

МІНІСТЕРСТВО ОСВІТИ І НАУКИ УКРАЇНИ
ДЕРЖАВНИЙ ВИЩИЙ НАВЧАЛЬНИЙ ЗАКЛАД
«ДОНЕЦЬКИЙ НАЦІОНАЛЬНИЙ ТЕХНІЧНИЙ УНІВЕРСИТЕТ»
Кафедра прикладної математики та інформатики

СХЕМОКОНСПЕКТ ЛЕКЦІЙ

з дисципліни «Теорія та практика комп'ютерних експериментів» для аспірантів спеціальності
122 «Комп'ютерні науки» всіх форм навчання

ЛУЦЬК-2023

УДК 004.94

С 92

Схемоконспект лекцій з дисципліни «Теорія та практика комп'ютерних експериментів» для аспірантів спеціальності 122 «Комп'ютерні науки» всіх форм навчання [Електронний ресурс] / укл. О.М. Любименко. – Луцьк : ДонНТУ, 2023. – 126 с.

Навчальна дисципліна «Теорія та практика комп'ютерних експериментів» є актуальною і затребуваною в контексті підготовки сучасних фахівців у галузі комп'ютерних наук.

Запропаний студентам для опанування схемоконспект лекцій з цієї дисципліни має за мету забезпечити їм певну допомогу під час вивчення ними ключових положень базових тем курсу. Схематична форма конспекту акцентує увагу студентів на закріпленні основних фахових знань, формуванні відповідних умінь і в цілому отриманні ними кваліфікаційних компетентностей для аспірантів денної та заочної форми навчання спеціальності 122 «Комп'ютерні науки».

Укладач
інформатики

Любименко О.М., доц., к.ф.-м.н., доц. кафедри прикладної математики та

Рецензент

Штепа О.А. доц., к.т.н., доц. кафедри електронної техніки

Відповідальний за випуск

В.о. завідувача кафедри

Маслова Н.О., доц., к.т.н., доц. кафедри прикладної математики та інформатики

Затверджено навчально-методичним відділом ДонНТУ, протокол № 7 від 25. 04. 2023 року

Розглянуто на засіданні кафедри прикладної математики та інформатики, протокол № 4 від 18. 04. 2023 року

© Донецький національний технічний університет, 2023

ЗМІСТ

ВСТУП	6
ТЕМА 1. СУТНІСТЬ ЕКСПЕРИМЕНТУ, ЗАГАЛЬНІ ВИМОГИ ДО ПРОВЕДЕННЯ КОМП'ЮТЕРНИХ ЕКСПЕРИМЕНТІВ	7
1.1.Опис та визначення експерименту, комп'ютерного експерименту.	8
1.2.Класифікація експериментів.	9
1.3.Організація та постановка експерименту.	10
1.4.Структурна схема експерименту.	11
Тема 2. Огляд основних етапів експерименту	13
2.1. Етапи підготовки наукового експерименту .	15
2.2.Методика експерименту	16
2.3. План експерименту	17
2.4. Комп'ютерне моделювання	18
2.5. Етапи комп'ютерного моделювання	19
ТЕМА 3. ПЛАНУВАННЯ ЕКСПЕРИМЕНТУ	33
3.1 Історія теорії планування експерименту	34
3.2 Теорія планування експерименту	35
3.3 Активний експеримент	38

3.4 Пасивний експеримент	39
3.5 Критерії оптимальності планів	42
3.6 Плани експерименту	45
3.7 Кількість експериментів	49
ТЕМА 4. ЧИННИКИ ТА ФУНКЦІЙ ВІДГУКУ В ДОСЛІДЖЕННІ ІНФОРМАЦІЙНО-КОМП'ЮТЕРНИХ СИСТЕМ.	53
4.1. Планування експерименту	54
4.2. Залежні змінні (відгуки)	56
4.3. Незалежні змінні (фактори)	57
4.4. Вибір рівнів факторів	59
ТЕМА 5. ПОВНИЙ ФАКТОРНИЙ ЕКСПЕРИМЕНТ	64
5.1. Опис, визначення повний факторний експеримент	65
5.2. Матриці планування експерименту	71
5.3. Розрахунок коефіцієнтів рівняння регресії	74
5.4. Визначення статистичної значущості коефіцієнтів рівняння регресії	78
Тема 6. ДРОБОВИЙ ФАКТОРНИЙ ЕКСПЕРИМЕНТ	81
6.1. Опис, визначення дробового факторного експерименту	82
6.2. Розрахунок значення коефіцієнтів регресії	91

ТЕМА 7. ОБРОБКА РЕЗУЛЬТАТІВ КОМП'ЮТЕРНИХ ЕКСПЕРИМЕНТІВ, ОЦІНКА ЇХ ТОЧНОСТІ ТА ДОСТОВІРНОСТІ 93

7.1. Визначення статистичної значущості коефіцієнтів математичної моделі	94
7.2. Розрахунок дисперсії адекватності	96
7.3. Розрахунок дисперсії відновлення	102
7.4. Перевірка адекватності моделі	104

ТЕМА 8. МЕТРИКИ ЕФЕКТИВНОСТІ WEB СЕРВІСІВ. МЕТРИКИ ЕФЕКТИВНОСТІ БАЗ ДАНИХ. ЕКСПЕРИМЕНТИ ТА ОПТИМІЗАЦІЯ WEB СЕРВІСІВ. ОСОБЛИВОСТІ ВИКОНАННЯ ЕКСПЕРИМЕНТІВ ТА ОПТИМІЗАЦІЇ (Т'ЮНІНГУ) БАЗ ДАНИХ. 108

8.1. Метрики ефективності web сервісів.	109
8.2. Метрики ефективності баз даних.	112
8.3. Експерименти та оптимізація web сервісів.	119
8.4. Особливості виконання експериментів та оптимізації (т'юнінгу) баз даних	121

СПИСОК ВИКОРИСТАНОЇ ЛІТЕРАТУРИ 124

ВСТУП

Статистичні методи планування експерименту з'явилися на початку ХХ століття. Їх основи були закладені в роботах англійського статистика Рональда Фішера, в яких автор довів доцільність одночасного варіювання всіма факторами, що істотно впливають на параметр, що оптимізується, на відміну від широко поширеного до цього однофакторного експерименту. Найбільш ефективно практичне застосування планування екстремального експерименту, теоретичні основи якого заклали Г. Вох та К. Wilson, отримало в 40-60 роках ХХ століття при реалізації Манхеттенського проекту в США. Розвиток теорії та практики планування експерименту отримало у роботах Хантера, Д. Фінта, Дж. Кіфера, Х. Шенка, Д.К. Монтгомері, В.В. Налімова Ю.П. Адлера, І.П. Копилова та інших вчених.

Таким чином, для успішного вирішення наукових, виробничих, і технологічних проблем і завдань, інженеру - досліднику необхідно мати компетенції в галузі математичної теорії та практичних навичок планування експерименту та методів обробки експериментальних даних.

При вивченні курсу «Теорія та практика комп'ютерних експериментів», відповідно до робочої програми цієї навчальної дисципліни, передбачено опанування аспірантами 8 тем, які поділені на декілька змістовних тематичних питань. Під час лекційних занять з цього курсу аспірантам також пропонується виконання спеціальних самостійних завдань (контрольні питання), для поглибленого ознайомлення із деякими тематичними питаннями. Стисле, але вичерпне подання матеріалу забезпечить аспірантам можливість зосередитися на основних ключових поняттях, проблемах, методах, підходах при проведенні комп'ютерних експериментів, та в цілому сприятиме формуванню у них необхідних для третього кваліфікаційного рівня знань, умінь і компетентностей.

ТЕМА 1. СУТНІСТЬ ЕКСПЕРИМЕНТУ, ЗАГАЛЬНІ ВИМОГИ ДО ПРОВЕДЕННЯ КОМП'ЮТЕРНИХ ЕКСПЕРИМЕНТІВ

План :

1. Опис та визначення експерименту, комп'ютерного експерименту.
2. Класифікація експериментів.
3. Організація та постановка експерименту.
4. Структурна схема експерименту.

- 1.1. ОПИС ТА ВИЗНАЧЕННЯ ЕКСПЕРИМЕНТУ, КОМП'ЮТЕРНОГО ЕКСПЕРИМЕНТУ
- Що таке експеримент?
- Слово «експеримент» (від латинського experiment – проба, дослід) означає науково поставлений дослід, проведений у цілком певних умовах, що дає змогу стежити за перебігом відповідного процесу та відтворювати його при повторенні цих умов [1].
- Комп'ютерний експеримент — дослідження математичної моделі за допомогою комп'ютера, під час якого за одними параметрами моделі обчислюють інші її параметри й на цій основі роблять висновки про властивості об'єкта, описані математичною моделлю [1].

1.2 КЛАСИФІКАЦІЯ ЕКСПЕРИМЕНТІВ [1]

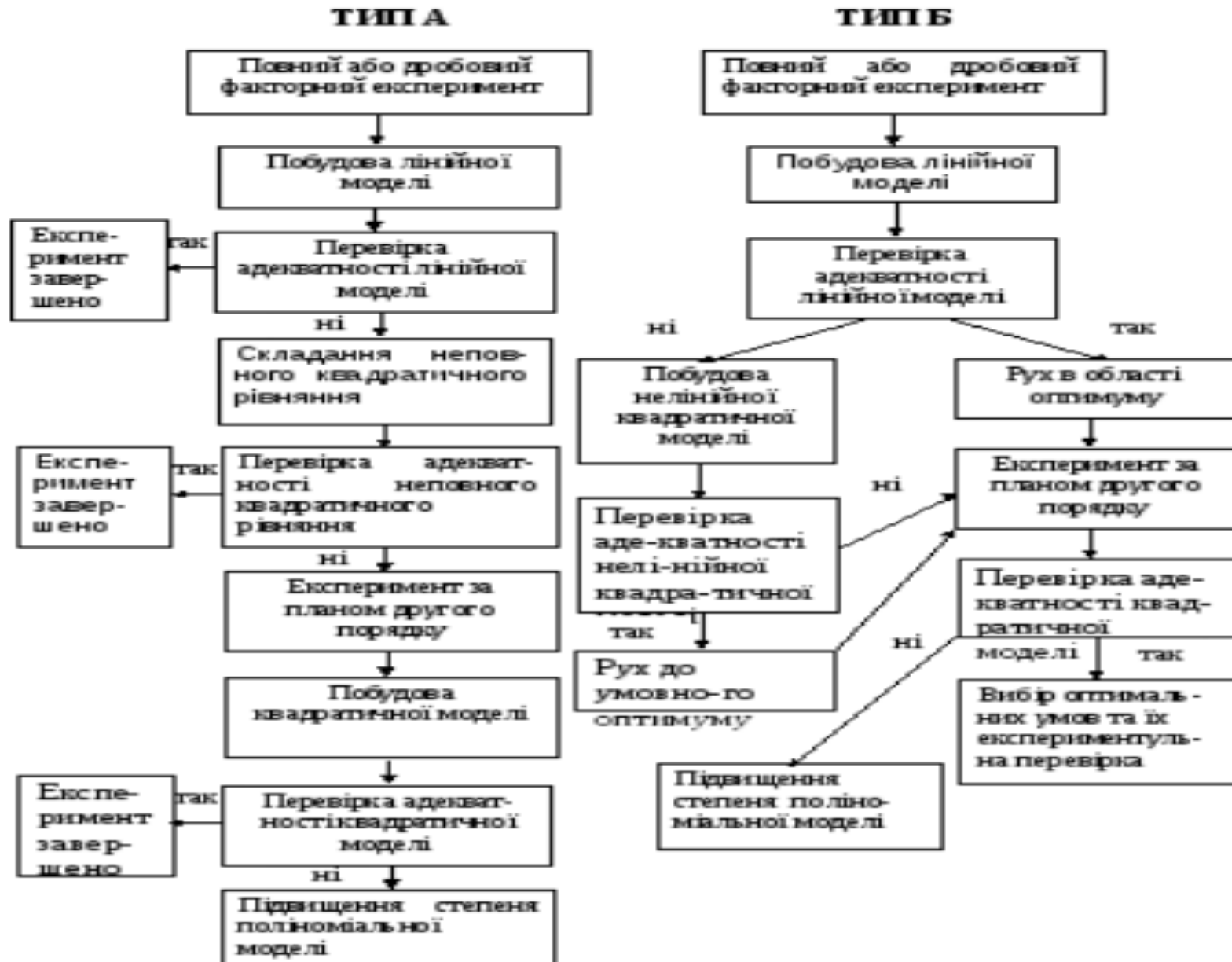


1.3 ОРГАНІЗАЦІЯ ТА ПОСТАНОВКА ЕКСПЕРИМЕНТУ

Експерименти в різних галузях науки розрізняють [1,2]:

- за способом формування умов (природних і штучних);
- за метою дослідження (трансформуючі, констатуючі, контролюючі, пошукові, розв'язувальні);
- за організацією проведення (лаборетбрні , натурні, польові, виробничі тощо);
- за структурою об'єктів і явищ, які вивчаються (прості, складні);
- за характером зовнішніх впливів на об'єкт дослідження (речовинні, енергетичні, інформаційні);
- за характером взаємодії засобу експериментального дослідження з об'єктом дослідження (звичайний та модельний);
- за типом моделей, досліджуваних у експерименті (матеріальний та мислений);
- за контрольованими величинами (пасивний і активний);
- за числом змінюваних факторів (однофакторний та багатофакторний);
- за характером об'єктів, які вивчаються (технологічні, соціометричні) тощо.

