МІНІСТЕРСТВО ОСВІТИ І НАУКИ УКРАЇНИ

**ХАРКІВСЬКИЙ НАЦІОНАЛЬНИЙ УНІВЕРСИТЕТ
МІСЬКОГО ГОСПОДАРСТВА імені О. М. БЕКЕТОВА**

О. О. Воронков

ОСНОВИ МОДЕЛЮВАННЯ СКЛАДНИХ СИСТЕМ

КОНСПЕКТ ЛЕКЦІЙ

(*для здобувачів першого (бакалаврського) рівня вищої освіти
зі спеціальності 193 − Геодезія та землеустрій)*

**Харків
ХНУМГ ім. О. М. Бекетова
2023**

Воронков О. О. Основи моделювання складних систем : конспект лекцій для здобувачів першого (бакалаврського) рівня вищої освіти зі спеціальності 193 − Геодезія та землеустрій / О. О. Воронков ; Харків. нац. ун-т міськ. госп-ва ім. О. М. Бекетова. – Харків : ХНУМГ ім. О. М. Бекетова, 2023. – 76 с.

Автор
канд. екон. наук, доц. О. О. Воронков

Рецензент
К. А. Мамонов, доктор економічних наук, професор, завідувач кафедри земельного адміністрування та геоінформаційних систем (Харківський національний університет міського господарства імені О. М. Бекетова)

Рекомендовано кафедрою земельного адміністрування та геоінформаційних систем, протокол № 1 від 05.09.2022

ЗМІСТ

ВСТУП

Дисципліна «Основи моделювання складних систем» є вибірковою дисципліною підготовки бакалавра за освітньо-професійною програмою «Геодезія, картографія та землеустрій» із спеціальності 193 – Геодезія та землеустрій». Обсяг курсу становить 150 академічних годин або 5 кредитів ЄКТС, у тому числі 16 годин лекційних та 30 годин практичних занять. Обсяг самостійної роботи студента становить 104 години, до якого включений час на виконання трьох завдань до самостійної роботи студента та тестування за теоретичним матеріалом змістових модулів дисципліни у системі *Moodle*. Навчальний матеріал курсу розділений на три змістових модуля: «Поняття та принципи моделювання складних систем», «Оптимізаційне моделювання складних систем» і «Статистичне моделювання складних систем», відповідно до яких виконується поточний контроль знань шляхом тестування. Підсумковий контроль знань (диференційований залік) проводиться у письмовій формі.

Метою вивчення навчальної дисципліни «Основи моделювання складних систем» є формування в здобувача системи знань про ознаки та системні властивості складних систем, принципи управління, види моделей, а також набуття навичок моделювання певних об’єктів та процесів, зокрема побудови й аналізу оптимізаційних та статистичних моделей.

У результаті вивчення курсу студенти повинні знати основні поняття та принципи побудови моделей складних систем, вміти виконувати збирання та підготовку вихідних даних для статистичного моделювання, визначати та формалізувати цілі та обмеження оптимізаційних моделей, використовувати результати моделювання для прийняття управлінських рішень.

Не викликає сумніву, що для сучасного фахівця будь-якої галузі корисними є знання із моделювання систем, адже фахівець, в якого наявне системне мислення, завжди матиме конкурентні переваги на ринку праці.

ЗМІСТОВИЙ МОДУЛЬ 1 ПОНЯТТЯ ТА ПРИНЦИПИ МОДЕЛЮВАННЯ СКЛАДНИХ СИСТЕМ

Тема 1 СКЛАДНІ СИСТЕМИ ТА ЇХ СПЕЦИФІКА

Поняття системи. Ознаки та системні властивості систем. Принципи управління системою. Характерні особливості складних систем.

1.1 Поняття системи

Перш ніж визначити, що таке складна система, розглянемо, чим являється система взагалі і чому виникла необхідність протиставлення простих та складних систем. Без цих вихідних понять неможливо розкрити внутрішній зміст складних систем.

Термін «система» ввійшов у побут сучасної людини, яка стикається з ним дуже часто – супутникова система, геоінформаційна система, комп'ютерна система, система освіти, система торгівлі, система супермаркетів, енергетична система, система водопостачання тощо. Виходить, що усі природні й штучні об'єкти та процеси, які нас оточують, є системами. Дійсно, кожен з них можна подати як систему за потребою. Така безмежна кількість різноманітних систем зумовлює певні труднощі для створення точного та вичерпного визначення поняття «система», отже їх існує дуже багато. Але ми зупинимось на загальнонауковому визначенні поняття «система», яке є коротким, точним та відбиває найхарактерніші риси будь-якої системи.

Системою називають сукупність взаємопов'язаних **елементів**, що об'єднані **спільною метою** функціонування.

Елементами системи зазвичай є будь-які її складники. Наприклад, у технічній системі складниками є певні частини її конструкції (якщо ця система автомобіль, або комп'ютер, або геодезичний прилад), які пов'язані один з одним, функціонують спільно та злагоджено, завдяки чому забезпечують потрібний результат. Саме цей результат виражає мету функціонування певної системи.

Розташування елементів та взаємозв'язки між ними визначають поняття **структура системи**. Структура системи забезпечує виконання системою своїх функцій, а отже досягнення мети.

У процесі функціонування система змінює один свій стан на інший (автомобіль рухається або стоїть, програмний засіб обчислює завдання або очікує команду користувача тощо). Зміна станів системи супроводжується зміною значень сукупності величин, що характеризують стани певної системи. Їх називають **параметрами стану**. Системи, що протягом певного часу переходять з одного стану в інший, тобто функціонують, називають **динамічними**. Динамічна система може мати множину можливих станів, послідовно переходячи з одного стану в інший, проте вона у певний момент часу завжди перебуває тільки в одному із станів. Стан системи визначають її складники та зв'язки між ними. Розрізняють три види зв'язків між елементами системи:

– механічні (обмін фізичними зусиллями між елементами);

– трофічні (обмін енергією);

– сигнальні (обмін сигналами, тобто інформацією).

Метою системи називають певний кінцевий стан, за якого параметри її вихідних характеристик отримують найкращі (потрібні або задані) значення.

Розглядання певного об'єкта або процеса як системи називають **системним підходом**. У сучасній науці системний підхід займає одно з найважливих місць. Потрібність у системному підході зазвичай виникає під час вирішення таких задач як задачі аналізу та синтезу.

Аналіз (з грец. розклад), тобто розчленування предмету пізнання. Аналіз є методом дослідження предмету шляхом уявного або реального розділення його на складники та вивчення їхніх властивостей окремо. За методом аналізу визначають склад системи, зв'язки елементів та їхні функції.

Синтез «з гр. поєднання», тобто поєднання абстрагованих сторін предмета пізнання та відображення його як конкретної цілісності. Це метод вивчення об'єкта у його цілісності, у єдиному і взаємному зв'язку його частин. За методом синтезу також створюють нову систему з потрібними функціями та властивостями, або вдосконалюють існуючу.

У наукових дослідженнях систем задачі синтезу та аналізу пов'язані одна з одною. Процес аналізу призначений для встановлення складників системи, їх зв'язків та функцій. Синтез дає змогу поєднати частини розділеної системи і пізнати предмет як єдине ціле. Корисним уявляється такий приклад. Людина збирається вперше їхати на автомобілі. Спочатку вона аналізує систему - визначає, де місце водія, які засоби управління водію доступні, яке призначення педалей та керма, коробки передач, які параметри режимів руху індикує панель приладів. Далі, синтезуючи функції складників, людина керує рухом системи.

Треба зауважити, що доцільність застосування системного підходу до вивчення об'єктів та процесів полягає у тім, що абстрактні системи, тобто системи, різні за фізичною природою та структурою (навіть такі як технічні та біологічні), часто мають ідентичні властивості та поведінку, а отже для їх моделювання можна використовувати той самий математичний апарат.

Системний підхід передбачає застосування методів **системного аналізу**, який є науковим методом пізнання. Він являє собою послідовність дій щодо встановлення структурних зв’язків між змінними або елементами досліджуваної системи, які спираються на комплекс загальнонаукових, експериментальних, природничих, статистичних та математичних методів. У зарубіжній науковій літературі поняття системного підходу та системного аналізу – синоніми, тобто однакові за змістом.

1.2 Ознаки та системні властивості систем

Якщо апріорна інформація про об’єкт дослідження мінімальна (майже відсутні відомості про внутрішню будову), систему подають як непрозорий «чорний ящик», виділений з навколишнього середовища (рис. 1.1). Ця максимально проста модель системи відображає дві важливі її властивості: цілісність і відокремленість від навколишнього середовища. Одночасно «чорний ящик» пов’язаний з навколишнім середовищем певними зв’язками. Зв’язки, що впливають на систему, називають **входами системи**. Оскільки система реалізує певну мету, вона чинить вплив на стан навколишнього середовища. Ці зв’язки називають **виходами системи**.

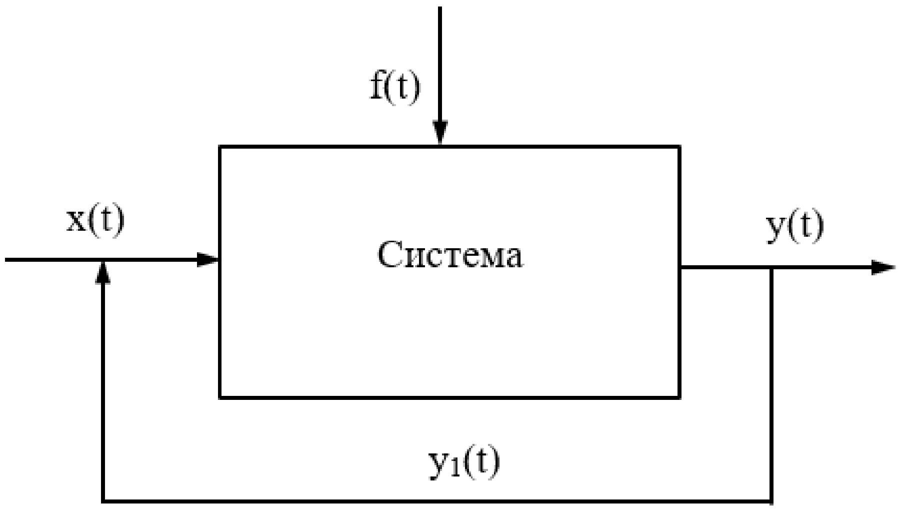


Рисунок 1.1 – Система «чорний ящик»:
$x(t)$ – вхід управління; $y(t)$ – вихід; $f(t)$ – вхід сторонніх випадкових впливів; $y_1(t)$ –зворотний зв’язок

Будь яка система характеризується такими властивостями (ознаками):

– границі системи, адже зрозуміло, що нескінченну систему неможливо уявити та аналізувати, тобто у такій системі немає змісту;

– відкритість, яка проявляється у зв’язках системи з навколишнім середовищем;

– структурованість системи, що проявляється у наявності елементів та зв'язків між ними. Зауважимо, що структура системи залежить від завдання дослідження;

– динамічність системи визначається здатністю змінювати з часом свій стан;

– ієрархічність системи – це така будова системи, за якою кожна її частина також є системою, що складається з своїх елементів, глибина поділу яких на складники визначається завданням дослідження. Зауважимо, що будь-який елемент системи так само можна розглядати як систему;

– емерджентність (з англійської «emergence» – виникнення, поява нового) системи полягає у наявності в неї таких властивостей, яких нема в її елементів, тобто поява нових властивостей;

– підпорядкованість функціонування всіх елементів системи спільній меті. Ця ознака формується в силу об'єктивної необхідності.

Кожна з розглянутих ознак для будь-якої системи є обов'язковою. Якщо в досліджуваної системи не дістає хоча б однієї з них, то це означає, що ми помилково визначили систему, зокрема, її границі. Тобто якщо в системи немає границь, або система ізольована від навколишнього середовища, або деякі з її складників не мають зв'язків з іншими, або якщо система немає загальної мети, то і змісту немає у вивченні такої сукупності елементів.

1.3 Принципи управління системою

Усі процеси та явища поза межами системи є **зовнішнім середовищем**. Серед них розрізнюють сукупність об'єктів, зміна властивостей яких впливає на систему, та сукупність об'єктів, властивості яких змінюються у результаті зміни стану системи. Отже система проявляє свої властивості у процесі взаємодії із зовнішнім середовищем.

У сукупності зовнішніх впливів на систему, що наведена на рисунку 1.1, виділяють корисний сигнал – сигнал управління $x(t)$ та випадкові впливи - сигнали $f(t)$, які є завадами. Сигнал на виході системи $y(t)$ реалізує мету системи. Ефективність функціонування системи визначається ступенем досягнення її мети. Очевидно, що випадкові впливи, тобто завади, спотворюють сигнал на виході системи $y(t)$, що зменшує її ефективність. Для підвищення ефективності на вхід системи подають сигнал управління $x(t)$.

Управлінням називають **цілеспрямований** вплив на систему з метою отримання найкращого ефекту. Отже управління являє собою таку організацію того або іншого процесу, що, змінюючи стани системи, забезпечує досягнення певних цілей.

Розрізняють два основних принципа управління – управління за збуренням та управління за відхиленням. Перший принцип – управління за збуренням – передбачає управління розімкненою системою, тобто системою, у структурі якої немає зворотного зв'язку. Цей принцип полягає у тім, що сигнал управління змінюється так, що він компенсує вплив завади. Перевага цього принципу управління полягає у тім, що вихідний сигнал може зовсім не відхилятися від потрібного значення, тобто ефективність функціонування системи може становити 100 %, але це тільки теоретично. На практиці у більшості випадків вичерпні дані про дію випадкових впливів f(t), тобто момент впливу, тривалість, величина та ін. не відомі. Перевага принципу управління за збуренням полягає у можливості заздалегідь врахувати усі випадкові впливи та їх компенсувати шляхом зміни, наприклад, сигналу управління. Але в цього принципа є недолік, адже не про всі випадкові впливи є інформація, отже більшість з них не вдається врахувати.

Другий принцип – управління за відхиленням – передбачає, що система замкнена зворотним зв'язком. Зворотний зв'язок передає дані про функціонування системи з її виходу на вхід, де сигнал зворотного зв'язку порівнюється з вхідним, який відповідно корегується, та на вхід надходить алгебраїчна сума сигналів. У більшості автоматичних систем застосовують від'ємний зворотний зв'язок, тобто сигнал зворотного зв'язку $y_1(t)$ віднімається від вхідного сигналу x(t). Тоді можливі три стани: $|y_1(t)| = |x(t)|$, $|y_1(t)| > |x(t)|$ та $|y_1(t)| < |x(t)|$. У першому випадку сигнал управління не змінюється ($|x(t)| - |y_1(t)| = 0$). У другому випадку сигнал управління зменшується, оскільки різниця стає від'ємною, а у третьому випадку – збільшується, оскільки різниця стає додатною. Перевага принципа управління за відхиленням – у можливості врахувати усі впливи та завади, а недолік – принципова наявність похибки на виході системи, оскільки зворотний зв'язок надає інформацію про похибку тільки після її виникнення.

Системи класифікують за низкою ознак: за природою утворення, за ступенем зв'язку із зовнішнім середовищем, за залежністю від часу, за керованістю, за місцем в ієрархії систем, за ступенем складності та ін.

1.4 Характерні особливості складних систем

Починаючи з середини XX століття, як результат технічного прогресу з'явились людино-машинні обчислювальні системи, розроблюються космічні проєкти та інші складні технічні задачі. Поряд з цим у соціальному управлінні створюються сучасні виробничі комплекси, що тягне за собою розвиток міст та комплексних заходів з охорони природи. Отже у науці відбувся перехід до таких наукових задач, як проблеми організації і функціонування складних об'єктів і людині доводиться звертатись до систем, границі і склад яких зовсім не очевидні і у кожному разі вимагають спеціальних досліджень. Для вирішення подібних питань і почали розроблятись складні системи.

У сучасній науці поняття **складна система** (з англ. «complexity» – «складність») дуже широко використовується. До категорії складних систем, перш за все, належать природні системи, зокрема живі організми – від людини до самих маленьких бактерій, які проявляють такі важливі властивості, як адаптація до навколишнього середовища, величезна різноманітність та надзвичайна складність внутрішньої структури. Складність цих систем зумовлена складністю механізмів взаємодії елементів. До складних систем належать, зокрема, так звані **організаційні системи**, під якими розуміють деякі організації, наприклад, виробничі підприємства.

Розглянемо коротко основні характерні риси складних систем.

Головною особливістю складних систем є те, що окрім відомих і чітко визначених факторів, що впливають на стан і динаміку системи, визнається вплив випадкових (**стохастичних**) факторів. Наукові та повсякденні факти свідчать, що не все, що відбувається навколо нас, можна точно пояснити. А отже доводиться користуватись категорією випадковості та через неї сприймати досліджувані процеси, визнаючи таким чином можливість зміни і переходу складної системи в абсолютно непередбачувані стани, тобто поведінка складної системи може опинитись непередбачуваною.

Інша особливість складної системи полягає у її **ієрархічності**. Складні системи можуть мати кілька рівнів ієрархії, а такі системи як біосфера або світова економіка налічують десятки рівнів. Це спричинено тим, що складники такої системи, які ми розглядаємо як її елементи, можуть у свою чергу бути також складними системами, тобто мати власну поведінку, стан і структуру, мати зворотні зв'язки, причому як від'ємні, так і додатні, а також спостережувані й контрольовані входи та виходи.

Великим і складним системам притаманні також властивості **цілісності** та е**мерджентності**. Цілісність системи означає, що всієї її частини слугують спільній цілі та сприяють формуванню найкращих (оптимальних) результатів у смислі прийнятого критерію ефективності. Емерджентність (поява нового) означає, що великі й складні системи мають такі властивості, які не властиві жодному з її елементів. Емерджентність або системні властивості кардинально відрізняють системи від не систем. Інакше кажучи, об'єднання підсистем з різною природою (наприклад, економічної, технічної, соціальної) і структурою в складні системи відбувається при їхньому взаємному впливі одна на одну, що й створює нову системну властивість, якої не має жодна з підсистем складної системи. Емерджентність відображає неможливість зведення властивостей системи до суми властивостей її компонентів.

Для розуміння поняття складної системи доцільно порівняти особливості складної та простої системи. Проста система зазвичай характеризується тим, що має однозначний зв’язок між зовнішніми впливами та вихідним параметром, тобто реакцією системи. Малі зміни в характері зовнішнього впливу викликають малі зміни в реакції системи, та навпаки – істотні за величиною зміни зовнішнього впливу викликають великі зміни в реакції системи. Як приклади поведінки таких систем можна навести рух на автомобілі, роботу з комп’ютером, або геодезичним приладом. Отже для будь-якого зовнішнього впливу поведінку простої системи можна передбачити.

Складні системи мають велику кількість елементів і характеризуються розгалуженою структурою, виконують складніші функції. Для складної системи характерні такі властивості як наявність певної кількості зворотних зв’язків, у складній системі малі зміни зовнішніх впливів викликають суттєві, часто якісні, зміни в реакції системи, непередбачуваність та ін.

Контрольні запитання

1. Що означає термін «система»? Наведіть визначення.
2. Поясніть сутність понять «елементи системи», «структура системи».
3. Що розуміють як стани системи та параметри стану?
4. Поясніть сутність поняття «мета системи» та наведіть приклади.
5. Поясніть сутність понять «аналіз» та «синтез».
6. Що являє собою системний підхід? Охарактеризуйте його призначення та основні риси.
7. Поясніть сутність моделі «чорний ящик».
8. Перелічіть системні ознаки та охарактеризуйте їх.
9. Які елементи системи визначають її границі щодо навколишнього середовища?
10. Чи є зворотний зв’язок системи її елементом, або він є елементом навколишнього середовища?
11. Якою ознакою визначається наявність структури системи?
12. Поясніть сутність поняття «управління системою». Що є метою управління?
13. Охарактеризуйте принцип управління за збуренням. У чому полягає перевага та недолік цього принципу управління?
14. Охарактеризуйте принцип управління за відхиленням. У чому полягає перевага та недолік цього принципу управління?

15. Поясніть, чим спричинено поняття «складна система» та які системи до них належать.

16. Охарактеризуйте властивості складних систем, які відрізняють їх від звичайних.

Тема 2 СПОСОБИ МОДЕЛЮВАННЯ СКЛАДНИХ СИСТЕМ

Сутність процесу моделювання. Види моделей. Характеристика етапів моделювання. Методи моделювання складних систем. Формальна математична модель об’єкта моделювання. Поняття статичної і динамічної моделі. Загальна схема створення математичної моделі.

2.1 Сутність процесу моделювання

Модель – це матеріальний або уявний об’єкт, який у процесі дослідження заміщує об’єкт-оригінал так, що його безпосереднє вивчення дає нові знання про об’єкт-оригінал. Отже, модель необхідна:

– для того, щоб зрозуміти, як влаштований конкретний об’єкт: яка його структура, основні властивості, закони розвитку і взаємодія з навколишнім середовищем;

– для того, щоб навчитися управляти об’єктом (або процесом) і визначити найкращі засоби управління при заданих меті і критеріях;

– для того, щоб прогнозувати прямі та побічні наслідки реалізації заданих засобів і форм впливу на об’єкт.

Моделі можуть бути описовими (словесними або вербальними), предметними (матеріальними), графічними, структурними, функціональними, інформаційними, логічними, математичними та ін. Один об’єкт можна представити різними моделями, а різні об’єкти можна описати однією моделлю. Модель об’єкта має відповідати меті аналізу, а отже відбивати характерні властивості об’єкта дослідження, необхідні досліднику.

Моделюванням є процес побудови, вивчення та застосування моделей, а отже він охоплює усі етапи від створення моделі до реалізації отриманих рішень щодо об’єкта-оригінала. Головна особливість моделювання полягає у тому, що цей метод використовує об’єкти-замісники, тобто інструментом для пізнання певного об’єкта є модель.

Необхідність застосування методу моделювання зумовлена, перш за все, тим, що існує багато об’єктів, або проблем щодо цих об’єктів, які піддавати дослідженням не можна або недоцільно, або дослідження вимагає багато часу та коштів.

Процес моделювання вимагає участі у ньому трьох таких компонентів:

– суб'єкта, тобто дослідника;

– об'єкта дослідження;

– моделі об'єкта дослідження.

Для дослідження складних систем найчастіше застосовують математичні моделі. Математична модель являє собою певну систему математичних співвідношень і логічних виразів, якими зазвичай є певні функції, рівняння, нерівності та ін., які відображають істотні властивості об'єкта дослідження. Переваги математичної моделі полягають у низькій вартості їх створення, швидкому отриманні результатів дослідження, можливості шляхом розрахунків провести експерименти та перевірити правильність побудови моделі. Також треба відзначити, що характерні риси деяких явищ та процесів, що принципово відрізняються за фізичною природою, можна описати однаковими математичними співвідношеннями. Це ще одна перевага математичного моделювання, отже воно ґрунтується на аналогії процесів і явищ, різних за фізичною природою, але однакових за математичним описом.

