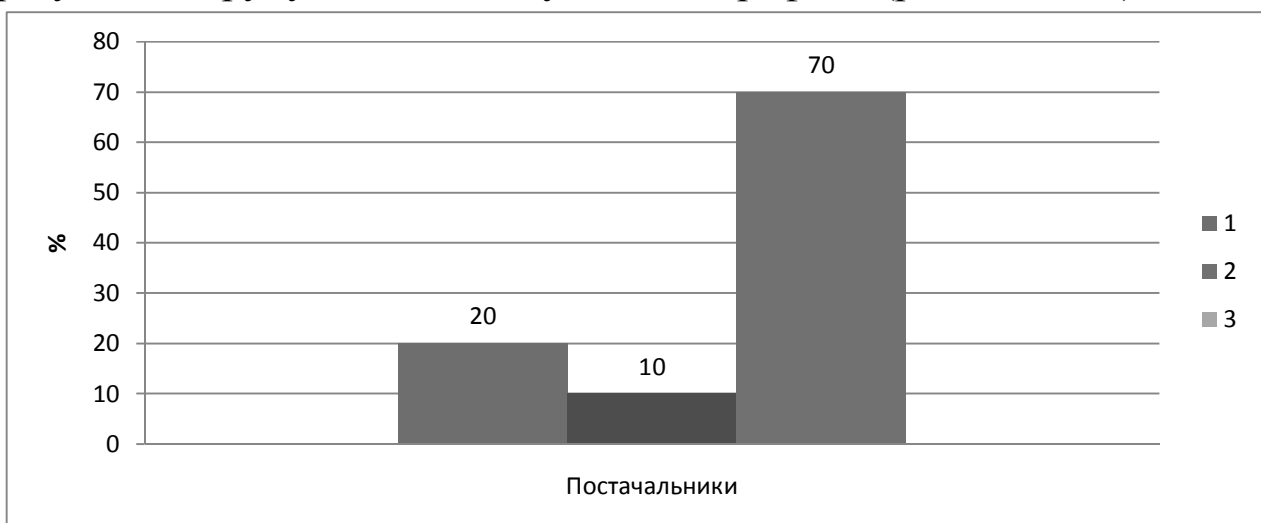


Рис. 8.3. Розподіл дефектів продукції залежно від постачальника сировини

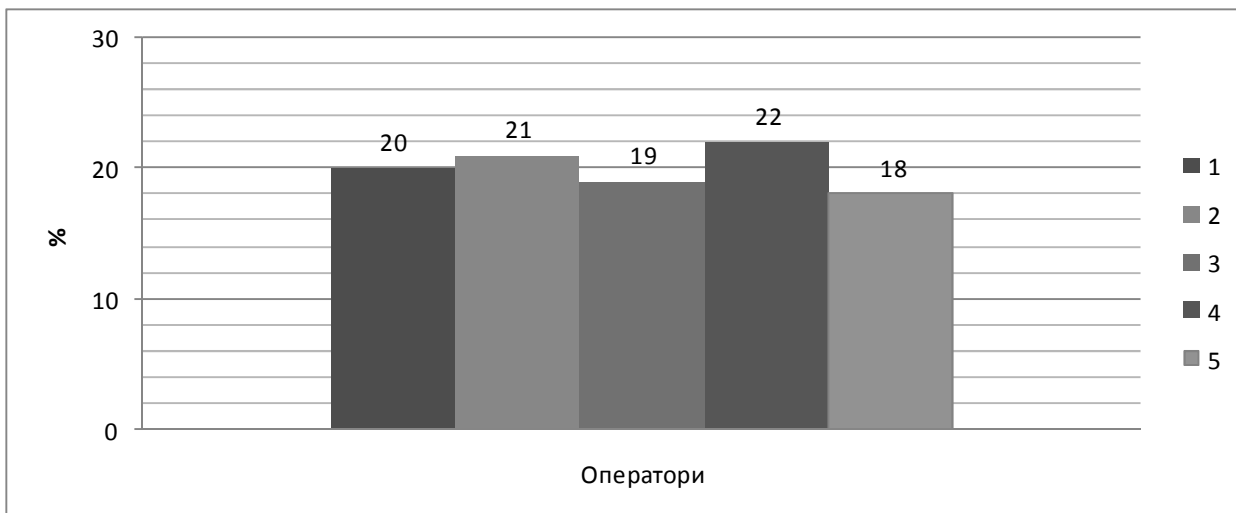


Рис. 8.4. Розподіл дефектів залежно від оператора

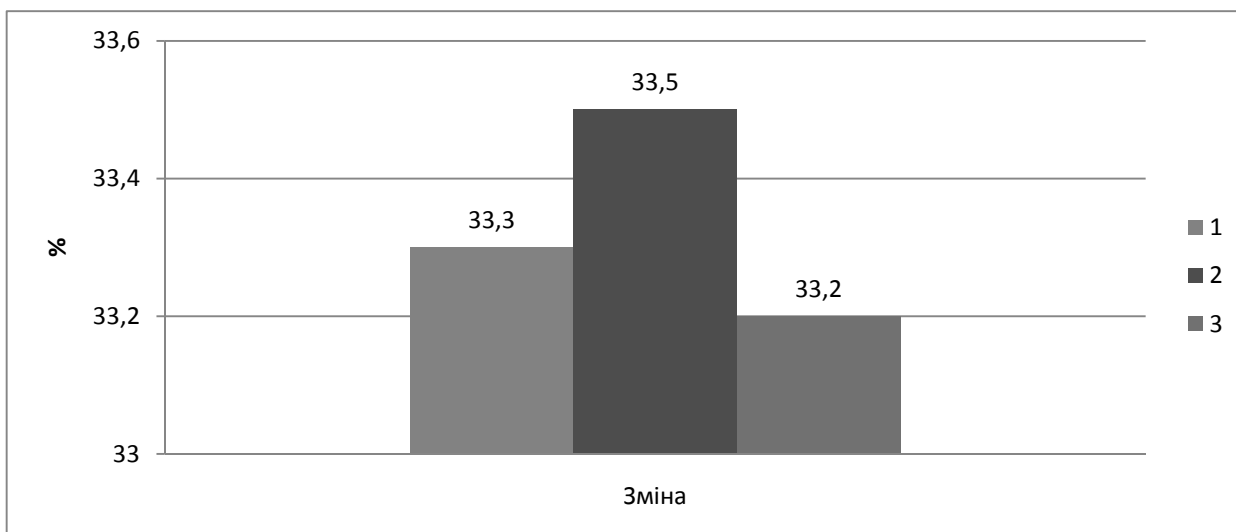


Рис. 8.5. Розподіл дефектів залежно від зміни

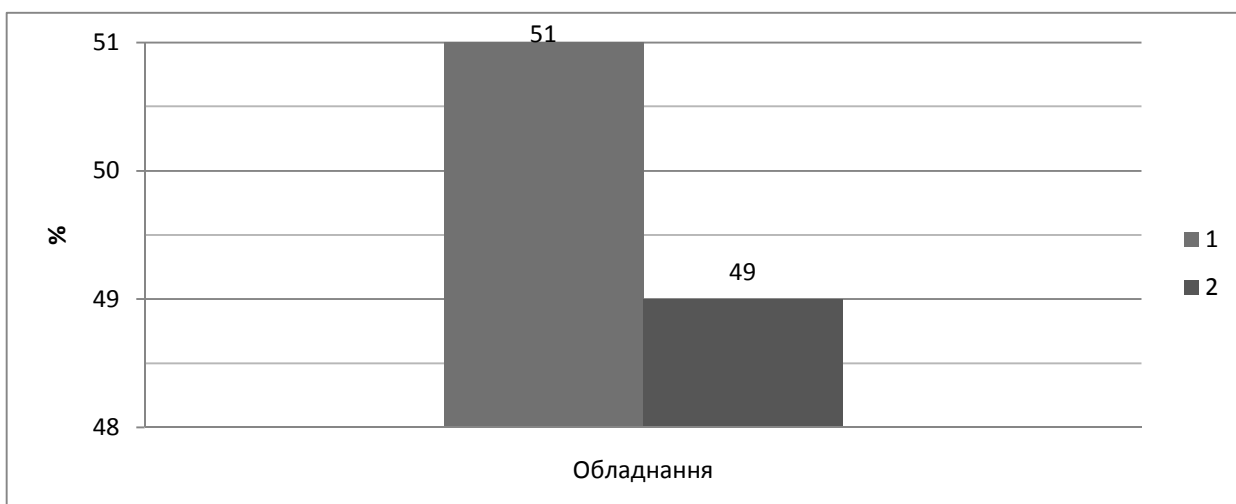


Рис. 8.6. Розподіл дефектів залежно від типу обладнання

Аналіз графіків дозволяє зробити висновок, що не має особливих відмінностей між виробниками чи змінами, в які працюють робітники, не впливає й тип обладнання, яке використовують для виробництва. При цьому 70 % дефектів трапляється з вини постачальника сировини, з чого можна зробити висновок, що для підвищення якості продукції та запобігання браку необхідно змінити постачальника.

У ситуаціях, коли після розшарування даних неможливо визначити наочно основний фактор у вирішенні проблеми, необхідно проводити більш глибокий аналіз даних [14].

8.5. Гістограма

Наступний інструмент управління якістю – метод **гістограми**. Цей метод дозволяє представити статистичні дані в графічному вигляді – вигляді стовпчикової діаграми. Гістограма найчастіше застосовується для вибору напрямку роботи, розробки та впровадження коригувальних заходів щодо встановлених цілей. Цей інструмент управління якістю може виконувати наступні аналітичні функції:

- виконувати позиційне порівняння (при визначенні співвідношень розглянутих об'єктів);
- виконувати порівняння в часі, аналізуючи динаміку значень показників;
- виконувати певне порівняння, виділяючи певну кількість об'єктів у розглянутій області значень.

Метод гістограми застосовується з метою попередньої оцінки однорідності експериментальних даних і для наочного уявлення характеру розподілу випадкових величин у вибірці.

Аналіз і представлення даних виконується наступним чином:

- до кожного інтервалу включаються спостереження, що лежать у межах від нижньої межі інтервалу до верхньої;
- значення, що потрапили на межі інтервалів, включаються наступним чином: у перший інтервал включається і нижня, і верхня межа інтервалів, а в усі інші інтервали включається лише верхня межа (якщо немає конкретних вимог щодо включення даних у той або інший інтервал);

– здійснюється аналіз форми та розташування гістограми щодо меж поля допуску. Це дає багато інформації про досліджуваний процес і без виконання розрахунків.

Процес аналізу якості за допомогою гістограми може здійснюватись як в індивідуальній формі, так і в командній.

Існують наступні **варіації структури гістограми**: дзвіноподібний розподіл, розподіл з двома піками, розподіл типу плато, розподіл грибінчастого типу, скошений розподіл, усічений розподіл, розподіл з ізольованим піком, розподіл із піком на краю.

Ефективність використання методу гістограми оцінюється аналізом повторної її побудови через деякий час після впровадження заходів та їх порівняння з результатами попереднього дослідження.

Перевірка якості виготовленої продукції здійснюється в ході аналізу характеру гістограми. Якщо графік має форму нормального розподілу, то це свідчить про стабільність процесу. На практиці часто зустрічаються випадки, коли форма розподілу відхиляється від нормального вигляду. Це свідчить про порушення в процесі та вказує на необхідність застосування керуючих впливів.

Розглянемо побудову гістограми на конкретному прикладі. У таблиці 8.5 надано результати вимірювань розмірів умовної партії із 60 одиниць заготовок на верстаті протягом доби. Нормативне значення заготовки знаходиться в межах від 160 до 170 см.

Таблиця 8.5

Розмір заготовки, см.

150	170	162	168	161	163	164	153	154	155
161	171	160	170	158	175	166	155	162	164
163	179	159	171	176	167	156	164	156	165
180	154	172	177	175	176	152	152	157	163
165	162	178	173	177	174	151	153	151	159
162	164	173	174	178	179	157	179	161	160

Для побудови гістограми серед наведених у таблиці 8.5 даних знайдемо максимальне та мінімальне значення (x_{\max} , x_{\min}) – у нашому прикладі 150 і 180 відповідно. Після цього для групування статистичного матеріалу необхідно вибрати кількість класів (інтервалів значень), які необхідно утворити. Число класів не повинно бути надто великим (ряд розподілу при цьому стає невиразним і частоти в ньому схильні до незакономірних коливань), але не повинно бути й надто малим (властивості розподілу

описуються статистичним рядом надто грубо). Практикою встановлено, що якщо кількість даних знаходиться в межах 40-100, цільових перевірок бажано формувати 6-9 класів, при вибірці у 100-500 об'єктів – 8-12 класів, при 500-1000 об'єктах – 10-16 класів. Оскільки в нашому випадку вибірка склала 60 одиниць заготовок, то доречно сформувати від 6 до 9 класів. Щоб обрати оптимальну їх кількість, скористаємось формулою Стерджеса:

$$n = 1 + 3,322 \lg N, \quad (8.1)$$

де N – обсяг сукупності,

n – кількість груп (класів).

Після визначення кількості груп вирішимо питання про вибір величини інтервалу. Залежно від ступеня коливання групувальної ознаки та характеру розподілу одиниць статистичної сукупності можна застосовувати рівні або нерівні інтервали. При більшості досліджень вчені використовують принцип саме рівності інтервалу. Тоді визначення величини інтервалу (h) при групуванні буде здійснюватися за формулою:

$$h = \frac{x_{\max} - x_{\min}}{n} \quad (8.2)$$

У нашому випадку $h = \frac{180 - 150}{6} = 5$ см. Таким чином, ми отримуємо 6 груп з наступними інтервалами:

150-155; 155-160; 160-165; 165-170; 170-175; 175-180.

Далі необхідно знайти кількість спостережень, що потрапили до кожної групи. Рекомендовано дотримуватися наступного правила: якщо число знаходиться на межі двох інтервалів, воно потрапляє в той інтервал, де було отримано вперше. Відповідно до цього правила в перший клас від 150 до 155 потрапило $m_1 = 11$ результатів спостережень, до другого класу від 155 до 160 потрапило $m_2 = 9$ результатів спостережень, в третій клас – $m_3 = 16$ результатів спостережень тощо. Ці дані доцільно занести в наступну таблицю (табл. 8.6).

За частотами можна побудувати гістограму. Для цього на кожному інтервалі потрібно побудувати стовпчик (прямокутник), висота якого дорівнює (пропорційна) частоті потрапляння спостережень у даний інтервал [5].

Інтервальний ряд значень довжини заготовки

Довжина заготовки, см	Дані контролю	Частота, од.
150-155	Ш Ш Ш Ш	11
155-160	Ш Ш Ш Ш І	9
160-165	Ш Ш Ш Ш Ш Ш Ш Ш Ш Ш	16
165-170	Ш Ш Ш І	5
170-175	Ш Ш Ш Ш І	9
175-180	Ш Ш Ш Ш Ш Ш Ш Ш Ш Ш	10

Приклад побудови гістограми наведено на рис. 8.7.

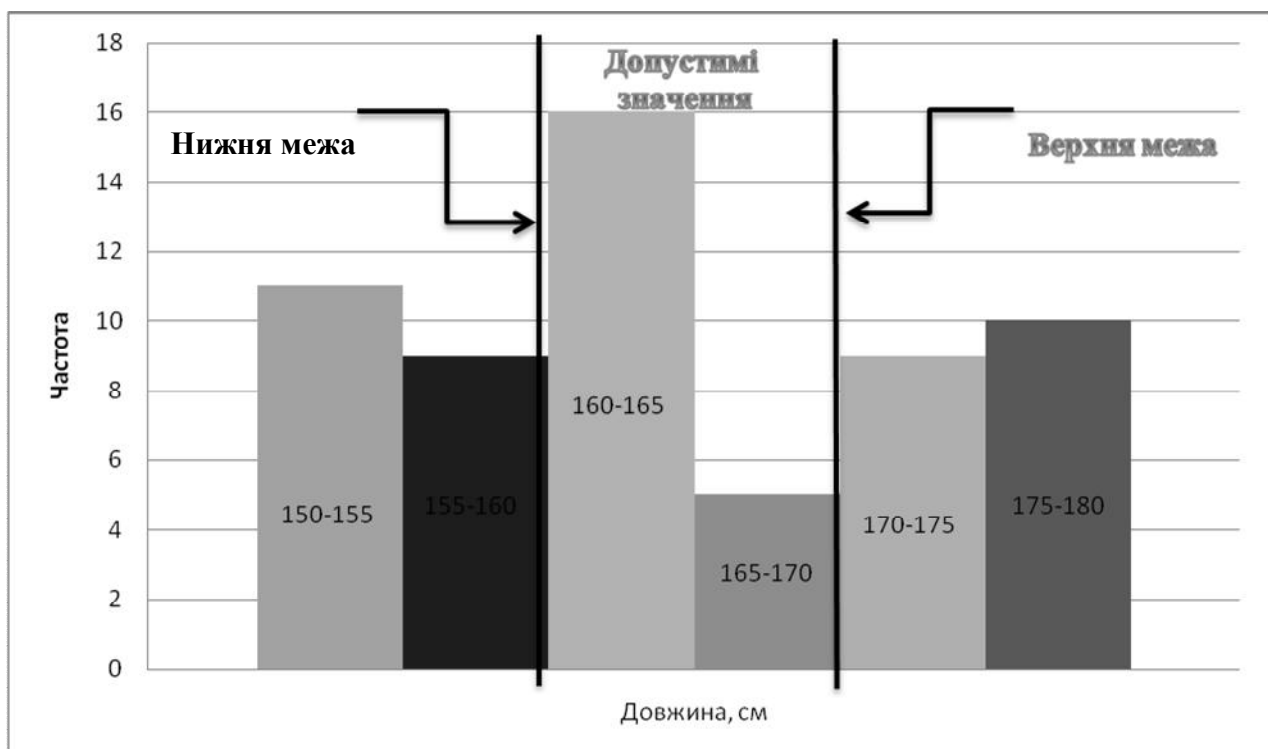


Рис. 8.7. Гістограма довжини заготовки

Як бачимо, гістограма являє собою стовпчиковий графік. Наносячи на графік допустимі значення параметра можемо визначити, як часто потрапляє цей параметр у допустимий діапазон. Проаналізувавши побудовану гістограму, можемо зробити висновок, що процес виробництва заготовок є неконтрольованим і неякісним, оскільки більшість виготовлених елементів знаходиться за межами мінімально та максимально допустимих значень.