1.4 СТРУКТУРНА СХЕМА ЕКСПЕРИМЕНТУ



Контрольні питання

1. Надайте опис та дайте визначення експерименту, комп'ютерного експерименту?
2. Надайте класифікацію експериментів?
3. Наведіть етапи підготовки наукового експерименту?
4. Опишіть методологію та організацію постановки експерименту?
5. Наведіть структурну схему експерименту?

Тема 2. Огляд основних етапів експерименту

План :

1. Етапи підготовки наукового експерименту .
2. Методика експерименту
3. План експерименту
4. Комп'ютерне моделювання
5. Етапи комп'ютерного моделювання

викладач: Любименко Олена Миколаївна
olena.liubymenko@donntu.edu.ua

• Повторення теми 1 : Структурна схема експериментів

Експеримент — це дослід, який проводиться з об'єктом або моделлю, що виявляється у впливі на досліджуваний об'єкт за допомогою спеціальних інструментів і приладів [1]..

Експеримент відрізняється від спостереження активною взаємодією з досліджуваним об'єктом і має такі особливості [1]:

- дослідник сам викликає досліджуване явище, а не чекає, коли воно з'явиться;
- можна змінювати умови протікання процесу, що вивчається [1];
- в експерименті можна поперемінно виключати окремі умови з метою встановлення закономірних зв'язків [1];
- експеримент дає змогу варіювати кількісне співвідношення умов і здійснювати математичну обробку даних [1].



Комп'ютерний експеримент — дослідження математичної моделі за допомогою комп'ютера, під час якого за одними параметрами моделі обчислюють інші її параметри й на цій основі роблять висновки про властивості об'єкта, описані математичною моделлю [1].

2.1 Етапи підготовки наукового експерименту

- перевіряти, та методику експериментальних робіт [1];
- визначити способи і прийоми впливу на об'єкт дослідження [1];
- забезпечити умови для виконання експериментальних робіт [1];
- розробити шляхи і прийоми фіксування ходу і результатів експерименту [1];
- підготувати засоби експерименту (прилади, установки, моделі тощо) [1];
- забезпечити експеримент необхідним обслуговуванням[1].

2.2 МЕТОДИКА ПРОВЕДЕННЯ ЕКСПЕИМЕНТУ

Методика – це сукупність обдуманих і фізичних операцій, які розміщені у визначеній послідовності для досягнення поставленої мети дослідження[1]/

Під час розроблення методики проведення експерименту необхідно передбачати [1]:

- – попереднє цілеспрямоване спостереження за об`єктом або явищем, що вивчається, з метою визначення вихідних даних (гіпотез, обрання змінних факторів);
- – створення умов, у яких можливе експериментування (добір об`єктів для експериментальної дії, усунення впливу випадкових факторів);
- – визначення області інтересу для змінних факторів та меж вимірювання;
- – можливість систематичного спостереження за розвитком явища і точного опису фактів;
- проведення систематичної реєстрації замірів і оцінок фактів різними засобами і способами;
- – створення складних ситуацій з метою підтвердження або спростування раніше отриманих даних;
- – перехід від емпіричного вивчення з логічним узагальненням до аналізу та теоретичного оброблення отриманих фактичних даних.

2.3 ПЛАН ЕКСПЕРИМЕНТУ

Перед кожним експериментом складається його план (програма виконання), який включає такі етапи [1]:

- мету, завдання та обґрунтування об'єму експерименту;
- вибір змінних факторів;
- визначення кількості дослідів та послідовності зміни факторів;
- вибір кроку зміни факторів, визначення інтервалів між майбутніми експериментальними точками;
- обґрунтування вибору засобів для вимірювання;
- опис проведення експерименту;
- обґрунтування вибору способів оброблення та аналізу результатів експерименту.

2.4 КОМП'ЮТЕРНЕ МОДЕЛЮВАННЯ

Процес комп'ютерного моделювання складається з декількох етапів, кожний з яких передбачає певні дії [2]

№	Етап	Дії
1	Постановка завдання та його аналіз	1. Визначити, з якою метою створюється модель. 2. Уточнити, які результати та в якій формі слід отримати. 3. Визначити, які вихідні дані потрібні для створення моделі
2	Побудова інформаційної моделі	1. Визначити властивості моделі та виявити взаємозв'язки між ними. 2. Оцінити, які з властивостей є істотними при побудові моделі, а якими можна знехтувати. 3. Математично описати залежність між властивостями моделі
3	Розробка методу та алгоритму реалізації комп'ютерної моделі	1. Вибрати чи розробити метод отримання вихідних результатів. 2. Скласти алгоритм отримання результатів за вибраними методами. 3. Перевірити правильність алгоритму
4	Розробка комп'ютерної моделі	1. Вибрати засоби програмної реалізації алгоритму на комп'ютері. 2. Розробити комп'ютерну модель. 3. Перевірити правильність створеної комп'ютерної моделі
5	Проведення комп'ютерного експерименту	1. Розробити план дослідження. 2. Провести експеримент на основі створеної комп'ютерної моделі. 3. Проаналізувати отримані результати. 4. Створити висновки щодо властивостей створеної моделі

2.5 Етапи комп'ютерного моделювання

Комп'ютерне моделювання, тобто створення та дослідження моделі з використанням комп'ютерних програм, можна поділити на кілька етапів. Розглянемо ці етапи на прикладі такої задачі [2].

Задача 1.

**Використовуючи
комп'ютер,
розробити проект
шкільного
спортивного
майданчика.**



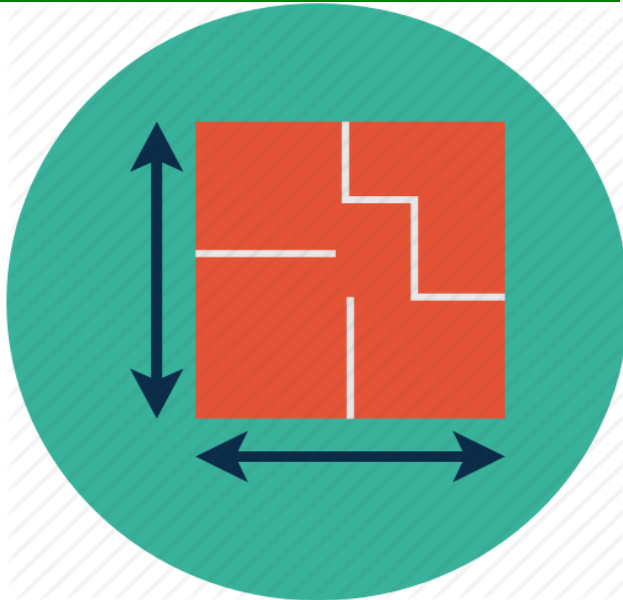
Етапи комп'ютерного моделювання

Для розв'язування цієї задачі спочатку потрібно [2] з'ясувати,

Розміри майданчика

Які об'єкти слід на ньому розмістити

В якому вигляді потрібно цей проект надати



баскетбольний і волейбольний майданчики, тенісний корт, майданчик для футболу, гімнастичні тренажери тощо

наприклад, у вигляді плану (креслення) такого майданчика, фінансового кошторису тощо

Етапи комп'ютерного моделювання

Далі потрібно створити план майбутнього майданчика, визначити, які властивості його об'єктів буде відображено на плані, які значення вони набуватимуть. Деякі із цих значень є стандартними, а деякі треба визначити самим [2].



Етапи комп'ютерного моделювання

Наприклад, якщо на майданчику потрібно встановити [2],

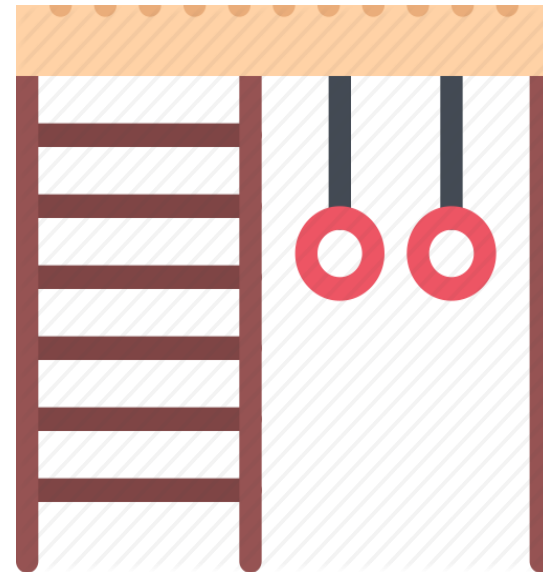
Баскетбольні щити

то їх розміри, висота нижнього краю від землі та висота кільця від землі є стандартними



Шведську стінку

довжину перекладин та їх кількість у шведській стінці можемо визначити самотійно



Етапи комп'ютерного моделювання

На цьому етапі також слід встановити залежності між об'єктами, наприклад відстані між ними, та формули, за якими можуть бути обчислені результати, зокрема вартості [2].

Цей етап розв'язування задачі називається **побудова інформаційної моделі**.



Етапи комп'ютерного моделювання

Далі потрібно вибрати засіб, який буде використано для створення цього проекту. У нашому випадку це можуть бути прикладні програми [2]:

Графічний редактор



Спеціальний пакет програм для креслення



Табличний процесор для обчислень



А можна скласти власну програму

яка за вхідними даними задачі на основі інформаційної моделі запропонує один чи кілька варіантів проектів такого майданчика, виконає потрібні обчислення

Цей етап називається *вибір програмних засобів для розв'язуванні задачі.*

Етапи комп'ютерного моделювання

Якщо ми вирішимо використати **прикладні програми**, то далі потрібно [2]:

1. *Вибрати ті засоби цих програм, які нададуть можливість створити потрібний проект.*
2. *Скласти алгоритм його створення, враховуючи вибрані засоби програми.*
3. *Виконати складений алгоритм і отримати комп'ютерну модель проекту майданчика.*

Ці етапи називаються **вибір засобів у середовищі прикладної програми, складання алгоритму, виконання алгоритму й отримання комп'ютерної моделі об'єкта** [2].

Етапи комп'ютерного моделювання

*Якщо ж ми вирішимо скласти **власну програму** для розв'язування цієї задачі, то перш за все потрібно вибрати мову програмування. Після цього треба [2]:*

1. Скласти алгоритм (алгоритми) розв'язування задачі.

2. Скласти програму вибраною мовою програмування відповідно до складених алгоритмів, розробити інтерфейс користувача програми.

3. Запустити складену програму на виконання, увести вхідні дані й одержати кінцеві результати (у даній задачі — комп'ютерну модель проекту майданчика).