У цій дисципліні ми вивчатимемо тільки математичні моделі.

2.2 Характеристика етапів моделювання

У силу великої розмаїтості складних систем їх математичне моделювання практично не піддається науковій формалізації. Принцип побудови математичної моделі істотно визначається природою конкретної досліджуваної системи. Проте в моделюванні прийнято виділяти деяку послідовність етапів:

– постановка проблеми та її якісний аналіз;

– побудова моделі;

– математичний аналіз моделі;

– аналіз отриманих рішень;

– перенесення результатів дослідження моделі на оригінал.

Кожен з етапів передбачає використання системного аналізу, який є науковим методом пізнання та являє собою послідовність дій із встановлення структурних зв'язків між змінними (параметрами) або елементами досліджуваної системи. Він спирається на комплекс загальнонаукових, експериментальних, природничих, статистичних та математичних методів.

На першому етапі під час постановки проблеми та її аналізу головним є чітке формулювання проблеми, визначення припущень та питань, які вимагають відповідей. Цей етап полягає у визначенні найважливих властивостей моделюємого об'єкта та відкидання другорядних, вивчення структури об'єкта, залежностей між його елементами, а також формулювання гіпотез, що пояснюють поведінку об’єкта. Під час реалізації першого етапу моделювання найбільш трудомісткою є задача підготовки вихідної інформації. Моделювання пред’являє жорсткі вимоги до системи інформації, оскільки реальні можливості отримання інформації обмежують вибір моделей. Одночасно треба брати до уваги, окрім можливості підготовки інформації за обмежений час, ще й вартість інформації, яка має не перевищувати ефект від її використання.

У процесі підготовки інформації звертаються до методів теорії імовірностей та математичної статистики, наприклад, для організації вибіркових обстежень, оцінювання вірогідності даних, визначення імовірних значень параметрів та ін.).

Другий етап – побудова математичної моделі – полягає у формалізації проблеми шляхом описання її конкретними математичними залежностями, такими як деякі функції, рівняння, нерівності та ін.). Перед усім обирають тип математичної моделі та вивчають можливості її застосування, а далі уточнюють конкретний перелік змінних, параметрів моделі та форми зв’язків.

Оскільки модель відбиває тільки деякі властивості об’єкта досліджування, що цікавлять дослідника, для одного об’єкта можна створити декілька різних моделей, які моделюють його різні характерні риси. Будь-яка модель вимагає перевірки її **адекватності**, оскільки можна замінювати об’єкт моделлю тільки якщо вони в достатній мірі схожі. Адекватність (латинською «adaequatus» – прирівняний) – це відповідність моделі об’єкту або процесу, що моделюється.

Очевидно, що модель втрачає сенс, якщо вона тотожна оригіналу, оскільки перестає бути моделлю, також у разі надмірної відмінності від оригіналу. Отже, адекватність є поняттям дещо умовним, оскільки повної відповідності моделі об’єкту бути не може. Але властивості моделі мають адекватно відбивати істотні для дослідження властивості об’єкту. Дослідження моделі не має сенсу, поки не вирішено питання про її адекватність, тобто чи правильно відображає модель досліджувану систему.

Третій етап процесу моделювання полягає у дослідженні моделі як самостійного об’єкта. Найчастіше застосовують таку форму дослідження як проведення модельних експериментів, за яких змінюють умови функціонування моделі та систематизують дані про її поведінку. Мета математичного аналізу моделі – визначення її загальних властивостей (рішень). На цьому етапі застосовують тільки математичні методи дослідження. Найбільш важливим питанням є доказ існування рішень в побудованої моделі, зокрема визначають, чи є рішення єдиним, які змінні можуть міститись в ньому, якими співвідношеннями вони пов’язані, у яких межах вони змінюються та за яких умов. Зауважимо, що аналітичне дослідження є переважним порівняно з емпіричним, тобто числовим. Оскільки аналітичне дослідження передбачає опис системи за допомогою аналітичних виразів, наприклад рівнянь або системи рівнянь, що дозволяє досліджувати поведінку системи за різних значень її параметрів та узагальнювати і прогнозувати результати.

Для отримання числового рішення виконують розрахунки певних параметрів моделі на підставі попередньо розроблених алгоритмів. Але при дослідженні складних систем такі завдання, як правило, мають велику розмірність, а отже потребують опрацювання великих масивів інформації. Крім того, рішення складних систем можуть опинитись багатоваріантними, тобто мати багато альтернативних рішень. Зазвичай дослідження за числовими методами можуть доповнити результати аналітичного дослідження, одночасно воно єдино здійснюване для багатьох моделей, оскільки далеко не завжди вдається побудувати аналітичні залежності для складних систем.

Аналіз числових результатів моделювання застосовують для оцінювання їх правильності й повноти, а також можливості їх практичного використання.

Поки ще не створено чіткої теорії перевірки правильності математичних моделей. Математичні методи перевірки можуть виявляти некоректні побудови моделей (коли доводиться невирішуваність моделі або не підтверджуються прийняті статистичні гіпотези) й тим самим звужувати клас потенційно правильних моделей.

Останній етап полягає у перенесенні отриманих щодо моделі відомостей на об’єкт дослідження, причому знання про модель корегують, враховуючи властивості об’єкта, які були змінені під час побудови моделі, або зовсім відкинуті. Підставою для перенесення певного отриманого на моделі результата на оригінал є його пов’язаність ознаками подібності оригіналу та моделі. На цьому етапі виконують практичну перевірку знань, що отримані за допомогою моделі, а також застосовують їх для створення узагальнюючої теорії об’єкта, для перетворення об’єкта та управління об’єктом.

2.3 Методи моделювання складних систем

Системний аналіз є сукупністю методологічних засобів для розроблення та обґрунтування рішень із складних проблем, як наприклад соціальних, економічних, наукових, технічних та ін. Відмінність (особливість) методів системного аналізу полягає у тому, що прийняття рішень доводиться здійснювати в умовах невизначеності, зумовленій впливом факторів, які не підлягають кількісному оцінюванню. Методи системного аналізу направлені на формування альтернативних варіантів рішення проблеми та порівняння їх за деякими критеріями ефективності.

Найважливіші принципи системного аналізу такі:

– процес прийняття рішень має починатись з чіткого формулювання кінцевих цілей;

– необхідне виявлення можливих альтернативних шляхів досягнення мети.

Центральною у системному аналізі є побудова узагальненої моделі, що відбиває всі взаємні зв'язки реальної ситуації, які можуть проявитися під час реалізації рішення. Цю модель досліджують для:

– визначення близькості кожного результата з альтернативних варіантів до бажаної мети;

– порівняння витрат ресурсів для кожного варіанта;

– оцінювання ступеня чутливості моделі до зовнішніх впливів.

Системний аналіз спирається на низку прикладних математичних методів, зокрема на:

– методи дослідження операцій, що полягають у розробленні і практичному застосуванні методів оптимального управління організаційними системами;

– метод експертних оцінок, який застосовують, коли недостатньо наявної інформації або взагалі відсутні дані про об'єкт. У цих випадках звертаються до компетентних фахівців-експертів для одержання відсутньої інформації, а потім спеціальним чином її обробляють;

– метод критичного шляху, що вирішує задачі упорядкування та координації, та ін.

Методологічні засоби розв'язання проблем за допомогою системного аналізу визначають залежно від складності та особливостей досліджуваної проблеми. Наприклад, якщо має місце одна чітко виражена мета та ступінь її досягнення треба оцінити за одним критерієм, застосовують методи математичного програмування. Коли ступінь досягнення мети треба оцінити за кількома критеріями, застосовують методи теорії корисності, що дозволяють впорядкувати критерії та визначити важливість кожного з них. Якщо проблема визначена взаємодією декількох осіб або систем, кожна з яких переслідує свої цілі, застосовують методи теорії ігор [6, 10]. Для дослідження зв'язків між елементами системи або її параметрами застосовують кореляційний аналіз, аналіз невизначеності та інші способи.

Контрольні запитання

1. Поясніть, що таке модель та яка мета моделювання.
2. Перечисліть види моделей, які вам відомі, та наведіть приклади.
3. Поясніть, яку модель називають математичною.
4. Перелічіть етапи моделювання та їхню послідовність.

5. Охарактеризуйте сутність етапів моделювання.

6. У чому полягає етап формалізації проблеми дослідження? Яка інформація потрібна для реалізації цього етапа?

7. Яким вимогам має задовольняти модель? Поясніть термін «адекватність моделі».

8. Охарактеризуйте етап дослідження моделі та які задачі вирішують на цьому етапі.

9. Поясніть, які рішення називають альтернативними.

10. Охарактеризуйте сутність системного аналізу та його особливості.

11. Які методи моделювання та дослідження застосовує системний аналіз?

12. Для дослідження яких явищ застосовують кореляційний аналіз?

ЗМІСТОВИЙ МОДУЛЬ 2 ОПТИМІЗАЦІЙНЕ МОДЕЛЮВАННЯ СКЛАДНИХ СИСТЕМ

Тема 3 ОСОБЛИВОСТІ І СФЕРИ ВИКОРИСТАННЯ ОПТИМІЗАЦІЙНИХ МЕТОДІВ

Поняття оптимізації. Математична постановка оптимізаційних задач. Вибір критерію оптимізації. Класифікація оптимізаційних моделей і методів розв’язання та види оптимізаційних моделей.

3.1 Поняття оптимізації

Шляхом дослідження систем визначають структуру певної системи, її властивості та закони взаємодії з навколишнім середовищем. Отриману з досліджень інформацію застосовують для управління системою. Управління завжди передбачає досягнення деякої мети та отже нерозривно пов’язане з метою. Але завжди під час управління прагнуть отримати найкращий результат, що вимагає визначення певних управлінських рішень, кількість яких може бути досить великою. Для вибору з усіх альтернативних рішень найкращого їх порівнюють одне з одним. Ознак, за якими треба порівнювати альтернативні рішення може бути багато, наприклад, одні з них вимагають великих витрат, інші – менших, одні потребують великого часу, інші – меншого, одні з них вимагають залучення кваліфікованих працівників, інші – некваліфікованих. Ознаки, за якими порівнюють альтернативні рішення називають **критеріями**. Критерії призначені для оцінювання ефективності управління, яка визначається ступенем досягнення мети управління. Якщо задача управління полягає у досягненні однієї мети, то для порівняння рішень потрібен один критерій. Таку задачу називають однокритеріальною. Якщо у задачі декілька цілей – багатокритеріальною.

Одним з найпоширених методів управління є оптимальне управління, тобто таке, що дає оптимальний результат, за який вважають найкращий результат, але не взагалі найкращий, а тільки такий, що задовольняє певним умовам, які називають **обмеженнями**.

Методи оптимального управління складними системами широко застосовують також в геодезії та землеустрої. Ви вже знайомі з курсу «Математична обробка геодезичних вимірів» з методом найменших квадратів. Це оптимізаційний метод, він використовує оптимізаційну модель для пошуку найкращого (найточнішого) результату врівноваження геодезичних побудов. Як приклади можна навести ще такі задачі як врівноваження лінійно-кутових геодезичних мереж, оптимальне проєктування рельєфу, оптимальна розбивка земельної ділянки на менші ділянки, визначення оптимальних висот зовнішніх геодезичних знаків та багато інших задач геодезичних задач.

До землевпорядних задач оптимального управління належать, наприклад: оптимальний перерозподіл земель у схемі землеустрою района; оптимізація перерозподілу земель у схемі землеустрою району, встановлення розмірів та структури землеволодінь та землекористувань сільськогосподарських і несільськогосподарських підприємств при міжгосподарському землеустрої, формування сировинних зон промислових підприємств із переробки сільськогосподарської продукції, обґрунтування сільськогосподарського освоєння, меліорації та трансформації земельних угідь у робочих проектах, пов’язаних з використанням та охороною землі, будова території багаторічних насаджень (садів, ягідників, виноградників), сінокосів та пасовищ, оптимізація показників агроекономічного обґрунтування проектів землеустрою тощо.

У багатьох випадках процес управління вимагає витрати будь-яких ресурсів: часу, матеріалів, електроенергії, кваліфікації персоналу. Тому під час вибору способу управління потрібно брати до уваги не тільки досягнення мети, а також кількість та вартість ресурсів, які доведеться затратити для досягнення цієї мети. Відомо, що ресурси, у тому числі й використовувані для управління, завжди обмежені (завжди не дістає грошей, часу та кваліфікованого персоналу). Саме це обмежує управлінські рішення.

Отже, математично задача оптимального управління полягає у визначенні такого рішення $x_{опт}$, що перетворює критерій ефективності на екстремум (мінімум або максимум) та задовольняє системі обмежень ресурсів. Такий клас задач називають задачами оптимізації за наявності обмежень, а рішення $x_{опт}$ (найкраще за певних умов) – **оптимальним**. Такі задачі оптимізації у свій час (середина XX століття) призвели до перегляду класичних методів оптимізації і створенню нових методів, відомих як **методи програмування**. Цей факт спричинений тим, що задача оптимізації за наявності обмежень на відміну від багатьох інших математичних задач припускає не один розв'язок, а множину різних розв'язків. Отже виникає додаткова задача вибору з цієї множини розв'язків такого, що з певної точки зору є найкращим. Можна навести багато прикладів подібних задач. Наприклад, з одного міста до іншого можна проїхати, користуючись різними видами транспорту: залізничним, повітряним, водним, автобусним, автомобільним. Додатково окремою задачею можна вважати вибір найбільш вигідного виду транспорту з погляду часу проїзду, або вартості, зручності та ін.

Зміст математичного програмування складають спеціальні методи пошуку екстремуму. Поняття програмування тут вживають як планування, на відміну від програмування для ЕОМ.

3.2 Математична постановка оптимізаційних задач. Вибір критерію оптимізації

Задачі, мета яких полягає у знаходженні найкращого (оптимального) з погляду певного критерію або критеріїв варіанту використання наявних ресурсів, називають оптимізаційними. Оптимізаційні задачі вирішують за допомогою оптимізаційних моделей за методами математичного програмування.

Наведемо такий приклад оптимізаційної задачі як будівництво ділянки залізничної магістралі. Для її спорудження є певна кількість ресурсів: людей, будівельних машин і механізмів, ремонтних майстерень, вантажних автомобілів тощо. Потрібно так спланувати будівництво, тобто призначити черговість робіт, розподілити машини і людей за ділянками шляху, забезпечити ремонтні роботи, щоб воно було завершене у мінімально можливий термін.

Структура оптимізаційної моделі включає:

– цільову функцію;

– області припустимих рішень;

– систему обмежень, що визначають цю область.

Цільова функція у самому загальному вигляді, у свою чергу, так само складається з трьох елементів:

– керованих змінних;

– некерованих змінних;

– форми функції (виду залежності між змінними).

Розглядаючи певну довільну систему рівнянь, яка містить m рівнянь та n невідомих, можна виділити три типи задач:

– якщо m = n, це алгебраїчна задача, в якої один розв’язок;

– якщо m > n, це задача перевизначена, що, як правило, не має розв’язків;

– якщо m < n, це задача невизначена, в якої нескінченно багато розв’язків.

Найчастіше доводиться мати справу із задачами третього типу, тобто з такими, що мають нескінченну множину розв’язків.

Математичний вираз, що дає кількісну оцінку ступеня досягнення мети управління, називають **критерієм ефективності** управління. Найкращим або **оптимальним** способом управління буде такий, за якого критерій ефективності досягає **мінімального** або **максимального** значення. При виборі, наприклад, режиму польоту за критерій якості управління можна прийняти або вираз для кількості палива, що витрачається на одиницю шляху, або шлях, на який витрачається одиниця палива. Найбільш економічному, тобто оптимальному, режиму відповідатиме або мінімальне (у першому випадку), або максимальне (у другому випадку) значення критерію.

Математичне формулювання оптимізаційної задачі полягає у визначенні:

– мети управління, яку виражають через критерій ефективності;

– певних обмежень, що зазвичай є системою алгебраїчних рівнянь або нерівностей, які виражають обмеженість ресурсів або інших величин, використовуваних при управлінні.

Рішення про спосіб управління, що задовольняє всім поставленим обмеженням і перетворює на мінімум (максимум) критерій ефективності, називають **оптимальним рішенням**.

Отже, сутність всіх оптимізаційних задач збігається до пошуку такого розв’язку $\bar{X}_{\text{опт}}$, що перетворює на екстремум критерій ефективності, виражений як функція від елементів прийнятого розв’язку $\bar{X}_{\text{опт}}$. Цей вираз називають **цільовою функцією** F(x).

Потрібно відзначити, що методи розв’язання оптимізаційних задач є не аналітичною, а алгоритмічною формою розв’язання задач. Вони не дають формулу, яка виражає остаточний результат, а лише вказують обчислювальну процедуру, що призводить до розв’язання задачі. Тому оптимізаційні методи стають ефективними при використанні обчислювальної техніки.

Відзначимо, що до оптимізаційних задач, як правило, незастосовні методи класичного аналізу для відшукання умовних екстремумів. Це зумовлене такими специфічними їхніми особливостями:

– коли на елементи розв’язку $\bar{X}_{\text{опт}}$ накладені обмеження, екстремум найскоріше досягається не в точках, де похідні дорівнюють нулю, а на границі області обмежень;

– у практичних задачах число змінних і число обмежень настільки велике, що пошук екстремуму шляхом визначення похідних є не ефективним;

– у багатьох задачах математичного програмування цільова функція не має похідних (наприклад, задана тільки для цілочислових значень аргументів).

У зв'язку з цим метою математичного програмування є створення аналітичних методів з визначення розв'язку або ефективних обчислювальних способів одержання наближеного розв'язку оптимізаційної задачі.

Математичне моделювання процесів управління складними системами є, з одного боку, дуже важливим і складним, а з іншого боку – таким, що практично не піддається науковій формалізації. Неодноразові спроби, виділити загальні принципи створення математичних моделей призводили або до декларування рекомендацій самого загального характеру, які важко застосовувати для розв'язання конкретних проблем, або, навпаки, до появи рекомендацій, що можна застосувати у дійсності тільки до вузького кола задач. Тому доцільно вивчати техніку математичного моделювання на конкретних прикладах.

Спільною для оптимізаційних задач є проблема пошуку найбільшого або найменшого (**оптимального**) значення певної функції, що відбиває **мету управління** системою, або, як ще кажуть, **цільової функції**. Пошук оптимального значення здійснюється на певній підмножині припустимих значень змінних, що описують стан цієї системи, яку називають **множиною припустимих планів**.

3.3 Класифікація оптимізаційних моделей і методів розв'язання

Нехай на певній множині *D* визначено функцію f(x). Нагадаємо, що точку x^*, що належить D ($x^* \in D$), називають **точкою глобального максимуму**, якщо для будь-якого $x \in D$ виконується нерівність $f(x) \le f(x^*)$. У цьому випадку значення $f(x^*)$ називають **глобальним максимумом функції**. Точку $\dot{x}$ називають **точкою локального максимуму**, якщо існує певне оточення цієї точки, у будь-якій точці якого значення функції менші за $\dot{x}$ ($f(x) \le f(\dot{x})$). Аналогічно визначають **глобальний** і **локальний мінімуми**. Узагальнюючим поняттям для максимуму й мінімуму є такий термін, як **екстремум** (оптимум).

Задачі математичного програмування дуже різноманітні. Їх математичне моделювання практично не піддається науковій формалізації через те, що принцип побудови математичної моделі істотно залежить від конкретної природи досліджуваної системи. Проте у цих задачах прийнято виділяти певну послідовність етапів дослідження, зокрема:

– постановка задачі;

– словесне формулювання задачі з визначенням мети її розв'язання та обмежень, що впливають на неї (побудова вербальної моделі);

– формалізація задачі – побудова адекватної математичної моделі. На цьому етапі цільову функцію F(x) виражають як залежність від розв’язку $\bar{X}$, а обмеження записують як систему рівностей та нерівностей;

– розв’язання задачі на базі математичної моделі;

– перевірка отриманих результатів на їхню адекватність природі досліджуваної системи, можливе корегування первісної моделі;

– розробка рекомендацій на підставі отриманого розв’язку.

Введемо низку визначень.

Розв’язком (або **планом**) називають певний вибір параметрів, що залежать від нас. Параметри, сукупність яких утворює розв’язок, називають **елементами розв’язку**. Як елементи розв’язку можуть бути числа, вектори, функції, фізичні ознаки та ін.:

$$x = (x_1, x_2, \ldots, \ldots, x_j, \ldots, x_n).$$

Система обмежень за ресурсами формує множину припустимих розв’язків (планів) D. Той факт, що розв’язок x належить множині припустимих розв’язків D, записують в такий спосіб:

$$x \in D.$$

Оптимальним розв’язком або **оптимальним планом** називають таке рішення, що перетворює цільову функцію F(x) на максимум або мінімум.

Отже, у найбільш загальному вигляді оптимізаційну задачу формулюють так: при заданих обмеженнях знайти таке рішення $x = \bar{X}_{\text{опт}}$, що перетворює цільову функцію F(x) на максимум або мінімум:

$$F^* = \underbrace{extr}_{x \in D}\{F(x, \alpha)\}$$

де α – система обмежень задачі.

Залежно від вигляду цільової функції F(x) і системи обмежень α виокремлюють такі методи розв’язання оптимізаційних задач.

Лінійне програмування. Застосовують, якщо в моделі цільова функція $F(x)$ є лінійною, а множина D, на якій шукають її екстремум, задана системою лінійних рівнянь і нерівностей.

Нелінійне програмування. Тут є нелінійними цільова функція та обмеження. У нелінійному програмуванні виділяють такі класи задач:

– **опукле програмування** – коли цільова функція є опуклою (якщо розглядають задачу її мінімізації) і опуклою є множина, на якій вирішується екстремальна задача;

– **квадратичне програмування** – коли цільова функція є квадратичною, а обмеження – лінійні рівності або нерівності.

Дискретне програмування. Даний метод використовують, коли на елементи рішення x накладено вимогу дискретності, наприклад, цілочисельності. Така вимога істотно ускладнює розв’язання задачі, оскільки застосування стандартних прийомів (вирішити задачу як аналогову, а потім заокруглити результат до цілого значення) неможливо.

Динамічне програмування. Це метод, що дозволяє шляхом покрокової оптимізації певних проміжних цільових функцій отримати загальний результуючий оптимум. У задачах динамічного програмування цільова функція $F(x)$ є адитивною або мультиплікативною функцією змінних х.

Стохастичне програмування. Даний вид програмування використовують, коли параметри умов або елементи розв’язку є випадковими величинами, що зумовлене невизначеністю, яка породжує ризикованість прийнятих рішень. У стохастичному програмуванні труднощі виникають не тільки під час розробки методів розв’язання задач, а й під час їхньої постановки.

Евристичне програмування. Застосовують для розв’язання задач, в яких точний оптимум знайти алгоритмічним шляхом неможливо через величезне число варіантів. У такому разі відшукують не оптимальний, а досить гарний з погляду практики розв’язок.