Основною перевагою методу гістограми є його наочність і відносна простота використання, однак цей метод має ряд *недоліків*: у зв'язку з великою кількістю видів гістограм, часом важко визначити,

в якому випадку на її форму впливають різні чинники, а коли відбуваються помилки в зібраних даних або побудові.

8.6. Діаграма розкиду

Метод **діаграми розкиду (розсіювання)** застосовується у виробництві на різних стадіях життєвого циклу продукції з метою виявлення залежності між показниками якості та основними факторами виробництва. Діаграма розкиду дозволяє оцінити на підставі графічного представлення без математичної обробки експериментальних даних характер і тісноту зв'язку значень двох змінних (кореляційне відношення), якими можуть бути: характеристика якості процесу та фактор, що впливає на хід процесу; дві різні характеристики якості; два фактора, що впливають на одну характеристику якості.

Цей метод оцінки якості застосовується на підприємстві для розробки та коригування виробничих цілей. Він *дозволяє*:

- виявити та наочно представити взаємну залежність груп даних;
- оцінити вплив однієї групи даних на розподіл іншої;
- визначити оптимальні умови протікання процесу (поліпшення протікання процесу);
- без математичної обробки експериментальних даних на підставі графічного зображення оцінити характер і тісноту зв'язку значень двох змінних.

Деякі типові варіанти зображення діаграми розкиду схематично наведені на рис. 8.8.

На **рис. 8.8 (а)** чітко проглядається пряма кореляція між параметрами x та y . У цьому випадку при здійсненні контролю за причинним фактором x можна управляти й значенням параметра y .

На **рис. 8.8 (б)** наведений приклад оберненої (негативної) кореляції між параметрами. При збільшенні x характеристика y зменшується. Якщо причинний фактор x знаходиться під контролем, характеристика y залишається стабільною.

На **рис. 8.8 (в)** також наведено приклад прямої кореляції. При збільшенні параметру x параметр y теж буде зростати, але розкид y великий по відношенню до певного значення x . Тому таку кореляцію називають легкою або незначною. При наявності такого розкиду за допомогою контролю причинного фактору x можна до певної міри

контролювати характеристику продукції y , але необхідно також мати на увазі й інші фактори, що певною мірою впливають на y .

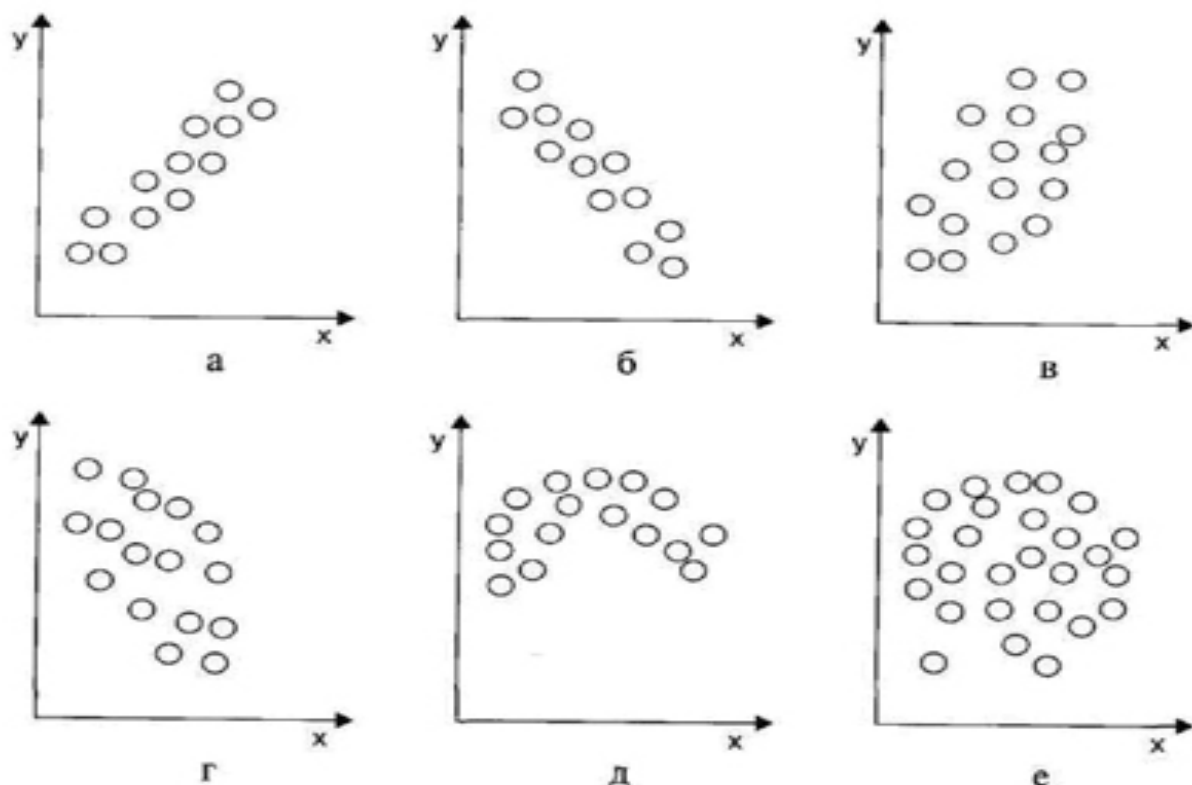


Рис. 8.8. Типові приклади діаграми розкиду

На **рис. 8.8 (г)** зображений випадок легкої оберненої кореляції між факторами, коли при збільшенні x характеристика y зменшується, але при цьому є присутнім великий розкид значень y , відповідних фіксованим значенням x . Варто пам'ятати, що крім фактора x на y буде впливати ще певне коло факторів.

На **рис. 8.8 (д)** зображена криволінійна кореляція між x та y . Якщо діаграму розкиду можна розділити на ділянки, які мають прямолінійний характер, то проводять такий поділ і досліджують кожну ділянку окремо, як прямолінійну кореляцію.

На **рис. 8.8 (е)** показаний приклад відсутності кореляції, коли не існує ніякої вираженої залежності між x та y . У цьому випадку необхідно продовжити пошук факторів, що корелюють з y , виключивши з цього пошуку фактор x .

Існує наступний алгоритм побудови діаграми розкиду.

Етап 1. Збирають приблизно 25-30 пар даних, між якими необхідно виявити взаємозв'язок і оформлюють ці дані у вигляді таблиці.

Етап 2. Знаходять максимальні та мінімальні значення для фактору x та результату y . На графіку обирають шкали на горизонтальній і вертикальній осях так, щоб обидві довжини робочих частин вийшли приблизно однаковими (для полегшення читання діаграми). По горизонтальній осі відкладають фактор x , по вертикальній – значення результату y . На кожній осі беруть не більше ніж 3-10 градусій і використовують для полегшення читання округлені числа.

Етап 3. Побудова графіка. Якщо в різних спостереженнях зустрічаються однакові значення, то можна намалювати одну точку поряд із іншою або поставити умовне позначення, наприклад « \emptyset ».

Етап 4. Нанесення необхідних відміток на діаграму.

Наприклад, побудуємо діаграму розкиду для оцінки впливу умовного фактору x на величину значення y . Для проведення експерименту було обрано 30 примірників. Значення x та відповідні їх значення y наведені в табл. 8.7.

Таблиця 8.7

Значення x та y для 30 примірників продукції

№	Фактор x	Значення y	№	Фактор x	Значення y
1.	8,0	2620	16.	4,5	2220
2.	7,0	2520	17.	12,0	2790
3.	7,5	2530	18.	11,0	2950
4.	5,3	2520	19.	14,0	3200
5.	5,0	2440	20.	13,2	2840
6.	2,5	2400	21.	16,0	3100
7.	5,5	2450	22.	2,7	2230
8.	10,1	2620	23.	2,5	2300
9.	6,0	2560	24.	6,9	2800
10.	5,0	2410	25.	4,4	2500
11.	6,5	2520	26.	9,2	2980
12.	9,0	2640	27.	6,5	2900
13.	9,0	2700	28.	3,0	2050
14.	1,0	2340	29.	1,0	2020
15.	10,5	2760	30.	1,0	2000

Побудована за цими даними діаграма розкиду наведена на рис. 8.9.

Як свідчать дані побудованого графіку, існує пряма залежність між значеннями x та y : чим більше значення x , тим більше значення y .

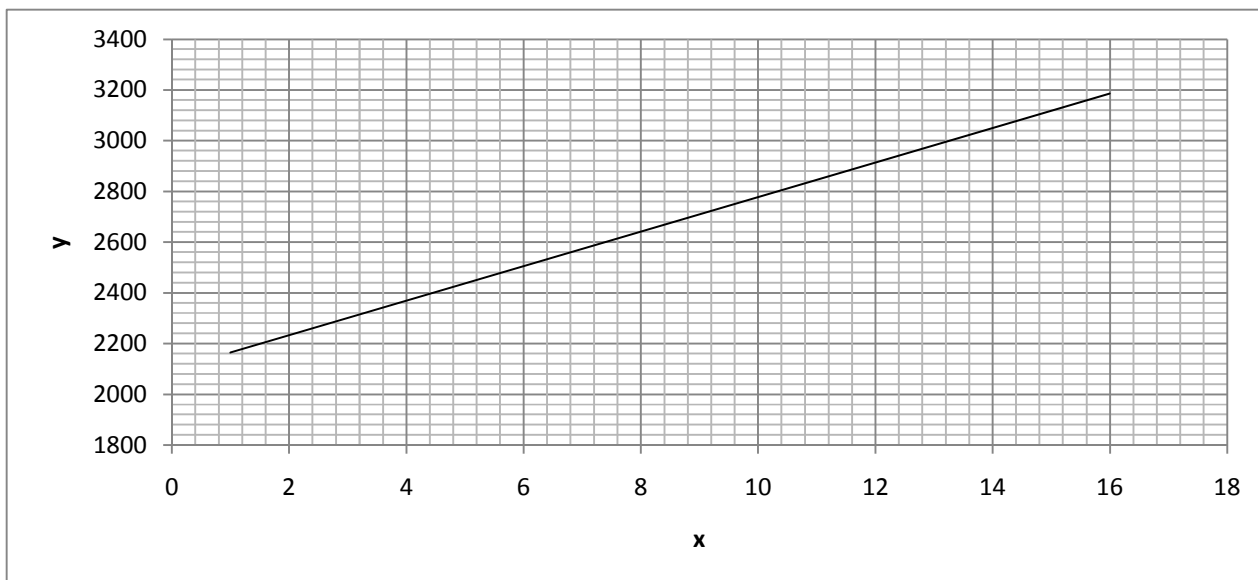


Рис. 8.9. Діаграма розкиду

Головною перевагою методу діаграми розкиду є можливість застосування критерію Стюдента під час оцінки достовірності (тісноти кореляційного зв'язку) даних. Для побудованого прикладу $R^2=0,780$, що свідчить про достаньо тісний взаємозв'язок.

Недоліком методу є те, що можна здійснювати контроль тільки одного (будь-якого) з двох параметрів, а також вузька область співставлення параметрів.

8.7. Діаграма Ішікави

Причинно-наслідкова **діаграма Ішікави** являє собою графічне упорядкування факторів, що впливають на об'єкт аналізу. Для побудови діаграми процес виготовлення продукції, що впливає на її якість, розглядається як взаємодія впливу **5 факторів (5M)**: людини (man), машини (machine), матеріалу (material), методу (method) та вимірів (measures).

Сама діаграма зазвичай має вигляд риба'ячого скелету. По-перше, при побудові малюють горизонтальну стрілку, яка буде відображати об'єкт аналізу. До горизонтальної стрілки підводять великі первинні стрілки, що умовно позначають головні фактори або їх групи, які впливають на об'єкт аналізу. Далі до кожної первинної стрілки підводять менші стрілки другого порядку, до яких, в свою чергу, підводять стрілки третього порядку тощо. Це триває до тих пір, поки на діаграму не будуть нанесені всі фактори, що роблять помітний вплив на якість продукції в конкретній ситуації. Кожна зі

стрілок, нанесена на схему, в залежності від її положення являє собою або причину або наслідок: попередня стрілка по відношенню до подальшої завжди виступає як причина, а наступна – як наслідок.

Головне завдання при побудові діаграми – забезпечення правильного підпорядкування у взаємозалежності факторів, а також чітке її оформлення.

Деталізована діаграма Ішікави може служити основою для складання плану взаємопов'язаних заходів, що забезпечують комплексне вирішення поставленої при аналізі задачі. Умовний приклад такої діаграми наведений на рис. 8.10.

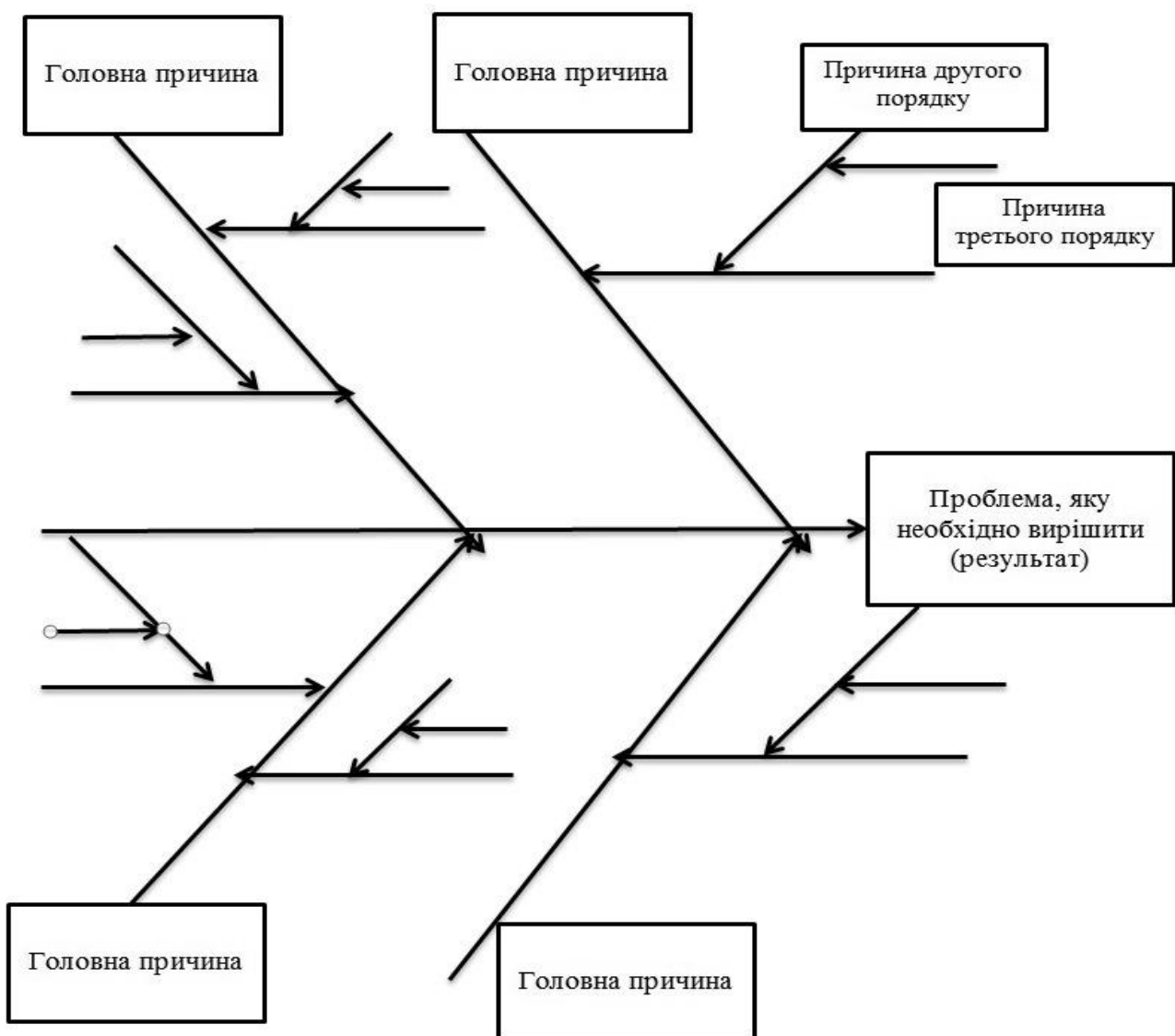


Рис. 8.10. Діаграма Ішікави

Побудова діаграми складається із наступних етапів:

Етап 1. Визначення мети побудови діаграми. На цьому етапі в якості об'єкту дослідження вибирають або виробничу проблему, або один із показників якості, щодо якого були замічені дефекти. Бажано обирати фактори, для яких існують цифрові значення.

Етап 2. Виявлення основних факторів впливу. При складанні списку факторів, що впливають на появу дефекту, не можна відкидати жоден із них, хоч би незначним він не здавався. Якщо на початковій стадії побудови діаграми якийсь фактор, що знижує якість, випав з поля зору, то, швидше за все, він не з'явиться в подальшій роботі. Фактори, вплив яких на даний дефект малоймовірний, можуть бути відкинуті при подальшому аналізі діаграми, але на етапі її побудови вони повинні бути представлені всі без винятку, щоб було ясно, що вони вже прийняті до уваги на якомусь етапі аналізу.