Етапи комп'ютерного моделювання

*Ці етапи називаються **вибір мови програмування, складання алгоритму розв'язування задачі, складання програми, виконання програми й отримання комп'ютерної моделі об'єкта [2].***



Етапи комп'ютерного моделювання

Після того як одержимо результат,

*з використанням
прикладних програм*

або

*з використанням
спеціально створеної
програми*

*Його потрібно дослідити на відповідність умові
поставленої задачі.*

*Якщо всі вимоги умови виконуються, то
можна вважати, що задачу розв'язано
правильно й одержали правильний її розв'язок,
у нашому випадку — проект спортивного
майданчика [2].*



Етапи комп'ютерного моделювання

Якщо ж ні, потрібно шукати помилки на попередніх етапах розв'язування задачі (вони називаються логічними помилками), виправляти їх і знову шукати правильний розв'язок задачі, тобто такий її розв'язок, який повністю відповідає умові задачі [2].

Цей етап називається дослідження отриманих результатів на реальність і на відповідність умові.



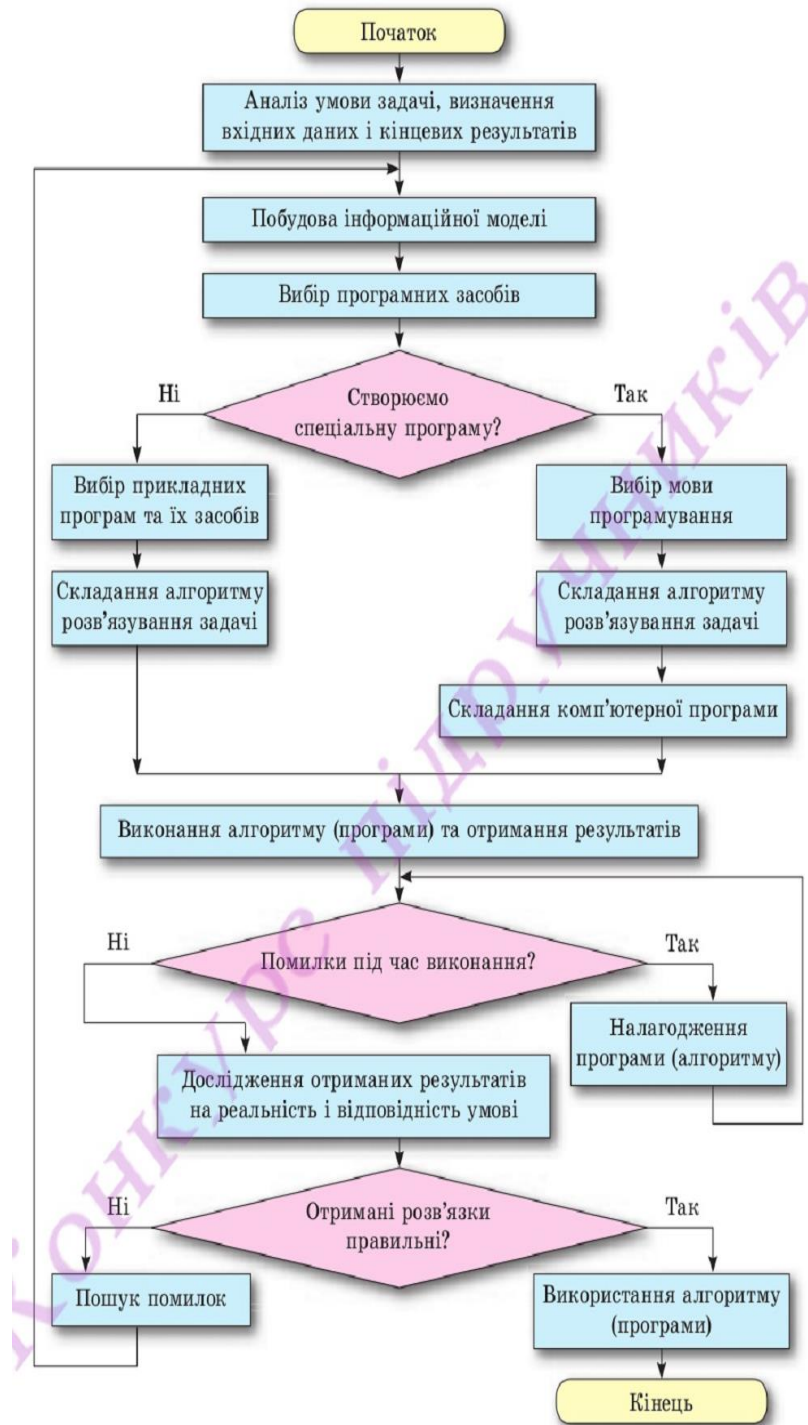
Етапи комп'ютерного моделювання

Звертаємо вашу увагу, що дослідження отриманих результатів найчастіше проводиться не на одному наборі вхідних даних, а цілій серії таких наборів (тестові набори даних), які підбираються так, щоб якомога повніше, з усіх сторін, у різних ситуаціях дослідити отримані результати на їх реальність і відповідність умові задачі [2].



Етапи комп'ютерного моделювання

З розглянутого прикладу можна виділити такі основні етапи комп'ютерного моделювання [2].



КОНТРОЛЬНІ ПИТАННЯ

- 1. Наведіть етапи підготовки наукового експерименту?
- 2. Опишіть методику проведення експерименту ?
- 3. Що таке план експерименту ?

ТЕМА 3. ПЛАНУВАННЯ ЕКСПЕРИМЕНТУ

План

3.1 Історія теорії планування експерименту

3.2 Теорія планування експерименту

3.3 Активний експеримент

3.4 Пасивний експеримент

3.5 Критерії оптимальності планів

3.6 Плани експерименту

3.7 Кількість експериментів

викладач: Любименко Олена Миколаївна
olena.liubymenko@donntu.edu.ua

3.1 ІСТОРІЯ ПЛАНУВАННЯ ЕКСПЕРИМЕНТУ

- Ініціатором використання методів планування експерименту і їх фундатором вважається американський дослідник Рональд Фішер [1].
- Іншим відомим автором перших робіт з планування експерименту є Френк Йетс. Ідеї планування експерименту формувалися в працях Дж. Бокса, Дж. Кифера та інших [1].
- На території сучасної України ця теорія розвивається в наукових колективах, заснованих А.Г. Бондар та В.А. Вознесенським [1].
- Методи планування експерименту закладені в спеціалізованих програмних пакетах, зокрема StatGrapfics, Statistica, Stat-Sens та інших [1].

3.2 ТЕОРІЯ ПЛАНУВАННЯ ЕКСПЕРИМЕНТУ

Тепер розглянемо більш наближений до технологій приклад – зважування трьох об'єктів A , B і C . Традиційно схема зважування виглядає так: спочатку робиться нульове зважування (нульова точка терезів), а потім по черзі зважується кожний з об'єктів (табл. 0.1).

Таблиця 0.1.

Традиційна схема зважування трьох об'єктів

Номер досліду	Об'єкти			Результат зважування
	A	B	C	
1	-1	-1	-1	y_0
2	+1	-1	-1	y_1
3	-1	+1	-1	y_2
4	-1	-1	+1	y_3

Значення в комірці +1 означає, що об'єкт присутній на терезах -1, що відсутній. Вага кожного об'єкту оцінюється за результатами двох дослідів, наприклад:

$$A = y_1 - y_0,$$

тоді дисперсія результатів зважування запишеться у вигляді

$$\sigma^2(A) = \sigma^2(y_1 - y_0) = 2\sigma^2(y),$$

де $\sigma(y) = \sqrt{\sigma^2(y)}$ – похибка зважування.

Запропонуємо іншу схему зважування (табл. 0.2).

Таблиця 0.2.

Ефективна схема зважування				
Номер дослідів	Об'єкти			Результати
	А	В	С	
1	+1	-1	-1	y_1
2	-1	+1	-1	y_2
3	-1	-1	+1	y_3
4	+1	+1	+1	y_4

З таблиці видно, що в перших трьох дослідів послідовно зважуються об'єкти A , B , C , а в останньому досліді – зважуються всі три об'єкти разом, тобто нульове зважування не проводиться. Тоді вага кожного об'єкту визначається по формулах з врахуванням $y_4 = y_1 + y_2 + y_3$:

$$A = \frac{y_1 - y_2 - y_3 + y_4}{2},$$

$$B = \frac{-y_1 + y_2 - y_3 - y_4}{2},$$

$$C = \frac{-y_1 - y_2 + y_3 + y_4}{2}.$$

Знайдемо дисперсію результатів по новій схемі зважування:

$$\sigma^2(A) = \sigma^2\left[\frac{y_1 - y_2 - y_3 + y_4}{2}\right] = \frac{4\sigma^2(y)}{4} = \sigma^2(y).$$

- ▶ **Планування експерименту** – розділ математичної статистики, який займається формуванням оптимальних матриць експерименту, починаючи від оптимального вибору досліджуваних факторів і власне вибору плану експерименту і до методів аналізу одержаних експериментальних результатів. Планування експериментів забезпечує отримання максимуму надійної інформації при обмежених ресурсах [1].
- ▶ **План експерименту** – це сукупність даних, що визначають кількість, умови і порядок проведення експериментальних досліджень [1].
- ▶ **Метою планування експерименту** є знаходження таких умов і правил проведення дослідів, за яких вдасться одержати надійну і достовірну інформацію про об'єкт дослідження з найменшими витратами, а також представити цю інформацію в зручній і зрозумілій формі з кількісними оцінками одержаних значень і їх точності [1].

Це означає забезпечення:

- ▶ - правильного визначення часткової структури рівняння регресії;
- ▶ - стійкості коефіцієнтів регресії;
- ▶ - оцінки коефіцієнтів регресії з максимальною точністю

3.3 АКТИВНИЙ ЕКСПЕРИМЕНТ

В теорії планування:

активним, називається експеримент, який проводиться за спеціальною матрицею експерименту, яка відповідає певним критеріям оптимальності [1],

Активний експеримент дозволяє:

- мінімізувати загальну кількість дослідів [1];
- одночасно варіювати всіма змінними та оптимально використовувати факторний простір [1];
- організувати експеримент так щоб виконувались багато вихідніх передумов регресійного аналізу [1];
- використовувати математичний апарат та отримувати математичні моделі, що мають найкращі властивості порівняно з моделями, побудованими за результатами пасивного експерименту [1];
- численні чинники, що заважають перетворити на випадкові величини шляхом рандомізації умов дослідів [1].

3.4 ПАСИВНИЙ ЕКСПЕРИМЕНТ

Пасивним експеримент, незалежно від того, що дослідник планував експеримент, користуючись якимись своїми міркуваннями [1].

Єдиним виключенням є випадок повного факторного експерименту (ПФЕ), який є окремим видом багатофакторного регулярного плану і утворюється при повному переборі всіх варіантів комбінацій рівнів варіювання факторів [1].

- Необхідність планування експерименту або формування вибірки признається для всіх статистичних методів [1].
- Найбільш розвинена теорія планування регресійних та дисперсійних експериментів, тому іноді вона подається як частина відповідно регресійного та дисперсійного аналізу [1].

В регресійному аналізі планування експериментів, як правило, використовується для моделей, лінійних відносно параметрів: модель $\hat{y}$ має вигляд алгебраїчної суми довільних функцій (загальний вираз лінійний відносно невідомих параметрів b_i).

$$\hat{y} = \sum_{i=1}^{k'} b_i f_i(x_1, x_2, \dots, x_{k'}),$$

де k' – загальна кількість ефектів (головних і взаємодій), введених в математичну модель.

Використання степеневих моделей не обґрунтовано з математичної точки зору. Це зв'язано з наступними причинами [1] :

1. матриці для нелінійної регресії, як правило, погано обумовлені, що порушує ряд передумов регресійного аналізу та робить його використання некоректним [1];
2. складність інтерпретації для нелінійних відносно параметрів багатофакторних моделей;
3. за Швирковим В.В. у репрезентативній представительной представницької вибірці зв'язок між змінними лінійний (змінні можуть бути довільними функціями), а будь-яка нелінійність викликана впливом неврахованих причин, тому використання нелінійного відносно параметрів регресійного аналізу теоретично не обґрунтовано [1].