Контрольні запитання

1. Охарактеризуйте сутність та особливості оптимізаційних задач.
2. Поясніть поняття «оптимальне рішення», критерій ефективності», «система обмежень».
3. Охарактеризуйте структуру оптимізаційної моделі, які елементи вона має містити.
4. Що являє собою цільова функція оптимізаційної задачі? Яке її призначення?
5. З яких причин до оптимізаційних задач не застосовують класичні методи пошуку умовного екстремуму функції? Які класичні методи пошуку екстремуму функції вам відомі?
6. Які загальні етапи розв’язання оптимізаційних задач прийнято виділяти?
7. Охарактеризуйте етап постановки оптимізаційної задачі. Які його особливості?
8. Поясніть, яку модель називають вербальною моделлю.
9. Охарактеризуйте етап формалізації оптимізаційної задачі. Яка його сутність, призначення та особливості?
10. Поясніть, у чому полягає адекватність математичної моделі системи та як перевірити чи є модель адекватною.

11. Охарактеризуйте інформацію, що містить числовий розв’язок оптимізаційної задачі.

12. За якими ознаками класифікують оптимізаційні моделі і методи розв’язання?

13. Дайте визначення понять: план, припустимий план, оптимальний план, розв’язок оптимізаційної задачі.

14. На чому ґрунтується класифікація моделей і методів розв’язання оптимізаційних задач? Які класи моделей і методів виділяють у математичному програмуванні?

15. Що являє собою множина припустимих розв’язків оптимізаційної задачі?

16. У якому випадку для моделювання можна скористатись лінійною оптимізаційною моделлю?

17. Які обов’язкові елементи містить оптимізаційна модель?

ТЕМА 4 МЕТОДИ РОЗВ’ЯЗАННЯ ОПТИМІЗАЦІЙНИХ ЗАДАЧ

Загальна форма задачі лінійного програмування (ЗЗЛП). Основні властивості ЗЗЛП та її перша геометрична інтерпретація. Геометрична інтерпретація множини припустимих рішень. Канонічна форма задачі лінійного програмування (КЗЛП). Симплекс-метод.

4.1 Загальна форма задачі лінійного програмування

Серед задач математичного програмування найпростішими є задачі лінійного програмування, для яких характерним є таке:

– цільова функція $F(x)$ лінійно залежить від елементів розв’язку $x_1, x_2, \dots, x_n$;

– обмеження, що накладають на елементи розв’язку, є лінійними рівностями та нерівностями щодо $x_1, x_2, \dots, \dots, x_j, \dots, x_n$.

Такі задачі часто зустрічаються на практиці.

Математичну модель задачі лінійного програмування завжди записують у **загальній** (ЗЗЛП) і **канонічній формах** (КЗЛП).

У загальному вигляді задачу лінійного програмування (ЗЛП) формулюють в такий спосіб:

знайти найбільше або найменше значення лінійної функції

$$L = \sum_{j=1}^{n} c_j x_j \to extr. \qquad (4.1)$$

На деякій множині D, де $x \in D$ відповідає системі обмежень,

$$\left.\begin{aligned} a_{11}x_1 + a_{12}x_2 + \dots + a_{1n}x_n &\le b_1 \\ a_{21}x_1 + a_{22}x_2 + \dots + a_{2n}x_n &\le b_2 \\ \dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots& \\ a_{m1}x_1 + a_{m2}x_2 + \dots + a_{mn}x_n &\le b_m \\ x_j \ge 0, \quad j = \overline{1,n}& \end{aligned}\right\} \tag{4.2}$$

Відзначимо, що в системі перші декілька обмежень є нерівностями, а решта – рівняннями. Щодо напрямку знака нерівності вважат0имемо, що ліва частина менша або дорівнює правій частині. Домогтися цього можна, помноживши на (-1) обидві частини нерівностей із протилежним знаком.

Вибір типу шуканого екстремуму цільової функції також не принциповий, оскільки задача пошуку максимуму функції $L = \Sigma c_j x_j$ еквівалентна задачі пошуку мінімуму функції $-L = \Sigma(-c_j x_j)$.

Задачу лінійного програмування, записану в такій формі, називають загальною задачею лінійного програмування (ЗЗЛП).

4.2 Основні властивості загальної задачі лінійного програмування та її перша геометрична інтерпретація

Введемо визначення: точку v опуклої множини *V* називають її кутовою (крайньою) точкою, якщо вона не є внутрішньою точкою ні для якого відрізка, кінці якого належать множині *V*. Кутові точки опуклого багатогранника є його вершинами, а сам він – опуклою оболонкою своїх вершин.

У випадку, коли ЗЗЛП містить дві змінні x_1 та x_2, її можна графічно зобразити на координатній площині й одержати розв’язок графічним методом. Графічне розв’язання ЗЗЛП є ілюстративним, але основний зміст і термінологія розповсюджуються на задачі великої розмірності.

Розглянемо приклад. Нехай цільова функція представлена виразом

$$L = x_1 + 3x_2 \to max,$$

а обмеження задані системою нерівностей:

$$\left.\begin{aligned} x_1 + x_2 &\le 6 \\ x_1 - x_2 &\le 2 \\ x_1 &\le 3 \\ x_1 \ge 0,\ x_2 &\ge 0 \end{aligned}\right\}.$$

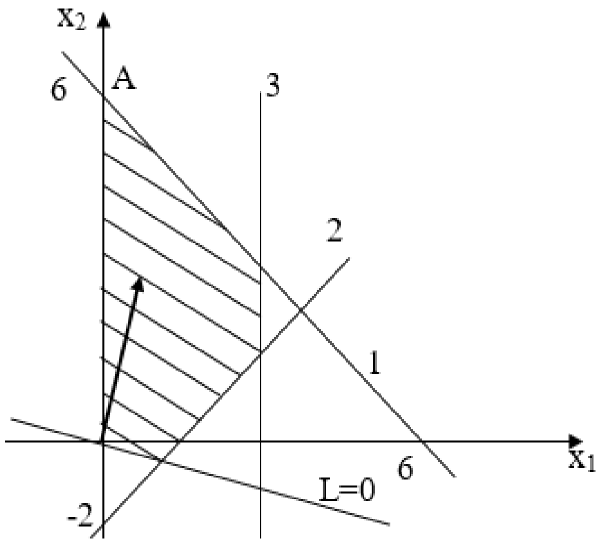


Рисунок 4.1 – Геометрична інтерпретація ЗЗЛП

Зобразимо пару значень x_1 і x_2 точкою на координатній площині $x_1 0 x_2$ з координатами (x_1, x_2), що показано на рисунку 4.1. Кожна нерівність визначає певну напівплощину. Перетинання трьох напівплощин є множиною припустимих планів D, оскільки кожна точка на його площині належить одночасно кожній із трьох напівплощин, а отже задовольняє обмеженням ЗЗЛП. Помітимо, що припустимих розв'язків – нескінченна множина.

Для визначення оптимального плану задачі, тобто такого розв'язку (x_1, x_2), що перетворює цільову функцію на максимум, скористаємось визначеннями:

– градієнтом функції *f(x)* називають вектор

$$\nabla f(x) = \left(\frac{\partial f(x)}{\partial x_1}; \dots; \frac{\partial f(x)}{\partial x_n}\right),$$

який вказує напрямок найскорішого зростання функції f(x);

– лінією рівня функції f(x) називають множину точок з області її визначення, у яких функція приймає одне й те саме фіксоване значення.

У нашому прикладі $\nabla L = (1; 3)$. Лінії рівня L перпендикулярні напрямку градієнта. Побудуємо опорну пряму $L = 0$, що проходить через початок координат, і будемо переміщати її у напрямку ∇L. Очевидно, що переміщати лінію рівня в напрямку зростання цільової функції має сенс тільки в межах області припустимих розв'язків. Точкою, у якій цільова функція дістане максимального значення, у нашому прикладі є точка А з координатами (0, 6). Отже, отриманий оптимальний план задачі

$$\bar{X}_{\text{опт}} = (0, 6),$$

за якого цільова функція приймає максимальне значення

$$L = 1 \cdot 0 + 3 \cdot 6 = 18.$$

Теоретично можливі також такі окремі випадки розв'язку ЗЗЛП:

– цільова функція L не обмежена зверху, тобто не має максимуму (рис. 4.2);

– лінія рівня збігається із стороною багатокутника області припустимих розв'язків (рис. 4.3). У цьому випадку всі точки, що лежать на стороні множини D, є оптимальними планами й кажуть, що має місце **альтернативний оптимум**.

На розглянутих рисунках припустимі плани ЗЗЛП мають вигляд опуклої багатогранної множини. Таке подання множини припустимих планів називають першою геометричною інтерпретацією ЗЗЛП.

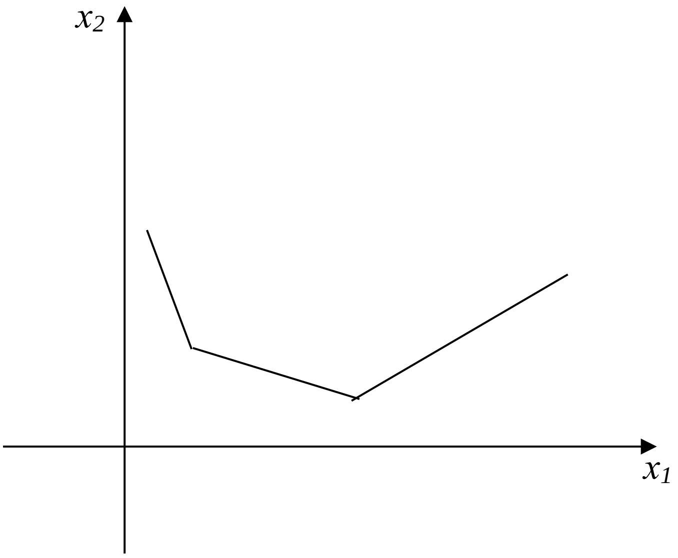


Рисунок 4.2 – Цільова функція L не обмежена зверху

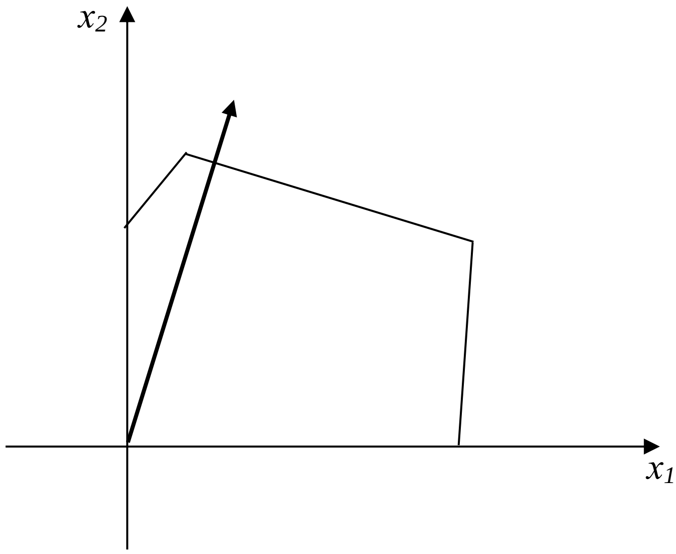


Рисунок 4.3 – Альтернативний оптимум

Основні теореми лінійного програмування відбивають фундаментальні властивості задач лінійного програмування і полягають в основі методів їх розв'язання. Ці теореми узагальнюють на випадок задач із довільною кількістю змінних ті властивості, які ми спостерігали у двовимірному випадку.

Теорема 4.1. Якщо цільова функція L приймає максимальне значення в деякій точці множини припустимих планів D, то вона приймає це значення й у деякій кутовій точці даної множини.

Теорема 4.2. Якщо цільова функція L приймає максимальне значення в кількох точках множини D, то вона приймає це саме значення в будь-якій точці, що є їх опуклою комбінацією.

4.3 Канонічна форма задачі лінійного програмування (КЗЛП)

Термін «канонічний» означає «твердо встановлений, прийнятий за зразок». Канонічною називають задачу лінійного програмування, якщо всі її обмеження є рівняннями. Це звичайна для нас форма подання системи рівнянь, у якій кількість невідомих дорівнює кількості рівнянь. Розглянута вище система обмежень (4.2) містить як рівняння, так і нерівності, причому кількість невідомих x_i дорівнює n, а кількість рівнянь – m, тобто $n \neq m$.

Переважна більшість методів розв'язання задач лінійного програмування призначена для канонічних задач. Тому початковий етап розв'язання будь-якої загальної ЗЛП завжди пов'язаний з приведенням її до еквівалентної КЗЛП.

Для переходу від ЗЗЛП до КЗЛП обмеження-нерівності перетворюють на рівняння шляхом додавання фіктивних невід'ємних змінних, кількість яких у самому простому випадку може дорівнювати кількості обмежень. Одночасно ці змінні включають до цільової функції з коефіцієнтом 0, щоб вони не впливали на її значення:

$$x_{n+i},\ i = \overline{1,k}.$$

Змінні, на які не накладено умову невід'ємності, подають як різниці двох нових невід'ємних змінних.

Додатково треба зазначити, що вибір типу шуканого екстремуму (максимуму або мінімуму) є відносним, тобто задача пошуку максимуму функції

$$L(x) = \sum_{j=1}^{n} c_j x_j$$

еквівалентна задачі пошуку мінімуму функції

$$-L(x) = \sum_{j=1}^{n} -c_j x_j.$$

У результаті проведених перетворень одержують канонічну задачу:

Знайти найбільше або найменше значення функції

$$L = c_1 x_1 + c_2 x_2 + \cdots + c_n x_n + 0 \cdot x_{n+1} + \cdots + 0 \cdot x_{n+k} \tag{4.3}$$

на певній множині D, де $x \in D$ відповідає системі обмежень

$$\left.\begin{array}{c} a_{11}x_1 + a_{12}x_2 + \ldots + a_{1n}x_n + x_{n+1} = b_1 \\ a_{21}x_1 + a_{22}x_2 + \ldots + a_{2n}x_n + x_{n+2} = b_2 \\ \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots \\ a_{m1}x_1 + a_{m2}x_2 + \ldots + a_{mn}x_n + x_{n+k} = b_m \\ x_j \geq 0, \quad j = \overline{1, n+k} \end{array}\right\}. \tag{4.4}$$

Можна показати, що отримана задача еквівалентна вихідній задачі, але перехід від загальної форми задачі лінійного програмування до канонічної призводить до зростання її розмірності. Канонічну форму задачі зручно представляти у традиційному матричному вигляді: знайти максимум

$$\bar{C}^T \bar{X} \tag{4.5}$$

за умови

$$\bar{A}^T \bar{X} = \bar{B}, \quad \bar{X} \geq 0, \tag{4.6}$$

де $\bar{A}$, $\bar{B}$ та $\bar{X}$ – вектори, відповідно:

$$\bar{A}_1 = \begin{pmatrix} a_{11} \\ a_{21} \\ \ldots \\ a_{m1} \end{pmatrix}; \quad \bar{A}_2 = \begin{pmatrix} a_{12} \\ a_{22} \\ \ldots \\ a_{m2} \end{pmatrix}; \quad \bar{A}_n = \begin{pmatrix} a_{1n} \\ a_{2n} \\ \ldots \\ a_{mn} \end{pmatrix}; \quad \bar{B} = \begin{pmatrix} b_1 \\ b_2 \\ \ldots \\ b_m \end{pmatrix},$$

$$\bar{X} = (x_1, x_2, \ldots, x_n); \quad \bar{C}^T = (c_1, c_2, \ldots, c_n). \tag{4.7}$$

Стовпці матриці $\bar{A}$ є векторами простору припустимих розв'язків. Отже кожному припустимому плану КЗЛП (n-мірному вектору $\bar{X}$) відповідає лінійна комбінація стовпців $\bar{A}$, що дорівнює стовпцю $\bar{B}$. Лінійна комбінація – це лінійний вираз, що пов'язує змінні. У нашому випадку мова йде про n-мірний вектор $\bar{X}$, компоненти якого $x_1, x_2, \dots, x_n$, а лінійною комбінацією стовпців є сума добутків:

$$a_{11}x_1 + a_{12}x_2 + \dots + a_{1n}x_n = b_1 .$$

Вектори $\overline{A}_j$, $j = \overline{1,n}$ називають **векторами вимог** задачі. Вектор $\bar{B}$ називають **вектором обмежень**. При цьому компоненти будь-якого припустимого плану $x \in D$ – x_j – є коефіцієнтами розкладання вектора обмежень $\bar{B}$ задачі за векторами вимог $\bar{A}_j$.

Приведення ЗЛП до канонічної форми залишає її нетривіальною задачею, оскільки кількість невідомих у системі обмежень зазвичай перевищує кількість рівнянь. Як відомо з алгебри, така система не має розв'язку. Це так, але треба пояснити, що така система не має одного єдиного розв'язку, а має необмежену кількість розв'язків, а отже є невизначеною. Розв'язків надто багато, але нам потрібні тільки такі, що забезпечують екстремум цільової функції.

Отже число рівнянь, які задають множину D, менша або дорівнює числу змінних задачі (m ≤ n). Якщо це не так, то або система рівнянь $\bar{A} \cdot \bar{X} = \bar{B}$ є несумісною, або вона містить надлишкові (лінійно залежні) рівняння.

Уведемо поняття базисних розв'язків канонічної задачі лінійного програмування. Оскільки система обмежень КЗЛП містить m рівнянь та n змінних, причому $m \le n$, то з m рівнянь можна однозначно обчислити тільки m невідомих, а решта невідомих, кількість яких дорівнює $n - m$, є зайвими. Але якщо їхні значення невідомі, то отримати m розв'язків неможливо, тому цим зайвим невідомим треба довільно надати будь-які числові значення. Отже у такій системі m змінних називають **базисними**, а інші (n - m) змінних – **вільними**.

Найпростіше дорівняти вільні (зайві) змінні до нуля та обчислити базисні змінні. Якщо коефіцієнти лінійної комбінації виявляться невід'ємними, то ми отримаємо **базисний припустимий план** x, в якого відмінні від нуля не більше за m елементів, а решта змінних дорівнюють нулю.

Базисним розв'язком системи m лінійних рівнянь з n змінними називають розв'язок, в якому всі n вільних змінних дорівнюють нулю.

Базисними розв'язками можуть бути різні групи з n змінних. У загальному випадку число таких груп не перевершує числа сполучень з n елементів по m C_n^m. Система з m лінійних рівнянь з n змінними ($m < n$) є **невизначеною**, оскільки кожному довільному набору вільних змінних відповідає один базисний розв'язок системи.

Опорним планом КЗЛП називають припустимий базисний розв’язок, компоненти якого перевищують нуль.

Базисний розв’язок, у якому хоча б одна з базисних змінних дорівнює нулю, називають **виродженим**.

Властивості базисних планів задачі лінійного програмування визначає наступна теорема.

Теорема 4.3. Кожний припустимий базисний план є кутовою точкою множини припустимих планів D.

Справедливим є й зворотне твердження: якщо *x* - кутова точка множини *D*, то вона є припустимим базисним планом задачі лінійного програмування.

4.4 Симплекс-метод

На підставі розглянутих теорем про властивості лінійних екстремальних задач можна побачити, що пошук їх розв’язків збігається до послідовного перебору кутових точок множини припустимих планів. Проте такий перебір для реальних багатомірних завдань на практиці нездійсненний або вкрай неефективний. Наприклад, число базисних планів у задачі з 10-ма змінними та 30-ма обмеженнями становить

$$C_{30}^{10} = \frac{30!}{10!\cdot 20!} = 1489411.$$

Класичним методом розв’язання задач лінійного програмування є симплекс-метод, що є методом послідовного поліпшення плану. Метод розроблений у 1947 році американським математиком Джорджем Данцигом.

Симплекс-метод є послідовним перебором кутових точок області припустимих розв’язків, за якого значення цільової функції зростає від одної ітерації до наступної (від однієї кутової точки до іншої). Отже, для використання симплекс-методу на його першому кроці потрібен будь-який перший базисний план, який називають **опорним планом**. Критерій оптимальності в симплекс-методі реалізують шляхом визначення спеціальних оцінок Δ_j для небазисних векторів-стовпців матриці $\bar{A}$, щодо поточного базису (симплекс-різниць). Симплекс-різниці обчислюють за формулою

$$\Delta_j = L_j - c_j, \tag{4.8}$$

де L_j – індекси векторів, що відповідають поточному базису:

$$L_j = \sum_{i=1}^{m} c_j \, a_{ij}. \tag{4.9}$$

Критерій оптимальності припустимого базисного плану формулюють так: план є оптимальним, якщо для всіх $j = \overline{1,n}$ $\Delta_j \geq 0$, і неоптимальним у протилежному випадку, тобто якщо існує таке $j = \overline{1,n}$, що $\Delta_j < 0$.

Якщо симплекс-різниці показують неоптимальність плану, переходять до наступного базису. Для цього один стовпець виводять з базису, а інший вводять. Щоб забезпечити покращення значення цільової функції (щоби за кожним кроком вона зростала, якщо шукаємо її максимум, або зменшувалась, якщо мінімум) в базис вводять вектор-стовпець, який має від'ємну оцінку. Якщо таких стовпців кілька, то для введення рекомендують вибирати стовпець, який має максимальний за модулем добуток оцінки Δ_j на відношення $\Theta_r = \min\left\{\frac{b_i}{a_{ij}}\right\}$. Одночасно на цьому кроці потрібно ухвалити рішення щодо того, який стовпець треба вивести з базису. Зробити це потрібно так, щоб знову сформований базис опинився припустимим. Припустимість базису, до якого здійснюють перехід, забезпечується певним правилом виводу стовпця з поточного базису.

У випадку, коли стовпець, що претендує на введення до базису, не містить додатних компонентів ($a_{ij} < 0$), це означає, що цільова функція задачі не обмежена на множині припустимих значень, тобто може досягати як завгодно великого значення, отже оптимальний план відсутній.

Після переходу до нового базису заново формують матрицю $\overline{A}$ та перевіряють новий план на оптимальність. Зазвичай для забезпечення раціональності й наочності обчислювального процесу алгоритм симплекс-методу оформляють як послідовність таблиць. Ці симплекс-таблиці не тільки прозоро подають хід отримання оптимального розв'язку, а ще містять вичерпну інформацію для подальшого аналізу оптимального плану задачі та для отримання його можливих інтерпретацій. Але у загальному випадку для розв'язання оптимізаційних задач зручно скористатись програмним продуктом MS Excel, який має надбудову «Пошук рішення». Для її встановлення треба війти у меню *Файл – Параметри – Управління – Надбудови* і встановити потрібний прапорець. Процедура «Пошук рішення» дозволяє обчислювати як лінійні так і нелінійні оптимізаційні моделі шляхом вибору метода рішення із списку. Нелінійний метод узагальненого понижуючого градієнта (ОПГ) призначений для розв'язання гладких нелінійних задач, симплекс-метод – для лінійних задач, еволюційний метод – для негладких нелінійних задач. Для ефективного використання процедури «Пошук рішення» корисно дотримуватись певних правил підготовки інформації.