Етап 3. Виявлення вторинних факторів. Безумовно, вплив головних чинників у свою чергу визначається тим, що вони самі залежать від якихось інших факторів. Тому після того, як визначені головні фактори, виявляються вторинні фактори, що впливають на кожен окремий фактор з «5М». Вторинні фактори також наносяться на діаграму. Фактори другого порядку можуть визначатися факторами третього порядку тощо. Групування факторів другого і наступних порядків зазвичай носить лише умовний характер і залежить від поставленої мети і умов аналізу.

Інформація про показники якості для побудови діаграми повинна бути зібрана з усіх доступних джерел: журналу реєстрації операцій, журналу реєстрації даних поточного контролю, повідомлень робітників виробничої дільниці тощо. Дуже важливо простежити кореляційну залежність між причинними факторами (параметрами процесу) і показниками якості.

Одним із інструментів побудови такої діаграми є програма *Statistics*, а в ній функція *Industrial Statistica & Six Sigma / Process Analysis*.

Основною перевагою діаграми Ішікави є те, що вона дає наочне уявлення не тільки про ті фактори, що впливають на досліджуваний об'єкт, але й про причинно-наслідкові зв'язки між цими факторами. Недоліком можна вважати її громіздкий характер і складність побудови без наявності спеціального програмного забезпечення.

8.8. Контрольна карта

На відміну від попередніх методів контролю якості, **контрольна карта** дає можливість фіксувати стан процесу в часі. Цей метод був запропонований У. Шухартом у 1924 р. для візуалізації статистичних характеристик того чи іншого виробничого процесу. Контрольна карта являє собою засіб оперативного управління виробництвом, оскільки дає можливість втрутитися в процес до того моменту, як він вийде з-під контролю.

Будь-яка контрольна карта – це графік, в якому по осі абсцис зазвичай відкладають параметр часу, а по осі ординат – характеристику якості. Також на карті завжди є присутніми три лінії: по центру розташовують лінію, яка характеризує потрібне номінальне значення показника (**НЗ**), дві інші лінії, одна з яких знаходиться нижче НЗ і описує мінімальні контрольні значення (**НКМ – нижня контрольна межа, LCL – Lower Control Level**), а та що знаходиться вище, описує максимальні контрольні значення (**ВКМ – верхня контрольна межа, UCL – Upper Control Level**).

Ці межі розраховують з урахуванням прийнятого розподілу значень контрольованого параметра і додаткової ймовірності отримання помилкового попереджувального сигналу про розлад операції. Довірчий інтервал указує, всередині яких меж очікується справжнє значення статистичної характеристики.

Межі регулювання окреслюють область значень регульованої вибіркової характеристики, що відповідає задовільному налагодженню технологічної операції (якщо контрольований параметр заданий односторонньою нормою, то на контрольну карту наноситься тільки одна межа регулювання). Для кращого сприйняття контрольної карти її центральну лінію та межі доцільно позначати різними кольорами.

На рис. 8.11 і рис. 8.12 зображені приклади простих контрольних карт для умовного параметру якості.

Якщо контрольована характеристика якості виходить за межі НКМ або ВКМ, то це свідчить про порушення стабільності процесу виробництва або його неконтрольованість. У цьому випадку необхідно виявити причини порушень і прийняти певні заходи.

Контрольна карта є відмінним засобом збереження інформації шляхом представлення історії якості дослідженого виробничого процесу: хто отримав брак, на якому обладнанні, коли було прийнято рішення зупинити процес виробництва або відновити його.

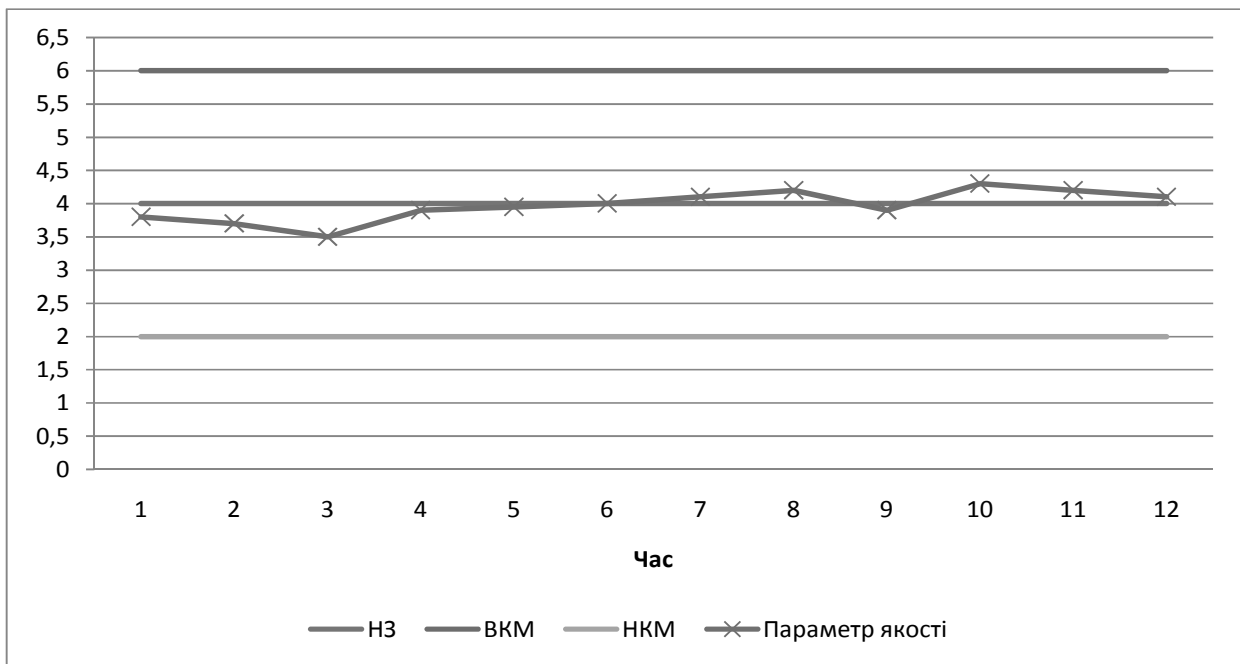


Рис. 8.11. Контрольна карта параметру якості: процес підлягає контролю

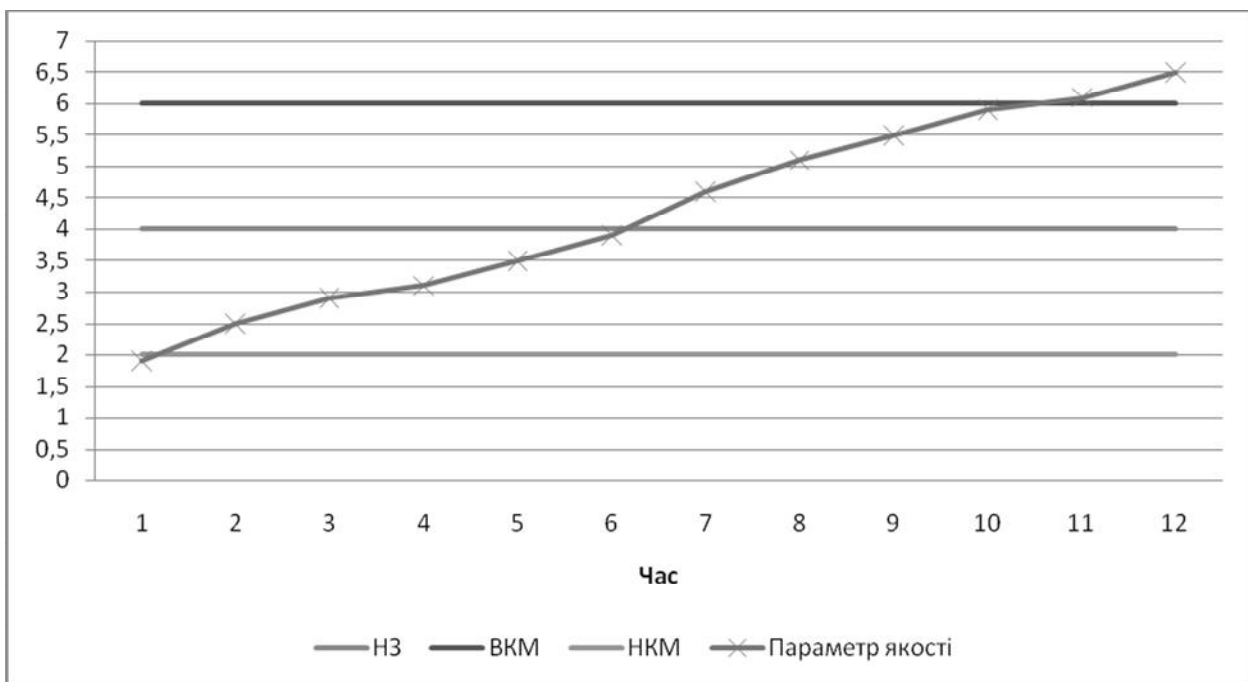


Рис. 8.12. Контрольна карта параметру якості: процес не підлягає контролю

- Найбільш розповсюдженими є наступні контрольні карти:
- середніх арифметичних і розмахів;
 - медіан і розмахів;
 - питомої ваги дефектної продукції;

- числа дефектних одиниць продукції;
- числа дефектів на одиницю продукції;
- індивідуальних значень;
- числа дефектів [4, 21].

Робота з контрольною карою полягає в тому, що за даними спостереження за значеннями контрольованого параметра встановлюється, чи знаходиться цей параметр в межах регулювання, і на основі цього приймається рішення про те, чи налагоджена технологічна операція або розлагоджена.

Рішення про розлагодження операції приймається при виході хоча б одного спостереження, зафіксованого на карті у вигляді точки, за кордон регулювання. Однак, ще до виходу точок за межі регулювання, контрольна карта дасть можливість робити судження про майбутні порушення технологічної операції за такими *ознаками*:

- поблизу кордонів регулювання з'являється кілька послідовних значень контрольованого параметра;

- значення розподіляються по один бік від центральної лінії, тобто середнє значення зміщується відносно центру налаштування (про наявність систематичного відхилення свідчить, наприклад, розташування поспіль семи значень вище або нижче середньої лінії, а також розташування 10 з 11, 12 з 14, 14 з 17 і 16 з 20 значень по один бік від середньої лінії);

- значення контрольованого параметра сильно розкидані;

- намічається тенденція наближення значень контрольованого параметра до однієї з меж регулювання.

Таким чином, *основною метою* контрольних карт є визначення неприродних змін в даних з повторювальних процесів і надання їм критеріїв для визначення відсутності статистичної керованості. Процес знаходиться в статистично керованому стані тільки в ситуаціях, коли неприродні зміни викликані випадковими причинами. Під час визначення цього прийняттого рівня змін будь-яке відхилення від нормативних значень прийнято вважати результатом дій особливих причин, які необхідно виявити, а потім знищити або послабити.

Метод контрольних карт допомагає визначити, чи дійсно процес став статистично керованим на правильному рівні та чи залишиться він на ньому. Даний метод дозволяє підтримувати управління та високий ступінь однорідності найважливіших характеристик

продукції шляхом безперервного запису інформації щодо якості продукції в процесі виробництва. Використання контрольних карт і їх якісний аналіз призводять до кращого розуміння та вдосконалення процесів виробництва.

Питання до самоконтролю

1. Дайте визначення виробничому процесу.
2. Коли з'явився статистичний метод контролю якості?
3. Чому до 1950-х років статичний метод контролю якості не був популярним та широко використовуваним?
4. За що присуджуються премії Демінга в Японії?
5. Перелічіть сім японських методів статистичного контролю якості?
6. Хто був основоположником статистичного методу контролю якості в Японії?
7. Що таке контрольні листки?
8. Назвіть основні види контрольних листків.
9. Коли був розроблений метод діаграми Парето?
10. Охарактеризуйте два види діаграми Парето.
11. Про що свідчить «Правило 80/20»?
12. В яких інших сферах, окрім контролю якості, можна використовувати діаграму Парето?
13. Дайте визначення методу розшарування.
14. За якими ознаками можна провести розшарування?
15. Які аналітичні функції виконує метод аналізу якості «гістограма»?
16. Перелічіть переваги та недоліки використання методу гістограми.
17. Що таке діаграма розкиду?
18. Які типові види діаграми розкиду існують?
19. Дайте характеристику основним видам діаграми розкиду.
20. Охарактеризуйте етапи побудови діаграми розкиду.
21. Які переваги та недоліки притаманні методу «діаграма розкиду»?
22. Що таке діаграма Ішікави?

РОЗДІЛ 9. ПОКАЗНИКИ ЕФЕКТИВНОСТІ ДІЯЛЬНОСТІ ПІДПРИЄМСТВА

9.1. Сутність показників ефекту та ефективності

Умови функціонування в ринковій економіці зумовлюють необхідність розвитку статистичного аналізу економічної діяльності в першу чергу на мікрорівні, тобто на рівні окремих підприємств, оскільки саме підприємства складають основу цієї економіки.

Аналіз на мікрорівні наповнюється абсолютно конкретним змістом, пов'язаним з повсякденною діяльністю, як підприємства, так і його колективу: менеджерів, простих робітників, директорів і власників.

Сьогодні нагальною є потреба проведення аналізу бізнес-планів підприємства, його маркетингових досліджень, можливостей виробництва і збуту продукції, внутрішніх і зовнішньоекономічних ситуацій, що впливають на діяльність, співвідношення попиту і пропозиції, конкретних споживачів і постачальників, витрат живої і матеріалізованої праці. Однак, одним із найбільш важливих елементів аналізу виступає дослідження *ефективності діяльності підприємства*.

Цей аналіз є необхідним не тільки керівнику підприємства для оцінки фінансового стану, а й цілому колу інших осіб, які беруть безпосередню участь у господарській практиці:

- керівникам маркетингових і рекламних відділів, які на основі отриманих даних розробляють стратегію просування товару на ринок;
- інвесторам, яким необхідно прийняти рішення про формування портфелю цінних паперів підприємства;
- аудиторам, яким необхідно перевірити звітність і діяльність підприємства для надання відповідних рекомендацій по веденню бухгалтерського обліку;
- кредиторам, які приймають рішення про видачу кредитів підприємству для впевненості, що їх кошти будуть повернуті з відсотками.

Таким чином, поняття *«економічний ефект»* і *«економічна ефективність»*, про які ми будемо говорити в даному розділі, відносяться до числа найважливіших категорій ринкової економіки. Ці поняття тісно пов'язані між собою.

Економічний ефект передбачає будь-який корисний результат, виражений у вартісній оцінці. Зазвичай у якості корисного результату виступають прибуток або економія витрат і ресурсів. Економічний ефект – *величина абсолютна*, що залежить від масштабів виробництва і економії витрат.

Економічна ефективність – це співвідношення між результатами господарської діяльності і витратами живої і матеріалізованої праці, ресурсами. Економічна ефективність залежить від економічного ефекту, а також від витрат і ресурсів, які викликали даний ефект. Таким чином, економічна ефективність – *величина відносна*, що отримується в результаті зіставлення ефекту з витратами і ресурсами.

Зазвичай аналізуються обидва показники, що характеризують успішність економічної діяльності підприємства, оскільки окремо показники ефекту та ефективності не можуть дати повної і всеосяжної оцінки діяльності підприємства. Наприклад, на підприємстві може бути така ситуація, коли досягнутий значний економічний ефект, виражений в отриманому прибутку при відносно низькій економічній ефективності. І навпаки, виробництво може характеризуватися високим рівнем ефективності при невеликій величині економічного ефекту. Ефект може бути як позитивним, так і негативним; ефективність – або дорівнювати нулю, або бути позитивною.

Систематичний і всебічний аналіз ефективності діяльності підприємства дозволяє:

- швидко, якісно та професійно оцінювати результативність господарської діяльності як підприємства в цілому, так і його структурних підрозділів;

- точно та своєчасно знаходити й ураховувати фактори, що впливають на одержуваний прибуток від певного виду діяльності, виду виготовленої продукції або послуг;

- визначати витрати на виробництво (витрати виробництва) і тенденції їх зміни, що є необхідним для розробки цінової політики підприємства;

- знаходити оптимальні шляхи вирішення проблем підприємства і отримання прибутку в найближчій і віддаленій перспективах.

Оцінку діяльності підприємства і його економічної ефективності неможливо зробити одним будь-яким показником. Різноманіття властивостей і ознак різних видів виробничо-господарської та

комерційної діяльності підприємства обумовлює й різноманіття показників. При цьому проблема їх використання полягає в тому, що жоден з них не виконує роль універсального показника, за значенням якого однозначно можна було б робити висновки щодо успіху або невдачі в бізнесі. Крім того, ринкові відносини припускають, що в кожній сфері бізнесу повинні бути свої показники (часто ніде більше не вживані). Так, у багатьох галузях харчової промисловості, наприклад, консервній і цукровій, велике значення має ступінь переробки сировини, використання вторинної сировини тощо.

Тому на практиці завжди використовують систему показників, які пов'язані між собою й оцінюють або відображають різні сторони діяльності підприємства.

Усі показники, виходячи з вимог ринку, можна розділити на дві групи:

- **оціночні показники**, які характеризують досягнутий чи можливий рівень розвитку або результатів діяльності;
- **витратні показники**, які відображають рівень витрат по здійсненню різних видів діяльності.