- План експерименту встановлює чисельні значення чинників та умови проведення дослідів.
- Фіксоване значення фактора називається рівнем фактора. Безліч точок, які відображають чисельні значення факторів, можна представити так:

$$X = \begin{bmatrix} x_{11} & x_{12} & x_{13} & \dots & x_{1k} \\ x_{21} & x_{22} & x_{23} & \dots & x_{2k} \\ \dots & \dots & \dots & \dots & \dots \\ x_{n1} & x_{n2} & x_{n3} & \dots & x_{nk} \end{bmatrix},$$

- Кількість рядків відповідає числу дослідів n ($i = 1, 2, \dots, n$). Далі, у плані-матриці експериментів, число рядків позначатимемо літерою N .
- Кількість стовпців відповідає кількості факторів k , які впливають вихідний параметр чи параметр оптимізації ($j = 1, 2, \dots, k$).

3.5 КРИТЕРІЇ ОПТИМАЛЬНОСТІ ПЛАНІВ

Критерії оптимальності планів та способи організації активного експерименту можна поділити на три групи [1].

- До першої групи відносять критерії, пов'язані з точністю оцінок коефіцієнтів регресії [1].
- *Ортогональність* дозволяє оцінювати всі коефіцієнти регресії незалежності один від одного і спрощувати або ускладнювати моделі, виключаючи або додаючи нові коефіцієнти без перерахунку вже знайдених. Ортогональність плану експерименту забезпечує мінімальну кількість обчислень [1].
- *Д-оптимальність* забезпечує мінімум узагальненої дисперсії всіх оцінок коефіцієнтів. При складанні планів, забезпечуючи мінімум узагальненої дисперсії лише частини оцінок коефіцієнтів, план називається усіченим Д-оптимальним [1].
- *А-оптимальність* забезпечує мінімум середньої дисперсії оцінок коефіцієнтів.
- *Е-оптимальність* не дає можливості деяким оцінкам коефіцієнтів мати надто великі дисперсії.

- До другої групи відносять критерії, які визначають точність передбачення відгуку за допомогою побудованої моделі [1].
- *Ротатабельність* планів забезпечує однакову точність передбачення для точок, так само віддалених від центру плану в будь-якому напрямку [1].
- *G-оптимальні* плани гарантують відсутність в галузі експерименту точок, що мають дуже низьку точність оцінки прокліку [1].
- *Q-оптимальні* плани мінімізують середню дисперсію предказання [1].

- До III групи відносять критерії, пов'язані зі стратегією експерименту [1].
- *Насиченість* плану (міра насиченості плану) забезпечує мінімум числа дослідів та визначається кількістю коефіцієнтів моделі [1].
- *Кроковий принцип планування* або композиційність планів передбачає реалізацію експерименту частинами, крок за кроком [1].
- Рішення про продовження експерименту та вибір методу на кожному наступному кроці приймають лише за результатами попереднього експерименту. Композиційні плани передбачають поступове ускладнення моделі, що синтезується [1].

3.6 ПЛАНИ ЕКСПЕРИМЕНТУ

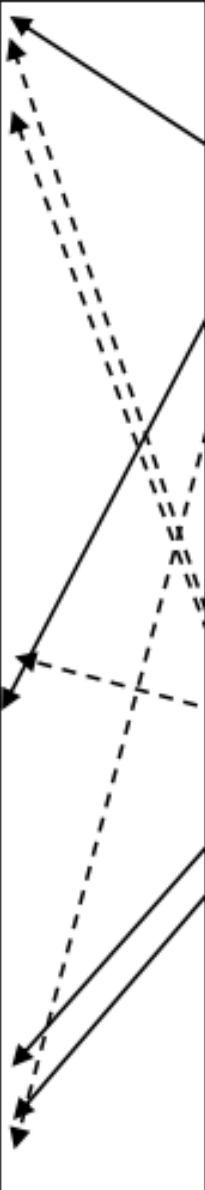
- Класичні плани мали однакову кількість рівнів варіювання для всіх факторів і відповідно називалися планами 1-го, 2-го та іноді більш високих порядків (наприклад, $3k$, $4k$) [1].
- Вершиною класичного планування експерименту стали багатофакторні регулярні плани, розроблені Бродським В.З. [1]/
- **НЕДОЛІКОМ** класичних планів експерименту є те, що вони є оптимальними тільки для певної структури рівняння регресії, яку необхідно знати що до проведення експерименту та побудови плану [1].
- Якщо фактична структура буде відрізнятися від тієї, яку визначив дослідник до побудови плану, то план не буде оптимальним [1].
- У зв'язку з цим були створені **робастні плани експерименту**, які залишаються оптимальними чи близькими до оптимальних незалежно від того, яка буде структура рівняння регресії [1].

робастні плани експерименту [1]

- залишаються оптимальними чи близькими до оптимальних незалежно від того, яка буде структура рівняння регресії [1].
- Для робастних планів характерним є те, що точки плану розміщені рівномірно в багатofакторному просторі, а також мінімальна корельованість між собою будь-яких ефектів (як головних, так і взаємодій) [1].
- В робастних планах головним є забезпечення найкращих умов для визначення структури рівняння регресії [1].

Наслідком виконання цієї умови є забезпечення стійкості коефіцієнтів регресії та виконання ряду передумов регресійного аналізу [1].

Таблиця 1. Місце робастних планів у загальній класифікації

Ідентифікація	Для оцінки коефіцієнтів регресії	<i>D</i> -оптимальність		Робастні (на основі багатофакторних регулярних планів)
		<i>A</i> -оптимальність		
		<i>E</i> -оптимальність		
		Мінімакс дисперсії оцінки коефіцієнтів		
		Мінімум суми відносних помилок оцінок		
	Для оцінки поверхні відгуку	Ортогональність		Робастні (на основі псевдовипадкових чисел)
		<i>G</i> -оптимальність		
		<i>Q</i> -оптимальність		
		<i>Ротатабельність</i>		
		Уніформність		
Специфікація		Дискримінуючі		
		Відсіюючі		
Оптимізація		Динамічні		
		Статичні		

ІСНУЄ ТРИ ВИДИ РОБАСТНИХ ПЛАНІВ:

- 1. робастні плани експерименту на базі БФРП (рис. 1);
- 2. плани на основі псевдовипадкових чисел (ЛП рівномірно розподілені послідовності, числа Холт

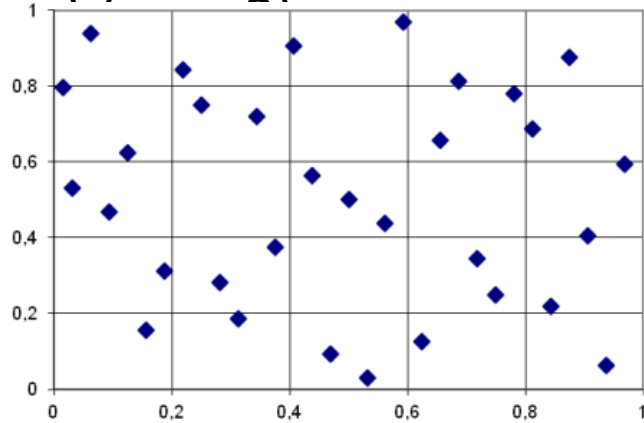


Рис. 2. Діаграма розсіювання для плану на основі ЛП_r рівномірно розподілених послідовностей для 32 дослідів

- 3. плани узагальненої свастики (рис. 3).

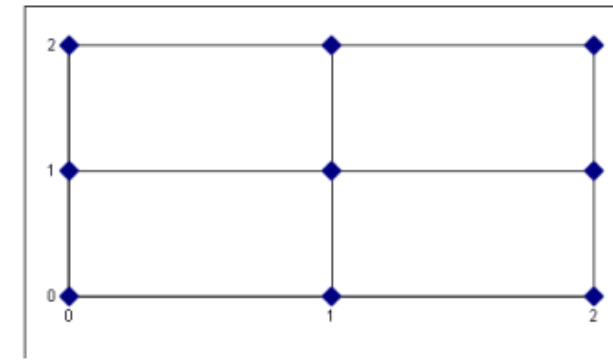


Рис. .1. Діаграма розсіювання для БФРП 3//9

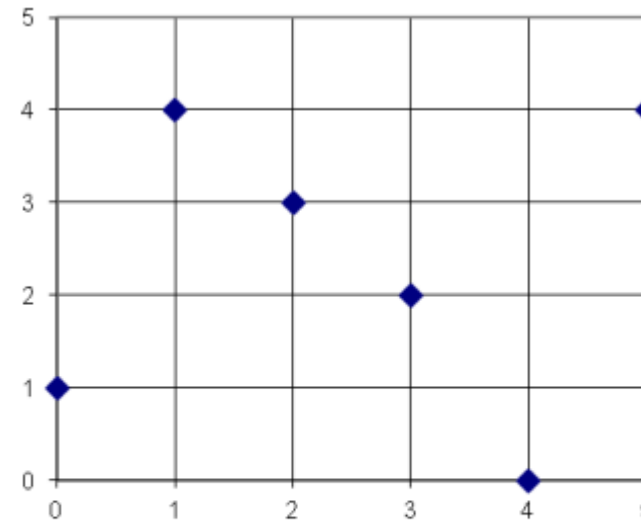


Рис .3. Діаграма розсіювання для плану
“узагальненої свастики” –
2 фактори, 6 дослідів

3.7 КІЛЬКІСТЬ ЕКСПЕРИМЕНТІВ

- Верхня межа числа експериментів визначається максимально можливими доступними ресурсами [1].
- Нижня залежить від очікуваної складності моделі, тобто максимально можливої кількості членів моделі [1].

Часто використовувані в класичній теорії планування експерименту насичені плані (кількість головних ефектів дорівнює числу експериментів) в загальному випадку непридатні, за винятком планів повного експерименту.

Чим більше факторів – тим більше можливих регресорів. Теоретичне ідеальна кількість експериментів, яка рівна кількості всіх можливих регресорів (повний факторний експеримент) в більшості реальних випадків неможлива, так як кількість регресорів вимірюється п'яти і шестизначними числами.

відмічено Саттерзвайтом і Парето, кількість регресорів, достатня для опису реальної емпіричної залежності складає незначну частку від загальної кількості. Число дослідів $N_{\text{розр}}$ для плану експерименту для побудови моделі розраховується за наступною формулою:

$$N_{\text{розр}} = (1,5 \dots 2) \left(1 + \sum_{i=1}^k s_i \right) ,$$

де k – число незалежних змінних (факторів), s_i – степінь поліному, достатня для апроксимації залежності відгуку від даної окремої незалежної змінної.

- Послідовність дій при визначенні кількості експериментів [1].
- 1. Визначається очікувана складність моделі (число факторів, степені поліному).
- 2. Обчислюється кількість експериментів.
- 3. Визначаються необхідні ресурси.
- 4. Визначається можливість отримання цих ресурсів.
- 5. Якщо це неможливо, то перехід до формалізації

Максимально необхідна кількість дослідів для проведення експериментальних досліджень визначається за формулою:

$$N_{\max} = N_{\text{розр}} n_1 + N_{\text{контр}} n_2 + P N_{\text{розр}} , \quad (2)$$

де $N_{\text{розр}}$ – необхідна кількість експериментів, розрахована за формулою (1); $N_{\text{контр}}$ – число контрольних експериментів; n_1, n_2 – кратність дублювання експериментів в навчальній (за якою будується модель) і контрольній (за якою модель перевіряється) вибірках відповідно; P – очікувана частка бракованих експериментів, які потрібно буде переробити.

Звичайно $n_1 = 2; 3$, $n_2 = 2$, $N_{\text{контр}} = 0,25N_{\text{розр}}$, $P = 0,1$. Ви можете прийняти $N_{\max} = N_{\text{розр}}$, але при цьому і взяти відповідальність за негативні наслідки такого кроку. Відсутність контрольних дослідів не дозволяє перевірити прогностичні властивості моделі. Про повторні досліди вже писалося раніше.

РЕЗУЛЬТАТ ФОРМАЛІЗАЦІЇ

Загальні результати формалізації [1].

- Опис незалежних змінних.
- Опис залежних змінних.
- При обмеженнях на можливі комбінації незалежних змінних необхідно описати простір допустимих значень.