Контрольні запитання

1. Які оптимізаційні задачі належать до лінійних?
2. Сформулюйте задачу лінійного програмування.
3. Дайте визначення для наступних понять: план, припустимий план, оптимальний план, розв’язок задачі.
4. Як відрізняються загальна задача лінійного програмування та канонічна?
5. Чи завжди загальну задачу лінійного програмування можна привести до канонічного вигляду?
6. Поясніть, як загальну ЗЛП перетворюють на канонічну. Який зміст мають додаткові змінні? Як враховується вплив додаткових змінних на цільову функцію?
7. Яку точку опуклої множини називають кутовою?
8. В чому полягає перша геометрична інтерпретація задачі лінійного програмування?
9. Поясніть, як обмеження формують область припустимих рішень ЗЛП. Які рішенні вважають припустимими?
10. Який план ЗЛП називають базисним? Назвіть ознаки базисного плану.
11. Як пов’язані базисні плани з кутовими точками області визначення задачі лінійного програмування?
12. Як із системи обмежень канонічної ЗЛП визначити кількість базисних змінних?
13. Який план задачі лінійного програмування називають виродженим?
14. Сформулюйте критерій оптимальності припустимого базисного плану, застосовуваний у симплекс-методі.
15. За яких умов роблять висновок про необмеженість цільової функції в розв’язуваній задачі?

Тема 5 ІНТЕРПРЕТАЦІЯ І АНАЛІЗ ОПТИМАЛЬНИХ ПЛАНІВ ОПТИМІЗАЦІЙНИХ МОДЕЛЕЙ

Характеристика задач аналізу оптимальних планів. Поняття та зв'язок прямої та двоїстої задач. симетричність відносини подвійності, Двоїсті оцінки. Основні теореми подвійності та їхній зміст. Післяоптимізаційний аналіз оптимізаційної моделі. Аналіз параметричної стійкості розв'язків задачі лінійного програмування.

5.1 Характеристика задач аналізу оптимальних планів

Оптимальні рішення, отримані за будь-яким методом оптимізації, являють собою вектор $\bar{X}_{опт}$, тобто низку чисел, які відповідають на запитання що треба зробити та скільки для досягнення мети задачі та яке максимальне (або мінімальне) значення набуде критерій оптимальності, виражений як цільова функція. Але отримана відповідь не є остаточною, оскільки реальні задачі, що потребують оптимальних рішень, не тільки складні та великі за розмірністю, а ще є динамічними. Їхні параметри можуть часто змінюватись, наприклад, можуть змінюватись обсяги ресурсів $\bar{B}$, як одного, так і кількох одночасно, коефіцієнти цільової функції $\bar{C}$, що впливає на величину критерію ефективності. Отже, оптимальні рішення, отримані за будь-яким методом оптимізації, та що являють собою вектор $\bar{X}_{опт}$, підлягають подальшому аналізу. Під час аналізу вирішують низку завдань:

– визначають, які ресурси є дефіцитними та істотно впливають на критерій ефективності, а які з них недефіцитні та не впливають на критерій ефективності;

– визначають рентабельні та нерентабельні види діяльності;

– оцінюють вплив двоїстих оцінок на значення цільової функції;

– визначають основні змінні, що потрапили у базис, та основні змінні, що не потрапили до базису;

– аналізують додаткові змінні, що потрапили у базис, та додаткові змінні, що не потрапили до базису;

– визначають припустимі межі змінення кількості ресурсів та оцінюють їхній вплив на оптимальний план;

– оцінюють межі можливих максимальних (або мінімальних) значень цільової функції залежно від зміни кількості ресурсів;

– оцінюють межі нових оптимальних планів залежно від зміни кількості ресурсів;

– визначають припустимі межі змінення коефіцієнтів цільової функції та оцінюють їхній вплив на її максимальне (або мінімальне) значення.

Завдання аналізу оптимальних планів оптимізаційних моделей потребують обчислення так званих «тіньових оцінок», які є оптимальним розв’язком двоїстої задачі. Двоїста задача разом с прямою задачею складають пару симетричних задач, математичні моделі яких та їхні оптимальні розв’язки пов’язані один з одним. Пару двоїстих задач застосовують для аналізу оптимальних розв’язків.

Отже, звернемось до теорії подвійності та двоїстим оцінкам в аналізі розв’язків лінійних оптимізаційних моделей.

5.2 Пряма та двоїста задачі як пара сполучених задач лінійного програмування

Задачі лінійного програмування зазвичай розглядають як задачі про розподіл ресурсів $b_1, b_2, \dots, b_m$ між різними споживачами, якими можуть бути, зокрема, будь-які технологічні процеси $\bar{A}_1,\ \bar{A}_2, \dots,\ \bar{A}_n$.

Будь-який припустимий розв’язок задачі лінійного програмування $x_1, x_2, \dots, x_n$ дає певний розподіл ресурсів, що вказує частку кожного з них, яка буде витрачена для реалізації деякого технологічного процесу.

Загальну задачу лінійного програмування формулюють в такий спосіб: знайти

$$L = c_1x_1 + \ c_2x_2 + \dots + \ c_nx_n \to max\ , \tag{5.1}$$

де *x* задовольняє обмеження

$$\left.\begin{array}{c} a_{11}x_1 + a_{12}x_2 + \ \dots + a_{1n}x_n \le b_1 \\ a_{21}x_1 + a_{22}x_2 + \ \dots + a_{2n}x_n \le b_2 \\ \dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots.. \\ a_{m1}x_1 + a_{m2}x_2 + \ \dots + a_{mn}x_n \le b_m \\ x_j \ge 0 \end{array}\right\}, \tag{5.2}$$

де a_{ij} – витрати i-го ресурсу на j-й технологічний процес. Зміст обмежень – сумарні витрати i-го ресурсу b_i мають не перевищувати кількість цього ресурсу, наявного у розпорядженні.

Для будь-якого технологічного процесу актуальним є запитання про те, щоб витрати ресурсів були мінімальними. Позначимо u_i ціну одиниці i-го ресурсу. Тоді можна записати вимогу, щоб вартість ресурсів була мінімальною:

$$L' = b_1u_1 + \ b_2u_2 + \dots + \ b_nu_n \to min,$$

а величина витрат на кожний вид технологічного процесу не перевищувала би його вартості:

$$\left.\begin{array}{c} a_{11}u_1 + a_{12}u_2 + \ \dots + a_{1m}u_m \ge c_1 \\ a_{21}u_1 + a_{22}u_2 + \ \dots + a_{2m}u_m \ge c_2 \\ \dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots.. \\ a_{n1}u_1 + a_{n2}u_2 + \ \dots + a_{nm}u_m \ge c_n \\ u_i \ge 0;\ \ i = \overline{1, m} \end{array}\right\}.$$

Отже, отримали задачу: знайти такий розв'язок $\overline{U}_{\text{опт}}$, що перетворює на мінімум цільову функцію

$$L' = \sum_{i=1}^{m} b_i u_i \to min \qquad (5.3)$$

і задовольняє обмеження

$$\left.\begin{array}{c} a_{11}u_1 + a_{12}u_2 + \ \dots + a_{1m}u_m \geq c_1 \\ a_{21}u_1 + a_{22}u_2 + \ \dots + a_{2m}u_m \geq c_2 \\ \dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots.. \\ a_{n1}u_1 + a_{n2}u_2 + \ \dots + a_{nm}u_m \geq c_n \\ u_i \geq 0; \ \ i = \overline{1,m} \end{array}\right\}. \qquad (5.4)$$

Цю задачу називають **двоїстою** щодо вихідної задачі, яку називають **прямою** задачею.

Порівняємо умови прямої та двоїстої задач у матричній формі:

Пряма	**Двоїста**
Знайти таке $\bar{X}_{\text{опт}}$, що	Знайти таке $\overline{U}_{\text{опт}}$, що
$\bar{C}^T \bar{X}_{\text{опт}} \to max$	$\bar{B}^T \overline{U}_{\text{опт}} \to min$
при обмеженнях	при обмеженнях
$\bar{A}\,\bar{X} \leq \bar{B},\ \bar{X} \geq 0.$	$\bar{A}^T\,\overline{U} \geq \bar{C},\ \overline{U} \geq 0.$

Загальна схема побудови двоїстої задачі така. Якщо задано загальну задачу лінійного програмування, де множина припустимих планів D визначається системою рівнянь і нерівностей (5.2), то двоїстою щодо неї називають загальну задачу лінійного програмування, де D' визначається системою рівнянь і нерівностей (5.4).

При переході від прямої задачі лінійного програмування до двоїстої:

– тип оптимуму змінюється на протилежний, тобто якщо пряма задача є задачею максимізації, то двоїста буде задачею мінімізації й навпаки;

– вектор коефіцієнтів цільової функції $\bar{C}$ та стовпець обмежень $\bar{B}$ міняються місцями. Тобто коефіцієнти цільової функції прямої задачі $c_1, c_2, \dots, c_n$ стають вільними членами обмежень двоїстої задачі, а вільні члени обмежень прямої задачі $b_1, b_2, \dots, b_m$ стають коефіцієнтами цільової функції двоїстої задачі;

– матрицю обмежень двоїстої задачі одержують транспонуванням матриці обмежень прямої задачі $\bar{A}$;

– множину індексів змінних, на які накладено умову невід'ємності в прямій задачі (наприклад, $x_j \geq 0$ або $u_i \geq 0$), визначають номера обмежень, що мають форму нерівностей у двоїстій задачі $a_j u \geq c_j$ (або $a_i x \leq b_i$);

– множину номерів обмежень, що мають форму нерівностей у прямій задачі (наприклад, $a_i x \leq b_i$ або $a_j u \geq c_j$), визначає множина індексів змінних, на які накладено умову невід'ємності, у двоїстій задачі ($u_i \geq 0$ або $x_j \geq 0$).

З наведених властивостей пари задач випливає важлива властивість – **симетричність відносини подвійності**, тобто задача, двоїста щодо двоїстої, збігається з прямою (вихідною) задачею. Отже, наявна пара взаємно двоїстих задач.

5.3 Основні теореми подвійності та їхній зміст

Фундаментальні властивості двоїстих задач лінійного програмування формулюють як твердження, що наведені нижче. Їх зазвичай називають **теоремами подвійності**. Наведемо ці теореми без доказів.

Теорема 5.1 (перша теорема подвійності). Якщо $\bar{X}_0$ та $\bar{U}_0$ – припустимі плани для пари двоїстих задач, тобто, якщо

$$\bar{A}\bar{X}_0 \leq \bar{B} \qquad \text{і} \qquad \bar{A}^T\bar{U}_0 \geq \bar{C}\ ,$$

$$\text{то} \qquad \bar{C}^T\bar{X}_0 \leq \bar{B}^T\bar{U}_0,$$

тобто значення цільової функції прямої задачі ніколи не перевищують значень цільової функції двоїстої задачі.

Теорема 5.1 зрозуміло, правильна й для оптимальних планів взаємно двоїстих задач: $\bar{U}_{\text{опт}} \cdot \bar{B}^T \geq \bar{X}_{\text{опт}} \cdot \bar{C}^T$, де $\bar{X}_{\text{опт}}$ і $\bar{U}_{\text{опт}}$ – деякі оптимальні плани задач.

Теорема 5.2 (**друга теорема подвійності**). Якщо для деяких припустимих планів $\bar{X}_0$ і $\bar{U}_0$ взаємно двоїстих задач виконується рівність

$$\bar{C}^T\bar{X}_0 \leq \bar{B}^T\bar{U}_0,$$

то $\bar{X}_0$ і $\bar{U}_0$ є оптимальними планами цих задач.

Тобто, якщо для деяких припустимих планів $\bar{X}_0$ і $\bar{U}_0$ прямої та двоїстої задач їхні цільові функції дорівнюють одна одній, то $\bar{X}_0$ і $\bar{U}_0$ – оптимальні плани пари сполучених задач.

Теорема 5.3. Якщо цільова функція в прямій задачі не обмежена зверху, то двоїста до неї задача не має припустимих планів.

Наступне твердження, відоме як **теорема рівноваги**, використовують при перевірці оптимальності планів задач лінійного програмування.

Теорема 5.4. Нехай то $\bar{X}_{\text{опт}}$ і $\bar{U}_{\text{опт}}$ – оптимальні плани канонічної та двоїстої до неї задач. Якщо j-та компонента плану $\bar{X}_{\text{опт}}$ строго додатна ($\bar{X}_{\text{опт}j} > 0$), то відповідне j-е обмеження двоїстої задачі виконується як рівність: $a_{1j}u_1 + \cdots + a_{mj}u_m = c_j$; якщо j-та компонента плану $\bar{X}_{\text{опт}}$ має нульове значення ($\bar{X}_{\text{опт}j} = 0$), то j-е обмеження двоїстої задачі виконується як нерівність $a_{1j}u_1 + \cdots + a_{mj}u_m > c_j$.

Практичне значення теорем подвійності полягає в тому, що вони дозволяють замінити процес розв’язання прямої задачі на розв’язання двоїстої, який у деяких випадках може виявитися простішим. Наприклад, задачу, область припустимих значень якої описується двома рівняннями, що пов’язують шість змінних (m = 2, n = 6), не можна вирішити графічним методом. Проте цей метод можна застосовувати для розв’язання двоїстої до неї задачі, що має тільки дві змінні.

Оптимальний розв’язок двоїстої задачі можна одержати з симплекс-таблиці, отриманої на фінальній ітерації процедури симплекс-методу, а також із звіту «Стійкість» процедури «Пошук рішення».

5.4 Післяоптимізаційний аналіз оптимізаційної моделі

Аналіз обмежень дефіцитних і недефіцитних ресурсів. Розв’язання прямої задачі дає оптимальний план виготовлення продукції, а розв’язання двоїстої задачі – систему оптимальних оцінок ресурсів, використованих для вироблення цієї продукції. У різних джерелах компоненти оптимального плану двоїстої задачі називають **двоїстими оцінками** або **тіньовими цінами**.

На основі теорем подвійності для пари задач лінійного програмування формулюють деякі важливі з щодо інтерпретації оптимальних планів слідства. Зокрема, з теореми 5.4 випливає, що якщо при реалізації оптимального плану прямої задачі і-е обмеження виконується як строга нерівність, то оптимальне значення відповідної двоїстої змінної дорівнює нулю:

$$a_{i1}x_{\text{опт}1} + \cdots + a_{in}x_{\text{опт}n} < b_i, \text{ то } u_{\text{опт}i} = 0.$$

Це означає, що якщо деякий ресурс b_i є в надлишковій кількості (не витрачається повністю при реалізації оптимального плану), то і-е обмеження стає несуттєвим, і тіньова оцінка такого ресурсу дорівнює нулю. Отже, тіньові оцінки характеризують **дефіцитність ресурсів**.

Статус ресурсів прямої задачі можна визначити трьома способами. Перший – підстановкою $\bar{X}_{\text{опт}}$ до системи обмежень прямої задачі. Якщо обмеження виконується як рівність, то відповідний ресурс є дефіцитним, у протилежному випадку – недефіцитним.

Другий спосіб – за допомогою додаткових змінних прямої задачі. Якщо додаткова змінна в оптимальному плані дорівнює нулю, то відповідний ресурс дефіцитний, а якщо відмінна від нуля – ресурс недефіцитний.

Третій спосіб – за допомогою двоїстих оцінок. Якщо $u_i \neq 0$, то зміна (збільшення або зменшення) обсягів i-го ресурсу призводить до відповідної зміни доходу підприємства, і тому такий ресурс є дефіцитним. Якщо $u_i = 0$, то i-й ресурс недефіцитний.

Отже, додатну двоїсту оцінку мають лише ті види ресурсів, які повністю витрачаються при оптимальному плані виробництва. Тому двоїсті оцінки визначають **дефіцитність** використовуваного ресурсу.

Оцінка рентабельності виробленої продукції. Оцінка рентабельності продукції, що випускається підприємством, ґрунтується на теоремі рівноваги (теорема 5.4) і виконується за допомогою двоїстих оцінок і обмежень двоїстої задачі, що характеризують кожен вид продукції. З цієї теореми випливає, що кількість продукції, витрати на виробництво якої перевищують дохід, в оптимальному плані дорівнює нулю.

Якщо при підстановці $\bar{U}_{\text{опт}}$ до системи обмежень двоїстої задачі вартість ресурсів, що витрачаються на одиницю продукції (ліва частина), перевищує ціну цієї продукції (права частина), то виробництво такої продукції для підприємства недоцільно. Тобто у цьому випадку цей вид продукції є нерентабельним. Якщо співвідношення виконується як рівність, продукція є рентабельною.

Інакше кажучи, коли при реалізації оптимального плану двоїстої задачі j-е обмеження виконується як строга нерівність, то оптимальне значення відповідної змінної $x_{\text{опт}j} = 0$ в оптимальному плані прямої задачі має дорівнювати нулю:

$$a_{1j}u_{\text{опт}1} + \dots + a_{mj}u_{\text{опт}m} > c_j$$

$$a_{1j}u_{\text{опт}1} + \dots + a_{mj}u_{\text{опт}m} > c_j, \text{ то } x_{\text{опт}j} = 0.$$

Цей факт виражає **принцип рентабельності виробництва**. З огляду на економічний зміст двоїстих оцінок $u_{\text{опт}1}, \dots, u_{\text{опт}m}$, вираз $a_{1j}u_{\text{опт}1} + \dots + a_{mj}u_{\text{опт}m}$ можна інтерпретувати як питомі витрати на j-й технологічний процес. Отже, якщо ці витрати перевищують прибуток від реалізації одиниці j-го продукту, то виробництво j-го продукту є нерентабельним і має не бути присутнім в оптимальному виробничому плані $x_{\text{опт}j} = 0$.

Зауважимо, що двоїсті оцінки не можна однозначно ототожнювати із цінами (хоча такі спроби іноді вживали на початковій стадії становлення дослідження операцій як науки).

Аналіз стійкості розв'язків лінійних оптимізаційних моделей. Традиційна інтерпретація двоїстої задачі лінійного програмування ґрунтується на моделі найпростішої задачі виробничого планування. У цій задачі кожний j-й елемент вектора $\bar{X}$ розглядають як план випуску продукції певного виду в натуральних одиницях, c_j – ціна одиниці продукції j-го виду, $\bar{A}_j$ – вектор, що визначає технологію витрат наявних m ресурсів на виробництво одиниці продукції j-го виду, $\bar{B}$ – вектор обмежень на обсяги цих ресурсів.

Припустимо, що для певних значень $\bar{A}_j$, $\bar{B}$ і c_j знайдений оптимальний план $\bar{X}_{\text{опт}}$ прямої задачі, який максимізує сумарний дохід $\underbrace{max}_{x \in D}\{c\bar{X}\} = c\bar{X}_{\text{опт}}$, і визначені оптимальні оцінки сировини, тобто оптимальний план двоїстої задачі $\bar{U}_{\text{опт}}$. З виразу цільової функції двоїстої задачі

$$L'_{\text{опт}} = b_1 u_{\text{опт}1} + \; b_2 u_{\text{опт}2} + \dots + \; b_n u_{\text{опт}n} \to min$$

видно, що величина двоїстої оцінки $\bar{U}_{\text{опт}}$ показує, наскільки зросте максимальне значення цільової функції прямої задачі при збільшенні кількості сировини відповідного виду на одну одиницю. Отже, змінні двоїстої задачі $u_{\text{опт}1}, u_{\text{опт}2}, \dots, u_{\text{опт}m}$ за своїм змістом є оцінками потенційної можливості одержання додаткового прибутку за рахунок збільшення обсягу відповідного ресурсу в умовах оптимального функціонування керованого об'єкта.

Виникає питання про те, як буде змінюватись оптимальний план $\bar{X}_{\text{опт}}$ при зміні компонентів вектора обмежень $\bar{B}$ і, зокрема, при яких варіаціях $\bar{B}$ оптимальний план $\bar{X}_{\text{опт}}$ залишиться оптимальним. Ця задача одержала назву проблеми **стійкості оптимального плану**. Очевидно, що дослідження стійкості $\bar{X}_{\text{опт}}$ має й практичне значення, тому що в реальному виробництві обсяги доступних ресурсів b_i можуть істотно коливатися після ухвалення планового розв'язку $\bar{X}_{\text{опт}}$.

Коли вектор обмежень $\bar{B}$ отримує приріст ΔB, виникають відповідні варіації для оптимального плану прямої задачі $\bar{X}_{\text{опт}}(B + \Delta B)$, а також для значення цільової функції $L[\bar{X}_{\text{опт}}(B + \Delta B)]$. Припустимо, що приріст ΔB такий, що він не призводить до зміни оптимального базису задачі, тобто $\bar{X}_{\text{опт}}(B + \Delta B) \geq 0$. Введемо функцію $F(b)$, яка повертає оптимальне значення цільової функції задачі для різних значень вектора обмежень$\bar{B}$

$$F(b) = \underbrace{max}_{x \in D(b)} \{c\bar{X}\} = c\bar{X}_{\text{опт}} \,. \tag{5.5}$$

Розглянемо відношення її приросту $\mathrm{F(b + \Delta b) - F(b)}$ до приросту аргумента Δb. Якщо для деякого і спрямувати $\Delta b \to 0$, одержимо

$$\lim_{\Delta b_i \to 0} \frac{F(b+\Delta b) - F(b)}{\Delta b_i} = \frac{\partial F(b)}{\partial b_i}. \tag{5.6}$$

З огляду на те, що відповідно до теореми 5.2 цільові функції пари сполучених задач при оптимальних планах дорівнюють одна одній, запишемо

$$F(b) = \sum_{j=1}^{n} c_j \, x_{\text{опт}j} = \; \sum_{i=1}^{m} b_i u_{\text{опт}i}. \tag{5.7}$$

Підставимо (5.3) до (5.2) і одержимо вираз

$$\frac{\partial F(b)}{\partial b_i} = \frac{\partial\left(\sum_{i=1}^{m} b_i u_{\text{опт}i}\right)}{\partial b_i} = u_{\text{опт}i} \,. \tag{5.8}$$

Ми показали, що тіньові оцінки ресурсів є похідними цільової функції. На цей випадок у лінійному програмуванні є теорема.

Теорема 5.5. В оптимальному плані двоїстої задачі значення змінної $u_{\text{опт}i}$ чисельно дорівнює частинній похідній цільової функції $L'_{\text{опт}}$ за відповідним аргументом b_i.

Звідси випливає інтерпретація оптимальних змінних двоїстої задачі:

Кожний елемент оптимального плану двоїстої задачі $u_{опт1}, \ldots, u_{оптm}$ можна розглядати як граничну (миттєву) оцінку внеску i-го ресурсу в сумарний дохід $L_{опт}$ при оптимальному розв'язку прямої задачі $x_{опт1}, \ldots, x_{оптn}$.

Інакше кажучи, $u_{оптi}$ дорівнює приросту доходу, що виникає при збільшенні ресурсу і на одиницю за умови оптимального використання ресурсів.