Такий розподіл є досить умовним, адже він залежить від мети проведення аналізу. Наприклад, показник «витрати виробництва» в одному випадку може розглядатися як оціночний, що характеризує досягнутий рівень витрат праці, а в іншому випадку (наприклад, при плануванні) – визначатися як витратний, що дозволяє встановити кількість витрат праці на виготовлення продукції або надання послуг. Те ж саме можна сказати про значущість показників, яка багато в чому залежить від характеру діяльності підприємства. Наприклад, показник прибутку, незважаючи на всю його важливість, далеко не всім сторонам (учасникам) є важливим в однаковій мірі: орендодавця більше цікавить рух ліквідності в компанії, а акціонерів цікавить не тільки розмір дивідендів, а й курс акцій, який залежить від темпів зростання обсягу їх продажів.

Залежно від мети аналізу показники можуть виражатися у формі *абсолютних, відносних і середніх* величин. Виділяють також *структурні та приростні* показники.

Абсолютні показники бувають вартісними і натуральними. В умовах ринкових відносин першорядне значення надається вартісним, що обумовлено сутністю товарно-грошових відносин. Абсолютні показники відображають рівень розвитку підприємства, досягнутий за певний період часу. Ними є: оборот (обсяг продажів),

валова і чиста виручка, валовий і чистий прибуток, величина дивідендів, рівень витрат виробництва та реалізації продукції, основні і оборотні виробничі засоби, статутний капітал, суми заборгованості тощо.

Відносні показники розраховуються як співвідношення абсолютних показників, що характеризують частку одного показника в іншому, або як співвідношення різнорідних показників. Процедура їх оцінки полягає в порівнянні звітних значень з базисними (плановими), середніми за попередній період, звітними за попередні періоди, середньогалузевими, показниками конкурентів. До них відносяться: прибуток на одиницю вартості основних засобів, витрат або статутного капіталу; продуктивність; фондоозброєність праці та багато інших показників.

Структурні показники характеризують частку окремих елементів у загальній сумі.

Приростні показники відображають зміну показника за певний період часу. Вони можуть бути подані у відносному або абсолютному вираженні. Такими є, наприклад, зміна статутного капіталу за рік, прибутку за рік та ін.

Отже, під час оцінки ефективності діяльності підприємства ми будемо мати справу з різноманітними й різнорідними показниками, причому в одному і тому ж випадку деякі з них можуть поліпшуватися, а інші – погіршуватися. Наприклад, збільшення прибутку при продажах в кредит (у випадку затримки з виплатами) веде одночасно й до зменшення готівки. В умовах ринку до числа контрольованих показників входять: виручка від реалізації, обсяг продажів, розмір капіталу, сума чистого прибутку, величина активів, число акціонерів, величина виплачуваних дивідендів, частка експорту в обороті тощо.

Як ми вже відмічали раніше, ефективність діяльності підприємства визначається за допомогою низки показників. Оцінити ефективність його роботи тільки за результатами економічних досягнень, як то прибутку, отриманого в результаті діяльності, є недостатнім. Величезний прибуток можна отримати як на основі жорсткої експлуатації працівників, так і на основі сучасних методів організації виробництва з використанням соціально-психологічних факторів. Особливо важливим є вивчення другого, гуманного шляху досягнення ефективності, що можна зробити комплексно, за показниками двох блоків.

Перший блок включає в себе економічні показники, які ми й будемо далі вивчати. До них можна віднести:

– показники ефективності праці, за якими оцінюють особливості використання робочої сили та виявляють, чи досягнута мета за допомогою мінімальних трудових витрат;

– показники використання основних засобів, що оцінюють їх ефективність;

– показники використання оборотних засобів, які характеризують ефективність використання оборотного капіталу підприємства;

– показники рентабельності, що слугують мірою прибутковості в ході всього кругообігу: від придбання сировини та обладнання до продажу виготовленої продукції;

– показники енергоємності, які характеризують витрати підприємства на виготовлення продукції;

– показники екологічності, що характеризують ступінь безпеки роботи на підприємстві та міру забруднення навколишнього середовища. В умовах сьогодення забруднення навколишнього середовища є одним із важливих показників неефективності промислового підприємства.

До показників *другого блоку* оцінки ефективності підприємства, які характеризують моральну сторону виробництва, відносять:

– трудову, духовну та громадську активність співробітників;

– задоволеність трудовою діяльністю;

– стабільність функціонування підприємства;

– спрацьованість колективу підприємства, що характеризує стійкість і міцність міжособистісних взаємодій.

9.2. Система показників ефективності господарської діяльності підприємства

У даному параграфі ми більш детально зупинимося на показниках першого блоку і почнемо наш розгляд із ***показників ефективності праці***.

Останнім часом у публікаціях з управління персоналом і економіки підприємства спостерігається така тенденція: ефективний персонал визнається важливим ресурсом, який забезпечує стійке існування та розвиток підприємства. Для того щоб діяльність персоналу підприємства була більш результативною і

використовувалася досить ефективно, необхідно здійснювати оцінку ефективності праці із розрахунком економічних показників і винагородою за їх досягнення.

Відзначимо, що оцінка ефективності праці є однією з найважливіших функцій управління персоналом, а її роль у системі управління підприємством полягає в тому, що саме на її основі управлінський суб'єкт приймає відповідні рішення, які безпосередньо впливають на об'єкт – персонал підприємства. Від того, наскільки ця інформація якісна і надійна, залежить ефективність прийнятого рішення [22, с.47].

Основні показники, які використовуються для оцінки ефективності праці на підприємстві, особливості їх розрахунку та стисла характеристика наведені в табл. 9.1.

Таблиця 9.1

Показники ефективності праці

Показник	Розрахунок	Характеристика
Виробіток (q)	$q = \frac{Q}{T},$ де Q – обсяг виготовленої продукції; T – час (година, день, тиждень тощо) або середньооблікова чисельність робітників.	Виробіток характеризує кількість продукції, виготовленої за одиницю робочого часу або на одного середньооблікового працівника. Це найбільш поширений і універсальний показник праці. Залежно від одиниці виміру робочого часу розрізняють показники виробітку: <ul style="list-style-type: none"> - за одну відпрацьовану людино-годину; - за один відпрацьований людино-день; - на одного середньооблікового робітника за рік, квартал або місяць або на одного працівника за ті ж періоди часу.
Темп зростання (Tr) (приросту Tpr) продуктивності праці	$Tr = \frac{q_1}{q_0},$ $Tpr = \left(\frac{q_1}{q_0} - 1\right)100,$ де q_1, q_0 – продуктивність праці поточного та базисного періоду відповідно.	Темпи зростання та приросту продуктивності праці характеризують відносне зростання показника в поточному періоді порівняно з базисним.

Продовження таблиці 9.1

Показник	Розрахунок	Характеристика
Трудомісткість (t)	$t = \frac{T}{Q}$	<p>Трудомісткість характеризує витрати робочого часу на виробництво одиниці продукції. Цей показник має низку переваг перед показником виробітку, оскільки встановлює пряму залежність між обсягом виробництва і трудовими витратами. Залежно від складу витрат, що включаються в трудомісткість продукції, розрізняють наступні її види:</p> <ul style="list-style-type: none"> - технологічна трудомісткість (витрати праці основних робітників); - трудомісткість обслуговування виробництва (витрати праці допоміжних робітників); - виробнича трудомісткість (витрати праці основних і допоміжних робітників); - трудомісткість управління виробництвом (витрати праці керівників, фахівців і службовців); - повна трудомісткість (витрати праці всього промислово-виробничого персоналу).
Темп росту (Tp) (приросту Tnp) продуктивності праці	$Tp = \frac{q_1}{q_0},$ $Tnp = \left(\frac{q_1}{q_0} - 1\right)100,$ <p>де q_1, q_0 – продуктивність праці поточного та базисного періоду відповідно.</p>	<p>Темпи росту та приросту продуктивності праці характеризують відносне зростання показника в поточному періоді порівняно з базисним.</p>
Частка приросту виготовленої продукції за рахунок підвищення продуктивності праці (d)	$d = \frac{\Delta Q(q)}{\Delta Q},$ <p>де $\Delta Q(q), \Delta Q$ – абсолютний приріст продукції відповідно за рахунок зміни продуктивності та загальний.</p>	<p>Показник характеризує питому вагу абсолютного приросту обсягу виробленої продукції в загальному обсязі абсолютного приросту продукції. Чим більшою є значення частки, тим більш ефективним є виробництво.</p>

Продовження таблиці 9.1

Показник	Розрахунок	Характеристика
Коефіцієнт використання корисного (максимально можливого) фонду часу (K)	$K = \frac{T_m}{T_k} 100,$ <p>де T_m, T_k – максимально можливий і календарний фонд часу відповідно.</p>	<p>Максимально можливий фонд робочого часу – це максимальний робочий час, який може бути відпрацьований на підприємстві у звітному періоді відповідно до трудового законодавства. Календарний фонд – це увесь фонд робочого і неробочого часу на підприємстві при даній чисельності працівників. Дані коефіцієнти показують на скільки відсотків був використаний максимально можливий або календарний фонд часу.</p>
Відносне вивільнення (економія) робітників ($E_{від}$)	$E_{від} = Q_1 : q_0 - Q_1 : q_1,$ $E_{від} = T_0 \cdot I_Q - T_1,$ <p>де 0 і 1 відносяться до базисного і звітного періоду відповідно.</p>	<p>Показник характеризує, скільки робітників були вивільнені з процесу виробництва за рахунок збільшення продуктивності праці.</p>

Основними засобами підвищення значень показників і, відповідно, ефективності праці є:

- покращення мотивації персоналу;
- покращення робітничих практик і організації праці;
- удосконалення процесів виробництва шляхом покращення операційних процесів;
- оптимізація взаємодії допоміжних технологічних процесів з основними;
- використання практик «Мотивація та надання можливостей» і «Розвиток лідерства та управління змінами»;
- використання практики організації робочого місця й робочого процесу «5 S»;
- введення практики «Управління процесами виробництва», що включає: участь і спільну відповідальність операторів, мінімізацію часу на переорієнтацію обладнання, розробку карти потоків, створення ланцюгів доданої вартості, управління інформацією тощо;
- прогнозування та планування діяльності тощо.

Упровадження лише деяких із цих засобів, згідно з концепцією «Філософії виробництва світового класу», вже у найближчому майбутньому продемонструє суттєві результати.

Наступним блоком показників оцінки ефективності, які ми розглянемо, є **показники оцінки ефективності використання основних засобів**. Основні засоби є частиною засобів виробництва підприємства, що беруть участь у процесі виробництва протягом тривалого часу, зберігаючи при цьому свою натуральну форму, їх вартість переноситься на продукцію поступово, по мірі використання.

Ефективність використання основних засобів підприємства характеризується показниками завантаження, які відображають ступінь використання виробничих потужностей за часом і обсягами продукції, що випускається. У фінансовому аналізі найбільш часто застосовуються показники оцінки ефективності використання машин і устаткування, як найбільш активної частини основних засобів підприємства (табл. 9.2).

Таблиця 9.2

Показники ефективності використання основних засобів

Показник	Розрахунок	Характеристика
Фондовіддача (загальна), Φ_3	$\Phi_3 = \frac{Q}{\overline{OB\Phi}}$ де Q – загальний обсяг виготовленої (реалізованої) продукції протягом року; $\overline{OB\Phi}$ – середньорічна вартість основних виробничих засобів	<p>Показник характеризує скільки гривень виготовленої (реалізованої) продукції припадає в середньому на одну гривню основних виробничих засобів. Він не має загальноприйнятого «нормального» значення, оскільки сильно залежить від галузевих особливостей. Наприклад, у фондоемних виробництвах частка основних засобів у активах підприємства велика, тому показник буде нижчим. Якщо розглядати показник фондовіддачі в динаміці, то його зростання свідчить про підвищення інтенсивності (ефективності) використання устаткування. Для того, щоб підвищити фондовіддачу потрібно або збільшити виручку при використанні вже наявного устаткування, або позбутися від зайвого устаткування, зменшивши таким чином його вартість.</p>

Продовження таблиці 9.2

Показник	Розрахунок	Характеристика
Фондовіддача активної частини основних виробничих засобів (Φ_a)	$\Phi_a = \frac{Q}{\overline{OB\Phi_a}}$ <p>де $\overline{OB\Phi_a}$ – середньорічна вартість активної частини основних виробничих засобів</p>	Показник характеризує скільки гривень виготовленої (реалізованої) продукції припадає в середньому на одну гривню активної частини основних виробничих засобів.
Фондоємність продукції (Φ_e)	$\Phi_e = \frac{\overline{OB\Phi}}{Q}$	<p>Показник характеризує скільки гривень основних виробничих засобів припадає в середньому на одну гривню виготовленої (реалізованої) продукції.</p> <p>Фондоємність показує ступінь використання основного капіталу у виробництві продукції, вона знаходить широке застосування при створенні динамічної моделі балансу між великими галузями.</p> <p>Виробництво продукції на багатьох підприємствах залежить від поставок комплектуючих від інших підприємств-виробників, випуск продукції яких також пов'язаний із використанням основних засобів. У цьому випадку виділяють таке поняття, як повна фондоємність, яка включає в себе пряму фондоємність (характеризує величину вкладеного капіталу на придбання основного технологічного устаткування для конкретного підприємства, ефективність роботи якого оцінюється) та непряму фондоємність (капітал, вкладений у придбання основних засобів виробництва підприємства-партнера).</p> <p>Якщо у підприємства, для якого визначається критерій ефективності, кілька партнерів, то розраховується непряма фондоємність кожного з них.</p>

Продовження таблиці 9.2

Показник	Розрахунок	Характеристика
Фондоозброєність праці (Φ_o)	$\Phi_o = \frac{\overline{OB\Phi}}{T},$ де T – середньооблікова чисельність робітників (працівників).	Показник характеризує вартість основних засобів, що припадає в середньому на одного робітника (працівника) підприємства, тобто відображає ступінь забезпеченості персоналу основними засобами виробництва. Зростання фондоозброєності праці відображає заміщення живої праці технікою, ліквідацію ручних процесів, підвищення ступеня механізації та комплексної автоматизації виробництва.
Коефіцієнт інтенсивності використання устаткування (K_I)	$K_I = \frac{Q}{\Pi},$ де Π – установлена потужність устаткування.	Коефіцієнт характеризує повноту використання потужності устаткування, фактичне виробництво продукції залежно від потенційних його можливостей з урахуванням специфіки виробництва. Інтенсивне завантаження основних засобів призводить до зниження собівартості продукції (за рахунок скорочення всіх постійних витрат) і зростання продуктивності праці.

Будь-яке підприємство прагне до підвищення ефективності використання своїх основних виробничих засобів. Це є запорукою збільшення виробництва продукції, що в кінцевому підсумку веде до збільшення доходів, а, отже, до підвищення рівня рентабельності. Тому проблема максимальної ефективності використання основних засобів повинна бути однією з ключових для будь-якого підприємства. Крім того при ефективному використанні основних засобів зменшується потреба в них, що веде до економії, тобто мінімізації виробничих витрат.

Для підвищення ефективності використання основних засобів і зменшення ступеня їх зносу підприємство може використовувати наступні шляхи:

- збільшення змінності роботи устаткування;
- збільшення ступеня використання потужності устаткування;
- скорочення простоїв устаткування я;

- підвищення кваліфікації користувачів устаткування;
- підвищення якості організації догляду за основними засобами;
- своєчасне проведення капітального ремонту, а також планово-попереджувальних заходів;
- підвищення якості трудової дисципліни;
- оптимальний склад верстатного парку;
- забезпечення виробництва необхідною кількістю робітників;
- забезпеченість матеріалами та інвентарем.

Ще однією важливою умовою підвищення економічної ефективності використання основних засобів є вдосконалення їх структури, яка визначається економічними та природними умовами, спеціалізацією, а також загальним рівнем розвитку матеріально-технічної бази на підприємстві. Дотримання певних пропорцій між елементами основних засобів, обумовлених їх роллю у виробництві – запорука ефективного функціонування.

Третя група показників, що характеризують ефективність діяльності підприємства, представлена показниками **ефективності використання оборотних засобів**. **Оборотні виробничі засоби** підприємства – це частина засобів виробництва, які повністю споживаються в кожному циклі виробництва, переносять свою вартість на вироблену продукцію і цілком відшкодовуються після кожного виробничого циклу.

У зв'язку з тим, що оборотні засоби формують основну частку ліквідних активів будь-якого підприємства, їх величина повинна бути достатньою для забезпечення його ритмічної та рівномірної роботи і, як наслідок, отримання прибутку. Використання оборотних засобів у операційній діяльності має здійснюватися на рівні, що мінімізує час і максимізує швидкість обігу оборотного капіталу і перетворення його в реальну грошову масу для подальшого фінансування та придбання нових оборотних засобів. Потреба у фінансуванні пропорційно залежить від швидкості обертання активів.

У практиці аналізу, для оцінки ефективності використання оборотних засобів, використовують різноманітні коефіцієнти оборотності, які можуть бути визначені як для всього оборотного капіталу підприємства в цілому, так і за окремими його складовими (табл. 9.3).