Завдання для планування експерименту [1].

- Завдання для конструювання плану (вид плану і його опис).
- Опис обмежень на факторний простір.
- Таблиці для побудови робочої матриці.

Вимоги до умов проведення експерименту [1]

- Фіксовані незалежні змінні.
- Контрольовані змінні.
- Методики розрахунку і встановлення (якщо фактори не прямого вимірювання).
- Методики вимірювання чи визначення відгуку.

Вимоги до моделі [1] : Призначення моделі, виходячи з мети її побудови.

КОНТРОЛЬНІ ПИТАННЯ

1. Опишіть теорію планування експерименту ?
2. Дайте визначення активного експерименту?
3. Дайте визначення пасивного експерименту?
4. Наведіть критерії оптимальності планів?
5. Охарактеризуйте плани експерименту?
6. Від чого залежить кількість експериментів?

ТЕМА 4. ЧИННИКИ ТА ФУНКЦІЙ ВІДГУКУ В ДОСЛІДЖЕННІ ІНФОРМАЦІЙНО-КОМП'ЮТЕРНИХ СИСТЕМ.

План.

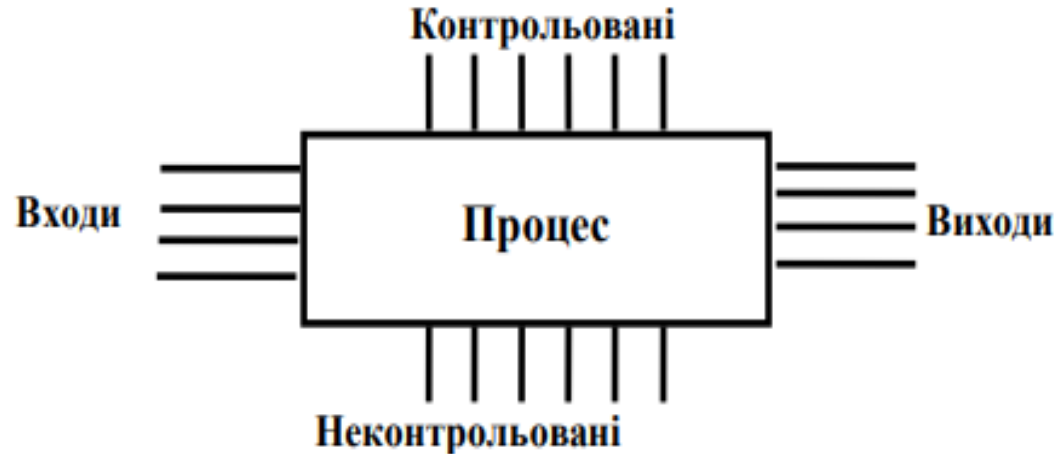
1. Планування експерименту
2. Залежні змінні (відгуки)
3. Незалежні змінні (фактори)
4. Вибір рівнів факторів

викладач: Любименко Олена Миколаївна
olena.liubymenko@donntu.edu.ua

4.1 ПЛАНУВАННЯ ЕКСПЕРИМЕНТУ

Планування експерименту включає наступні пункти [1] :

1. Збір і аналіз апріорної інформації, обґрунтування, розуміння факту необхідності експерименту, його матеріальне забезпечення [1];
2. Вибір факторів і відгуку або відгуків [1];
(**вибір факторів** - це встановлення основних і другорядних параметрів, що впливають на досліджуваний процес [1].
відгуки - це реакції на дію - факторів (вхідних змінних) вихідних характеристик досліджуваного явища або процесу) [1].
3. Вибір плану (побудова матриці планування, число реплік, спосіб пошуку екстремуму або рандомізації, вибір методики обробки результатів та аналіз даних експерименту; формулювання висновків і рекомендацій) [1];
4. Власне експеримент



В концепції чорного ящика постулюється, що всі моделі, які з однаковою точністю описують зв'язок входи-виходи ізоморфні.

- Входи – це ті фактори зовнішнього середовища, як і ми можемо змінювати за своїм розсудом. До них часто застосовують термін незалежні змінні [1].
- Виходи – (залежні змінні, відгуки) фактори дії системи на зовнішнє середовище [1]
- Контрольовані некеровані фактори – фактори, які ми змінювати не можемо, але значення яких доступні для контролю (вимірювання) [1].
- Неконтрольовані некеровані фактори – фактори, які ми не можемо ні змінювати, ні контролювати [1].

4.2 ЗАЛЕЖНІ ЗМІННІ (ВІДГУКИ):

- 1. Мати фізичну сутність і достатньо повно з точки зору поставленої прикладної мети характеризувати процес [1].
- 2. Бути відтворюваними, тобто при повторенні експериментів в номінально однакових умовах результати повинні співпадати з точністю до випадкової помилки [1].
- 3. Кожному набору вхідних змінних повинно відповідати одне (з точністю до випадкової помилки) значення відгуку [1].
- 4. Повинна вимірюватися не в номінальній шкалі (бажано) [1].

4.3 НЕЗАЛЕЖНІ ЗМІННІ (ФАКТОРИ) :

- 1. Бути керованими, тобто ми повинні мати можливість встановлювати необхідні їх значення [1];
- 2. не повинні залежати від інших змінних;
- 3. повинні бути детермінованими величинами [1];
- 4. повинні бути однозначні: одному значенню змінної (при інших однакових умовах) повинно відповідати одне (з точністю до випадкової помилки) значення відгуку; Вхід Процес и Виходи Контрольовані некеровані фактори Неконтрольовані некеровані [1]/
- 5. множина незалежних змінних повинна бути повна з точки зору прикладної мети дослідження. Тобто, вибраних факторів (можливо з урахуванням контрольованих некерованих) повинно бути достатньо для опису поведінки системи [1].

При наявності кількох відгуків множини незалежних змінних можуть співпадати не повністю. Це повинно бути відображено у відповідних описах [1].

- *Контрольовані некеровані змінні.* Враховуються в тому випадку, коли вони суттєво впливають на відгук. Якщо ми збираємося їх враховувати, то необхідна така ж інформація, як і для незалежних змінних [1].
- *Зафіксовані незалежні змінні.* До них відносяться ті змінні, які суттєво впливають на відгук, але в даному дослідженні змінюватись не будуть. Обов'язково потрібен їх перелік, опис і обґрунтування фіксації, виходячи з прикладної мети дослідження [1].

4.4 ВИБІР РІВНІВ ФАКТОРІВ

- Фактор може мати безперервний або дискретний характер зміни [1].
- Планування експерименту починають з вибору центру плану, тобто точки, відповідної початковому значенню всіх використовуваних в експерименті факторів ($X_{10}, X_{20}, \dots, X_{k0}$), в околі якої в подальшому ставиться серія планованих дослідів [1].

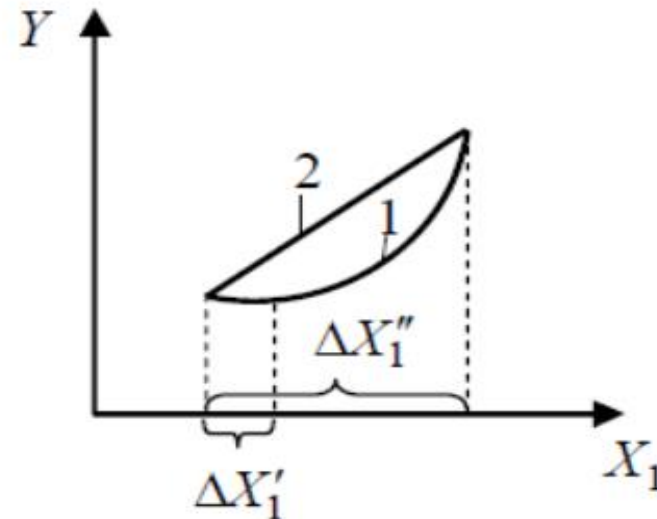


Рис. 2. Вид досліджуваної функції (1) і два варіанти кроку експерименту: $\Delta X_1'$ - занижена величина; $\Delta X_1''$ - завищена величина.

Для зручності обробки результатів дослідів, проводиться перетворення значень керованих змінних (тобто факторів X_i , які враховуються в експерименті) до безрозмірним величинам (x_{ijb})

$$x_{ijb} = \frac{X_{ij} - X_{i0}}{\Delta X_i} \quad (.1)$$

де X_{i0} — базове або початкове значення i -го фактора в центрі плану; ΔX_i — значення інтервалу варіювання i -го фактора, X_{ij} — поточний j -ве значення i -го фактора.

- Формалізація мети планування виражається у вигляді деякої функції, яку називають цільовою функцією.

Розробку моделі процесу слід проводити за принципом «від простого до складного». Відповідно до цього принципу починають з пропозиції, що модель досліджуваного процесу є лінійно. і має вигляд полінома 1-го порядку. При цьому необхідно враховується вплив на функцію відгуку не тільки кожного розглянутого в експерименті фактора окремо, але і їх взаємодію

$$Y = b_0 + \sum_{i=1}^k b_i X_i + \sum_{i \neq j} b_{ij} X_i X_j \dots \quad (.2)$$

У цьому рівнянні враховані лише парні взаємодії $X_i X_j$.

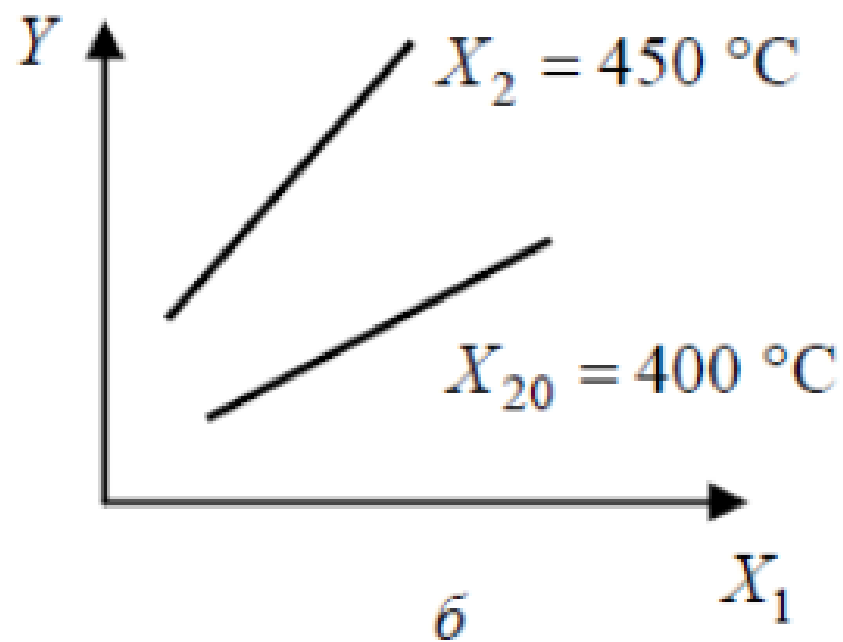
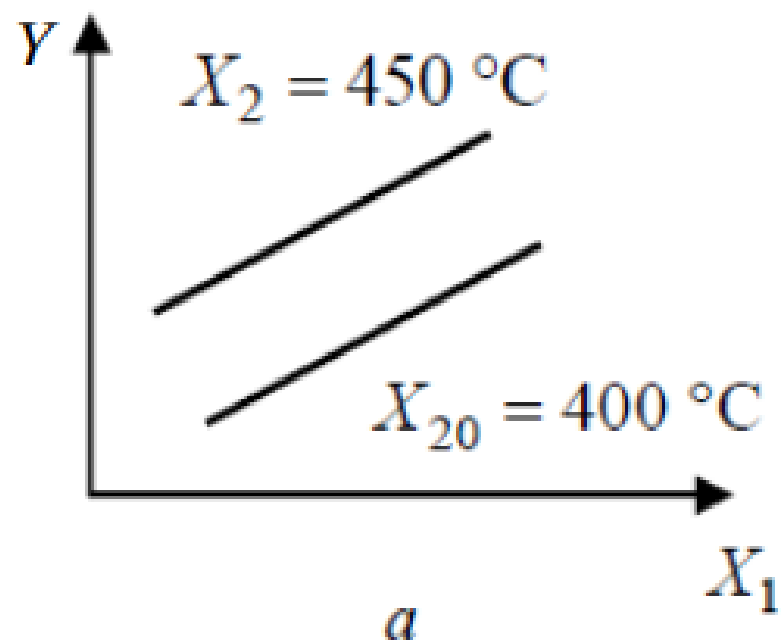


Рис. 3. Приклад відсутності взаємодії факторів X_1 і X_2 (а)
і наявності взаємодії факторів X_1 і X_2 (б).