Отже, якщо знайдено оптимальний план прямої задачі, можна провести аналіз стійкості двоїстих оцінок щодо змін компонентів вектора $\bar{B}$. Це дозволяє оцінити стійкість оптимального плану двоїстої задачі щодо зміни обмежень прямої задачі та ступінь впливу зміни $\bar{B}$ на максимальне значення цільової функції, а також визначити найдоцільніший варіант можливих змін $\bar{B}$.

План двоїстої задачі не змінюється для всіх значень $b_i + \Delta b_i$, за яких стовпець вектора $\bar{B}_{опт}$ не містить від'ємних чисел, тобто коли серед компонентів вектора

$$\bar{B}_{опт} = \begin{vmatrix} b_1 + \Delta b_1 \\ b_2 + \Delta b_2 \\ \ldots\ldots\ldots\ldots \\ b_m + \Delta b_m \end{vmatrix} \tag{5.9}$$

немає від'ємних. $\bar{B}_{опт}$ – матриця, складена з компонентів векторів базису, що визначає оптимальний план задачі.

Припустимі інтервали зміни для і-го ресурсу можна визначити із звіту «Про стійкість» процедури MS Excel «Пошук рішення». Оптимальний план двоїстої задачі залишається незмінним, якщо Δb_1 належить вказаному у цьому звіті інтервалу. Але це є справедливим тільки за умови, що змінюється обсяг одного ресурса, а кількість інших ресурсів незмінна.

Стосовно прямої задачі, можна обчислити межі, у яких лежить можливий дохід підприємства, а також оптимальний план прямої задачі при зміні кількості кожного ресурсу окремо.

У випадку одночасної зміни обсягу всіх або кількох ресурсів для визначення нового оптимального плану використовують спеціальний прийом, що ґрунтується на одному з основних співвідношень обчислювальної процедури симплекс-методу:

$$\bar{X}_{опт} = \bar{D}^{-1} \cdot \bar{B} \tag{5.10}$$

де $\bar{D}$ – матриця, що складається з базисних векторів оптимального плану, компоненти яких узяті з початкового опорного плану;

$\bar{B}$ – вектор обмежень.

5.5 Аналіз параметричної стійкості розв’язків задачі лінійного програмування

Задача дослідження параметричної стійкості полягає у вивченні границь коливання цін на продукцію керованого підприємства, за яких прийнятий план випуску продукції залишається оптимальним. Отже, питання параметричної стійкості оптимального плану задачі передбачає варіацію коефіцієнтів цільової функції, тобто складників вектора $\bar{C}^T$.

Зміст проблеми стійкості оптимального плану стосовно варіацій цільової функції проілюстрований за допомогою першої геометричної інтерпретації. На рисунку 5.1 наведена множина припустимих планів D. Як видно з рисунку, цільова функція L досягає екстремального значення в точці x^*, а зміні її коефіцієнтів від c до c’ або c" на рисунку відповідає поворот лінії рівня відносно кутової точки x^*. Активним обмеженням (тобто таким, що перетворюються на рівність) у точці x^* відповідають лінії 1 і 2. Доти, поки при повороті, спричиненому зміною вектора c, лінія рівня цільової функції не виходить за межі, утворені лініями обмежень множини, x^* залишається оптимальним планом. Як показано на рисунку 5.1, цей план не змінюється при переході від c до c’ і, навпаки, при переході від c до c" лінія рівня цільової функції

$$L(x) = c"x$$

перетинає лінію 2, що спричинить зміну оптимального базисного плану, яким тепер стане точка x’. Використовуючи умови оптимальності плану задачі лінійного програмування

$$\Delta_j = L_j - c_j \geq 0, \qquad (5.11)$$

можна одержати кількісні оцінки для меж коливань коефіцієнтів цільової функції, за яких не відбувається зміна оптимального плану. Припустимо, що варіації зазнав певний елемент

$$c'_r = c_r + \Delta c_r.$$

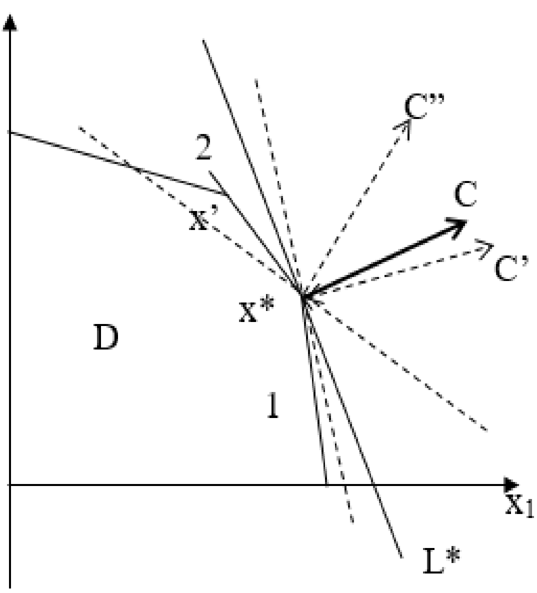


Рисунок 5.1 – Графічна інтерпретація параметричної стійкості оптимального плана

Можливі два випадки:

– стовпець r не входить до оптимального базису. Тоді для незмінності оптимального плану необхідно й достатньо виконання умови

$$\Delta_r' = L_r - c_r' \geq 0,$$

звідси можна одержати значення для припустимої варіації:

$$\Delta c_r \leq L_r - c_r; \qquad (5.12)$$

– стовпець r входить до оптимального базису. У цьому випадку для збереження оптимальності поточного плану буде потрібним виконання для всіх небазисних стовпців умов (5.11) або

$$\Delta_j = \sum_{i=1}^{m} c_{j\text{баз}} \cdot a_{ij} - c_j \geq 0, \qquad (5.10)$$

оскільки у цьому випадку зміни відбуваються також у стовпчику $C_{баз}$ симплекс-таблиці, а це, у свою чергу, стосується всіх ненульових оцінок Δ_j.

Отже, у цьому випадку припустима варіація має задовольняти умови

$$\Delta c_r \leq \sum_{i=1}^{m} c_{j\text{баз}} \cdot a_{ij} - c_j \;. \qquad (5.11)$$

Повернемося до числового приклада й визначимо межі зміни параметрів цільової функції, при яких знайдений план x^* = (83; 36; 0; 0; 21) залишається оптимальним. У цій задачі інтерес представляють варіації коефіцієнтів c_1 і c_2, які стоять при базисних змінних в оптимальному плані.

Запишемо умови (5.11) для коефіцієнта c_1:

$$\begin{cases} \Delta_3 = L_3 - c_3 = (6+\Delta c_1)*4/7 + 5*(-3/7) + 0*11/28 - 0 = \dfrac{24}{7} + \dfrac{4}{7}\Delta c_1 - \dfrac{15}{7} \\ \Delta_4 = L_4 - c_4 = (6+\Delta c_1)*(-3/7) + 5*4/7 + 0*11/7 - 0 = -\dfrac{18}{7} - \dfrac{3}{7}\Delta c_1 + \dfrac{20}{7} \end{cases} \Rightarrow$$

$$\begin{cases} \Delta_3 = \dfrac{24}{7} + \dfrac{4}{7}\Delta c_1 - \dfrac{15}{7} \geq 0 \\ \Delta_4 = -\dfrac{18}{7} - \dfrac{3}{7}\Delta c_1 + \dfrac{20}{7} \geq 0 \end{cases} \Rightarrow \begin{cases} \Delta c_1 \geq -\dfrac{9}{4} \\ \Delta c_1 \leq \dfrac{2}{3} \end{cases};$$

$$-\frac{9}{4} \leq \Delta c_1 \leq \frac{2}{3};$$

$$3{,}75 \leq c_1 \leq 6{,}67 \,.$$

Аналогічно визначимо варіацію коефіцієнта c_2:

$$\begin{cases} \Delta_3 = L_3 - c_3 = 6*4/7 + (5+\Delta c_2)*(-3/7) + 0*11/28 - 0 = \dfrac{9}{7} - \dfrac{3}{7}\Delta c_2 \\ \Delta_4 = L_4 - c_4 = 6*(-3/7) + (5+\Delta c_2)*4/7 + 0*(-11/7) - 0 = \dfrac{2}{7} + \dfrac{4}{7}\Delta c_2 \end{cases} \Rightarrow$$

$$\begin{cases} \Delta_3 = \dfrac{9}{7} - \dfrac{3}{7}\Delta c_2 \geq 0 \\ \Delta_4 = -\dfrac{2}{7} + \dfrac{4}{7}\Delta c_2 \geq 0 \end{cases} \Rightarrow \begin{cases} \Delta c_2 \leq 3 \\ \Delta c_2 \geq \dfrac{1}{2} \end{cases}.$$

$$\frac{1}{2} \leq \Delta c_2 \leq 3$$

$$5{,}5 \le c_2 \le 8.$$

Наведений приклад дослідження чутливості оптимального плану щодо зміни параметрів задачі є простим. Існують і складніші задачі, в яких, наприклад, досліджуються спільні варіації параметрів різних типів. Вони складають предмет спеціального розділу дослідження операцій, що одержав назву параметричного програмування.

Контрольні запитання

1. Поясніть суть подвійності в лінійному програмуванні.
2. Складіть просту економіко-математичну модель і запишіть до неї двоїсту. Дайте економічну інтерпретацію двоїстих оцінок.
3. Скільки змінних і обмежень має двоїста задача щодо прямої задачі?
4. Поясніть економічний зміст першої теореми подвійності.
5. Поясніть економічний зміст другої теореми подвійності.
6. У чому полягає економічний зміст третьої теореми подвійності?
7. Сформулюйте правила побудови двоїстих задач.
8. Як на підставі оптимального розв’язку прямої задачі одержати оптимальний розв’язок двоїстої задачі?
9. У чому полягає економічна інтерпретація прямої та двоїстої задач лінійного програмування?
10. Як визначити, чи є обмежений ресурс дефіцитним?
11. Як визначити, що продукція є рентабельною або нерентабельною?
12. У чому полягає економічний зміст змінних двоїстої задачі?
13. Який зміст вкладають у поняття «параметрична стійкість»?
14. Сформулюйте умови для припустимих змін цільової функції задачі, за яких її оптимальний план залишається незмінним.
15. Охарактеризуйте способи визначення статусу ресурсів прямої задачі, які вам відомі.
16. Поясніть, як визначити інтервали стійкості двоїстих оцінок щодо зміни запасів дефіцитних ресурсів?
17. Поясніть, як визначити оптимальний план виробництва продукції та зміну доходу підприємства при збільшенні або зменшенні обсягу ресурсів?
18. Як розрахувати інтервали можливої зміни ціни одиниці кожного виду продукції?

Тема 6 ТРАНСПОРТНА ЗАДАЧА

Транспортна задача в матричній постановці та її властивості. Методи побудови опорного плану. Метод потенціалів. Вироджена задача. Альтернативні рішення розподільних задач. Сіткова постановка транспортної задачі. Створення геоінформаційної моделі транспортної мережі.

6.1 Транспортна задача в матричній постановці та її властивості

У лінійному програмуванні існує клас задач, структура яких дозволяє створити спеціальні методи розв'язання, що вигідно відрізняються від методів розв'язання задач загального характеру. Таким є клас розподільних задач, до яких належить, зокрема, транспортна задача.

Вербальну модель транспортної задачі формулюють так. Нехай є $\boldsymbol{m}$ **пунктів відправлення** ПВ (постачальників) A_1, A_2,... A_m, у яких знаходиться однорідна продукція у кількостях a_1, a_2,...,a_m відповідно. Нехай є $\boldsymbol{n}$ **пунктів призначення** ПП (споживачів) B_1, B_2,..., B_n, що подали заявки відповідно на b_1, b_2,...,b_n одиниць вантажу. Сума всіх заявок дорівнює сумі всіх запасів.

Відомі вартості c_{ij} перевезення одиниці продукції з i-го пункту відправлення до j-го пункту призначення. Вважають, що вартість перевезення кількох одиниць вантажу пропорційна їх числу.

Потрібно скласти такий план перевезень (звідки, куди та скільки одиниць везти), щоб усі заявки були виконані, а загальна вартість всіх перевезень була мінімальною.

Поставимо цю задачу як задачу лінійного програмування. Позначимо x_{ij} – кількість одиниць вантажу, що відправляють з i-го ПВ А до j-го ПП В.

Сукупність чисел x_{ij} називають «**планом перевезень**», а самі величини x_{ij} – «**перевезеннями**». Ці невід'ємні змінні мають задовольняти наступним умовам:

– сумарна кількість вантажу, що направляють з кожного ПВ до усіх ПП, має дорівнювати запасу вантажу в даному пункті $\sum_{j=1}^{n} x_{ij} = a_i$;

– сумарна кількість вантажу, що доставляють до кожного ПП з усіх ПВ, має дорівнювати заявці, поданій даним пунктом $\sum_{i=1}^{m} x_{ij} = b_i$;

– сумарна вартість всіх перевезень, тобто сума величин x_{ij}, помножених на їхні вартості c_{ij}, має бути мінімальною $f(x) = \sum_{i=1}^{m} \sum_{j=1}^{n} c_{ij} x_{ij} \rightarrow min.$

Отже, задача збігається до визначення такого плану перевезень певного продукту з пунктів його виробництва до пунктів споживання

$$x = (x_{11}, \dots, x_{1n}, x_{21}, \dots, x_{2n}, \dots, x_{i1}, \dots, x_{in}, \dots, x_{m1}, \dots, x_{mn}),$$

який мінімізує цільову функцію

$$f(x) = \sum_{i=1}^{m} \sum_{j=1}^{n} c_{ij} x_{ij} \to min \tag{6.1}$$

на множині припустимих планів:

$$\sum_{j=1}^{n} x_{ij} = a_i \,; \sum_{i=1}^{m} x_{ij} = b_i; \;\; x_{ij} \geq 0; i = \overline{1,m}; j = \overline{1,n} \tag{6.2}$$

за умови дотримання балансу (сума запасів дорівнює сумі заявок):

$$\sum_{i=1}^{m} a_i = \sum_{j=1}^{n} b_j. \tag{6.3}$$

Якщо умова (6.3) виконується, то задачу називають **збалансованою** або **закритою**, а інакше задачу називають **незбалансованою** або **відкритою**. Рівність суми запасів і суми потреб є умовою існування розв’язку транспортної задачі.

Транспортна задача належить до класу задач лінійного програмування і тому має всі властивості лінійних оптимізаційних задач, але одночасно в неї є певні додаткові корисні властивості, які дозволили розробити спеціальні (простіші) методи для її розв’язання.

Число лінійно незалежних серед рівнянь в умовах-обмеженнях транспортної задачі (ТЗ) дорівнює

$$m + n - 1,$$

отже число базисних змінних також дорівнює

$$m + n - 1.$$

Загальне число змінних x_{ij} у транспортній задачі дорівнює $m \cdot n$, а число вільних змінних

$$k = m \cdot n - (m + n - 1) = (m - 1) \cdot (n - 1).$$

Відомо, що в задачі лінійного програмування оптимальний розв’язок досягається в одній з вершин ОПР, в опорній точці, де принаймні k змінних дорівнюють нулю. Виходить, у нашому випадку для оптимального плану принаймні $(m-1) \cdot (n-1)$ перевезень мають дорівнювати нулю (з відповідних ПВ у відповідні ПП нічого не перевозиться).

Будь-який план перевезень називають **припустимим**, якщо він задовольняє умови-обмеження (всі заявки задоволені, всі запаси вичерпані).

Припустимий план називають **опорним**, якщо в ньому відмінні від нуля не більше за $m + n - 1$ базисних перевезень, а інші $(m-1) \cdot (n-1)$ перевезень дорівнюють нулю.

План називають **оптимальним**, якщо він, серед всіх припустимих планів, призводить до мінімальної сумарної вартості перевезень (L=min).

В силу особливої структури ТЗ для її розв’язання не застосовують систему рівнянь. Всі операції із знаходження оптимального плану збігаються до маніпуляцій безпосередньо з таблицею, де у певному порядку записані умови транспортної задачі: перелік ПВ та ПП, заявки та запаси, а також вартості перевезень c_{ij}. У процесі заповнення цієї таблиці в її клітках проставляють самі перевезення x_{ij}. Транспортна таблиця містить m рядків і n стовпців. Рядки транспортної таблиці відповідають пунктам виробництва (в останній клітці кожного рядка зазначений обсяг запасу продукту a_i), а стовпці – пунктам споживання (остання клітка кожного стовпця містить значення заявки b_j). У правому верхньому куті кожної клітки ставлять вартість c_{ij} перевезення одиниці продукту із A_i до B_j, а центр клітки залишають вільним, щоб розміщати у ньому саме перевезення x_{ij}. Клітки, які містять нульові перевезення ($x_{ij} = 0$), називають вільними, а ненульові - зайнятими ($x_{ij} > 0$).

6.2 Методи побудови опорного плану

За аналогією з іншими задачами лінійного програмування розв'язання транспортної задачі починають з побудови припустимого базисного плану. Найпростіший спосіб його знаходження ґрунтується на так званому **методі північно-західного кута**. Суть методу полягає в послідовному розподілі всіх запасів, наявних у першому, другому і т.д. пунктах виробництва, до першого, другого і т.д. пунктів споживання. Кожний крок розподілу полягає у спробі повного вичерпання запасів у черговому пункті виробництва або у спробі повного задоволення потреб у черговому пункті споживання. На кожному кроці q величини поточних нерозподілених запасів позначають a_i^q, а поточних незадоволених потреб – b_j^q. Побудову припустимого початкового плану, відповідно до методу північно-західного кута, починають з лівого верхнього кута транспортної таблиці. Для чергової клітки, розташованої в рядку i та стовпці j, розглядають значення нерозподіленого запасу в i-му пункті виробництва й незадоволеної заявки в j-му пункті споживання, з них вибирають мінімальне та призначають як обсяг перевезення між даними пунктами: $x_{ij} = \min\{a_i^q, b_j^q\}$. В результаті цього значення нерозподіленого запасу і незадоволеної потреби у відповідних пунктах зменшуються:

$$a_i^{(q+1)} = a_i^q - x_{ij}\ ; \qquad b_j^{(q+1)} = b_j^q - x_{ij}\ .$$

Очевидно, що на кожному кроці виконується хоча б одна з рівностей: $a_i^{(q+1)} = 0$ або $b_j^{(q+1)} = 0$. Якщо справедливо $a_i^{(q+1)} = 0$, це означає, що весь запас i-го пункту виробництва вичерпаний і необхідно перейти до розподілу запасу в пункті виробництва i + 1, тобто переміститися до наступної клітки униз за стовпцем. Якщо $b_j^{(q+1)} = 0$, то повністю задоволена заявка для j-го пункту, після чого виконують перехід на клітку, розташовану праворуч за рядком. Знову обрана клітка стає поточною, і для неї повторюють всі перелічені операції.

Ґрунтуючись на умові балансу запасів і заявок (6.3), неважко довести, що за кінцеве число кроків буде отриманий припустимий план. У силу тієї самої умови число кроків алгоритму не може перевищувати $m + n - 1$, тому завжди залишатся вільними (нульовими) $mn - (m + n - 1)$ кліток. Отже, отриманий план є базисним. Не виключено, що на певному проміжному кроці поточний нерозподілений запас опиниться рівним поточній незадоволеній заявці $\left(a_i^q = b_j^q\right)$. У цьому випадку перехід до наступної клітки відбувається в діагональному напрямку (одночасно змінюються поточні пункти виробництва й призначення), а це означає «втрату» одного ненульового компонента в плані або, **виродженість** побудованого плану.

Іншій метод побудови початкового опорного плану – це метод мінімальної вартості. При застосуванні цього методу з усієї таблиці вартостей вибирають найменшу та у відповідній клітці записують найменше з чисел a_i та b_j. З розгляду виключають або рядок, що відповідає постачальнику, запаси якого вичерпані, або стовпець, що відповідає споживачу, потреби якого повністю задоволені, або рядок і стовпець, якщо вичерпані запаси постачальника і задоволені потреби споживача. У решті кліток таблиці знову вибирають найменшу вартість, і процес розподілу запасів продовжують, поки усі запаси не будуть розподілені, а потреби задоволені.

Третій спосіб знаходження початкового опорного плану - метод подвійної переваги. При застосуванні цього методу у кожному стовпці відзначають позначкою v клітку з найменшою вартістю, потім те саме роблять у кожному рядку. У результаті деякі клітки мають позначку vv. У них міститься мінімальна вартість як за стовпцем, так і за рядком. У відзначених клітках поміщають максимально можливі обсяги перевезень, виключаючи кожен раз з розгляду відповідні стовпці та рядки. Далі розподіляють перевезення по клітках з позначкою v. У частині таблиці, що залишилась, перевезення розподіляють за методом найменшої вартості.

6.3 Метод потенціалів

Одним з методів розв’язання транспортної задачі є **метод потенціалів**. Вперше він був запропонований у 1949 р. Л. В. Канторовичем і М. К. Гавуріним. Пізніше на базі загальних ідей лінійного програмування аналогічний метод був запропонований Дж. Данцигом.

Так само як транспортна задача є окремим випадком задачі ЛП, так і метод потенціалів можна трактувати як різновид симплексних процедур. Він є ітеративним процесом, на кожному кроці якого розглядають певний поточний базисний план, перевіряють його оптимальність і за необхідністю здійснюють перехід до «кращого» базисного плану.

Алгоритм методу потенціалів починається з вибору певного припустимого базисного плану. Якщо початковий опорний план має $m+n-1$ додатних перевезень, то його називають **невиродженим**. Якщо опорний план має менше за $m+n-1$ додатних перевезень, то його називають **виродженим**.

У цьому початковому опорному плані кожному пункту ставлять у відповідність певне число, яке називають його **попереднім потенціалом**. Якщо даний план **невироджений** (число ненульових базисних кліток дорівнює $m+n-1$), то за ним можна так визначити потенціали u_i і v_j, щоб для кожної базисної клітки (тобто для тієї, в якій $x_{ij} > 0$) виконувалася умова

$$v_j - u_i = c_{ij}, \text{ якщо } x_{ij} > 0. \quad (6.4)$$

Якщо план ТЗ є оптимальним, то йому відповідає система з $m+n$ чисел, які задовольняють умовам:

$$u_i + v_j = c_{ij} \quad \text{для} \quad x_{ij} > 0,$$

$$u_i + v_j \le c_{ij} \quad \text{для} \quad x_{ij} = 0, \quad i = \overline{1,m}; j = \overline{1,n}.$$

де u_i і v_j - потенціали постачальників і споживачів відповідно.

Отже, якщо хоча б для однієї вільної клітки

$$\Delta_{ij} = u_i + v_j - c_{ij} > 0, \quad (6.5)$$

то план не оптимальний і вимагає поліпшення. **Якщо в оптимальному плані транспортної задачі вільна клітка має нульову оцінку, то це вказує на наявність альтернативного оптимального рішення**. Переміщення ресурсу до цієї клітки дає альтернативний оптимум.