Таблиця 9.3

Показники ефективності використання оборотних засобів

Показник	Розрахунок	Характеристика
Коефіцієнт оборотності оборотних засобів ($K_{об}$)	$K_{об} = \frac{Q}{\overline{ЗОБ}}$, де $\overline{ЗОБ}$ – середні залишки оборотних засобів за період	Коефіцієнт показує кількість обертів здійснюють оборотні кошти протягом аналізованого періоду. Він характеризує саму сутність процесу виробництва або торгівлі, яка полягає в наступному циклі: «Кошти – Товар для реалізації – Кошти». Швидкість цього перетворення, а точніше, скільки разів за період відбувається один оберт із коштів у товар і знову в кошти і є оборотністю. Чим вона вища, тим швидше окупаються вкладені кошти, тим швидше підприємство отримує прибуток й ефективніше працює.
Коефіцієнт закріплення оборотних засобів ($K_з$)	$K_з = \frac{\overline{ЗОБ}}{Q}$	Коефіцієнт характеризує скільки гривень оборотних коштів припадає в середньому на гривню реалізованої продукції. Чим меншим є коефіцієнт закріплення, тим ефективніше використовуються оборотні кошти.
Тривалість одного обороту ($T_{об}$)	$T_{об} = \frac{365(90)}{K_{об}}$	Показник характеризує за який термін на підприємство повертаються його оборотні кошти у вигляді виручки від реалізації продукції. Це один із найважливіших показників, який показує, через скільки днів підприємство побачить вкладені в оборот кошти у вигляді грошової виручки і зможе їх використовувати. Виходячи з цього, можна планувати і здійснення платежів, і розширення виробництва.
Матеріалоємність продукції (M_e)	$M_e = \frac{\sum M}{Q}$, де М – сума матеріальних витрат, включених у собівартість реалізованої продукції	Показник характеризує сумарні витрати матеріалів на виробництво одиниці продукції. Розраховують декілька показників матеріалоємності: питому, структурну та абсолютну.

Продовження таблиці 9.3

Показник	Розрахунок	Характеристика
		Остання з них дозволяє визначити норму витрат ресурсів на виробництво одиниці готової продукції, її чисту вагу та ступінь витрат запасів. Структурна матеріалоемність показує частку окремих виробів у загальному показнику матеріалоемності. А питомий тип цього показника є структурним варіантом, який був приведений до натуральної загальної одиниці вимірювання. Його використовують тільки для виробництва однієї групи продукції.
Матеріаловіддача (M_6)	$M_6 = \frac{Q}{\sum M}$	Показник характеризує обсяг виготовленої продукції на одну гривню матеріальних ресурсів. Чим більшим є показник матеріаловіддачі, тим менше необхідно матеріалу для випуску одного і того ж обсягу продукції. Збільшити матеріаловіддачу можна за допомогою впровадження прогресивних технологій і техніки у виробництво, посилення контролю за ощадливим використанням матеріалів.

У практиці управління підприємства розрізняють безліч способів підвищення ефективності використання оборотних засобів, до основних з яких відносять:

- зниження обсягів дебіторської заборгованості;
- ритмічне виконання планів господарської діяльності;
- удосконалення організації збуту продукції підприємства;
- упровадження прогресивних методів продажу продукції;
- упорядкування збору та зберігання порожньої тари, прискорення повернення тари постачальникам;
- удосконалення розрахунків з постачальниками і покупцями;
- покращення претензійної роботи;

– прискорення обертання грошових коштів за рахунок поліпшення інкасації торгової виручки, суворого лімітування залишків грошових коштів у касах торгових підприємств, у дорозі, на розрахунковому рахунку в банку;

– приведення до мінімуму запасів господарських матеріалів, малоцінних і швидкозношуваних предметів, інвентарю, спецодягу на складі, скорочення підзвітних сум, витрат майбутніх періодів тощо.

Ефективне використання оборотних засобів вивільняє грошові кошти, що направляються на розширення виробництва і зростання добробуту працівників підприємств.

Прискорення обертання оборотних засобів дозволяє вивільнити значні суми, і таким чином, збільшити обсяг виробництва без додаткових фінансових ресурсів, а вивільнені кошти використати відповідно до потреб підприємства.

Наступна група показників ефективності – **узагальнюючі показники** – характеризують ефективність господарської діяльності в цілому або ефективність використання сукупності ресурсів. До цієї групи відносять наступні типи показників (табл. 9.4).

Таблиця 9.4

Узагальнюючі показники ефективності

Показник	Розрахунок	Характеристика
Витрати на одну гривню реалізованої продукції (h)	$h = \frac{zQ}{pQ},$ де zQ – витрати на виробництво і реалізацію продукції; pQ – обсяг реалізованої продукції.	Показник характеризує скільки гривень витрат припадає у середньому на гривню реалізованої продукції.
Рентабельність продукції (R_n)	$R_n = \frac{\Pi}{zQ},$ де Π – прибуток від реалізації продукції.	Показник характеризує скільки гривень прибутку отримує в середньому підприємство з кожної гривні, витраченої на виробництво і реалізацію продукції. Його розраховують по підприємству в цілому і по окремих підрозділах чи видах продукції. Негативна динаміка показника свідчить про необхідність перегляду цін або посилення контролю за витратами на виробництво та реалізацію продукції.

Продовження таблиці 9.4

Показник	Розрахунок	Характеристика
		Збільшення цього показника в динаміці при незмінній величині витрат свідчить про підвищення обсягу реалізації, а отже, про збільшення прибутку, і навпаки.
Рентабельність виробництва (загальна) R_e	$R_e = \frac{БПр}{\overline{ОВФ} + \overline{ОВЗ}},$ <p>де БПр – балансовий прибуток; $\overline{ОВФ}$ – середня вартість основних виробничих засобів; $\overline{ОВЗ}$ – середня вартість оборотних виробничих засобів.</p>	Показник характеризує прибутковість виробничої діяльності за певний період часу. Рентабельність виробництва співставляє величину отриманого прибутку і розміру коштів, які дозволили його отримати, показує суму прибутку в розрахунку на одну гривню витрачених виробничих засобів. Чим менше засобів використано для отримання певної суми прибутку, тим вище рентабельність виробництва, а отже, вище ефективність діяльності підприємства.
Рентабельність продажів (R_{np})	$R_{np} = \frac{П}{Q}$ <p>де П – чистий прибуток підприємства (або прибуток від реалізації продукції)</p>	Показник характеризує суму прибутку, яку отримує підприємство з однієї гривні реалізованої продукції. Він дозволяє охарактеризувати найголовніше для підприємства – реалізацію основної продукції. Крім цього оцінюється частка собівартості в процесі продажів. Знаючи рентабельність продажів, підприємство може контролювати цінову політику і витрати. Варто зауважити, що різні підприємства виробляють товари за допомогою різних стратегій і технік, що викликає відмінність рівнів рентабельності. Але навіть якщо показники виручки, операційні витрати, а також прибуток до вирахування податків рівні у двох компаній, їх показник рентабельності продажів буде відрізнятися. Виною тому безпосередній вплив величини процентних виплат на повний обсяг чистого прибутку.

Показники рентабельності є найважливішими показниками ефективності діяльності підприємства. Будучи загальними показниками, вони найбільш повно і всебічно характеризують ефективність його діяльності в цілому.

Перейдемо до наступної групи показників. Вимірювання та оцінка **енергоефективності** також є необхідною частиною аналізу ефективності діяльності підприємства. Для того, щоб оцінити енергоефективність діяльності суб'єкта підприємництва виділяють два кола показників: *вартісні* та *натуральні*. Вартісні для оцінки або вимірювання енергоефективності є найбільш зручними та універсальними, але лише за умови незмінності цінової і тарифної політики. Однак у сучасних економічних умовах при визначенні системи показників енергоефективності доцільно приділяти не менше уваги натуральному вимірюванню. Основні показники енергоефективності діяльності підприємства наведені в табл. 9.5.

Таблиця 9.5

Показники оцінки енергоефективності

Показник	Розрахунок	Характеристика
Енергоекономічний рівень виробництва (E_{pv})	$E_{pv} = \frac{Qp}{E_c}$ <p>де Qp – обсяг виробленої продукції у вартісному вираженні; E_c – сумарний обсяг спожитої енергії на технологічні цілі в грн.</p>	Показник характеризує випуск продукції у вартісному вираженні на одну гривню використаної енергії. Він дозволяє оцінити рівень реалізації енергозберігаючих технологій, економічних теплових схем, енергозберігаючого обладнання тощо.
Втрати енергії (B)	$B = E_{пер} - E_{спож}$ <p>де $E_{пер}$ – підведена первинна енергія; $E_{спож}$ – спожита корисна енергія</p>	За оцінками спеціалістів, сумарні втрати при виробництві і транспортуванні теплової енергії складають до 60%. При виробленні та транспортуванні електроенергії втрати істотно нижчі і в середньому складають 20%. Якщо підприємству вдасться знизити рівні цих втрат, то його виробництво стане в декілька разів ефективніше.

Продовження таблиці 9.5

Показник	Розрахунок	Характеристика
Питома енергоємність продукції (ϖ)	$\varpi = \frac{E_c}{Q_p}$	Показник характеризує витрати енергії (енергоресурсів і енергоносіїв) на виробництво одиниці продукції.
Інтегральний коефіцієнт корисного використання енергії та енергоносіїв ($ККВ$)	$ККВ = \eta_{pn} \cdot \eta_y,$ де η_{pn} – середньозважений ККД розподілу і перетворення енергії на підприємстві; η_y – середньозважений за часткою в загальному енергоспоживанні ККД споживчих установок.	Показник характеризує ефективність процесу перетворення та передачі енергії підприємства (установки, підрозділу, обладнання).
Енергоозброєність (E_{oz})	$E_{oz} = \frac{P_e}{ОВФ},$ де P_e – сумарна потужність енергоустановок підприємства; $ОВФ$ – сумарна вартість основних виробничих засобів.	Показник характеризує зв'язок витрат живої праці з виробничим споживанням механічної та електричної енергії, що замінює застосування фізичної сили робітників. Підвищення значення показника є однією з основних умов науково-технічного прогресу на виробництві та зростання продуктивності праці.
Коефіцієнт екстенсивного використання устаткування (K_E)	$K_E = \frac{T_\phi}{T_K},$ де T_ϕ – час фактичної роботи устаткування, T_K – календарний фонд часу устаткування.	Коефіцієнт характеризує, яка частина усього можливого часу була фактично використана для роботи енергетичного устаткування.

Продовження таблиці 9.5

Показник	Розрахунок	Характеристика
Коефіцієнт інтенсивного використання устаткування (K_I)	$K_I = \frac{\overline{P}_\phi}{P_{E\phi}},$ <p>де \overline{P}_ϕ – середня фактична потужність двигуна, $P_{E\phi}$ – потужність ефективна максимально тривала.</p>	Коефіцієнт показує, як використовувалася протягом роботи потужність устаткування, оскільки будь-яке обладнання може працювати зі змінним навантаженням.
Інтегральний коефіцієнт використання устаткування (K_{II})	$K_{II} = \frac{E_\phi}{E_{MM}},$ <p>де E_ϕ – фактично вироблена (спожита) енергія, E_{MM} – енергія максимально можлива, яка могла бути вироблена (спожита) за умови повного використання ефективної максимально тривалої потужності протягом усього календарного фонду.</p>	Показник дає зведену характеристику використання устаткування, він дає уявлення, яка частина максимально можливої кількості енергії була фактично виготовлена (спожита) протягом звітного періоду.

Для того, щоб підвищити рівень енергоефективності на підприємстві, можна здійснити наступні заходи:

- провести утеплення зовнішніх огорожувальних конструкцій будівель підприємства;
- здійснити утеплення суміщених покрівель або горищних перекриттів;
- упровадити автоматизовані вузли управління теплоспоживанням будівель;
- упровадити індивідуальне регулювання тепловіддачі кожного опалювального приладу за допомогою термостатів;

- використовувати альтернативні джерела енергії;
- збільшити ефективність використання енергоресурсів на одиницю продукції підприємства;
- зменшити фінансові витрати за рахунок зниження плати за енергоресурси, паливо;
- отримувати додатковий прибуток за рахунок зменшення плати за енергоресурси та збільшення регіональних і місцевих бюджетів за рахунок додаткових надходжень податків.

Остання група показників економічного блоку оцінки ефективності діяльності підприємства представлена **показниками оцінки екологічності виробництва продукції**. Предметом екологічної оцінки ефективності є вплив діяльності на навколишнє середовище. Під цим впливом ми розуміємо ті зміни у навколишньому середовищі, які повністю або частково є результатом запланованої діяльності підприємства. Екологічна оцінка дозволяє виявляти той вплив, який, можливо, був невірно врахований національними стандартами, встановленими для окремих середовищ і джерел впливу.

Підвищення екологічної ефективності на підприємстві сприяє поліпшенню якості його продукції, підвищує продуктивність праці робітників, знижує негативний вплив на навколишнє середовище, що у свою чергу, позбавляє підприємство від сплати великої кількості штрафів державі.

Показники екологічної ефективності поділяються на три групи.

1. Показники ефективності управління підприємством, що демонструють ефективність його політики.

2. Показники ефективності функціонування, що відображають екологічні аспекти, пов'язані з функціонуванням підприємства, включаючи його діяльність, продукцію, послуги.

3. Показники стану навколишнього середовища, які дають інформацію про стан довкілля в місті, регіоні, де розташоване підприємство.

До показників *першої групи* відносять:

- кількість днів навчання з питань екології на одного працівника на рік;
- кількість аварійних ситуацій, що мали місце протягом року;

– абсолютна величина ризику, якого вдалось уникнути внаслідок запобігання аварії (розраховується як добуток імовірності аварії на величину фінансового збитку).

– Показниками *другої групи* є:

– кількість сировинних матеріалів, повторно використовуваних у виробничому процесі;

– кількість пакувальних матеріалів повторно використовуваних, або що припадають на одиницю продукції;

– середні витрати палива на одиницю пробігу транспортного засобу;

– кількість транспортних засобів, оснащених технологічними пристроями очищення вихлопів.

До *третьої групи* показників відносять показники стану навколишнього середовища, що містять інформацію про місцеві, регіональні, національні або глобальні умови навколишнього середовища. До глобальних показників можна віднести показники, які відображають вплив діяльності підприємства на товщину озонового шару, середню глобальну температуру і розмір рибних ресурсів в океані. Для національних і регіональних показників використовують комплексний індекс забруднення атмосфери, що враховує декілька домішок, і найбільшу виміряну разову концентрацію домішок, поділену на гранично допустиму концентрацію.

Варто відмітити, що під час вибору показників, що описують екологічну ефективність діяльності підприємства, необхідно відштовхуватись від національних і міжнародних екологічних стандартів. Незалежно від наявності або відсутності на підприємстві системи управління навколишнім середовищем, підприємство повинно планувати оцінювання екологічної ефективності відповідно до встановлених критеріїв так, щоб отримані значення обраних показників можна було зіставляти з цими критеріями. Для оцінки критеріїв екологічної ефективності враховують:

– поточні та минулі характеристики;

– вимоги законодавства;

– визнані регламенти, стандарти та кращу практику;

– дані про ефективність і інформацію, розроблену промисловістю і організаціями інших галузей;

– результати розгляду керівництвом і аудитів;

- думки зацікавлених сторін;
- наукові дослідження.

Наведені показники пропонується використовувати в якості базової системи показників, у яку залежно від обставин можна включати додаткові приватні критерії екологічної ефективності, пов'язані з іншими аспектами діяльності.

Для того, щоб підвищити екологічну ефективність підприємства, необхідно виконувати цілу низку корисних дій. Перш все, потрібно проводити заходи щодо забезпечення, моніторингу та контролю чистоти повітря. До цих дій відносяться заходи з технологічної модернізації виробничих ділянок, заміна застарілого устаткування та поліпшення виробничого процесу, використання сировини, що обумовлює менші викиди, реконструкція вентиляційних систем, оснащення обладнання підприємства газоочисними установками.

Наступне коло заходів передбачає раціональне використання водних ресурсів, їх моніторинг і контроль. Основними діями в цьому напрямку виступають: встановлення очисних споруд, фільтрів, насосів і створення водооборотних систем охолодження технологічного устаткування.

Найбільшу увагу необхідно приділити управлінню промисловими відходами. Для цього підприємству варто вести контроль і облік кількості та складу відходів, організувати їх безпечний роздільний збір, організувати тимчасове безпечне зберігання відходів, переробляти, очищувати та повторно використовувати відходи у виробництві, розробляти та втілювати заходи по зменшенню кількості відходів [3, 8, 18].

9.3. Приклад аналізу показників ефекту та ефективності

У попередніх параграфах було розглянуто групи показників, які використовують для аналізу ефективності діяльності підприємства. Варто відмітити, що на перших ступенях фінансового аналізу є зовсім не обов'язковим розраховувати їх всі – на перших етапах дослідження деяких із них буде більш ніж достатньо. Тим не менше, важливо використовувати як відносні, так і абсолютні показники одночасно, оскільки зростання абсолютних значень іноді зовсім не означає суттєве покращення ситуації, оскільки відносні показники можуть зменшуватися, і навпаки.

Розглянемо таку ситуацію на первісному аналізі двох підприємств, що виготовляють промислову продукцію. У табл. 9.6 наведені показники, що характеризують результати діяльності кондитерського підприємства «Summer meadow» за два роки (рядки 1-6 таблиці).