Контрольні питання

1. Що таке планування експерименту?
2. Опишіть залежності змінні (відгуки)?
3. Наведіть незалежні змінні (фактори)?
4. Опишіть вибір рівнів факторів?

ТЕМА 5. ПОВНИЙ ФАКТОРНИЙ ЕКСПЕРИМЕНТ

План.

1. Опис, визначення повний факторний експеримент
2. Матриці планування експерименту
3. Розрахунок коефіцієнтів рівняння регресії
4. Визначення статистичної значущості коефіцієнтів рівняння регресії

викладач: Любименко Олена Миколаївна
olena.liubymenko@donntu.edu.ua

5.1 Опис, визначення повний факторний експеримент

За двофакторного експерименту безліч значень унімодальної функції оптимізації, які вона приймає при зміні змінних факторів, можна представити у вигляді поверхні з однією згладженою вершиною (рис.). Таку поверхню, що відповідає виходу досліджуваного процесу, називають *поверхнею відгуку*, а точку з максимальним значенням однокстремальної функції — *оптимумом*.

Передбачається, що існує деякий аналітичний зв'язок між факторами й відгуком процесу. При плануванні прагнуть визначити цю залежність, тобто побудувати математичну модель процесу. Математично завдання планування експерименту полягає в тому, щоб знайти рівняння поверхні відгуку:

$$\eta = \varphi(x_1, x_2, \dots, x_n),$$

де η - вихід процесу, тобто параметр оптимізації; x_i - фактори, які варіюють при проведенні експерименту.

Для більшості реальних завдань вид поверхні відгуку заздалегідь невідомий, тому при експериментальному пошуку оптимальних умов функцію η представляють у вигляді статичного ряду:

$$\eta = \beta_0 + \sum \beta_i x_i + \sum \beta_{ij} x_i x_j + \sum \beta_{ijk} x_i x_j x_k + \dots$$

1

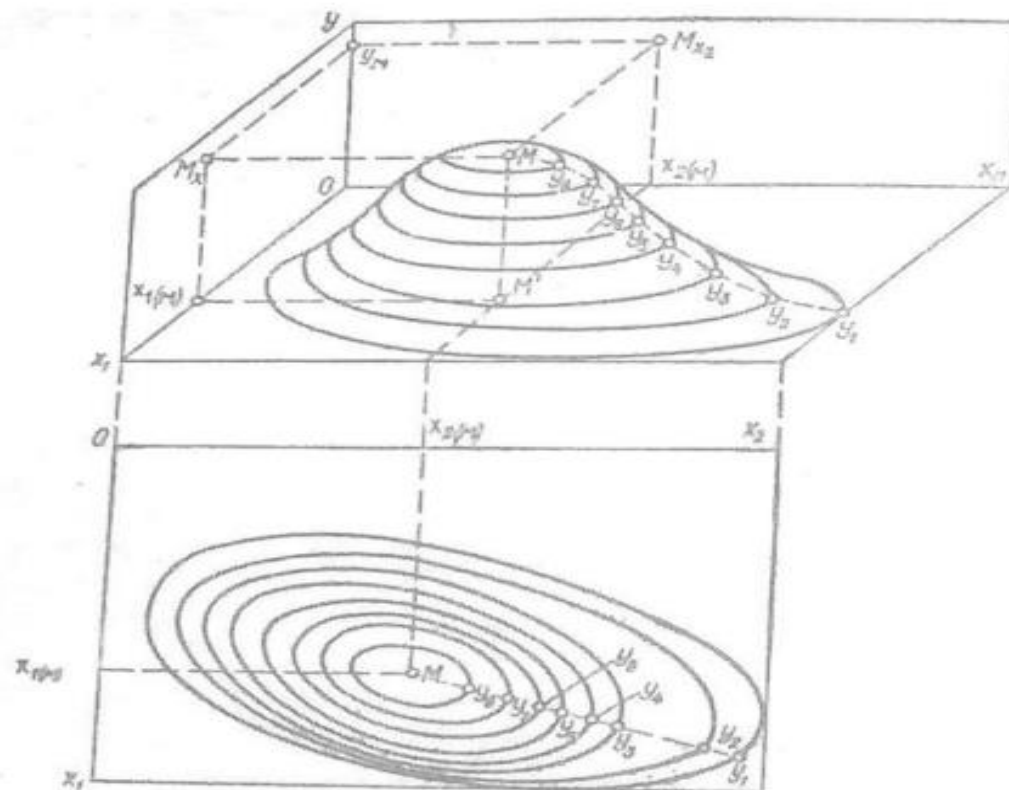


Рис. 1. - Геометричне зображення поверхні відгуку у двухфакторному експерименті: $v_{3,5}$ - ізолінії рівного виходу функції $v = f(x_1, x_2)$;

М - точка оптимуму

Повний факторний експеримент (ПФЕ)

застосовується в тих випадках, коли є припущення, що технологічний процес у досліджуваній галузі визначення факторів може бути досить точно описаний лінійними рівняннями виду [1].

рівняння регресії, одержуване на підставі результатів експериментів, має вигляд:

$$y = b_0 + \sum b_i x_i + \sum b_{ij} x_i x_j + \sum b_{ijk} x_i x_j x_k + \dots \quad 2$$

де y - вибіркова оцінка функції відгуку.

$b_0, b_0, b_0, b_i, b_{ij}, b_{123}$ - коефіцієнти;

x_i, x_j, x_1, x_2, x_3 - незалежні змінні (фактори), які можна варіювати при постановці експерименту.

Знаходження моделі методом ПФЕ складаються з наступних етапів:

- 1) планування експерименту [1];
- 2) власна експеримент [1];
- 3) перевірка відтворюваності (однорідності вибірових дисперсій) [1];
- 4) одержання математичної моделі об'єкта з перевіркою статистичної значимості вибірових коефіцієнтів регресії [1];
- 5) перевірка адекватності математичного опису [1].

ПФЕ- це експеримент, в якому реалізують всі можливі поєднання аналізованих рівнів варіювання факторів [1].

Причому фактори варіюють на двох рівнях: min -1, max +1. [1]

Число дослідів складає [1]

$$N = 2^k n$$

де k - кількість факторів;

n — число повторень експерименту.

Побудова плану повного факторного експерименту для знаходження лінійної моделі.

а. Нехай $n=1$ у кожній точці плану, а $k=2$, тоді

$N=2^2 \cdot 1=4$ дослідів достатньо, щоб отримати лінійну модель:

$$y = b_0 + b_1x_1 + b_2x_2,$$

Якщо $n=2$ у кожній точці плану, тоді $N=2^2 \cdot 2=8$.

Для $k=3$, $n=1$: $N=2^3 \cdot 1=8$ дослідів, щоб отримати лінійну модель:

$$y = b_0 + b_1x_1 + b_2x_2 + b_3x_3.$$

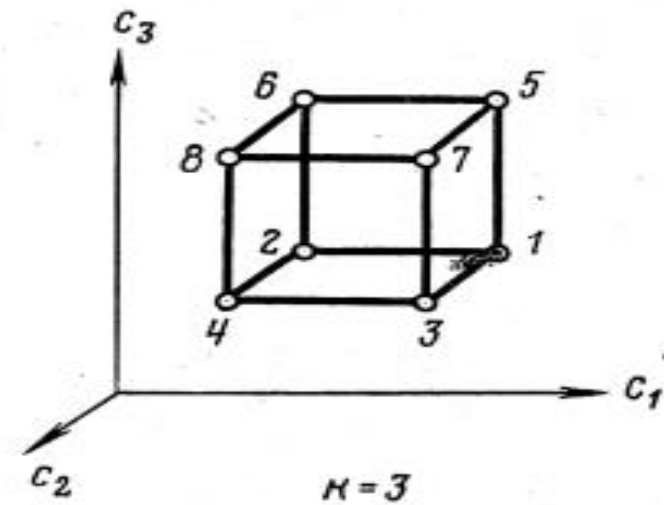
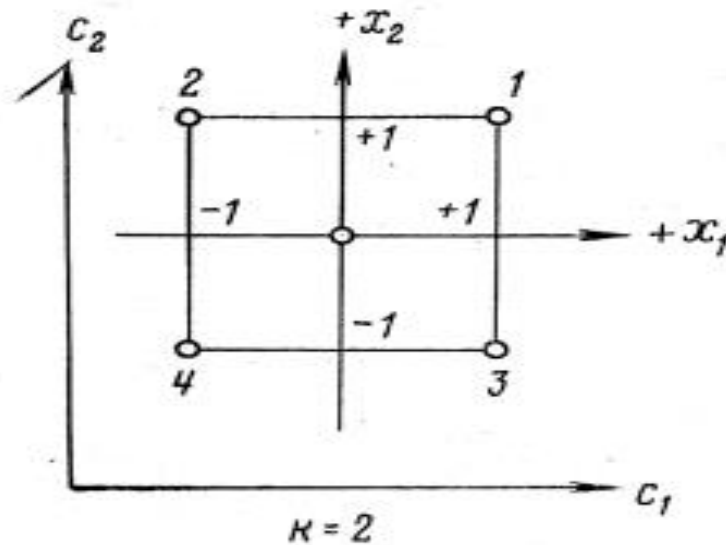
Якщо $n=3$ у кожній точці плану, тоді $N=2^3 \cdot 3=24$.

- б. Планування ведеться на 2-х рівнях: верхній (+1) та нижній (-1). Для розрахунку дисперсії відновлення досліди можна повторити в нульовій точці (центрі плану) або у кожній точці плану [1].
- в. факторний простір при $k=2$ (а), при $k=3$ (б) має вигляд (рис. 2) [1].
- г. Тоді план 1-го порядку буде мати вигляд для $k=2$ (табл.)

5.2. МАТРИЦІ ПЛАНУВАННЯ ЕКСПЕРИМЕНТУ

Матриця планування експерименту 2^2

Номер досліду	X_1	X_2	y
1	+1	+1	y_1
2	-1	+1	y_2
3	+1	-1	y_3
4	-1	-1	y_4



Розташування точок плану 1-го порядку у факторному просторі при $k=2$ (а) та $k=3$ (б).

Матриця планування експерименту 2^3

Номер дослід	X ₁	X ₂	X ₃	y
1	+	+	+	y ₁
2	-	+	+	y ₂
3	+	-	+	y ₃
4	-	-	+	y ₄
5	+	+	-	y ₅
6	-	+	-	y ₆
7	+	-	-	y ₇
8	-	-	-	y ₈

Матриця експерименту складається в такий спосіб, щоб реалізувалися всілякі комбінації значень чинників [1].

При складанні матриці плану зазвичай спочатку використовують кодовані значення факторів, а потім переходять до натуральних значень факторів [1].

$$x_i = \frac{C_i - C_{oi}}{Z} \qquad C_i = C_{oi} - x_i Z$$

де C_{oi} і C_i – натуральні значення факторів на НУ та поточне значення відповідно;

x_i - Кодоване значення фактора (-1; 0; 1);

Z – інтервал варіювання у натуральному вираженні чинника, тобто.

5.3 РОЗРАХУНОК КОЕФІЦІЄНТІВ РІВНЯННЯ РЕГРЕСІЇ

Побудувавши матрицю планування, здійснюють експеримент. Отримавши експериментальні дані, розраховують значення коефіцієнтів регресії [1].

Таким чином, коефіцієнт b_0 (а) розраховується усереднення всіх значень показника якості [1] .

Значення вільного члена (b_0) беруть як середнє арифметичне всіх значень параметра оптимізації в матриці [1] :

$$a) \quad b_0 = \frac{\sum_{i=1}^N y_i}{N} ,$$

Де y_i - значення параметра оптимізації в i -му досвіді; N – кількість дослідів у матриці.