Оскільки система (6.4) містить $m+n-1$ рівнянь та $m+n$ невідомих, то один з потенціалів можна задати довільно (наприклад, дорівняти v_j або u_i, до нуля). Після цього інші невідомі u_i та v_j визначаються однозначно.

Економічна інтерпретація потенціалів така: u_i – це узяті з протилежним знаком середні витрати i-го постачальника на транспортування одиниці продукції до споживачів, v_j – це середні витрати j-го споживача на доставку одиниці продукції. Їх можна застосовувати для корегування оптимального рішення залежно від деяких додаткових умов. В результаті обчислень, застосовуючи систему рівнянь (6.4), отримують повну систему потенціалів для поточного опорного плана.

Далі обчислюють величини $\Delta_{ij} = v_j + u_i - c_{ij}$ для вільних кліток транспортної таблиці. Якщо всі $\Delta_{ij} \le 0$, план є оптимальним, у протилежному випадку, якщо існує хоча б одна клітка, для якої $\Delta_{ij} > 0$, його можна поліпшити. Причому, якщо для деякої клітки транспортної таблиці $\Delta_{ij} = 0$, то задача має альтернативний оптимальний план. Процес «поліпшення» плану полягає у визначенні клітки, що вводиться, й клітки, що виводиться. Кандидатом на введення може бути будь-яка клітка, у якій $\Delta_{ij} > 0$, оскільки після введення її до базису буде забезпечена рівність $v_j + u_i = c_{ij}$. Для визначеності рекомендується брати ту клітку, у якій оцінка Δ_{ij} максимальна.

Виведену клітку визначають за допомогою так званого «ланцюжка перетворення плану», що описує характер перерозподілу вантажних потоків. У відповідності із властивостями транспортної задачі для невиродженого базисного плану в поточній таблиці можна утворити замкнений ланцюжок, що складається з вертикальних і горизонтальних ланок. Однією з вершин ланцюжка є обрана вільна клітка, а інші – зайняті клітки.

Логіка алгоритму побудови ланцюжка така: «вийшовши» із клітки у горизонтальному напрямку, треба «зупинитися» у тій зайнятій клітці плану, з якої можна рухатись далі за вертикаллю. У побудованому ланцюжку, починаючи із клітки, що вводиться (яка вважається першою), позначають вершини: непарні – «+Θ», а парні «-Θ». Знаком «+» позначають ті клітки, у яких обсяги перевезень мають збільшитись. Знаком «-» – ті клітки, у яких перевезення зменшуються з метою збереження балансу. Серед множини кліток, позначених знаком «-», обирають клітку з найменшим значенням x_{ij}. Вона стає кандидатом на вивід, тому що зменшення обсягу перевезень на більшу величину може призвести до від'ємних значень x_{ij} в інших «мінусових» клітках. Потім проводять перерахування плану за ланцюжком: до обсягів перевезень у клітках, позначених знаком «+», додають обсяг Θ, а з обсягів кліток, позначених знаком «-», його віднімають. У результаті вводу однієї клітки й виводу іншої утворюється новий базисний план, для якого на наступній ітерації описані вище дії повторюють з метою перевірки нового плану на оптимальність. Тобто для знов отриманого плану повторюють дії стандартної ітерації: розраховують потенціали й оцінки для небазисних кліток транспортної таблиці. Коли отриманий план оптимальний, то всі оцінки для небазисних кліток $\Delta_{ij} \le 0$, тобто $u_i + v_j$ не перевищують відповідних цін c_{ij}. За цим планом обчислюють оптимальне (найменше) значення сумарних витрат на перевезення L^*.

Транспортну задачу, крім використання її з метою мінімізації витрат при транспортуванні вантажів, застосовують для оптимізації цілого ряду економічних процесів. Природним застосуванням транспортної задачі є **задача мінімізації часу транспортування**. Наприклад, на підприємстві є задана кількість виробничих ліній, що працюють із заданою продуктивністю. Є також задана кількість складів продукції. Відомий час транспортування від кожної лінії до кожного складу. Задача збігається до вибору таких маршрутів, при використанні яких час, затрачуваний на транспортування, буде мінімальним.

Іншою задачею є вибір оптимального варіанта з використання наявного устаткування. Задача полягає у виборі оптимального плану виробництва виробів при мінімізації часу, необхідного на їх виготовлення. Цю задачу формулюють в такий спосіб. Є задана кількість видів устаткування, на кожному з яких виготовляють певну задану кількість виробів. Відома продуктивність для кожного типу устаткування й кожного виробу. Необхідно скласти план, за якого підприємство затратить мінімальну кількість часу на виготовлення всіх виробів.

6.4 Випадок виродження

Зупинимось на ситуації виникнення виродженого плану. Якщо задача вироджена, то на якомусь етапі розв'язання може опинитись, що таблиця містить менше за $m+n-1$ заповнених кліток. Це, зокрема, може відбутися, якщо однакове мінімальне значення буде досягнуто відразу на кількох клітках, позначених знаком «-». Для подолання виродженості в транспортній задачі поточний план доповнюють необхідною кількістю нульових кліток (фіктивними перевезеннями) так, щоб вони дозволяли розрахувати повну систему потенціалів, і далі діяти відповідно до правил описаного вище алгоритму. Тобто, якщо не вистачає k заповнених кліток, то їх вважають фіктивно заповненими. Фактично такий прийом є аналогом методу збурювань для транспортної задачі як окремого випадку ЗЛП. До такого висновку легко прийти, якщо покласти, що фіктивні клітки, які додаються, містять певний малий обсяг ε.

6.5 Транспортна задача в сітковій постановці

Транспортну задачу зручно моделювати й вирішувати в термінах **сіток** і **потоків**. Основою цих моделей є орієнтовані або неорієнтовані графи. Розглянемо основні поняття та визначення.

Орієнтованим графом називають трійку (*I*, *D*, *G*), у якій *I* – непуста множина вершин; *D* – множина дуг; *G* – відображення, яке кожній дузі d ∈ D ставить у відповідність впорядковану пару вершин (i, j), де i, j∈*I*.

Неорієнтованим графом називають трійку (*I*, *D*, *G*), у якій *I* – непуста множина вершин; *D* – множина ребер; *G* – відображення, яке кожному ребру $d \in D$ ставить у відповідність невпорядковану пару вершин [i, j], де i, j∈ *I*.

Граф (*I*, *D*, *G*) називають **кінцевим**, якщо множини *I* і *D* кінцеві.

Геометрично граф можна подати у вигляді множини точок, що зображують вершини, та з'єднуючих їх ліній, що відповідають **дугам** (рис. 6.1). Очевидно, що з кожним орієнтованим графом можна однозначно пов'язати неорієнтований, замінивши дуги на **ребра**. Якщо будь-які дві вершини графа з'єднуються не більше чим однією дугою (ребром), то граф називають **простим** і його можна задати за допомогою пари (I, D). У цьому випадку кожна дуга (ребро) d повністю визначається парою вершин (i, j), що з'єднуються, і це умовно записують у вигляді: d = (i, j). Впорядкована пара вершин (i, j), що ставиться у відповідність певній дузі d, задає її орієнтацію: i називають **початком дуги**, а j - її **кінцем**. Саму дугу вважають **інцидентною** цим вершинам.

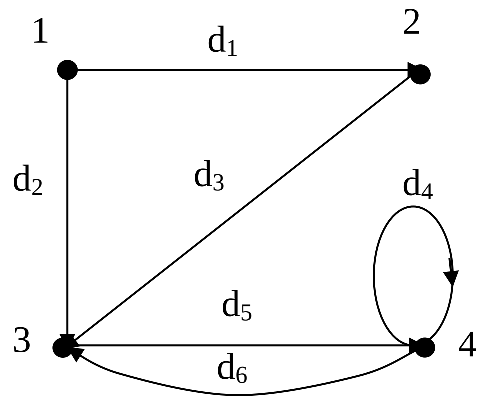


Рисунок 6.1 – Орієнтований граф

Шляхом довжини n в орієнтованому графі (I, D) називають впорядковану послідовність різних дуг (d_1, d_2,..., d_n), для яких початок кожної наступної збігається з кінцем попередньої. Кінцевий шлях, в якого початкова вершина збігається з кінцевою, називають **контуром**.

Для неорієнтованого графа аналогом поняття шлях є **ланцюг**, а контуру – **цикл**.

Якщо дві будь-які вершини неорієнтованого графа можна з'єднати ланцюгом, то його називають **зв'язним**. Орієнтований граф називають зв'язним, якщо йому відповідає зв'язний неорієнтований граф.

Зв'язний неорієнтований граф, що не містить циклів, називають **деревом**.

Розглянемо приклад. Є кінцевий граф (I, D, G), кожній вершині i якого зіставлене певне число b_i, яке називають інтенсивністю вершини. Граф (I, D, G), вершинам якого зіставлені значення інтенсивностей b_i, називають **мережею**. Якщо $b_i > 0$, то вершину i називають **джерелом**, якщо $b_i < 0$, то – **стоком**, а якщо $b_i = 0$, то – **нейтральною вершиною**. Множину джерел, стоків та нейтральних вершин позначають відповідно I^+, I^-, I^0.

Для визначеної вище мережі **потоком** називають таку сукупність величин, заданих на множині дуг, $X = \{x_d\}_{d \in D}$, що

$$\sum_{d \in D_i^+} x_d - \sum_{d \in D_i^-} x_d \ , \qquad i \in I \tag{6.6}$$

$$x_d \geq 0, \qquad d \in D, \tag{6.7}$$

де D_i^+ – множина дуг, що виходять з вершини i;

D_i^-. – множина дуг, що входять до неї.

Величину x_d називають **значенням потоку** за дугою d і за змістом інтерпретують як кількість продукту, що пропускають даною дугою.

Співвідношення (6.6) означає, що для будь-якої вершини мережі різниця вихідного та вхідного потоків дорівнює її інтенсивності.

На базі розглянутої термінології можна сформулювати багато різних задач. Розглянемо найвідоміші з них. Для кожної дуги $d \in D$ визначимо значення $c_d \geq 0$, що називають **вартістю переміщення одиниці продукту** за дугою, тоді сумарна вартість потоку X прийме вид

$$f(X) = \sum_{d \in D} c_d x_d \ . \tag{6.8}$$

Задачу мінімізації функції (6.8) за обмежень (6.6) – (6.7) зазвичай називають **лінійною сітковою задачею**. Очевидно, що вона є задачею лінійного програмування. Якщо додатково для кожної дуги мережі $d \in D$ визначити величини $r_d \geq 0$, що називають **пропускними здатностями**, то, додавши обмеження

$$0 \leq x_d \leq r_d, \quad d \in D, \tag{6.9}$$

одержимо задачу про потік у мережі з обмеженими пропускними здатностями.

Наведені формулювання задач навмисно подані у абстрактному вигляді, що дозволяє підкреслити їх універсальність. До очевидної сфери їх використання належить організація вантажоперевезень у транспортній мережі. У таких моделях вершини трактують як пункти, з’єднані мережею доріг, і характеризують потребами у певному продукті ($b_i < 0$) або його запасами ($b_i > 0$). Задачі з визначення плану, що мінімізує витрати на перевезення, які з математичної точки зору повністю ідентичні (6.6) – (6.9), також називають **транспортними задачами в сітковій постановці**.

Розглянемо задачу з визначення оптимального потоку X у деякій сітці (I, D, G), для якого

$$f(X) = \sum_{d \in D} c_d x_d \ \rightarrow min \tag{6.10}$$

при обмеженнях

$$\sum_{d \in D_i^+} x_d - \sum_{d \in D_i^-} x_d = b_i \ , \quad i \in I \tag{6.11}$$

$$0 \leq x_d \leq r_d, \quad d \in D, \tag{6.12}$$

де $r_d \geq 0$. Передбачається також, що сітка є збалансованою, тобто

$$\sum_{i \in I} b_i = 0 \ . \tag{6.13}$$

Для розв’язання транспортної задачі в сітковій постановці (6.10) – (6.12) також можна застосовувати метод потенціалів, який є узагальненням описаного вище методу потенціалів для транспортної задачі в матричній постановці.

Оскільки задача (6.10) – (6.12) є окремим випадком задачі лінійного програмування, її можна привести до канонічної форми. При цьому можна встановити, що ранг матриці задачі дорівнює $m - 1$, де m – кількість вершин у мережі.

6.6 Створення геоінформаційної моделі транспортної мережі

Сучасні ГІС забезпечують гнучкість моделювання, а також наявність аналітичних функцій, таких як приєднання даних, геоприв'язка, просторовий аналіз, геостатистичний аналіз, інтерполяція. Основою для формування моделей є джерела просторових аналогових або цифрових даних. Найбільше відомими в даний час є такі картографічні сервіси як Open Street Map, Scribble Maps, Build-A-Map, Яндекс.Карти, Google Maps, Animaps, Virtual Earth (Bing Maps). Веб-картографічний ресурс Open Street Map (OSM) містить детальну вільну і безкоштовну географічну карту. База даних OSM має дані про вулиці й дороги, а також деякі місцеві дані. Отримання доступу до даних OSM у ГІС-форматі інтегровано з програмним забезпеченням QGIS.

Процес пошуку, завантаження та використання даних OSM за допомогою QGIS доцільно починати з визначення границь потрібної місцевості, наприклад, Лозівського району Харківської області. Для цього у новому проєкті QGIS створюють векторний шар «Area». Наявність цього шару забезпечить обмеження інформації при створенні інших шарів отриманими границями шару «Area».

Створення векторних шарів моделі транспортної системи. Наступним кроком треба створити щонайменше два шари транспортної моделі – «Автомобільні дороги» та «Населені пункти». Для створення шарів доцільно скористатись модулем QGIS «QuickOSM». Він дозволяє сформувати запит шляхом використання опції «Швидкий запит». У таблиці атрибутів шару «Автомобільні дороги» потрібно створити такі поля:

– osm_id – ID номер об'єкту який присвоюється автоматично QGIS;
– Name – назва вулиці;
– Bridge – міст;
– Highway – тип дороги за класифікацією;
– longLR – протяжність дороги;
– speed – максимально припустима швидкість.

У таблиці атрибутів шару «Населені пункти» потрібно створити такі поля:

– osm_id – це ID номер об'єкту який присвоюється автоматично QGIS;
– fclass – тип населеного пункту;
– Name – назва населеного пункту;
– Population – чисельність населення.

Отже, створені векторні шари «Автомобільні дороги» та «Населені пункти» є складниками геоінформаційної моделі, кожен з яких забезпечений атрибутивними даними, що характеризують властивості елементів шару, потрібні для аналізу транспортної системи.

При проведенні аналізу необхідно визначити віддаленість від початкової точки до транспортних послуг (зупинок, павільйонів та ін.) і аналізувати не за прямим вектором напряму, а з урахуванням конкретних маршрутів, тобто за існуючою дорожньою мережею, якою рухається реальний пасажир.

Транспортні мережі в ГІС будують за такою схемою:

– пошук і підготовка даних для побудови мережі серед відкритих інтернет-джерел;

– організація та підготовка даних (формування бази географічних даних, перевірка топології);

– створення набору мережевих даних в базі геоданих;

– побудова дорожньої мережі.

Побудову мережі громадського транспорту починають зі створення графа дорожньої мережі. Граф доріг – це цифрова векторна карта, що містить топологічно пов’язані один з одним дуги і вузли. Розташування і властивості цих дуг і вузлів із заданою точністю і повнотою передають маршрути та організацію руху наземного транспорту. На етапі побудови графу у семантичні характеристики дуг і вузлів записують інформацію про зв’язки мережі доріг та атрибути для розв’язання пошукових завдань.

Засоби редагування графа доріг у ГІС призначені для уточнення графа в місцях багаторівневих розв’язок і для формування заборон поворотів. Користувач має можливість вручну видалити або додати вузли мережі, замінити дугу з двостороннім рухом на дугу з одностороннім рухом, провести розпаралелювання доріг, створити дуги й розвороти, сформувати на перехрестях заборони поворотів.

Пошук мінімального шляху між точками транспортної системи (населеними пунктами) здійснюється з урахуванням будь-яких характеристик, що записані в дуги мережі (тип доріг, швидкість руху, кількість проїзних смуг). Найкоротший маршрут можна знайти або за мінімальною довжиною шляху, або за мінімальним часом проходження маршруту. При знаходженні мінімального шляху є можливість виключення деяких дуг, наприклад, аварійних ділянок, з пошуку. Результати пошуку відображаються на карті у вигляді об’єкта карти – маршруту. Для аналізу транспортної мережі QGIS містить декілька модулів мережевого аналізу.

Аналіз графа дорожньої мережі. Створена модель дає змогу виконати мережевий аналіз, але до його початку потрібно попередньо провести певні процедури для підготовки даних. Одна з цих процедур – вибір потрібних для аналізу об’єктів, що реалізують шляхом запиту: **highway! = footway AND highway! =path**.

Далі для аналізу потрібно перетворити векторний шар «Автомобільні дороги» на граф. Усі подальші задачі аналізу транспортної системи розв'язують саме з цим графом. Вузли ліній векторного шару «Автомобільні дороги» утворюють множину вершин графа, дугами графа є відрізки ліній цього шару. Вузли, що мають однакові координати, вважають однією і тією самою вершиною графа. Отже, дві лінії, що мають спільний вузол, є пов'язаними одна з одною. При побудові графа можна «прив'язати» до векторного шару будь-яку кількість додаткових точок. Для кожної додаткової точки буде знайдено або найближчу вершину графа, або найближчу дугу. В останньому випадку дуга буде розбитою на дві частини, між якими буде додано нову спільну вершину. Можна скористатись атрибутами векторного шару «Автомобільні дороги» як властивостями дуг графа, наприклад, довжиною дуги.

Для характеристики транспортної мережі обчислюють низку чинників, одним з яких є діаметр графа. Топологічний діаметр графа - це параметр, що визначає форму транспортної мережі, зокрема степінь витягнутості та округлості мережі щодо усіх вершин графа. Для його розрахунку спочатку обчислюють матрицю «Джерело-Призначення». Одночасно ця матриця дозволяє визначити найкоротшу відстань між усіма парами населених пунктів, що також є вершинами графа.

Для задач мережевого аналізу QGIS має кілька модулів, один з найпростіших – модуль QNEAT3.

Контрольні запитання

1. Наведіть формулювання транспортної задачі.
2. Які специфічні властивості дозволяють виділити розподільні задачі в окремий клас з множини задач лінійного програмування?
3. Опишіть методи побудови припустимого плану транспортної задачі.
4. За якої умови транспортна задача має розв'язок?
5. Скільки ненульових елементів повинен містити невироджений базисний план транспортної задачі?
6. Наведіть визначення «плану перевезень» та охарактеризуйте умови, яким він має задовольняти.
7. Запишіть математичну форму транспортної задачі.
8. Запишіть математичну форму транспортної задачі як загальної задачі лінійного програмування та відзначте її особливості.
9. Сформулюйте критерій оптимальності для припустимого плану транспортної задачі.
10. Поясніть, на чому ґрунтується метод потенціалів.

11. З яких міркувань випливає критерій оптимальності припустимого плану транспортної задачі?

12. Перелічіть основні етапи методу потенціалів.

13. Які умови повинні бути дотримані при побудові ланцюжка перетворення плану в методі потенціалів?

14. Що треба робити при виникненні ситуації виродженості поточного плану в транспортній задачі?

15. Наведіть формулювання лінійної сіткової задачі.

16. Поясніть поняття орієнтованого та неорієнтованого графа, наведіть визначення.

17. Поясніть, що таке вершини, дуги та ребра графа.

18. Поясніть поняття початку і кінця дуги, шляху, ланцюга, контуру і циклу графа.

19. Покажіть, що транспортна задача в матричній постановці є окремим випадком транспортної задачі в сітковій постановці.

20. У чому полягає задача про найкоротший шлях?

ЗМІСТОВИЙ МОДУЛЬ 3 СТАТИСТИЧНЕ МОДЕЛЮВАННЯ СКЛАДНИХ СИСТЕМ

Тема 7 МЕТОДИ СТАТИСТИЧНОГО МОДЕЛЮВАННЯ СКЛАДНИХ СИСТЕМ

Процедура збирання та підготовки даних для статистичного моделювання. Статистичний ряд та його числові характеристики. Визначення виду статистичної моделі. Статистичний закон розподілу. Критерії перевірки статистичних гіпотез.

7.1 Процедура збирання та підготовки даних для статистичного моделювання

Головним джерелом інформації про закономірності у випадкових явищах є статистичні спостереження, які отримують шляхом проведення дослідів. Найчастіше досліди полягають у вимірюванні певних фізичних величин, що характеризують досліджуване явище. Зокрема це можуть бути сигнали на вході та виході системи або її окремих елементів. Методи збирання, систематизації, аналізу, інтерпретації та відображення результатів спостережень за випадковими явищами з метою виявлення їх закономірностей розробляє математична статистика. Отже, збір статистичних даних проводять за спеціальними правилами статистичного спостереження.

Слово **статистика** походить від італійського слова stato, яке означає державу. Відповідно термін «статистика» спочатку ототожнювався зі збором даних, корисних для держави. Звернемось до низки понять і термінів.

Сукупність предметів або явищ, об'єднаних певною загальною ознакою або властивістю, називають **об'єктом спостереження**. Кожен об'єкт статистичного спостереження складається з окремих елементів – **одиниць спостереження**. Для вивчення характерних властивостей об'єкта випадково відбирають з усієї сукупності об'єктів їх обмежене число.

Всю сукупність об'єктів, з яких проводять вибірку, називають **генеральною сукупністю**.

Вибірковою сукупністю, або просто **вибіркою**, називають сукупність випадково відібраних об'єктів. **Об'єм вибірки** - це число відібраних об'єктів генеральної сукупності, або інакше - число спостережень.

Результати статистичного спостереження є числовою інформацією, тобто **даними**. Отже, статистичні дані – це відомості про те, яких значень набули досліджувані фізичні величини, потрібні для аналізу.

Сукупність спостережених значень досліджуваної величини являє собою первинний статистичний матеріал. Її називають **простою статистичною сукупністю**. Її елементи зазвичай ніяк не впорядковані. За великої кількості спостережень проста статистична сукупність незручна для аналізу та мало наглядна. Тому експериментальні дані попередньо піддають опрацьовуванню.

Просту статистичну сукупність застосовують для характеристики однієї досліджуваної величини, зокрема для визначення її середнього значення та дисперсії, а також для побудови її статистичного закону розподілу. Для визначення залежності між певними величинами досліджують сумісно дві або більше величин.

Перш за все усі наявні значення досліджуваної величини розташовують у порядку зростання або зменшення та подають як таблицю, що називають **статистичним** (або варіаційним) **рядом**. У верхньому рядку статистичного ряду розташовують впорядковані значення досліджуваної величини x_i, а в нижньому кількість появ відповідного значення m_i у сукупності даних. Статистичний ряд характеризує змінення (варіювання) виміряної величини X:

x_1	x_2	...	x_i	...	x_n
m_1	m_2	...	m_i	...	m_n
p_1^*	p_2^*	...	p_i^*	...	p_n^*

Але частіше до статистичного ряду додають рядок, що містить статистичні частоти появи відповідних значень p_i^*, що обчислюють за співвідношенням:

$$p_i^* = \frac{m_i}{n}, \tag{7.1}$$

де n – обсяг вибірки, $n = \sum m_i$.