Таблиця 9.6

**Показники діяльності кондитерського підприємства
«Summer meadow», 2016-2017 рр.**

№	Показник	2016 р.	2017 р.	Абсолютний приріст	Темп приросту, %
<i>A</i>	<i>I</i>	<i>2</i>	<i>3</i>	<i>4</i>	<i>5</i>
1	Виручка від реалізації продукції (чистий дохід), тис. грн.	205811,8	156800,1	-49011,7	-23,8
2	Повна собівартість реалізованої продукції, тис. грн.	164893,7	121104,9	-43788,8	-26,6
3	Прибуток від реалізації продукції, тис. грн.	40918,1	35695,2	-5222,9	-12,8
4	Середньорічна вартість основних засобів, тис. грн.	29544,7	57524,1	27979,4	94,7
5	Середньооблікова кількість працівників, чол.	911	723	-188	-20,6
6	Фонд оплати праці, тис. грн.	41433,9	41698,2	264,3	0,6
7	Середньомісячна заробітна плата одного працівника, грн.	3790,15	4806,15	1016,0	26,8
8	Продуктивність праці, тис. грн./чол.	225,9	216,9	-9,0	-4,0
9	Фондоозброєність праці, тис. грн./чол.	32,4	79,6	47,2	145,7
10	Фондовіддача основних засобів, грн./грн.	6,97	2,73	-4,24	-60,8
11	Витрати на одну гривню реалізованої продукції, грн.	0,80	0,77	-0,03	-3,8
12	Рентабельність реалізованої продукції, %	24,8	29,5	4,7	19,0

На основі формул, наведених у табл. 9.1-9.5, нами було розраховано показники середньомісячної заробітної плати одного

працівника, продуктивності праці, фондоозброєності та фондівіддачі, витрати на одну гривню реалізованої продукції, а також рентабельність реалізованої продукції (рядки 7-12). Для того, щоб показати зміну цих показників у часі, ми доповнили таблицю розрахунками абсолютного приросту та темпу приросту (стовпчики 4-5).

Як свідчать дані табл. 9.6, за 2017 рр. діяльність підприємства характеризується зменшенням як виручки від реалізації продукції, так і прибутку, що є негативними змінами. Тобто, економічний ефект має від'ємне значення. Однак, якщо проаналізувати динаміку відносних показників, ми можемо побачити, що рентабельність продукції зростає, а витрати на одну гривню реалізованої продукції падають, що говорить про підвищення ефективності діяльності «Summer meadow».

Бувають і протилежні ситуації. У табл. 9.7 наведені такі ж самі показники для іншого підприємства – підприємства з виробництва верстатів «Machinery».

Таблиця 9.7

**Показники діяльності верстатобудівного підприємства
«Machinery», 2016-2017 рр.**

№	Показник	2016 р.	2017 р.	Абсолютний приріст	Темп приросту, %
<i>A</i>	<i>I</i>	<i>2</i>	<i>3</i>	<i>4</i>	<i>5</i>
1	Виручка від реалізації продукції (чистий дохід), тис. грн.	1911	4902	2991	156,5
2	Повна собівартість реалізованої продукції, тис. грн.	1242	3711	2469	198,8
3	Прибуток від реалізації продукції, тис. грн.	669	1191	522	78,0
4	Середньорічна вартість основних фондів, тис. грн.	36789,5	33458,0	-3331,5	-9,1
5	Середньооблікова кількість працівників, чол.	72	40	-32	-44,4
6	Фонд оплати праці, тис. грн.	2715	2132	-583	-21,5
7	Середньомісячна заробітна плата одного працівника, грн.	3142,36	4441,67	1299,31	41,3

Продовження таблиці 9.7

<i>A</i>	<i>I</i>	2	3	4	5
8	Продуктивність праці, тис. грн./чол.	26,5	122,6	96,0	362,6
9	Фондоозброєність праці, тис. грн./чол.	511,0	836,5	325,5	63,7
10	Фондовіддача основних засобів, грн./грн.	0,05	0,15	0,1	200,0
11	Витрати на одну грн. реалізованої продукції, грн.	0,65	0,76	0,11	16,9
12	Рентабельність реалізованої продукції, %	53,9	32,1	-21,8	-40,4

Як видно з табл. 9.7, за аналізований період виручка від реалізації продукції зросла більш ніж у 2,5 рази, суттєво збільшився й прибуток підприємства – на 78%. Однак, при цьому зростанні абсолютних показників спостерігалися й негативні зміни – зросли витрати на 1 грн. реалізованої продукції, а рентабельність реалізованої продукції зменшилась більш ніж на 40%.

Отже обидва приклади свідчать про те, що на практиці існують ситуації, коли позитивні значення ефекту супроводжуються негативними значеннями ефективності та навпаки. Очевидно, що відносні показники, як, наприклад, рентабельність реалізованої продукції, враховують не лише зміну ефекту (прибутку), а також зміну витрат, які сприяли отриманню такого ефекту: на першому підприємстві разом із скороченням прибутку зменшилася також і повна собівартість реалізованої продукції, а на другому підприємстві прибуток зріс на 78 %, а повна собівартість при цьому зросла майже у 3 рази.

Усе викладене вище підтверджує необхідність використання в аналізі як відносних, так і абсолютних показників. Крім того, щоб прояснити ситуацію і визначити її причини необхідно проводити більш детальний аналіз.

9.4. Спеціальні показники ефективності діяльності підприємства

Розглянуті показники ефекту та ефективності є узагальнюючими показниками, які використовують на більшості промислових підприємств. Однак, для підприємств різних галузей економіки

можуть використовуватися певні спеціальні показники. Наприклад, під час оцінки ефективності діяльності підприємств будівництва додатково розраховують механоозброєність виробництва та механоозброєність праці. У сфері оптової та роздрібної торгівлі аналізують показник прибутку в розрахунку на 1 м² торгівельної або складської площі та прибуток на одного продавця. Транспортні підприємства досліджують величину інерції в обсязі перевезень, величину виникнення нового обсягу перевезень, величину зникнення старого обсягу перевезень та інтенсивного збереження обсягу перевезень.

Таким чином, залежно від галузі діяльності підприємства, перелік показників може розширюватися. Розглянемо дві окремі галузі, показники ефективності діяльності яких мають велику частку специфічних величин. Це галузі *Інтернет-продажів* та *інвестиційної діяльності*.

Почнемо із першої з них – Інтернет-торгівлі. В особливу категорію показників ефективності можна виділити показники, що розраховують власники Інтернет-магазинів. Основними з них є такі.

1. **Відвідуваність сайту.** Доцільно вимірювати денні, тижневі та місячні показники, що дозволяє виявляти сплески відвідуваності сайту й аналізувати їх причини, наприклад, пов'язані з національними святами, зміною погоди та відкриттям нового сезону. Вимірюючи цей показник протягом тривалого періоду часу, керівництво магазину може встановити цільовий показник середньої кількості відвідувань і розробляти стратегію для її досягнення.

Встановити планове значення кількості відвідувачів сайту в день можна й на основі аналізу конкурентів, адже деякі сайти та портали містять відкриті лічильники статистики. Якщо відмінність між аналізованими значеннями є надто великою, варто приділити увагу дослідженню каналів, що використовують конкуренти для залучення трафіку, тобто виявити, де саме вони розміщують рекламу.

2. **Кількість переглядів товарних сторінок.** При оцінці ефективності функціонування сайту магазину необхідно аналізувати й кожну окрему сторінку товару. Наприклад, якщо магазин пропонує певні товари за акцією або за певною знижкою, а кількість переглядів сторінок товарів цього виду нічим не відрізняється від середніх значень, це може означати, що клієнтам важко знайти ці сторінки, а саме проведення акції виявляється неефективним. У цьому випадку варто, наприклад, окрім звичайних сторінок товарів, розробити на

сайті окремий підрозділ, присвячений акціям, знижкам і спеціальним пропозиціям.

3. **Середній час перебування на сайті та середня кількість переглянутих сторінок.** Ці два показники можуть бути дуже показовими, адже на їх основі можна розрахувати середній час завантаження однієї сторінки сайту магазину. Якщо сторінка є дуже важкою та довго грузиться, це може відштовхнути потенціального покупця тривалим часом очікування, в результаті чого він звернеться до конкурентів. Для того, щоб підвищити ефективність роботи сайту магазину, варто зробити його лайт-версію, або, в світі існуючих тенденцій, версію для мобільних носіїв.

4. **Сторінки виходу,** тобто сторінки, після перегляду яких відвідувач покидає сайт Інтернет-магазину. Дуже важливо знати, коли саме клієнт втрачає бажання знаходитись на сайті, тому що це допоможе виявити проблеми в його розробці. Наприклад, про проблеми із оформленням замовлення свідчать виходи зі сторінок реєстрації, кошика та безпосереднього оформлення списку бажаних товарів. У цьому випадку варто внести деякі зміни в роботу сайту для його ефективної роботи.

5. **Канали залучення відвідувачів.** Відстеження джерел залучення відвідувачів – важлива метрика для оцінки ефективності роботи Інтернет-магазину, оскільки каналів для залучення може бути дуже багато і економічна ефективність від їх використання теж різна. Потрібно відстежувати не просто джерела залучення відвідувачів, а й їх віддачу. Якщо один з каналів показує хороше залучення відвідувачів на сайті та зростання продажів – варто збільшувати бюджет на цей канал. А якщо канал не ефективний, можна поміркувати щодо того, як мінімізувати витрати або ж оптимізувати роботу, наприклад, видаливши неефективні ключові слова з рекламної кампанії в Rambler, Google тощо або переглянути seo-ядро, по якому просувається сайт.

6. **Показник повернення відвідувачів.** Інформативним є не тільки аналіз нових клієнтів, але й кількості тих відвідувачів, що повернулися на сайт Інтернет-магазину. Ця абсолютна величина дозволяє оцінити наскільки сайт є цікавим для цільової аудиторії. Відмітимо, що процедура повернення відвідувачів завжди є більш простою, ніж процедура залучення нових. Для цього використовують інструменти збору e-mail-адрес користувачів, технологію ремаркетингу і заклики до дії вступити до груп магазинів у

соціальних мережах. Чим довше магазин працює на ринку Інтернет-торгівлі, тим вище повинен бути показник повернення відвідувачів. Якщо ж показник повернення не зростає, то відвідувач не знайшов на сайті те, що шукав, або умови покупки виявилися непривабливими, що вказує на низьку ефективність роботи магазину. У той же час, низьке значення показника природно може бути у магазинів, що продають товари довготривалого використання. Наприклад, Інтернет-магазин, що спеціалізується на продажу електрогрилів. Після придбання одиниці товару необхідність у її покупці знову у клієнта зникає протягом усього строку служби цього приладу.

7. Дохід від покупця, який розраховується як різниця між прибутком та витратами магазину. Цей показник дає розуміння про те, наскільки успішним є Інтернет-магазин. Якщо показник надто низький, варто задуматися як скоротити витрати на рекламу.

8. Середній чек замовлення. Для того, щоб розрахувати цей показник необхідно поділити суму від продажів за певний проміжок часу на кількість замовлень. Це дозволяє оцінити наскільки у магазину високий або низький середній чек покупки. Виходячи з даних про загальну середню вартість чеку можна розробити стратегію щодо поліпшення цього показника, наприклад, пропонувати в кошику Інтернет-магазину придбати додатковий товар зі знижкою або загальну знижку за суму замовлення.

9. Кількість покинутих кошиків. Згідно із матеріалами дослідження інституту Baymard, середній відсоток покинутих кошиків складає 67,8 %. Причин для цього може бути дуже багато:

- вартість товарів у замовленні не відповідає вартості зазначеної в картці товару. Або, наприклад, у кошику з'являється додатковий рядок вартості доставки, яка збільшує вартість товару і відштовхує замовника;
- промо-код на знижку не працює;
- відвідувач не бачить чи існує доставка товару в його країну або регіон;
- на сторінці оформлення замовлення з'являються додаткові витрати, наприклад, податки;
- недостатня кількість варіантів оплати замовлення;
- технічні проблеми із заповненням платежів.

Для того, щоб оптимізувати роботу сайту, необхідно постійно перевіряти роботу кошика товарів, тестувати нові ідеї з оптимізації процесу оформлення замовлення.

Підвищує ефективність і робота з покинутими кошиками: можна відправляти листи з повідомленням про те, що оформлення замовлення не завершено і товари чекають покупця в кошику. Для роботи з кинутими кошиками можна використовувати Convead, він дозволяє налаштовувати продажі, створювати сегмент користувачів, які не завершили оформлення замовлення і відправляти їм листи.

Варто пам'ятати, що розглянуті показники потребують постійного аналізу. Їх потрібно порівнювати в динаміці та у просторі, тестувати нові ідеї для підвищення ефективності.

Розглянемо ще одну галузь економіки, яка характеризується наявністю специфічних показників, – *сферу інвестиційної діяльності*.

Система показників, яка використовується для оцінки ефективності інвестиційного проекту, наведена в табл. 9.8.

Таблиця 9.8

Показники ефективності інвестиційного проекту

Показник	Розрахунок	Характеристика
Термін окупності інвестицій (PP)	$PP = \frac{IC}{CF},$ <p>де IC – інвестиційний капітал (початкові витрати інвестора в об'єкт вкладення); CF – грошовий потік, який створюється об'єктом інвестицій.</p>	Показник характеризує період, за який окупляться початкові інвестиції у проект. Економічний зміст даного показника полягає в тому, що б показати термін, за який інвестор поверне назад свої вкладені кошти.
Чистий дисконтований дохід (NPV)	$NPV = \sum_{t=1}^n \frac{CF_t}{(1+r)^t} - \sum_{t=1}^n \frac{IC_t}{(1+r)^t},$ <p>де CF_t – грошовий потік у період часу t; IC_t – інвестиції в період часу t; r – ставка дисконтування; n – термін реалізації проекту.</p>	Коефіцієнт відображає зміну грошових потоків і показує різницю між дисконтованими грошовими доходами і витратами. Його використовують для вибору найбільш привабливого проекту. Коефіцієнт має певну градацію значень: - NPV < 0 – проект буде збитковим і його не варто приймати; - NPV = 0 – проект забезпечить рівень беззбитковості; - NPV > 0 – проект є привабливим.

Продовження таблиці 9.8

Показник	Розрахунок	Характеристика
Коефіцієнт рентабельності інвестицій (<i>ARR</i>)	$ARR = \frac{\overline{CF}}{IC},$ де \overline{CF} – середній грошовий потік.	Показник відображає прибутковість об'єкта інвестицій без урахування дисконтування.
Внутрішня норма доходності (<i>IRR</i>)	$IRR = r_i + \frac{NPV_1}{NPV_1 - NPV_2} \cdot (r_2 - r_1),$ де r_i – ставка дисконтування на період; r_1 та r_2 – обрані інвестором ставки дисконтування в певному діапазоні (наприклад, 10-15%; 15-20%); NPV_1 та NPV_2 – чистий приведений дохід відповідно при ставці дисконтування r_1 та r_2	Показник характеризує норму доходності вкладених інвестицій. На практиці значення <i>IRR</i> порівнюють із заданою нормою дисконту r . При цьому, якщо $IRR > r$, то проект забезпечує позитивну величину <i>NPV</i> і відсоток доходу, який складає ($IRR - r$).
Індекс прибутковості (доходності) інвестицій (<i>PI</i>)	$PI = \frac{\sum_{t=1}^n CF_t / (1+r)^t}{IC}$	Коефіцієнт характеризує прибутковість вкладеного капіталу. Він має таку градацію значень: - $PI < 1$ – проект себе не окупить; - $PI = 1$ – проект буде мати прибутковість, що дорівнює ставці дисконтування; $PI > 1$ – проект себе повністю окупить і принесе додаткові кошти.
Індекс прибутковості (доходності) дисконтованих інвестицій (<i>DPI</i>)	$DPI = \frac{\sum_{t=1}^n CF_t / (1+r)^t}{\sum_{t=1}^n IC_t / (1+r)^t}$	Даний коефіцієнт, як і попередній, характеризує прибутковість вкладеного капіталу. Він використовується у тому випадку, коли грошові витрати (інвестиції) здійснюються в декілька етапів, а не лише до початку отримання доходів.
Дисконтований термін окупності інвестицій (<i>DPP</i>)	$DPP = \min n \Rightarrow \sum_{t=1}^n \frac{CF_t}{(1+r)^t} \geq IC,$ де n – термін реалізації проекту.	Показник відображає період, що потрібний для повернення вкладених у проект інвестицій за рахунок грошового потоку з урахуванням ставки дисконтування.

Використання коефіцієнтів оцінки ефективності інвестиційних проектів дозволяє підприємству обрати найбільш привабливі об'єкти для вкладення коштів. Розглянуті показники дають можливість здійснити фінансове описання інвестицій, однак не розкривають причинно-наслідкових зв'язків з одержуваними доходами (складно оцінити венчурні проекти і стартапи). У той же час, простота розрахунків коефіцієнтів дозволяє вже на першому етапі аналізу виключити нерентабельні проекти.

Для того, щоб підвищити інвестиційну ефективність діяльності підприємства необхідно здійснювати такі заходи:

- обирати проекти, при яких суб'єкт господарювання отримає найбільший прибуток на вкладений капітал при мінімальних інвестиційних витратах;
- диверсифікувати портфель інвестицій для зниження ризиків;
- забезпечити ліквідність інвестицій;
- залучувати субсидії та пільгові кредити міжнародних фінансово-кредитних організацій і приватних іноземних інвесторів.

Питання для самоконтролю

1. Визначте економічну сутність показників ефекту та ефективності.
2. У чому відмінність між показниками ефекту та ефективності?
3. Окресліть коло осіб, зацікавлених в оцінці ефективності діяльності підприємства.
4. Що дозволяє зробити систематичний аналіз ефективності на підприємстві?
5. Класифікуйте показники ефективності.
6. На які два блоки можна розділити показники ефективності діяльності підприємства?
7. Назвіть групи економічних показників, що характеризують ефективність.
8. Які показники входять до групи оцінки ефективності персоналу?
9. Охарактеризуйте показники виробітку та трудомісткості.
10. Якими способами можна підвищити ефективність праці робітників?
11. Що розуміють під основними засобами підприємства?
12. Назвіть основні показники, що характеризують ефективність використання основних засобів.
13. Дайте визначення показникам фондоозброєності та фондівіддачі.
14. Що характеризує показник інтенсивності використання устаткування?
15. Які існують шляхи для підвищення ефективності використання основних засобів?
16. Що таке оборотні засоби підприємства?
17. Які показники оцінки ефективності оборотних засобів ви знаєте?