Лінійні коефіцієнти (θ_i) регресії розраховують за формулою

$$\text{б)} \quad \theta_i = \frac{\sum_{u=1}^N x_{iu} y_u}{N},$$

Де x_{iu} —кодоване значення фактора x_i у-му досвіді.

Коефіцієнти b_i (б) – це добуток значень фактора на значення показників якості, а також усереднений поділ на кількість рядків плану експерименту N [1].

в) Відповідно коефіцієнти при парних взаємодіях [1]

b_{ij} (в) – це усереднені значення впливу факторів та показників якості.

Коефіцієнти регресії, що характеризують парну взаємодію факторів, знаходять за формулою [1]:

$$b_{ij} = \frac{\sum_{u=1}^N x_{iu} x_{ju} y_u}{N}$$

При розрахунку коефіцієнтів у електронних таблицях Excel використовується функція *Лінійн* [1].

Розрахунок коефіцієнтів рівняння регресії можна як для кодованих значень чинника, так значень у натуральному вираженні, тобто. у наведених формулах *а, б, в* в якості x_{iu} та x_{ju} можна підставляти або -1 та +1, або натуральні значення факторів [1].

При цьому якщо розрахунок виконаний для *кодованих значень факторів*, то за величиною та знаком коефіцієнта можна судити про ступінь і характер впливу кожного фактора на показник якості [1].

5.4 ВИЗНАЧЕННЯ СТАТИСТИЧНОЇ ЗНАЧУЩОСТІ КОЕФІЦІЄНТІВ РІВНЯННЯ РЕГРЕСІЇ

Методика оцінки значущості полягає в наступному: коефіцієнт рівняння регресії значимо, якщо його абсолютна величина перевищує половину довірчого інтервалу, тобто, більше похибки щодо оцінки значення критерію оптимізації (показника якості) [1].

Розмір довірчого інтервалу (похибка в оцінці показника) дорівнює [1] :

$$\Delta b_i = \frac{t_{\alpha n} S_{\{y\}}}{\sqrt{Nn}}$$

де $t_{\alpha n}$ – табличне значення критерію Стюдента;

N – число рядків плану експерименту ;

n – кількість паралельних дослідів відповідно;

$S_{\{y\}}$ – стандартне відхилення відтворюваності.

$$S_{\{y\}} = \sqrt{S_y^2}$$

Приклад:

При реалізації повного факторного експерименту: $N=8$; $k=3$; $n=3$ було отримано рівняння регресії

$$y = 85,98 + 2,6 x_1 + 0,54 x_2 + 1,13 x_3 - 0,22 x_1 x_2 + 0,96 x_1 x_3 - 0,31 x_2 x_3$$

Дисперсія відтворюваності складає 0,65.

Оцінимо важливість коефіцієнтів.

Розв'язання

Визначаємо значення критерію Стюдента за довірчої ймовірності $\alpha=0,95$.

Визначаємо значення числа ступенів свободи

$$f = N(n-1) = 8(3-1) = 16$$

Для $\alpha = 0,95$ та $f = 16$ табличне значення критерію Стюдента дорівнює $t_{\alpha n} = 2,16$. Тоді похибка в оцінці показника якості становитиме:

$$\Delta b_i = \frac{t_{\alpha n} S_{\{y\}}}{\sqrt{Nn}} = \pm \frac{2,16 \cdot \sqrt{0,65}}{\sqrt{8 \cdot 3}} = \pm 0,355$$

Контрольні питання

1. Дайте визначення ПФЕ?
2. Опишіть етапи з яких складаються побудова моделі методом ПФЕ.
3. Який є метод для оцінки параметрів статистичної регресії. Та опишіть його.
4. Яким чином можна перевірити адекватність побудованої моделі?
5. Як визначити коефіцієнт регресії?

ТЕМА 6. ДРОБОВИЙ ФАКТОРНИЙ ЕКСПЕРИМЕНТ

План.

1. Опис, визначення дробового факторного експерименту
2. Розрахунок значення коефіцієнтів регресії

викладач: Любименко Олена Миколаївна
olena.liubymenko@donntu.edu.ua

6.1. ОПИС, ВИЗНАЧЕННЯ ДРОБОВОГО ФАКТОРНОГО ЕКСПЕРИМЕНТУ

При багатофакторному експерименті, особливо коли число факторів більше шести ($n > 6$), кількість дослідів планів ПФЕ 2^n ($N = 2^n$) стає надмірною. Переходять до *дробового факторного експерименту* (ДФЕ) - частини повного факторного експерименту (ПФЕ), або **дробовим реплікам** [1].

ДФЕ може містити половину, чверть дослідів від ПФЕ. До матриці ДФЕ висувають ті ж вимоги, що і до матриці ПФЕ.

Плани ДФЕ 2^{n-k} , де k - показник дрібності плану ПФЕ. При $k=1$ кількість дослідів у плані ДФЕ вдвічі менша ніж у плані ПФЕ, тому такі плани називають напівреплікою плану ПФЕ, при $k = 2$ – чвертьреплікою плану ПФЕ [1].

Кількість дослідів для $k \geq 3$ можна зменшити, ввівши означення змішаних ефектів.

Так для $k=3$ за планом ПФЕ кількість дослідів дорівнює при $n=1$:

$$N = p^k \cdot n = 2^3 \cdot 1 = 8$$

Перехід до напіврепліки ($1/2$) від ПФЕ приведе до плану (табл. 1) і кількості дослідів при $n=1$:

$$N = p^{k-q} \cdot n = 2^{3-1} \cdot 1 = 4$$

де $q=1$ – кількість лінійних факторів, змішаних з ефектами взаємодії.

Тобто замість $N=8$ дослідів (або $N=16$ при $n=2$) використано $N=4$ дослідів (або $N=8$ при $n=2$).

Таблиця 1 Напіврепліка ($\frac{1}{2}$) від плану ПФЕ для $k=3$
з генеруючим співвідношенням $x_3=x_1x_2$

i	x_0	x_1	x_2	$x_3=x_1x_2$	y_i
1	+1	+1	+1	+1	y_1
2	+1	-1	+1	-1	y_2
3	+1	+1	-1	-1	y_3
4	+1	-1	-1	+1	y_4

Після реалізації плану ДФЕ (табл. 10) отримаємо модель:

$$y = b_0 + b_1x_1 + b_2x_2 + b_3x_3$$

План напіврепліки ($\frac{1}{2}$) від ПФЕ для $k=3$ характеризується параметрами:

- генеруючим співвідношенням

$$x_3=x_1x_2$$

- визначальним контрастом

$$x_3^2=x_1x_2x_3: 1=x_1x_2x_3, \text{ або } J=x_1x_2x_3.$$

А напіврепліка ($1/2$) (табл. 11) від ПФЕ для $k=3$ характеризується:

- генеруючим співвідношенням

$$x_3 = -x_1x_2$$

- визначальним контрастом

$$x_3^2 = -x_1x_2x_3: I = -x_1x_2x_3, \text{ або } J = -x_1x_2x_3$$

Таблиця 11 Напіврепліка ($1/2$) від плану ПФЕ для $k=3$
з генеруючим співвідношенням $x_3 = -x_1x_2$

i	x_0	x_1	x_2	$x_3 = -x_1x_2$	y_i
1	+1	+1	+1	-1	y_1
2	+1	-1	+1	+1	y_2
3	+1	+1	-1	+1	y_3
4	+1	-1	-1	-1	y_4

Для виявлення змішаних ефектів необхідно виконати операцію множення на визначальному контрасті плану – ДФЕ (табл. 1):

$$\left. \begin{aligned} 1 &= x_1 x_2 x_3 \\ x_1 &= x_1^2 x_2 x_3 \Rightarrow x_1 = x_2 x_3 \\ x_2 &= x_1 x_2^2 x_3 \Rightarrow x_2 = x_1 x_3 \\ x_3 &= x_1 x_2 x_3^2 \Rightarrow x_3 = x_1 x_2 \end{aligned} \right\}$$

Тоді, статистичні оцінки генеральних коефіцієнтів вибірковими коефіцієнтами будуть такі:

$$\left. \begin{aligned} b_0 &\rightarrow \beta_0 \\ b_1 &\rightarrow \beta_1 + \beta_{23} \\ b_2 &\rightarrow \beta_2 + \beta_{13} \\ b_3 &\rightarrow \beta_3 + \beta_{12} \end{aligned} \right\}$$

Для $k=4$ можливо створення напіврепліки ($1/2$) від ПФЕ:

ПФЕ ($n=1$) $N=p^k n = 2^4 \cdot 1 = 16$ із змішаним коефіцієнтом $q=1$:

ДФЕ (n=1) ($\frac{1}{2}$ репліка) $N=p^{k-q} \cdot n = 2^{4-1} \cdot 1 = 8$ (табл. 3).

Таблиця 3.

Дробовий факторний експеримент [напіврепліка ($\frac{1}{2}$)] від повного факторного експерименту (k=4) з q=1 змішаним ефектом (з генеруючим співвідношенням $x_4=x_1x_2x_3$)

i	x_0	x_1	x_2	x_3	$x_4=x_1x_2x_3$	y_i
1	+1	+1	+1	+1	+1	y_1
2	+1	-1	+1	+1	-1	y_2
3	+1	+1	-1	+1	-1	y_3
4	+1	-1	-1	+1	+1	y_4
5	+1	+1	+1	-1	-1	y_5
6	+1	-1	+1	-1	+1	y_6
7	+1	+1	-1	-1	+1	y_7
8	+1	-1	-1	-1	-1	y_8

Напіврепліка ($\frac{1}{2}$) (табл. 3) характеризується параметрами:

- генеруюче співвідношення

$$x_4=x_1x_2x_3$$

- визначальний контраст

$$x_4^2=x_1x_2x_3x_4 \Rightarrow 1=x_1x_2x_3x_4$$

- статистичні оцінки:

$$\left. \begin{array}{l} x_1=x_2x_3x_4 \\ x_2=x_1x_3x_4 \\ x_3=x_1x_2x_4 \\ x_4=x_1x_2x_3 \end{array} \right\} \begin{array}{l} b_0 \rightarrow \beta_0 \\ b_1 \rightarrow \beta_1 + \beta_{234} \\ b_2 \rightarrow \beta_2 + \beta_{134} \\ b_3 \rightarrow \beta_3 + \beta_{124} \\ b_4 \rightarrow \beta_4 + \beta_{123} \end{array}$$

Дробова репліка високого ступеня дрібности (1/16 репліки від ПФЕ з $k=8$) зі $q=4$ лінійними ефектами, змішаними з ефектами взаємодії (табл. 4), має ($n=1$) таку кількість дослідів:

$$\text{ПФЕ } N=p^k n = 2^8 \cdot 1 = 256$$

$$\text{ДФЕ (1/16 репліки) } N=p^{k-q} n = 2^{8-4} \cdot 1 = 16$$

У цьому випадку:

- частинні генеруючі співвідношення:

$$\left. \begin{aligned} x_5 &= x_1 x_2 x_3 x_4 \\ x_6 &= x_1 x_2 x_3 \\ x_7 &= x_1 x_2 x_4 \\ x_8 &= x_1 x_3 x_4 \end{aligned} \right\}$$

- частинні визначальні контрасти:

$$\left. \begin{aligned} 1 &= x_1 x_2 x_3 x_4 x_5 \\ 1 &= x_1 x_2 x_3 x_6 \\ 1 &= x_1 x_2 x_4 x_7 \\ 1 &= x_1 x_3 x_4 x_8 \end{aligned} \right\}$$