Статистична частота p_i^* є статистичною імовірністю появи відповідного значення досліджуваної величини, а отже сума усіх статистичних частот має дорівнювати одиниці:

$$\sum_{i=1}^{n} p_i^* = 1. \tag{7.2}$$

Варіація досліджуваної величини може бути дискретною або безперервною. **Дискретною** називають варіацію, за якою окремі значення величини відрізняються одне від одного на певну величину (зазвичай ціле число). **Безперервною** називають варіацію, за якою значення величини можуть відрізнятися одне від одного на скільки завгодно малу величину.

Значення величини за безперервної варіації зазвичай поділяють на інтервали (групують за інтервалами). Кількість інтервалів має не бути надто великою, або малою, тому її визначають за певним правилом (правилом Старджеса). У цьому випадку будують **групований статистичний ряд**. У групованому статистичному ряді частоти p_i^* належать не до окремого значення величини, а до інтервалу значень, тобто p_i^* - статистична імовірність того, що відповідне значення потрапить у відповідний інтервал. Під час опрацьовування ряду зазвичай як значення інтервалу приймають його середину:

$x_1 \dots x_2$	$x_2 \dots x_3$	…	$x_{i-1} \dots x_i$	…	$x_{n-1} \dots x_n$
m_1	m_2	…	m_i	…	m_n
p_1^*	p_2^*	…	p_i^*	…	p_n^*

Найважливішою характеристикою статистичного ряду є середнє значення досліджуваної величини, яке визначають за формулою:

$$\bar{x} = \frac{\sum_{i=1}^{n} x_i}{n}, \tag{7.3}$$

Нагадаємо, що середнє значення досліджуваної величини є оцінкою її математичного сподівання, тобто найбільш імовірним її значенням, навколо якого розкидані усі результати вимірів. Середнє значення досліджуваної величини вимірюється у тих самих одиницях виміру, що й сама досліджувана величина.

Для характеристики варіації досліджуваної величини в математичній статистиці застосовують певні числа, до яких належать такі.

Варіаційний розмах R, або широта розподілення - різниця між найбільшим та найменшим значеннями варіаційного ряду:

$$R = x_{max} - x_{min}. \tag{7.4}$$

Варіаційний розмах застосовують для наближеного оцінювання варіації. **Середнє лінійне відхилення *d***, що є середнім арифметичним абсолютних значень відхилень досліджуваної величини від середнього значення:

$$d = \frac{\sum_{i=1}^{n}|x_i - \bar{x}|}{n}. \tag{7.5}$$

Іншою важливою характеристикою статистичного ряду є **дисперсія**, яку обчислюють за формулою:

$$D = \frac{\sum_{i=1}^{n}(x_i - \bar{x})^2}{n}. \tag{7.6}$$

Як видно з формули (7.6), дисперсія статистичного ряду вимірюється у квадраті одиниці виміру досліджуваної величини. Для зручності оцінювання розкиду вимірів щодо середнього значення досліджуваної величини застосовують **середнє квадратичне відхилення**, яке визначають як квадратний корінь з дисперсії:

$$\sigma = \sqrt{D}. \tag{7.7}$$

Отже, середнє квадратичне відхилення вимірюється у тих самих одиницях виміру, що й сама досліджувана величина Після первинного опрацювання експериментальних даних, яке полягає у формуванні варіаційного ряду, приступають до визначення виду статистичної моделі.

7.2 Визначення виду статистичної моделі

Тут постає питання, про яку статистичну модель системи йде мова. Чи не є такою моделлю варіаційний ряд? Ми впорядкували дослідні дані, визначили частоти окремих значень, обчислили такі числові характеристики ряду як середнє арифметичне, варіаційний розмах, дисперсію та ін. Так, варіаційний ряд з його характеристиками дійсно є статистичною моделлю однієї випадкової величини. Але його подальше опрацювання дозволяє отримати ще іншу цінну для аналізу та висновків інформацію.

Зокрема статистичним рядом можна скористатися для визначення закону розподілу випадкової величини. Графічно статистичний ряд часто виражають у вигляді гістограми на рисунку 7.1 (гістограмою називають графічне зображення, на якому досліджувана величина зображується як площина). На осі абсцис X для цього відкладають значення x_i, а на осі ординат Y – відношення частоти до величини інтервалу – $f^*(x) = \frac{p_i}{\Delta x_i}$. Тоді площа кожного прямокутника дорівнює статистичній імовірності потрапляння досліджуваної величини на інтервал Δx_i:

$$f^*(x)\Delta x_i = p_i,$$

але оскільки сума імовірностей статистичного ряду дорівнює одиниці (7.2), площа гістограми також дорівнює одиниці.

Гістограма являє собою статистичний диференційний закон розподілу, тобто статистичну щільність імовірності $f^*(x)$. Як відомо, закон розподілу є вичерпною імовірнісною характеристикою випадкової величини, адже дозволяє вирішувати будь-які питання щодо досліджуваної величини.

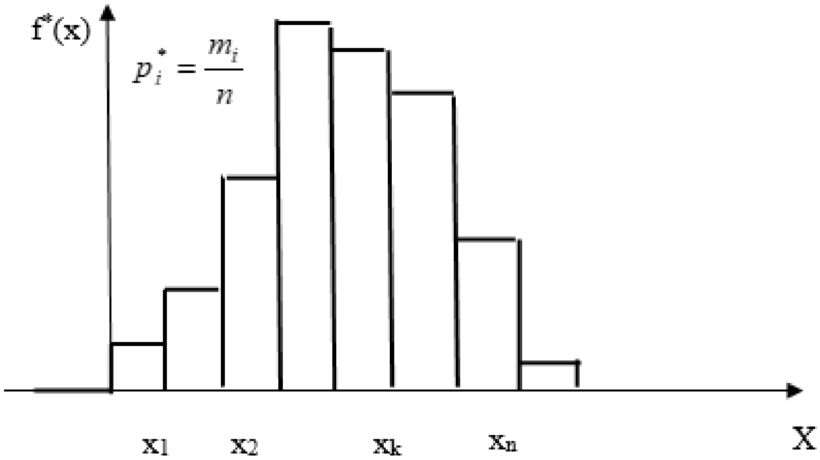


Рисунок 7.1 – Гістограма

Для зручності подальших обчислень на підставі гістограми отримують аналітичну залежність шляхом вирівнювання гістограми.

Задача вирівнювання статистичного ряду полягає у виборі теоретичної кривої, що виражає тільки істотні риси статистичного матеріалу. Для оцінювання ступеня відповідності статистичного ряду і отриманої аналітичної залежності застосовують низку різних критеріїв згоди.

Треба зауважити, що будь-який висновок, який отриманий за статистичними даними, вважають **гіпотезою**, тобто науковим припущенням. Кожну статистичну гіпотезу Н перевіряють, чи не суперечить вона первинним статистичним даним. Основну гіпотезу позначають Н0:, альтернативні гіпотези позначають відповідно Н1:, Н2:, Н3: і так далі. Для перевірки основної гіпотези застосовують один з низки розроблених критеріїв. Як правило, такий критерій є певною випадковою величиною, яка залежить від розбіжності, що неминуче виникає між теоретичною кривою, що згладжує статистичну, наприклад, гістограму, та статистичними точками гістограми. Отже, потрібно з’ясувати, чи зумовлені ці розбіжності випадковими обставинами, або вони пов’язані з поганим вибором теоретичної кривої, що згладжує. Для перевірки правдоподібності гіпотези обирають міру розбіжності, яка є випадковою величиною із власним законом розподілу та найчастіше залежить від числа ступенів свободи, наприклад, критерій Пірсона хі-квадрат χ^2, критерій Стьюдента t, критерій Фішера F. Обчисливши критерій, визначають імовірність обчисленого значення і, якщо ця імовірність велика, то вважають, що гіпотеза H0: не суперечить дослідним даним, тобто розбіжність зумовлена випадковими причинами. Якщо ж ця імовірність мала, то гіпотезу H0: відкидають як неправдоподібну.

Контрольні запитання

1. Поясніть сутність термінів «об'єкт спостереження», «одиниці спостереження», «генеральна сукупність», «вибірка», «об'єм вибірки».
2. Поясніть призначення та властивості статистичного ряду.
3. У яких випадках та з якою метою створюють групований статистичний ряд?
4. Охарактеризуйте сутність статистичної частоти, як її обчислюють та для чого застосовують.
5. Які числові характеристики статистичного ряду вам відомі та як їх застосовують для аналізу статистичних даних
6. Поясніть, як характеризують статистичні дані арифметичне середнє та дисперсія статистичного ряду.
7. Поясніть, з яких міркувань обирають вид статистичної моделі та наведіть приклади.
8. Чи є статистичний ряд моделлю статистичних даних та які висновки дозволяє зробити їх аналіз за допомогою статистичного ряду?
9. Поясніть, як на підставі статистичного ряду визначити закон розподілу випадкової величини.
10. Як на підставі статистичного ряду побудувати гістограму? Як гістограма пов'язана із щільністю розподілу та де на гістограмі імовірність певного значення випадкової величини?
11. З якою метою гістограму апроксимують гладкою аналітичною залежністю?
12. Поясніть поняття «статистична гіпотеза». Як позначають гіпотези?
13. Які властивості випадкових величин, що застосовують як критерії для перевірки статистичних гіпотез?

Тема 8 МОДЕЛЮВАННЯ ТА АНАЛІЗ ЗВ'ЯЗКІВ СКЛАДНОЇ СИСТЕМИ

Етапи статистичного моделювання. Основні поняття кореляційно-регресійного аналізу. Функціональна та стохастична залежності. Регресія. Згладжування експериментальних залежностей. Перевірка гіпотез о параметрах моделі, оцінка її статистичної значущості та адекватності. Значущість рівняння лінійної регресії. Середня помилка апроксимації. Стандартні помилки параметрів.

8.1 Етапи статистичного моделювання

Прийнято виділяти шість основних етапів статистичного моделювання: постановочний, апріорний, етап параметризації, інформаційний, етапи ідентифікації та верифікації моделі. Зупинимося докладніше на кожному з цих етапів і розглянемо проблеми, пов'язані з їхньою реалізацією.

1-й етап (*постановочний*). На даному етапі формують мету дослідження і визначають потрібний набір змінних моделі. Під час вибору змінних бажано кожну змінну теоретично обґрунтувати. При цьому рекомендують, щоб число їх було не дуже великим і, як мінімум, у кілька разів меншим за число спостережень.

2-й етап (*апріорний*). На даному етапі проводять аналіз сутності досліджуваного об'єкта, формування та опрацювання апріорної статистичної інформації.

3-й етап (*параметризація*). На цьому етапі здійснюють безпосередньо моделювання, тобто вибір загального виду моделі, виявлення зв'язків, які вона має врахувати. Основним завданням цього етапу є вибір виду функції φ(x), тобто специфікації моделі. Як специфікацію моделі розуміють математичну форму вираження виявлених зв'язків, встановлення складу екзогенних (лат. exo- зовнішній) і ендогенних (грец. endon – внутрішні) змінних, формулювання вихідних передумов і обмежень моделі. Від того, наскільки вдало вирішена проблема специфікації моделі, істотно залежить успіх всього моделювання.

4-й етап (*інформаційний*). На даному етапі здійснюють збір необхідної статистичної інформації - спостережуваних значень змінних

$$(x_{i1}, x_{i2}, \ldots, x_{ip};\ y_{i1}, y_{i2}, \ldots, y_{iq}), i = \overline{1, n}.$$

Як результати спостережень приймають дані, отримані як за участю дослідника, так і без його участі, тобто в умовах активного або пасивного експерименту.

5-й етап (*ідентифікація моделі*) полягає у здійсненні статистичного аналізу моделі та оцінюванні її параметрів.

6-й етап (*верифікація моделі*) передбачає перевірку істинності, тобто адекватності моделі. З'ясовують, наскільки вдало вирішені проблеми специфікації моделі, яка точність розрахунків, зроблених за даною моделлю, і, в остаточному підсумку, наскільки відповідає побудована модель реальному об'єкту або процесу, який моделюють. Наприклад, якщо модель побудована для прогнозу, то для її верифікації порівнюють отримані на основі побудованої моделі значення з реальними значеннями змінних у відповідні моменти часу.

Розглянутий поділ статистичного моделювання на окремі етапи має дещо умовний характер, тому що ці етапи можуть перетинатися, взаємно доповнювати один одного та ін.

8.2 Основні поняття кореляційно-регресійного аналізу

В теорії і практиці велику роль має дослідження залежностей, що характеризують взаємозв'язки між явищами і процесами, що об'єктивно відбуваються у складних системах. Взаємний зв'язок елементів є важливою ознакою системи, отже моделювання цих зв'язків завжди є актуальним. Для дослідження інтенсивності, виду і форми причинно-наслідкових співвідношень між явищами застосовують методи кореляційно-регресійного і дисперсійного аналізу.

Ознайомимось ще з кількома важливими поняттями та визначеннями. У загальному випадку розрізняють два типи залежностей між явищами: **функціональну** і **стохастичну**.

Залежність виду

$$Y = f(X), \tag{8.1}$$

у якої кожному значенню незалежної змінної X відповідає одне-єдине значення залежної змінної Y, називають **функціональною**. У функціональній залежності (8.1) для кожного значення незалежної змінної X існує цілком визначене значення залежної змінної Y.

Статистичною (стохастичною або імовірнісною) залежністю між незалежною та залежною змінними X та Y називають таку залежність, в якої одному значенню змінної X x_i може відповідати низка значень змінної Y: y_{i1}, y_{i2}, …, y_{ik}, що зазвичай спричинене впливом різних факторів на змінну Y, окрім змінної X, або помилками вимірів. У випадку статистичної залежності для кожного значення x_i визначають умовне середнє $\overline{y}_i$ (рис. 8.1).

Регресією Y на X (лат. regressio – «зворотний рух, повернення») називають умовне математичне сподівання випадкової величини Y $\overline{y}_i$ за умови, що X прийняла значення x_i. Лінію, яка з'єднує точки $\bar{y}_i$, називають **лінією регресії** (рис. 8.1).

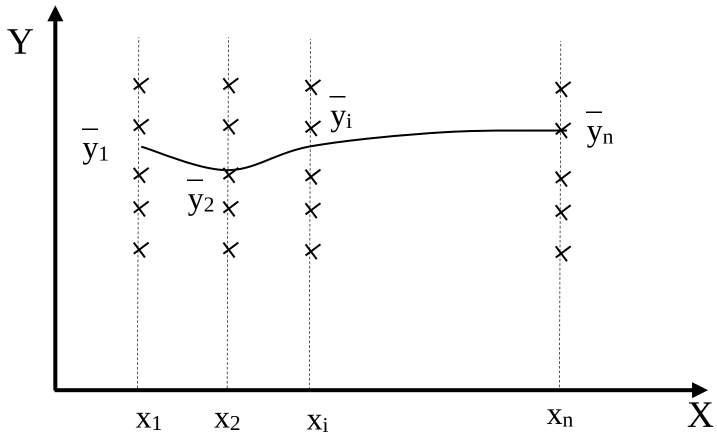


Рисунок 8.1 – Статистична залежність $y = \varphi(x)$

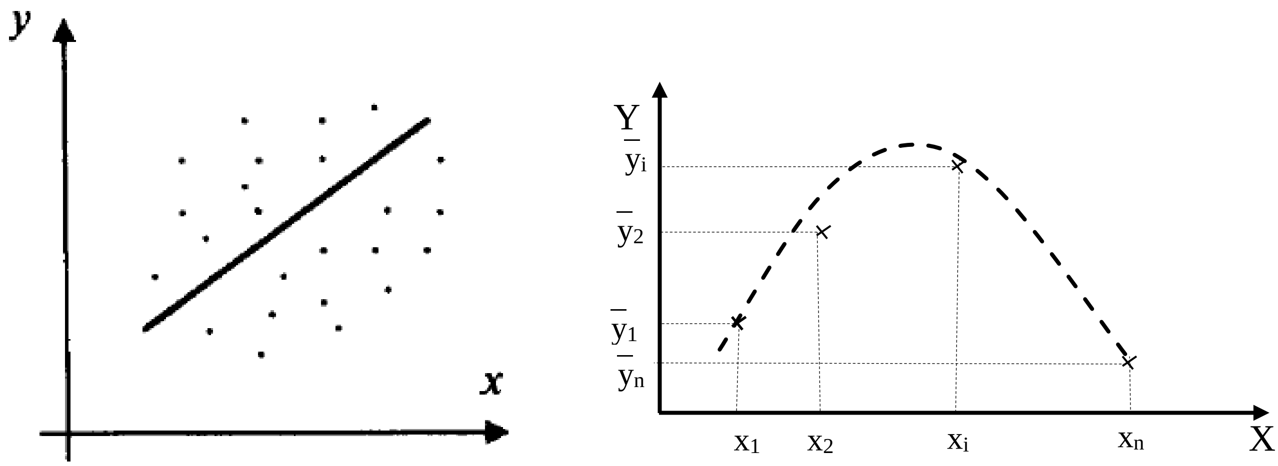
Отже, лінія регресії графічно відображує залежність від незалежної змінної X умовного математичного сподівання випадкової величини Y. Графічні зображення наглядно відбивають характер шуканої залежності, але їх незручно застосовувати для аналізу та прогнозування. Тому лінію регресії апроксимують аналітичним виразом, який називають **рівнянням регресії**:

$$\bar{y}_x = \varphi(x). \quad (8.2)$$

Розрізняють **парну** (або просту) регресію, якщо досліджують вплив на залежну змінну Y однієї незалежної змінної X, і **множинну** регресію, якщо досліджують вплив на залежну змінну Y множини незалежних змінних X_i.

Статистичну залежність ще називають **кореляційною**, якщо при зміні значень однієї величини X змінюється середнє значення іншої величини Y. Нагадаємо, що термін «кореляція» означає «зв'язок». Статистичну залежність називають кореляційною, оскільки вона дозволяє не тільки встановити наявність зв'язку двох та більше факторів, але й визначити ступінь цього зв'язку. Треба розуміти, що функціональна залежність дозволяє за умови, що значення X відоме, однозначно обчислити величину Y. Наявність же кореляційної залежності дозволяє лише встановити тенденцію зміни Y при зміні X.
Завданням кореляційно-регресійного аналізу є визначення форми залежності між змінними X та Y, а також оцінювання тісноти зв'язку між ними.

Обирають вид залежності (8.2) зазвичай або з теоретичних міркувань, якщо відомий вид теоретичної залежності $\bar{y}_x = \varphi(x)$, або графічно, для чого залежність зображують точками на координатній площині. Таке зображення статистичної залежності називають **полем кореляції** (рис. 8.2).



а б

Рисунок 8.2 – Поле кореляції

8.3 Згладжування експериментальних залежностей

При проведенні дослідів, метою яких є визначення залежності, наприклад, однієї фізичної величини X від іншої Y у результати вимірювань неодмінно вноситься похибка. Отже виміри містять випадкову складову. У зв'язку з цим виникає завдання, як за експериментальними даними щонайкраще відтворити шукану залежність (8.2).

Принципово через n експериментальних точок можна провести залежність, що точно проходить через усі точки, якщо в цієї залежності є n свобідних параметрів:

$$y = a_0 + a_1 x + a_2 x^2 + \dots + a_j x^j + \dots + a_{n-1} x^{n-1}, \quad j = \overline{0, n-1}.$$

Зауважимо, що будь-яка залежність містить, як мінімум дві змінні x та y, а вид конкретної залежності визначається числовими коефіцієнтами при цих змінних. Ці числові коефіцієнти називають **параметрами залежності**. Метод визначення кривої за n точками дозволяє вирішувати тільки детерміністські задачі (тільки функціональні залежності). Залежність, що визначена розглянутим методом відбиває, крім закономірних змін досліджуваної величини y, також і випадкові похибки, а отже вносить спотворення у вигляд шуканої залежності Y від X. Тому виникає задача згладжування експериментальної залежності, у процесі якого відкидаються погрішності вимірів та об'єктивніше відтворюється шукана закономірність.

Для вирішення таких задач зазвичай застосовують метод найменших квадратів, який дає можливість, якщо відомий клас залежності, так вибрати числові значення параметрів цієї залежності, щоби крива щонайкраще відбивала дослідні дані. Вибір типу аналітичної залежності зазвичай здійснюють за її зовнішнім виглядом або з теоретичних міркувань.

Метод найменших квадратів полягає у задоволенні вимоги найкращого узгодження кривої $y = \varphi(x)$ із шуканою залежністю, яка збігається до того, щоб сума квадратів відхилень експериментальних точок від згладжуючої кривої була мінімальною:

$$\sum_i \left(y_{i\text{р}} - y_i\right)^2 = min, \tag{8.3}$$

де y_i – експериментальні значення (отримані з дослідів);

$y_{i\text{р}}$ – розрахункові значення (розраховані за аналітичним виразом).

Доведено, що якщо усі вимірі виконували з однаковою точністю, та якщо погрішності вимірів підкоряються нормальному закону, то знайдена залежність буде найбільш імовірною з усіх можливих.

Наприклад, між точками поля кореляції на рисунку 8.2,а можна спробувати провести пряму лінію, а розташування отриманих статистичним шляхом точок на рисунку 8.2,б нагадує параболу, тоді для згладжування експериментальної залежності $\bar{\bar{y}}_x = \varphi(x)$ можна скористатись поліномом другого порядку:

$$y = a_0 + a_1 x + a_2 x^2, \tag{8.4}$$

але на практиці найчастіше застосовують лінійне рівняння регресії:

$$y = a_0 + a_1 x, \tag{8.5}$$

в якого коефіцієнт при змінній х a_1 називають **коефіцієнтом регресії**.

Після визначення вигляду апроксимуючої залежності (лінійна, степенева або інша функція) виникає завдання із визначення параметрів обраної залежності. Цими невідомими параметрами є коефіцієнти при змінних х та у, а значення самих змінних х та у відомі з дослідів, адже вони представлені як статистичні дані та містяться у статистичному ряді.

Враховуючи, що $y_{i\text{р}} = \varphi(x_i)$, вираз (8.3) можна переписати так:

$$\sum_{i=1}^{n} [y_i - \varphi(x_i)]^2 = min. \tag{8.6}$$

Невідомі параметри шуканої залежності

$$\varphi(x) = \varphi(x_i, a_1, a_2, \dots, a_j, \dots, a_m)$$

визначають, записавши умову (8.3) не тільки як функцію аргументів х та у, але і як функцію невідомих параметрів a_j.

$$\sum_{i=1}^{n} [y_i - \varphi(x_i)]^2 = \sum_{i=1}^{n} \left[y_i - \varphi(x_i, a_1, a_2, \dots, a_j, \dots, a_m)\right]^2 = min. \tag{8.7}$$

Умова (8.7) виконується, якщо усі часткові похідні суми квадратів відхилень за параметрами a_j дорівнюють нулю. Часткові похідні дають систему m+1 рівнянь з m+1 невідомими, розв'язання якої дає шукані параметри a_j, що задовольняють умові (8.7).