18. Який існує зв'язок між коефіцієнтом оборотності оборотних засобів і коефіцієнтом їх закріплення?
19. Охарактеризуйте показники матеріалоємності та матеріаловіддачі.
20. Назвіть основні узагальнюючі показники ефективності.
21. Що таке рентабельність продукції і як вона розраховується?
22. Охарактеризуйте показник рентабельності продажів.
23. Назвіть основні показники енергоефективності підприємства.
24. Опишіть показник втрат енергії та назвіть середньостатистичний відсоток цих втрат для різного виду енергії.
25. Що таке енергоекономічний рівень виробництва?
26. Як підвищити рівень енергетичної ефективності підприємства?
27. Які групи показників екологічної ефективності ви знаєте?
28. Назвіть основні методи контролю чистоти повітря.
29. Опишіть основні методи управління промисловими відходами.
30. Які додаткові показники ефективності використовують у сферах будівництва, оптової та роздрібною торгівлі та транспорту.
31. Назвіть показники ефективності діяльності Інтернет-магазину.
32. Охарактеризуйте показник середнього чеку замовлення.
33. Які специфічні показники використовують під час оцінки ефективності інвестицій?
34. Що таке чистий дисконтований дохід? Опишіть градацію значень.
35. Як розраховується індекс прибутковості інвестицій? Охарактеризуйте його граничні значення.
36. Що таке термін окупності інвестицій?

СПИСОК ВИКОРИСТАНОЇ ТА РЕКОМЕНДОВАНОЇ ЛІТЕРАТУРИ

1. Андерсен Бьёрн. Бизнес-процессы. Инструменты совершенствования / Пер. с англ. С.В. Ариничева / Науч. ред. Ю. П. Адлер. – М.: РИА «Стандарты и качество», 2003. – 272 с.
2. Варзунов А. В. Анализ и управление бизнес-процессами. Учебное пособие / А. В. Варзунов, Е. К. Торосян, Л. П. Сажнева. – СПб: Университет ИТМО, 2016. – 112 с.
3. Гальперин М. В. Экологические основы природопользования: учебник / М. В. Гальперин. – М.: Форум, 2014. – 256 с.
4. Герасимов Б. И. Управление качеством: учеб. пособие / Б. И. Герасимов, Н. В. Злобина, С. П. Спиридонов. – М.: Кнорус, 2007. – 272 с.
5. Драчев О. И. Статистические методы управления качеством / О. И. Драчев, А. А. Жилин. – Старый Оскол: ТНТ, 2011. – 148 с.
6. Елисеева И. И., Юзбашев М. М. Общая теория статистики: Учебник / Под ред. И.И. Елисеевой. – 5-е изд., перераб. и доп. – М.: Финансы и статистика, 2004. – 656 с.
7. Захарченко Н. И. Бизнес-статистика и прогнозирование в MS Excel / Н. И. Захарченко. – М.: Издательский дом "Вильямс", 2004. – 208 с.
8. Йоуни П. Основы экономики окружающей среды : учеб. пособ для вузов / П. Йоуни. – М.: Вузовская книга, 2011. – 296 с.
9. Кане М. М. Системы, методы и инструменты менеджмента качества / М. М. Кане, Б. В. Иванов, В. Н. Корешков. – СПб.: Питер, 2008. – 560 с.
10. Козлов А. С. Проектирование и исследование бизнес-процессов: учеб. пособие / А. С. Козлов. – 4-е изд., стер. – М. : Флинта, 2011. – 267 с.
11. Куприенко Н. В. Статистика. Методы анализа распределений. Выборочное наблюдение / Н. В. Куприенко. – СПб. : Изд-во Политехн. ун-та, 2009. – 138 с.
12. Образцова О. И. Статистика предприятий и бизнес-статистика / О. И. Образцова. – М.: НИУ ВШЭ, 2011. – 704 с.
13. Паклин Н. Б. Бизнес аналитика: от данных к знаниям / Н. Б. Паклин, В. И. Орешков. – СПб.: Изд. "Питер", 2009. – 624 с.
14. Панде П. Что такое «Шесть сигм»? Революционный метод управления качеством / П. Панде, Л. Холп; пер. с англ. – М.: Альпина Бизнес Букс, 2005. – 158 с.
15. Семенова К. Д., Карпенко Л. М. Статистика підприємств: Навч. посібник. / К.Д. Семенова, Л. М. Карпенко – Одеса: Атлант, 2009. – 348 с.
16. Сигел Ф. Эндрю. Практическая бизнес-статистика / Сигел Ф. Эндрю; пер. с англ. – М.: Издательский дом "Вильямс", 2008. – 1056 с.
17. Системы управления базами данных: пособие / авт.-сост. В. В. Скакун. – Минск: БГУ, 2007. – 116 с.
18. Фетисов Г. Г. Региональная экономика и управление / Г.Г. Фетисов, В. П. Орешин. – М.: ИНФРА, 2006. – 416 с.

19. Ханк Д. Э. Бизнес-прогнозирование / Д. Э. Ханк, Д. У. Уичерн, А. Дж. Райтс. – М. : Издательский дом "Вильямс", 2003. – 656 с.
20. Четыркин Е. М., Калихман И. Л. Вероятность и статистика / Е. М. Четыркин, И. Л. Калихман. – М.: Финансы и статистика, 1982. – 319 с.
21. Шухарт У. А. Экономический контроль качества произведенного продукта / У. А. Шухарт. Вэн Ноустренд К., Нью-Йорк, 1931. – 50 с.
22. Шекшня Е. В. Управление персоналом современной организации: учебник / Е. В. Шекшня. – М.: Изд-во «Интел-Синтез», 2007. – 336 с.
23. Эддоус М., Стэнсфилд Р. Методы принятия решений / Пер. с англ. под ред. член-корр. РАН И.И. Елисеевой. – М.: Аудит, ЮНИТИ, 1997. – 590 с.
24. Янковой А. Г. Математико-статистические методы и модели в управлении предприятием: учебное пособие / А. Г. Янковой. – Одесса, 2014. – 250 с.

ДОДАТКИ

Додаток А

Таблиця ймовірностей для стандартного нормального розподілу

Значення Z	Ймовірність	Значення Z	Ймовірність	Значення Z	Ймовірність	Значення Z	Ймовірність	Значення Z	Ймовірність	Значення Z	Ймовірність
-0,00	0,5000	-0,41	0,3409	-0,82	0,2061	-1,23	0,1093	-1,64	0,0505	-2,05	0,0202
-0,01	0,4960	-0,42	0,3372	-0,83	0,2033	-1,24	0,1075	-1,65	0,0495	-2,06	0,0197
-0,02	0,4920	-0,43	0,3336	-0,84	0,2005	-1,25	0,1056	-1,66	0,0485	-2,07	0,0192
-0,03	0,4880	-0,44	0,3300	-0,85	0,1977	-1,26	0,1038	-1,67	0,0475	-2,08	0,0188
-0,04	0,4840	-0,45	0,3264	-0,86	0,1949	-1,27	0,1020	-1,68	0,0465	-2,09	0,0183
-0,05	0,4801	-0,46	0,3228	-0,87	0,1922	-1,28	0,1003	-1,69	0,0455	-2,10	0,0179
-0,06	0,4761	-0,47	0,3192	-0,88	0,1894	-1,29	0,0985	-1,70	0,0446	-2,11	0,0174
-0,07	0,4721	-0,48	0,3156	-0,89	0,1867	-1,30	0,0968	-1,71	0,0436	-2,12	0,0170
-0,08	0,4681	-0,49	0,3121	-0,90	0,1841	-1,31	0,0951	-1,72	0,0427	-2,13	0,0166
-0,09	0,4641	-0,50	0,3085	-0,91	0,1814	-1,32	0,0934	-1,73	0,0418	-2,14	0,0162
-0,10	0,4602	-0,51	0,3050	-0,92	0,1788	-1,33	0,0918	-1,74	0,0409	-2,15	0,0158
-0,11	0,4562	-0,52	0,3015	-0,93	0,1762	-1,34	0,0901	-1,75	0,0401	-2,16	0,0154
-0,12	0,4522	-0,53	0,2981	-0,94	0,1736	-1,35	0,0885	-1,76	0,0392	-2,17	0,0150
-0,13	0,4483	-0,54	0,2946	-0,95	0,1711	-1,36	0,0869	-1,77	0,0384	-2,18	0,0146
-0,14	0,4443	-0,55	0,2912	-0,96	0,1685	-1,37	0,0853	-1,78	0,0375	-2,19	0,0143
-0,15	0,4404	-0,56	0,2877	-0,97	0,1660	-1,38	0,0838	-1,79	0,0367	-2,20	0,0139
-0,16	0,4364	-0,57	0,2843	-0,98	0,1635	-1,39	0,0823	-1,80	0,0359	-2,21	0,0136
-0,17	0,4325	-0,58	0,2810	-0,99	0,1611	-1,40	0,0808	-1,81	0,0351	-2,22	0,0132
-0,18	0,4286	-0,59	0,2776	-1,00	0,1587	-1,41	0,0793	-1,82	0,0344	-2,23	0,0129
-0,19	0,4247	-0,60	0,2743	-1,01	0,1562	-1,42	0,0778	-1,83	0,0336	-2,24	0,0125
-0,20	0,4207	-0,61	0,2709	-1,02	0,1539	-1,43	0,0764	-1,84	0,0329	-2,25	0,0122
-0,21	0,4168	-0,62	0,2676	-1,03	0,1515	-1,44	0,0749	-1,85	0,0322	-2,26	0,0119
-0,22	0,4129	-0,63	0,2643	-1,04	0,1492	-1,45	0,0735	-1,86	0,0314	-2,27	0,0116
-0,23	0,4090	-0,64	0,2611	-1,05	0,1469	-1,46	0,0721	-1,87	0,0307	-2,28	0,0113
-0,24	0,4052	-0,65	0,2578	-1,06	0,1446	-1,47	0,0708	-1,88	0,0301	-2,29	0,0110
-0,25	0,4013	-0,66	0,2546	-1,07	0,1423	-1,48	0,0694	-1,89	0,0294	-2,30	0,0107
-0,26	0,3974	-0,67	0,2514	-1,08	0,1401	-1,49	0,0681	-1,90	0,0287	-2,31	0,0104
-0,27	0,3936	-0,68	0,2483	-1,09	0,1379	-1,50	0,0668	-1,91	0,0281	-2,32	0,0102
-0,28	0,3897	-0,69	0,2451	-1,10	0,1357	-1,51	0,0655	-1,92	0,0274	-2,33	0,0099
-0,29	0,3859	-0,70	0,2420	-1,11	0,1335	-1,52	0,0643	-1,93	0,0268	-2,34	0,0096
-0,30	0,3821	-0,71	0,2389	-1,12	0,1314	-1,53	0,0630	-1,94	0,0262	-2,35	0,0094
-0,31	0,3783	-0,72	0,2358	-1,13	0,1292	-1,54	0,0618	-1,95	0,0256	-2,36	0,0091
-0,32	0,3745	-0,73	0,2327	-1,14	0,1271	-1,55	0,0606	-1,96	0,0250	-2,37	0,0089
-0,33	0,3707	-0,74	0,2296	-1,15	0,1251	-1,56	0,0594	-1,97	0,0244	-2,38	0,0087
-0,34	0,3669	-0,75	0,2266	-1,16	0,1230	-1,57	0,0582	-1,98	0,0239	-2,39	0,0084
-0,35	0,3632	-0,76	0,2236	-1,17	0,1210	-1,58	0,0571	-1,99	0,0233	-2,40	0,0082
-0,36	0,3594	-0,77	0,2206	-1,18	0,1190	-1,59	0,0559	-2,00	0,0228	-2,41	0,0080
-0,37	0,3557	-0,78	0,2177	-1,19	0,1170	-1,60	0,0548	-2,01	0,0222	-2,42	0,0078
-0,38	0,3520	-0,79	0,2148	-1,20	0,1151	-1,61	0,0537	-2,02	0,0217	-2,43	0,0075
-0,39	0,3483	-0,80	0,2119	-1,21	0,1131	-1,62	0,0526	-2,03	0,0212	-2,44	0,0073
-0,40	0,3446	-0,81	0,2090	-1,22	0,1112	-1,63	0,0516	-2,04	0,0207	-2,45	0,0071

Продовження додатку А

Значення Z	Імовірність	Значення Z	Імовірність	Значення Z	Імовірність	Значення Z	Імовірність	Значення Z	Імовірність	Значення Z	Імовірність
-2,46	0,0069	-2,88	0,0020	0,29	0,6141	0,71	0,7611	1,13	0,8708	1,55	0,9394
-2,47	0,0068	-2,89	0,0019	0,30	0,6179	0,72	0,7642	1,14	0,8729	1,56	0,9406
-2,48	0,0066	-2,90	0,0019	0,31	0,6217	0,73	0,7673	1,15	0,8749	1,57	0,9418
-2,49	0,0064	-2,91	0,0018	0,32	0,6255	0,74	0,7704	1,16	0,8770	1,58	0,9429
-2,50	0,0062	-2,92	0,0018	0,33	0,6293	0,75	0,7734	1,17	0,8790	1,59	0,9441
-2,51	0,0060	-2,93	0,0017	0,34	0,6331	0,76	0,7764	1,18	0,8810	1,60	0,9452
-2,52	0,0059	-2,94	0,0016	0,35	0,6368	0,77	0,7794	1,19	0,8830	1,61	0,9463
-2,53	0,0057	-2,95	0,0016	0,36	0,6406	0,78	0,7823	1,20	0,8849	1,62	0,9474
-2,54	0,0055	-2,96	0,0015	0,37	0,6443	0,79	0,7852	1,21	0,8869	1,63	0,9484
-2,55	0,0054	-2,97	0,0015	0,38	0,6480	0,80	0,7881	1,22	0,8888	1,64	0,9495
-2,56	0,0052	-2,98	0,0014	0,39	0,6517	0,81	0,7910	1,23	0,8907	1,65	0,9505
-2,57	0,0051	-2,99	0,0014	0,40	0,6554	0,82	0,7939	1,24	0,8925	1,66	0,9515
-2,58	0,0049	-3,00	0,0013	0,41	0,6591	0,83	0,7967	1,25	0,8944	1,67	0,9525
-2,59	0,0048	0,00	0,5000	0,42	0,6628	0,84	0,7995	1,26	0,8962	1,68	0,9535
-2,60	0,0047	0,01	0,5040	0,43	0,6664	0,85	0,8023	1,27	0,8980	1,69	0,9545
-2,61	0,0045	0,02	0,5080	0,44	0,6700	0,86	0,8051	1,28	0,8997	1,70	0,9554
-2,62	0,0044	0,03	0,5120	0,45	0,6736	0,87	0,8078	1,29	0,9015	1,71	0,9564
-2,63	0,0043	0,04	0,5160	0,46	0,6772	0,88	0,8106	1,30	0,9032	1,72	0,9573
-2,64	0,0041	0,05	0,5199	0,47	0,6808	0,89	0,8133	1,31	0,9049	1,73	0,9582
-2,65	0,0040	0,06	0,5239	0,48	0,6844	0,90	0,8159	1,32	0,9066	1,74	0,9591
-2,66	0,0039	0,07	0,5279	0,49	0,6879	0,91	0,8186	1,33	0,9082	1,75	0,9599
-2,67	0,0038	0,08	0,5319	0,50	0,6915	0,92	0,8212	1,34	0,9099	1,76	0,9608
-2,68	0,0037	0,09	0,5359	0,51	0,6950	0,93	0,8238	1,35	0,9115	1,77	0,9616
-2,69	0,0036	0,10	0,5398	0,52	0,6985	0,94	0,8264	1,36	0,9131	1,78	0,9625
-2,70	0,0035	0,11	0,5438	0,53	0,7019	0,95	0,8289	1,37	0,9147	1,79	0,9633
-2,71	0,0034	0,12	0,5478	0,54	0,7054	0,96	0,8315	1,38	0,9162	1,80	0,9641
-2,72	0,0033	0,13	0,5517	0,55	0,7088	0,97	0,8340	1,39	0,9177	1,81	0,9649
-2,73	0,0032	0,14	0,5557	0,56	0,7123	0,98	0,8365	1,40	0,9192	1,82	0,9656
-2,74	0,0031	0,15	0,5596	0,57	0,7157	0,99	0,8389	1,41	0,9207	1,83	0,9664
-2,75	0,0030	0,16	0,5636	0,58	0,7190	1,00	0,8413	1,42	0,9222	1,84	0,9671
-2,76	0,0029	0,17	0,5675	0,59	0,7224	1,01	0,8438	1,43	0,9236	1,85	0,9678
-2,77	0,0028	0,18	0,5714	0,60	0,7257	1,02	0,8461	1,44	0,9251	1,86	0,9686
-2,78	0,0027	0,19	0,5753	0,61	0,7291	1,03	0,8485	1,45	0,9265	1,87	0,9693
-2,79	0,0026	0,20	0,5793	0,62	0,7324	1,04	0,8508	1,46	0,9279	1,88	0,9699
-2,80	0,0026	0,21	0,5832	0,63	0,7357	1,05	0,8531	1,47	0,9292	1,89	0,9706
-2,81	0,0025	0,22	0,5871	0,64	0,7389	1,06	0,8554	1,48	0,9306	1,90	0,9713
-2,82	0,0024	0,23	0,5910	0,65	0,7422	1,07	0,8577	1,49	0,9319	1,91	0,9719
-2,83	0,0023	0,24	0,5948	0,66	0,7454	1,08	0,8599	1,50	0,9332	1,92	0,9726
-2,84	0,0023	0,25	0,5987	0,67	0,7486	1,09	0,8621	1,51	0,9345	1,93	0,9732
-2,85	0,0022	0,26	0,6026	0,68	0,7517	1,10	0,8643	1,52	0,9357	1,94	0,9738
-2,86	0,0021	0,27	0,6064	0,69	0,7549	1,11	0,8665	1,53	0,9370	1,95	0,9744
-2,87	0,0021	0,28	0,6103	0,70	0,7580	1,12	0,8686	1,54	0,9382	1,96	0,9750

Продовження додатку А

Значення Z	Імовірність	Значення Z	Імовірність	Значення Z	Імовірність
1,97	0,9756	2,39	0,9916	2,81	0,9975
1,98	0,9761	2,40	0,9918	2,82	0,9976
1,99	0,0233	2,41	0,9920	2,83	0,9977
2,00	0,9972	2,42	0,9922	2,84	0,9977
2,01	0,9778	2,43	0,9925	2,85	0,9978
2,02	0,9783	2,44	0,9927	2,86	0,9979
2,03	0,9788	2,45	0,9929	2,81	0,9975
2,04	0,9793	2,46	0,9931	2,87	0,9979
2,05	0,9798	2,47	0,9932	2,88	0,9980
2,06	0,9803	2,48	0,9934	2,89	0,9981
2,07	0,9808	2,49	0,9936	2,90	0,9981
2,08	0,9812	2,50	0,9938	2,91	0,9982
2,09	0,9817	2,51	0,9940	2,92	0,9982
2,10	0,9821	2,52	0,9941	2,93	0,9983
2,11	0,9826	2,53	0,9943	2,94	0,9984
2,12	0,9830	2,54	0,9945	2,95	0,9984
2,13	0,9834	2,55	0,9946	2,96	0,9985
2,14	0,9838	2,56	0,9948	2,97	0,9985
2,15	0,9842	2,57	0,9949	2,98	0,9986
2,16	0,9846	2,58	0,9951	2,99	0,9986
2,17	0,9850	2,59	0,9952	3,00	0,9987
2,18	0,9854	2,60	0,9953		
2,19	0,9857	2,61	0,9955		
2,20	0,9861	2,62	0,9956		
2,21	0,9864	2,63	0,9957		
2,22	0,9868	2,64	0,9959		
2,23	0,9871	2,65	0,9960		
2,24	0,9875	2,66	0,9961		
2,25	0,9878	2,67	0,9962		
2,26	0,9881	2,68	0,9963		
2,27	0,9884	2,69	0,9964		
2,28	0,9887	2,70	0,9965		
2,29	0,9890	2,71	0,9966		
2,30	0,9893	2,72	0,9967		
2,31	0,9896	2,73	0,9968		
2,32	0,9898	2,74	0,9969		
2,33	0,9901	2,75	0,9970		
2,34	0,9904	2,76	0,9971		
2,35	0,9906	2,77	0,9972		
2,36	0,9909	2,78	0,9973		
2,37	0,9911	2,79	0,9974		
2,38	0,9913	2,80	0,9974		

t-таблиця

Довірчий рівень								
Двосторонній		80%	90%	95%	98%	99%	99,8%	99,9%
Односторонній		90%	95%	97,5%	99%	99,5%	99,9%	99,95%
Рівень перевірки гіпотези								
Двостороння		0,20	0,10	0,05	0,02	0,01	0,002	0,001
Одностороння		0,10	0,05	0,025	0,01	0,005	0,001	0,0005
Для однієї вибірки n	Число ступенів свободи	Критичні значення						
2	1	3,078	6,314	12,706	31,821	63,657	318,309	636,619
3	2	1,886	2,920	4,303	6,965	9,925	22,327	31,599
4	3	1,638	2,353	3,182	4,541	5,841	10,215	12,924
5	4	1,533	2,132	2,776	3,747	4,604	7,173	8,610
6	5	1,476	2,015	2,571	3,365	4,032	5,893	6,869
7	6	1,440	1,943	2,447	3,143	3,707	5,208	5,959
8	7	1,415	1,895	2,365	2,998	3,499	4,785	5,408
9	8	1,397	1,860	2,306	2,896	3,355	4,501	5,041
10	9	1,383	1,833	2,262	2,821	3,250	4,297	4,781
11	10	1,372	1,812	2,228	2,764	3,169	4,144	4,587
12	11	1,363	1,796	2,201	2,718	3,106	4,025	4,437
13	12	1,356	1,782	2,179	2,681	3,055	3,930	4,318
14	13	1,350	1,771	2,160	2,650	3,012	3,852	4,221
15	14	1,345	1,761	2,145	2,624	2,977	3,787	4,140
16	15	1,341	1,753	2,131	2,602	2,947	3,733	4,073
17	16	1,337	1,746	2,120	2,583	2,921	3,686	4,015
18	17	1,333	1,740	2,110	2,567	2,898	3,646	3,965
19	18	1,330	1,734	2,101	2,552	2,878	3,610	3,922
20	19	1,328	1,729	2,093	2,539	2,861	3,579	3,883
21	20	1,325	1,725	2,086	2,528	2,845	3,552	3,850
22	21	1,323	1,721	2,080	2,518	2,831	3,527	3,819
23	22	1,321	1,717	2,074	2,508	2,819	3,505	3,792
24	23	1,319	1,714	2,069	2,500	2,807	3,485	3,768
25	24	1,318	1,711	2,064	2,492	2,797	3,467	3,745
26	25	1,316	1,708	2,060	2,485	2,787	3,450	3,725
27	26	1,315	1,706	2,056	2,479	2,779	3,435	3,707
28	27	1,314	1,703	2,052	2,473	2,771	3,421	3,690
29	28	1,313	1,701	2,048	2,467	2,763	3,408	3,674
30	29	1,311	1,699	2,045	2,462	2,756	3,396	3,659
31	30	1,310	1,697	2,042	2,457	2,750	3,385	3,646
32	31	1,309	1,696	2,040	2,453	2,744	3,375	3,633
33	32	1,309	1,694	2,037	2,449	2,738	3,365	3,622
34	33	1,308	1,692	2,035	2,445	2,733	3,356	3,611
35	34	1,307	1,691	2,032	2,441	2,728	3,348	3,601
36	35	1,306	1,690	2,030	2,438	2,724	3,340	3,591
37	36	1,306	1,688	2,028	2,434	2,719	3,333	3,582
38	37	1,305	1,687	2,026	2,431	2,715	3,326	3,574
39	38	1,304	1,686	2,024	2,429	2,712	3,319	3,566
40	39	1,304	1,685	2,023	2,426	2,708	3,313	3,558
Безкінечність		1,282	1,645	1,960	2,326	2,576	3,090	3,291

Критичні значення F -критерію Фішера при $\alpha = 0,05$

f2/df1	1	2	3	4	5	6	7	8	9	10	12	15	20	30	60	120	INF
1	161,45	199,50	215,71	224,58	230,16	233,99	236,77	238,88	240,54	241,88	243,91	245,95	248,01	250,10	252,19	253,25	254,31
2	18,513	19,000	19,164	19,247	19,296	19,329	19,353	19,371	19,385	19,396	19,413	19,429	19,446	19,462	19,479	19,487	19,496
3	10,128	9,5521	9,2766	9,1172	9,0135	8,9406	8,8867	8,8452	8,8123	8,7855	8,7446	8,7029	8,6602	8,6166	8,5720	8,5494	8,5264
4	7,7086	6,9443	6,5914	6,3882	6,2561	6,1631	6,0942	6,0410	5,9988	5,9644	5,9117	5,8578	5,8025	5,7459	5,6877	5,6581	5,6281
5	6,6079	5,7861	5,4095	5,1922	5,0503	4,9503	4,8759	4,8183	4,7725	4,7351	4,6777	4,6188	4,5581	4,4957	4,4314	4,3985	4,3650
6	5,9874	5,1433	4,7571	4,5337	4,3874	4,2839	4,2067	4,1468	4,0990	4,0600	3,9999	3,9381	3,8742	3,8082	3,7398	3,7047	3,6689
7	5,5914	4,7374	4,3468	4,1203	3,9715	3,8660	3,7870	3,7257	3,6767	3,6365	3,5747	3,5107	3,4445	3,3758	3,3043	3,2674	3,2298
8	5,3177	4,4590	4,0662	3,8379	3,6875	3,5806	3,5005	3,4381	3,3881	3,3472	3,2839	3,2184	3,1503	3,0794	3,0053	2,9669	2,9276
9	5,1174	4,2565	3,8625	3,6331	3,4817	3,3738	3,2927	3,2296	3,1789	3,1373	3,0729	3,0061	2,9365	2,8637	2,7872	2,7475	2,7067
10	4,9646	4,1028	3,7083	3,4780	3,3258	3,2172	3,1355	3,0717	3,0204	2,9782	2,9130	2,8450	2,7740	2,6996	2,6211	2,5801	2,5379
11	4,8443	3,9823	3,5874	3,3567	3,2039	3,0946	3,0123	2,9480	2,8962	2,8536	2,7876	2,7186	2,6464	2,5705	2,4901	2,4480	2,4045
12	4,7472	3,8853	3,4903	3,2592	3,1059	2,9961	2,9134	2,8486	2,7964	2,7534	2,6866	2,6169	2,5436	2,4663	2,3842	2,3410	2,2962
13	4,6672	3,8056	3,4105	3,1791	3,0254	2,9153	2,8321	2,7669	2,7144	2,6710	2,6037	2,5331	2,4589	2,3803	2,2966	2,2524	2,2064
14	4,6001	3,7389	3,3439	3,1122	2,9582	2,8477	2,7642	2,6987	2,6458	2,6022	2,5342	2,4630	2,3879	2,3082	2,2229	2,1778	2,1307
15	4,5431	3,6823	3,2874	3,0556	2,9013	2,7905	2,7066	2,6408	2,5876	2,5437	2,4753	2,4034	2,3275	2,2468	2,1601	2,1141	2,0658
16	4,4940	3,6337	3,2389	3,0069	2,8524	2,7413	2,6572	2,5911	2,5377	2,4935	2,4247	2,3522	2,2756	2,1938	2,1058	2,0589	2,0096
17	4,4513	3,5915	3,1968	2,9647	2,8100	2,6987	2,6143	2,5480	2,4943	2,4499	2,3807	2,3077	2,2304	2,1477	2,0584	2,0107	1,9604
18	4,4139	3,5546	3,1599	2,9277	2,7729	2,6613	2,5767	2,5102	2,4563	2,4117	2,3421	2,2686	2,1906	2,1071	2,0166	1,9681	1,9168
19	4,3807	3,5219	3,1274	2,8951	2,7401	2,6283	2,5435	2,4768	2,4227	2,3779	2,3080	2,2341	2,1555	2,0712	1,9795	1,9302	1,8780
20	4,3512	3,4928	3,0984	2,8661	2,7109	2,5990	2,5140	2,4471	2,3928	2,3479	2,2776	2,2033	2,1242	2,0391	1,9464	1,8963	1,8432
21	4,3248	3,4668	3,0725	2,8401	2,6848	2,5727	2,4876	2,4205	2,3660	2,3210	2,2504	2,1757	2,0960	2,0102	1,9165	1,8657	1,8117
22	4,3009	3,4434	3,0491	2,8167	2,6613	2,5491	2,4638	2,3965	2,3419	2,2967	2,2258	2,1508	2,0707	1,9842	1,8894	1,8380	1,7831
23	4,2793	3,4221	3,0280	2,7955	2,6400	2,5277	2,4422	2,3748	2,3201	2,2747	2,2036	2,1282	2,0476	1,9605	1,8648	1,8128	1,7570
24	4,2597	3,4028	3,0088	2,7763	2,6207	2,5082	2,4226	2,3551	2,3002	2,2547	2,1834	2,1077	2,0267	1,9390	1,8424	1,7896	1,7330
25	4,2417	3,3852	2,9912	2,7587	2,6030	2,4904	2,4047	2,3371	2,2821	2,2365	2,1649	2,0889	2,0075	1,9192	1,8217	1,7684	1,7110
30	4,1709	3,3158	2,9223	2,6896	2,5336	2,4205	2,3343	2,2662	2,2107	2,1646	2,0921	2,0148	1,9317	1,8409	1,7396	1,6835	1,6223
60	4,0012	3,1504	2,7581	2,5252	2,3683	2,2541	2,1665	2,0970	2,0401	1,9926	1,9174	1,8364	1,7480	1,6491	1,5343	1,4673	1,3893
120	3,9201	3,0718	2,6802	2,4472	2,2899	2,1750	2,0868	2,0164	1,9588	1,9105	1,8337	1,7505	1,6587	1,5543	1,4290	1,3519	1,2539
inf	3,8415	2,9957	2,6049	2,3719	2,2141	2,0986	2,0096	1,9384	1,8799	1,8307	1,7522	1,6664	1,5705	1,4591	1,3180	1,2214	1,0000

Вихідні дані для побудови регресійно-кореляційної моделі

Рік	Середньомісячна заробітна плата одного працівника, грн.	ВВП на душу населення, грн.	ВРП на душу населення, грн.	Прожитковий мінімум, грн.	Рівень безробіття, %	Рівень зайнятості, %	Обсяг інвестицій у основний капітал у розрахунку на одну особу, тис. грн.
	Y	X ₁	X ₂	X ₃	X ₄	X ₅	X ₆
2000	230	3462,7	2788	270	12,4	57,4	1,2
2001	311	4195,9	3764	311	11,7	57,8	1,6
2002	377	4681,9	4239	342	10,3	59,3	1,9
2003	462	5591,5	5114	342	9,7	57,5	2,5
2004	590	7272,9	7273	362	9,2	56,7	3,7
2005	806	9371,6	9372	423	7,8	57,7	4,5
2006	1041	11630,2	11630	472	7,4	57,9	6,0
2007	1351	15496,5	15496	532	6,9	58,7	9,0
2008	1806	20494,9	20495	626	6,9	59,3	11,1
2009	1906	19832,3	19832	701	9,6	57,7	7,5
2010	2239	23600,4	23600	875	8,8	58,5	7,4
2011	2633	28813,9	28488	953	8,6	59,2	10,3
2012	3026	30912,5	32002	1095	8,1	59,7	11,7
2013	3265	31988,7	33473	1218	7,7	60,3	5,9
2014	3480	35834,0	36904,0	1218	9,7	56,6	4,7
2015	4195	46210,2	47233,2	1218	9,5	56,7	6,4
2016	5183	55943,7	51211,3	1399	9,3	56,3	8,4

**Результати регресійно-кореляційного аналізу
(перший етап)**

ВЫВОД ИТОГОВ						
<i>Регрессионная статистика</i>						
Множественный R	0,999716571					
R-квадрат	0,999433223					
Нормированный R-квадрат	0,999093157					
Стандартная ошибка	45,14932948					
Наблюдения	17					
<i>Дисперсионный анализ</i>						
	<i>df</i>	<i>SS</i>	<i>MS</i>	<i>F</i>	<i>Значимость F</i>	
Регрессия	6	35945477,26	5990912,877	2938,937795	1,22707E-15	
Остаток	10	20384,61953	2038,461953			
Итого	16	35965861,88				
	<i>Коэффициенты</i>	<i>Стандартная ошибка</i>	<i>t-статистика</i>	<i>P-Значение</i>	<i>Нижние 95%</i>	<i>Верхние 95%</i>
Y-пересечение	56,10249248	884,3063388	0,063442373	0,950664485	-1914,25481	2026,45979
Переменная X 1	0,084320525	0,009781260	8,620620005	6,08453E-06	0,06252652	0,10611453
Переменная X 2	-0,015596377	0,011761626	-1,326039195	0,21431658	-0,04180291	0,01061016
Переменная X 3	1,121467644	0,186997874	5,997221374	0,000132599	0,70481041	1,53812487
Переменная X 4	-10,53046106	11,24558913	-0,936408128	0,371125968	-35,587195	14,5262729
Переменная X 5	-4,459647105	15,45068141	-0,288637568	0,778753327	-38,8859105	29,9666163
Переменная X 6	-3,352357616	6,838616561	-0,490209911	0,634562289	-18,5897448	11,8850296

**Результати регресійно-кореляційного аналізу
(останній етап)**

ВЫВОД ИТОГОВ						
<i>Регрессионная статистика</i>						
Множественный R	0,999651209					
R-квадрат	0,99930254					
Нормированный R-квадрат	0,999202903					
Стандартная ошибка	42,32926908					
Наблюдения	17					
Дисперсионный анализ						
	<i>df</i>	<i>SS</i>	<i>MS</i>	<i>F</i>	<i>Значимость F</i>	
Регрессия	2	35940777,14	17970388,57	10029,42255	8,02849E-23	
Остаток	14	25084,7383	1791,767021			
Итого	16	35965861,88				
	<i>Коэффициенты</i>	<i>Стандартная ошибка</i>	<i>t-статистика</i>	<i>P-Значение</i>	<i>Нижние 95%</i>	<i>Верхние 95%</i>
Y-пересечение	-285,3715225	30,00497185	-9,510807874	1,73236E-07	-349,72578	-221,01725
Переменная X 1	0,072511996	0,002687783	26,97836401	1,80095E-13	0,0667472	0,0782767
Переменная X 2	0,970005079	0,108326406	8,95446564	3,60203E-07	0,7376680	1,2023421

Навчальне видання

Семенова К. Д., Тарасова К. І.

БІЗНЕС-СТАТИСТИКА

ПІДРУЧНИК

Підписано до друку 30.07.2018. Формат 60x84/16.
Друк цифровий. Папір офсетний. Гарнітура Times.
Ум.-друк. Арк. 12,3. Наклад 300 прим.
Замовлення 12117

Видавець ФОП Гуляєва В.М.
Свідоцтво серія ДК № 6205

Друкарня ФОП Гуляєва В.М.
Київська обл., м. Обухів, мкрн. Сосновий, 2, офіс 2
(050) 497-89-01
drukaryk.com