Розглянемо приклад. Припустимо, що в результаті дослідів отримані такі експериментальні дані:

x_i	1	2	3
y_i	1	3	4

Побудуємо поле кореляції (рис. 8.3).

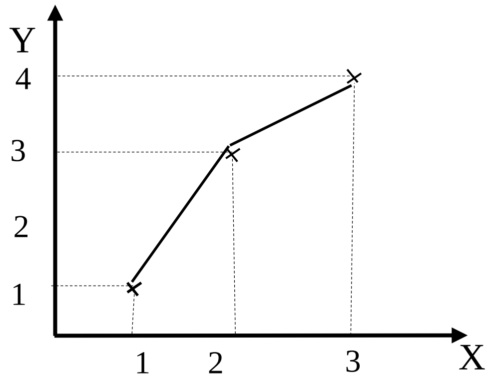


Рисунок 8.3 – Графік для прикладу

З графіка на рисунку 8.3 видно, що експериментальні точки не лежать на одній прямій, і що із зростанням незалежної змінної X залежна змінна Y має тенденцію до зростання.

Будемо вважати, що шукана залежність – лінійна: $y = a_0 + a_1 x$, підставимо її до виразу (8.7):

$$\sum_{i=1}^{n}(y_i - a_0 - a_1 x_i)^2 \rightarrow min.$$

Візьмемо часткові похідні за параметрами a_0 та a_1 і дорівняємо їх нулю. Оскільки невідомих параметрів два, ми отримаємо систему з двома рівняннями:

$$\left.\begin{aligned} \frac{\partial[\Sigma(y_i - a_0 - a_1 \cdot x_i)^2]}{\partial a_0} = 2 \cdot \Sigma(y_i - a_0 - a_1 \cdot x_i) \cdot (-1) = 0 \\ \frac{\partial[\Sigma(y_i - a_0 - a_1 \cdot x_i)^2]}{\partial a_1} = 2 \cdot \Sigma(y_i - a_0 - a_1 \cdot x_i) \cdot (-x_i) = 0 \end{aligned}\right\}.$$

Щоб згадати, як беруть похідні, треба просто спробувати їх узяти, і ви згадаєте:

– у першому рівнянні вираз $2 \cdot \Sigma(y_i - a_0 - a_1 \cdot x_i)$ є похідноюа від квадрата виразу під сумою (похідна від суми є сумою похідних, кожен доданок – у квадраті, отже множник 2 можна винести за знак суми; (-1) – це похідна від виразу у дужках за змінною a_0;

– у другому рівнянні аналогічно, тільки $(-x_i)$ – це похідна від виразу у дужках за змінною a_1.

Виконаємо перетворення, для чого відкриємо дужки:

оскільки справа 0, множники 2 та (-1) можна скоротити, тоді під знаком суми буде сума n доданків, отримуємо

$$\left.\begin{aligned} \Sigma y_i - \Sigma a_0 - \Sigma a_1 x_i = 0 \\ \Sigma y_i x_i - \Sigma a_0 x_i - \Sigma a_1 x_i^2 = 0 \end{aligned}\right\},$$

або

$$\left.\begin{array}{r}\Sigma a_0 + \Sigma a_1 x_i = \Sigma y_i \\ \Sigma a_0 x_i + \Sigma a_1 x_i^2 = \Sigma y_i x_i\end{array}\right\} \Longrightarrow \left.\begin{array}{r}n a_0 + a_1 \Sigma x_i = \Sigma y_i \\ a_0 \Sigma x_i + a_1 \Sigma x_i^2 = \Sigma y_i x_i\end{array}\right\}.$$

Обчислимо значення: $n = 3$; $\Sigma x_i = 6$; $\Sigma x_i^2 = 14$; $\Sigma y_i = 8$; $\Sigma y_i x_i = 19$ та підставимо їх до системи рівнянь:

$$\left.\begin{array}{r}3a_0 + 6a_1 = 8 \\ 6a_0 + 14a_1 = 19\end{array}\right\} \Rightarrow \left.\begin{array}{r}a_0 = -\frac{1}{3} \\ a_1 = \frac{3}{2}\end{array}\right\}.$$

Шукана залежність має вигляд: $y = \frac{-1}{3} + 1{,}5\,x$. Вона є найбільш імовірною з усіх лінійних залежностей.

Протабулюємо її та побудуємо графік (рис. 8.4):

y_i	-0,33	1,17	2,67	4,17	5,67	7,17
x_i	0	1	2	3	4	5

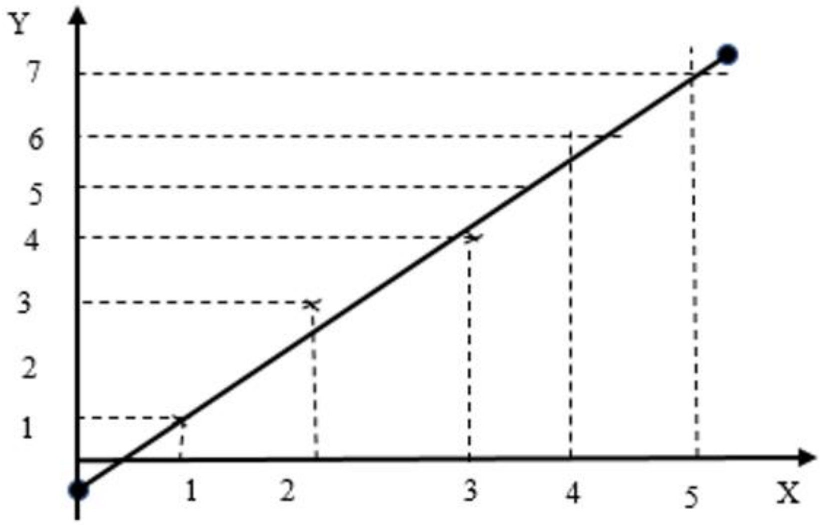


Рисунок 8.4 – Апроксимація даних лінійною залежністю

На рисунку 8.4 видно, як апроксимуюча лінія проходить поблизу експериментальних точок.

Також звернемо увагу на можливість прогнозування значень залежної змінної. Найчастіше як прогнозування розуміють передбачуваний розвиток певного процесу. Процесом називають зміни деякого явища у часі. Якщо вважати у розглянутому прикладі незалежну змінну X часом t, то значення залежної змінної Y характеризували би процес зміни Y у часі. Розрізняють прогнози короткострокові та довгострокові. Короткострокові прогнози – це прогнози на найближчий час. Для них можна застосовувати звичайні рівняння регресії. Для довгострокового прогнозування природних, економічних, політичних та ін. явищ розроблені спеціальні досить складні методи.

8.4 Перевірка гіпотез о параметрах моделі, оцінка її статистичної значущості та адекватності

Для визначення параметрів лінійної моделі $y = a_0 + a_1 x$, що є найбільш імовірною з усіх лінійних залежностей іноді зручно скористатись формулою

$$a_0 = \bar{y} - b\bar{x}\,, \quad a_1 = \frac{cov(x,y)}{\sigma_x^2}\,, \tag{8.8}$$

де $cov(x, y)$ – коваріація величин x і y, $cov(x, y) = \overline{yx} - \bar{y} \cdot \bar{x}$;

σ_x^2 – дисперсія незалежної величини х, $\sigma_x^2 = \overline{x^2} - \bar{x}^2$;

$\bar{x}$ – середнє значення незалежної величини x, $\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i$;

$\bar{y}$ – середнє значення залежної величини y, $\bar{y} = \frac{1}{n}\sum_{i=1}^{n} y_i$;

$\overline{yx}$ – середнє добутку x та y, $\overline{yx} = \frac{1}{n}\sum_{i=1}^{n} y_i x_i$;

$\overline{x^2}$ – середнє квадрата незалежної x, $\overline{x^2} = \frac{1}{n}\sum_{i=1}^{n} x_i^2$.

Коваріація – числова характеристика спільного розподілу двох випадкових величин, що дорівнює математичному сподіванню добутку відхилень цих випадкових величин від їхніх математичних сподівань. Дисперсія – характеристика випадкової величини, що визначають як математичне сподівання квадрата відхилення випадкової величини від її математичного сподівання. Математичне сподівання - сума добутків значень випадкової величини та відповідних імовірностей.

Параметр a_1 називають коефіцієнтом регресії. Його величина показує середню зміну залежної величини зі зміною незалежної величини на одну одиницю. Формально a_0 – значення у при $x = 0$.

Рівняння регресії завжди доповнюють показником тісноти зв’язку залежної величини Y та незалежної величини X. При використанні лінійної регресії як такий показник застосовують лінійний коефіцієнт кореляції r_{xy}, який можна розрахувати за формулами:

$$r_{xy} = a_1 \frac{\sigma_x}{\sigma_y} = \frac{cov(x,y)}{\sigma_x \sigma_y}. \tag{8.9}$$

Лінійний коефіцієнт кореляції розташовується у межах: $-1 \le r_{xy} \le 1$. Чим ближче абсолютне значення r_{xy} до одиниці, тим сильніше лінійний зв’язок у залежності, причому при $r_{xy} = \pm 1$ має місце строго функціональна залежність. Але необхідно мати на увазі, що близькість абсолютної величини лінійного коефіцієнта кореляції до нуля ще не означає відсутності зв’язку між ознаками. При іншій (нелінійній) специфікації моделі зв’язок між досліджуваними змінними може виявитися досить тісним.

Для оцінювання якості знайденої лінійної функції розраховують квадрат лінійного коефіцієнта кореляції r_{xy}^2 який називають коефіцієнтом детермінації.

Коефіцієнт детермінації характеризує частку дисперсії залежної величини y, що пояснюється регресією, у загальній її дисперсії:

$$r_{xy}^2 = 1 - \frac{\sigma_{зал}^2}{\sigma_y^2}, \quad (8.10)$$

де $\sigma_{зал}^2$ – залишкова дисперсія змінної y, $\sigma_{зал}^2 = \frac{1}{n}\sum_{i=1}^{n}(y - y_x)^2$;

σ_y^2 – дисперсія змінної y, $\sigma_y^2 = \frac{1}{n}\sum_{i=1}^{n}(y - \bar{y})^2 = \overline{y^2} - \bar{y}^2$.

Відповідно величина $1 - r_{xy}^2$ характеризує частку дисперсії y, викликану впливом інших, не врахованих в моделі, факторів.

Після того як рівняння лінійної регресії знайдене, проводять оцінку значущості як рівняння в цілому, так і окремих його параметрів.

Оцінка значущості рівняння лінійної регресії. Перевірити значущість рівняння регресії означає встановити, чи відповідає математична модель, що виражає залежність між змінними, експериментальним даним.

Щоб отримати загальне уявлення про якість моделі, визначають середню помилку апроксимації:

$$\bar{A} = \frac{1}{n}\sum_{i=1}^{n}\left|\frac{y - y_x}{y}\right| \cdot 100\%. \quad (8.11)$$

де y_i – статистичні значення y;

y_x – значення y, обчислені з отриманого рівняння регресії.

Середня помилка апроксимації має не перевищувати 8–10 %.

Для оцінювання значущості рівняння регресії в цілому висувають нульову гіпотезу H0: рівняння не відповідає даним статистичного ряду. Перевірку цього факту здійснюють за F-критерієм Фішера, для розрахунку якого виконують дисперсійний аналіз. В математичній статистиці дисперсійний аналіз розглядають як самостійний інструмент статистичного аналізу. Відповідно до основної ідеї дисперсійного аналізу, загальну суму квадратів відхилень змінної y від середнього значення $\bar{y}$ розкладають на дві частини – зумовлену регресією та зумовлену впливом неврахованих у моделі факторів:

$$\sum_i (y - \bar{y})^2 = \sum_i (y_x - \bar{y})^2 + \sum_i (y - y_x)^2, \quad (8.12)$$

де $\sum_i (y - \bar{y})^2$ – загальна сума квадратів відхилень статистичних значень залежної величини від її середнього;

$\sum_i (y_x - \bar{y})^2$ – сума квадратів відхилень, зумовлена регресією;

y_x – значення залежної величини, обчислені за рівнянням регресії;

$\sum_i (y - y_x)^2$– залишкова сума квадратів відхилень, зумовлена впливом неврахованих у моделі факторів.

Схема дисперсійного аналізу має вигляд, наданий таблицею 8.1, де n – кількість спостережень, m – кількість параметрів при змінній x. У знаменниках формул стоїть число ступенів свободи, що залежить від кількості результатів вимірів та кількості шуканих параметрів рівняння регресії.

Таблиця 8.1 – Схема дисперсійного аналізу

Компоненти дисперсії	Сума квадратів	Число ступенів свободи	Дисперсія на один ступінь свободи
Загальна	$\sum_i (y-\bar{y})^2$	$n-1$	$S^2_{заг} = \frac{\sum_i (y-\bar{y})^2}{n-1}$
Факторна	$\sum_i (y_x-\bar{y})^2$	m	$S^2_{факт} = \frac{\sum_i (y_x-\bar{y})^2}{m}$
Залишкова	$\sum_i (y-y_x)^2$	$n-m-1$	$S^2_{зал} = \frac{\sum_i (y-y_x)^2}{n-m-1}$

Визначення дисперсії на один ступінь свободи за формулами, що наведені у таблиці 8.1, призводить їх до порівняного вигляду. Зіставляючи зумовлену регресією й залишкову дисперсії, що розраховані на один ступінь свободи, одержимо величину F-критерію Фішера:

$$F = \frac{S^2_{факт}}{S^2_{заг}} . \tag{8.13}$$

Фактичне значення F-критерію Фішера (8.13) порівнюють з табличним значенням $F_{табл}(\alpha, k_1, k_2)$ при рівні значущості α і ступенях свободи $k_1 = m$ і $k_2 = n - m - 1$. У результаті, якщо фактичне значення F-критерію перевищує табличне, то признають статистичну значущість рівняння в цілому.

Для парної лінійної регресії $m = 1$, тому

$$F = \frac{S^2_{факт}}{S^2_{заг}} = \frac{\sum_i (y_x-\bar{y})^2}{\sum_i (y-y_x)^2} \cdot (n-2) . \tag{8.14}$$

Треба нагадати, що рівнем значущості α називають таку імовірність події, за якою цією подією можна зневажити, тобто дуже малу. Зазвичай обирають рівень значущості $\alpha = 0,1$ або $\alpha = 0,05$.

У свою чергу, величина F-критерія пов’язана з коефіцієнтом детермінації r^2_{xy}, і її можна розрахувати за такою формулою:

$$F = \frac{r^2_{xy}}{1-r^2_{xy}} \cdot (n-2) . \tag{8.15}$$

В парній лінійній регресії оцінюють значущість не тільки рівняння в цілому, але й окремих його параметрів. З цією метою висувають нульову гіпотезу Н0: параметр є незначущим. Для перевірки нульової гіпотези для кожного з параметрів визначають його стандартну помилку: m_{a_0} і m_{a_1}.

Стандартну помилку коефіцієнта регресії m_{a_1} визначають за формулою:

$$m_{a_1} = \sqrt{\frac{S^2_{зал}}{\sum_i (x-\bar{x})^2}} = \frac{S^2_{зал}}{\sigma_x \cdot \sqrt{n}} , \tag{8.16}$$

де $S^2_{зал}$ – залишкова дисперсія на один ступінь свободи, $S^2_{зал} = \frac{\sum_i (y-y_x)^2}{n-2}$.

Величину стандартної помилки разом з критерієм Стьюдента t, який є випадковою величиною з t-розподілом, тобто з розподілом Стьюдента, при n – 2 ступенях свободи застосовують для перевірки істотності коефіцієнту регресії і для розрахунку його довірчого інтервалу.

Для оцінювання істотності коефіцієнту регресії його величину порівнюють із його стандартною помилкою, тобто обчислюють фактичне значення t-критерію Стьюдента за формулою:

$$t_{a_1} = \frac{a_1}{m_{a_1}}, \tag{8.17}$$

яке далі порівнюють з табличним значенням за певним рівнем значущості α і числом ступенів свободи n – 2. Довірчий інтервал для коефіцієнта регресії визначають як $b \pm t_{\text{табл}\, m_{a1}}$.

Стандартну помилку параметра a_0 визначають за формулою:

$$m_a = \sqrt{S^2_{зал} \frac{\sum_i x^2}{n \sum_i (x-\bar{x})^2}} = S_{зал} \frac{\sqrt{\sum_i x^2}}{\sigma_x \cdot n} \,. \tag{8.18}$$

Процедура оцінювання істотності даного параметра така сама, як і розглянута вище для коефіцієнта регресії. Обчислюють t-критерій:

$$t_{a_0} = \frac{a_0}{m_{a_0}},$$

його величину порівнюють з табличним значенням при n – 2 ступенях свободи.

Значущість лінійного коефіцієнта кореляції перевіряють на основі величини помилки коефіцієнта кореляції m_r, яку визначають за формулою:

$$m_r = \sqrt{\frac{1-r^2}{n-2}}\,, \tag{8.19}$$

а фактичне значення t-критерію Стьюдента визначають як $t_r = \frac{r}{m_r}$.

Існує зв’язок між t-критерієм Стьюдента і F-критерієм Фішера, що виражається співвідношенням:

$$t_{\,a_1} = t_r = \sqrt{F}\,. \tag{8.20}$$

Контрольні запитання

1. Надайте характеристику основних етапів статистичного моделювання.
2. Поясніть сутність термінів «рівняння регресії» і «функціональна залежність».
3. На яких підставах обирають вид регресійної залежності?
4. Поясніть, з якою метою та за яким методом згладжують експериментальні залежності.
5. У чому полягає сутність методу найменших квадратів? Охарактеризуйте особливості методу та його переваг.
6. Поясніть поняття «статистична гіпотеза». Наведіть приклади.
7. Поясніть, з якою метою перевіряють статистичну гіпотезу та охарактеризуйте сутність її перевірки.
8. Як розраховують залишкову дисперсію та загальну дисперсію залежної величини y?
9. Поясніть, що таке коефіцієнт регресії, коефіцієнт кореляції та коефіцієнт детермінації.
10. Як визначають статистичну значущість рівняння регресії в цілому?
11. З якою метою визначають стандартну помилку коефіцієнта регресії? Який критерій для цього використовують?
12. Як обчислюють довірчий інтервал для коефіцієнта регресії?
13. З якою метою та як визначають значущість лінійного коефіцієнта кореляції?
14. У яких випадках для апроксимації статистичного ряду використовують лінійну модель?
15. Поясніть, що розуміють як кількість ступенів свободи. Як її визначають для зумовленої регресією та залишкової сум квадратів?
16. Поясніть призначення критерію Фішера.
17. Як оцінюють статистичну значущість рівняння регресії в цілому?
18. Як оцінюють значущість параметрів рівняння регресії?
19. Поясніть сутність середньої помилки апроксимації. Як її обчислюють?
20. Поясніть, що називають процесом, які його ознаки.
21. З якою метою застосовують прогнозування? Які види прогнозування вам відомі?
22. Поясніть підходи до короткострокового та довгострокового прогнозування.

СПИСОК РЕКОМЕНДОВАНИХ ДЖЕРЕЛ

1. Стеценко І. В. Моделювання систем [Електрон. ресурс] : навч. посіб / І. В. Стеценко ; М-во освіти і науки України, Черкас. держ. технол. ун-т. – Черкаси : ЧДТУ, 2010. – 399 с. – ISBN 978-966-402-073-9. – Електрон. текст дані. – Режим доступу: http://web.kpi.kharkov.ua/auts/wp-content/uploads/sites/67/2017/02/MOCSKachanov posobie.pdf, вільний (дата звернення: 12.11.2022). – Назва з екрана. .

2. Теорія систем і системний аналіз : навч. посіб. / А. Є. Ачкасов, В. А. Лушкін, В. М. Охріменко, Т. Б. Воронкова ; Харків. нац. ун-т міськ. госп-ва ім. О. М. Бекетова. – Харків : ХНУМГ, 2014. – 167 с. – ISBN 978-966-695-319-6.

3. Григорків В. С. Оптимізаційні методи та моделі : підручник / В. С. Григорків, М. В. Григорків. – Чернівці : Чернівецький нац. ун-т, 2016. – 400 с.

4. Оптимізаційні методи та моделі : навч. посіб. / В. А. Кривень, В. Б. Валяшек, Л. І. Цимбалюк, Г. В. Козбур. – Тернопіль : ТНТУ, 2015. – 83 с.

5. Таха Х. А. Введение в исследование операций / Х. А. Таха. – М. : Изд. дом «Вильямс», 2005. – 912 с.

6. Беспалов Н. А. Экономико-математические методы в топографо-геодезическом производстве / Н. А. Беспалов, А. И. Голубцов, А. А. Синдеев. – М. : Недра, 1983. – 320 с.

7. Моделі і методи соціально-економічного прогнозування = Models and methods of social and economic forecasting : підруч. для студ. вищ. навч. закл. / В. М. Геєць, Т. С. Клебанова, О. І. Черняк, В. В. Іванов, Н. А. Дубровіна ; Харків. нац. екон. ун-т. – Харків : ВД «Інжек», 2005. – 394 с.

8. Наконечний С. І. Економетрія : навч. посіб. / С. І. Наконечний, Т. О. Терещенко, Т. П. Романюк. – 4-те вид., доп. та перероб. – Київ : КНЕУ, 2006. – 528 с.

9. Вижва А. Статистичне моделювання випадкових процесів та двовимірних полів в аеромагнітометрії / А. Вижва, В. Демидов // Вісник Київського національного університету імені Тараса Шевченка : наук.-техн. зб. – Київ, 2012. – Вип. 56. – С. 52–55. (Серія «Геологія»).

12. Шевчук О. А. Статические статистические модели прогноза объема выполняемых работ / О. А. Шевчук // Системи обробки інформації. – Донецк, 2013. – Вип. 7 С. 114.

Електронне навчальне видання

ВОРОНКОВ Олексій Олександрович

ОСНОВИ МОДЕЛЮВАННЯ СКЛАДНИХ СИСТЕМ

КОНСПЕКТ ЛЕКЦІЙ

(для здобувачів першого (бакалаврського) рівня вищої освіти
зі спеціальності 193 – Геодезія та землеустрій)

Відповідальний за випуск *С. Г. Нестеренко*
За авторською редакцією
Комп’ютерне верстання *О. О. Воронков*

План 2023, поз. 7Л

Підп. до друку 10.05.2023. Формат 60 × 84/16.
Ум. друк. арк. 4,4.

Видавець і виготовлювач:
Харківський національний університет
міського господарства імені О. М. Бекетова,
вул. Маршала Бажанова, 17, Харків, 61002.
Електронна адреса: office@kname.edu.ua
Свідоцтво суб’єкта видавничої справи:
ДК № 5328 від 11.04.2017.