Таблиця 4. 1/16 репліка від ПФЕ ($k=8$; $q=4$; $n=1$)

i	x_0	x_1	x_2	x_3	x_4	x_5	x_6	x_7	x_8	y_i
1	+1	+1	+1	+1	+1	+1	+1	+1	+1	y_1
2	+1	+1	+1	-1	+1	-1	-1	+1	-1	y_2
3	+1	+1	-1	+1	+1	-1	-1	-1	+1	y_3
4	+1	+1	-1	-1	+1	+1	+1	-1	-1	y_4
5	+1	-1	+1	+1	+1	-1	-1	-1	-1	y_5
6	+1	-1	+1	-1	+1	+1	+1	-1	+1	y_6
7	+1	-1	-1	+1	+1	+1	+1	+1	-1	y_7
8	+1	-1	-1	-1	+1	-1	-1	+1	+1	y_8
9	+1	+1	+1	+1	-1	-1	+1	-1	-1	y_9
10	+1	-1	-1	-1	-1	+1	-1	-1	-1	y_{10}
11	+1	-1	+1	+1	-1	+1	-1	+1	+1	y_{11}
12	+1	+1	-1	-1	-1	-1	+1	+1	+1	y_{12}
13	+1	+1	-1	+1	-1	+1	-1	+1	-1	y_{13}
14	+1	-1	+1	-1	-1	-1	+1	+1	-1	y_{14}
15	+1	+1	+1	-1	-1	+1	-1	-1	+1	y_{15}
16	+1	-1	-1	+1	-1	-1	+1	-1	+1	y_{16}

За операцією множення двох, трьох та чотирьох співвідношень отримаємо узагальнений визначальний контраст:

$$\begin{aligned}
 1 &= x_1 x_2 x_3 x_4 x_5 = x_1 x_2 x_3 x_6 = x_1 x_2 x_4 x_7 = x_1 x_3 x_4 x_8 = x_4 x_5 x_6 = x_3 x_5 x_7 = x_2 x_5 x_8 = \\
 &= x_3 x_4 x_6 x_7 = x_2 x_4 x_6 x_8 = x_2 x_3 x_7 x_8 = x_1 x_2 x_5 x_6 x_7 = x_1 x_3 x_5 x_6 x_8 = x_1 x_6 x_7 x_8 = \\
 &= x_1 x_4 x_5 x_7 x_8 = x_2 x_3 x_4 x_5 x_6 x_7 x_8
 \end{aligned}$$

Оцінки за :

$$b_0 \rightarrow \beta_0$$

$$\begin{aligned}
 b_1 &\rightarrow \beta_1 + \beta_{2345} + \beta_{236} + \beta_{247} + \beta_{348} + \beta_{1456} + \beta_{1357} + \beta_{1258} + \beta_{13467} + \beta_{12468} + \\
 &+ \beta_{12378} + \beta_{2567} + \beta_{3568} + \beta_{678} + \beta_{4578} + \beta_{12345678}
 \end{aligned}$$

$$b_2 \rightarrow \beta_2 + \beta_{1345} + \dots\dots\dots$$

6.2 Розрахунок значення коефіцієнтів регресії

Розрахуємо значення коефіцієнтів регресії з використанням рівняння [1] :

$$b_j = \frac{1}{N} \sum_{i=1}^N x_{ji} y_i, \quad j = 0, 1, \dots, k$$

Оцінка дисперсії похибки коефіцієнтів регресії дорівнює: $s_{bi}^2 = s_{\text{восп}}^2 / N$

Значимість коефіцієнтів регресії перевіряємо з використанням критерію Стюдента:

$$t_i = \frac{|b_i|}{\sqrt{s_{bi}^2}}.$$

Контрольні питання

1. У чому суть дробового факторного експерименту?
2. Як будуватися матриця планування ДФЕ у кодованих змінних та фізичних змінних?
3. Що таке генеруюче співвідношення?
4. Як формується узагальнений визначальний контраст?
5. Як утворюється система спільних оцінок коефіцієнтів регресії?
6. У чому різниця планів ПФЕ та ДФЕ?

ТЕМА 7. ОБРОБКА РЕЗУЛЬТАТІВ КОМП'ЮТЕРНИХ ЕКСПЕРИМЕНТІВ, ОЦІНКА ЇХ ТОЧНОСТІ ТА ДОСТОВІРНОСТІ

План.

1. Визначення статистичної значущості коефіцієнтів математичної моделі
2. Розрахунок дисперсії адекватності
3. Розрахунок дисперсії відновлення
4. Перевірка адекватності моделі

викладач: Любименко Олена Миколаївна
olena.liubymenko@donntu.edu.ua

7.1 ВИЗНАЧЕННЯ СТАТИСТИЧНОЇ ЗНАЧУЩОСТІ КОЕФІЦІЄНТІВ МАТЕМАТИЧНОЇ МОДЕЛІ

а) за допомогою довірчого інтервалу:

- якщо $|b_j| \geq \Delta b_j$,

то коефіцієнт значущий порівняно з помилкою експерименту;

- якщо $|b_j| < \Delta b_j$,

то коефіцієнт незначущий і його можна викинути з рівняння.

Довірча ймовірність того, що генеральний коефіцієнт моделі попаде у довірчий інтервал $b_j \pm \Delta b_j$ становить:

$$p(b_j - \Delta b_j < \beta_j < b_j + \Delta b_j) = 1 - \alpha,$$

де довірчий інтервал розраховують за:

$$\Delta b_j = \frac{S_{\{y\}} t_T \{\alpha; f\}}{\sqrt{\sum_{i=1}^N n_i}}$$

де $\sum_{i=1}^N n_i = Nn$, якщо $n_1 = n_2 = \dots = n_i = \dots = n_N = n = \text{const}$;

$\sum_{i=1}^N n_i = N$, якщо $n_i = 1 = \text{const}$;

$S_{\{y\}}$ – середнє квадратичне значення помилки всього експерименту:

$$S_{\{y\}} = \sqrt{S_{\{y\}}^2}$$

α - рівень значущості: $\alpha=1-p$;

f – число ступенів вільностей;

t_T – табличне значення критерію Стюдента;

n_i – число повторних дослідів в i -точці плану.

б) за критерієм Стюдента, розраховуючи

$$|t_{jp}| = \frac{|b_j|}{S_{\{b_j\}}}.$$

Якщо $|t_{jp}| \geq t_T\{\alpha; f\}$, то коефіцієнт значущий,

а якщо $|t_{jp}| < t_T\{\alpha; f\}$ - незначущий,

де $S_{\{b_j\}}$ - середнє квадратичне відхилення при визначенні j-коефіцієнта.

Дисперсія помилки у визначенні коефіцієнта моделі:

$$S_{\{b_j\}}^2 = \frac{S_{\{y\}}^2}{Nn}$$

під час рівномірного дубляжу дослідів $n_i = n = \text{const}$;

$$S_{\{b_j\}}^2 = \frac{S_{\{y\}}^2}{\sum_{i=1}^N n_i}$$

під час нерівномірного дубляжу дослідів $n_i = \text{var}$;

$$\text{та } S_{\{b_j\}}^2 = \frac{S_{\{y\}}^2}{N},$$

якщо $n_i = n = 1$;

$S_{\{y\}}^2$ - дисперсія відновлення (помилки всього експерименту);

$$S_{\{b_j\}} = +\sqrt{S_{\{b_j\}}^2}$$

Довірча ймовірність того, що генеральний коефіцієнт моделі попаде в довірчий інтервал $(b_j \pm \Delta b_j)$ становить:

$$P(b_j - S_{\{b_j\}} t_T < \beta_j < b_j + S_{\{b_j\}} t_T) = 1 - \alpha,$$

Якщо N (або Nn) $\geq 10-20$, то можна прийняти, що $t_T\{\alpha; f\} \approx 2$, а $\Delta b_j = 2S_{\{b_j\}}$.

7.2 Розрахунок дисперсії адекватності

$$S_{ад}^2 = \frac{SS_{ад}}{f_{ад}}$$

1. Число рівнів вільностей:

а) без повторних дослідів $n_0=1$ $f_{ад}=N-\lambda$,

де λ - число коефіцієнтів моделі, включаючи і b_0 ;

2. Дисперсія адекватности:

а) $n=1$ (без повторних дослідів):

$$S_{ад}^2 = \frac{SS_{ад}}{f_{ад}} = \frac{\sum_{i=1}^N (y_i - \hat{y}_i)^2}{N - \lambda},$$

де y_i – результат експерименту в i -точці;

$\hat{y}_i$ - результат розрахунку у за моделлю в i -й точці.

б) Дублювання за всіма точками плану експерименту (окрім нульової точки):

- $n_i = \text{const} = n$ (рівномірне дублювання):

$$S_{\text{ад}}^2 = \frac{\sum_{i=1}^N (\bar{y}_i - \hat{y}_i)^2}{N - \lambda}$$

де $\bar{y}_i$ - середнє значення в i -й точці плану $\bar{y}_i = \frac{\sum_{u=1}^n y_{iu}}{n}$.

- $n_i \neq \text{const} = \text{var}$ (нерівномірне дублювання):

$$S_{\text{ад}}^2 = \frac{\sum_{i=1}^N \sum_{u=1}^n (\bar{y}_i - \hat{y}_i)^2}{N - \lambda}$$

в) повторні досліді лише в нульовій або іншій точці плану:

$$S_{ад}^2 = \frac{\sum_{i=1}^N (\bar{y}_i - \hat{y}_i)^2 - \sum_{i_0=1}^{n_0} (y_{0i} - \bar{y}_0)^2}{N - \lambda - (n_0 - 1)}$$

y_{0i} - результат в i -й нульовій точці;

$$\bar{y}_0 = \frac{\sum_{i=1}^{n_0} y_{0i}}{n_0} - \text{середня за нульовими точками};$$

n_0 – кількість повторних дослідів у нульовій точці плану.

г) Дублювання за всіма точками плану, включаючи і нульову точку:

- $n_i = \text{const} = n$:

$$S_{\text{ад}}^2 = \frac{\sum_{i=1}^N n(\bar{y}_i - \hat{y}_i)^2}{N - \frac{(k+2)(k+1)}{2} - (n_0 - 1)} \quad \text{- нелінійна модель}$$

$$S_{\text{ад}}^2 = \frac{\sum_{i=1}^N n(\bar{y}_i - \hat{y}_i)^2}{N - \lambda - (n_0 - 1)} \quad \text{- лінійна модель}$$

де $\bar{y}_i$ - середнє повторних дослідів в і-й точці плану;

k – кількість факторів.

- $n_i \neq \text{const} = \text{var}$:

$$S_{\text{ад}}^2 = \frac{\sum_{i=1}^N \sum_{u=1}^{n_0} (y_{iu} - \hat{y}_i)^2}{N - \lambda - (n_0 - 1)}.$$

7.3 Розрахунок дисперсії відновлення

1) Дублювання лише в нульовій або іншій точці плану:

$$S_{\varepsilon}^2 = S_{\{y\}}^2 = \frac{\sum_{i_0=1}^{n_0} (y_{i_0} - \bar{y}_0)^2}{n_0 - 1} = \frac{SS_0}{f_0}, \quad f_0 = n_0 - 1$$

2) Дублювання за всіма точками плану:

- $n_i = \text{const} = n$:

$$S_{\{y\}}^2 = \frac{\sum_{i=1}^N \sum_{u=1}^n (y_{iu} - \bar{y}_i)^2}{\sum_{i=1}^N (n_i - 1)} = \frac{SS_{\{y\}}}{N(n - 1)} = \frac{SS_{\{y\}}}{f_{\{y\}}}$$

- $n_i \neq \text{const} = \text{var}$:

$$S_{\{y\}}^2 = \frac{\sum_{i=1}^N (f_i S_i^2)}{\sum_{i=1}^N f_i}$$

де $f_i = n_i - 1$

$$S_i^2 = \frac{\sum_{u=1}^{n_i} (y_{iu} - \bar{y}_i)^2}{n_i - 1}$$

3) $n_i = 1$

$$S_{\{y\}}^2 = S_{\{\bar{y}\}}^2 = \frac{\sum_{i=1}^N (y_i - \bar{y})^2}{N - 1}, \quad f = N - 1$$

$n_i \neq 1 = \text{const}$

$$S_{\{\bar{y}\}}^2 = \frac{S_{\{y\}}^2}{n},$$

7.4 Перевірка адекватності моделі

1. Висувають нульову гіпотезу: H_0 : вибрана модель адекватна.
2. Перевірку H_0 -адекватності моделі – здійснюють за теоретичним (табличним) F_T критерієм Фішера. З цією метою розраховують статистику:

$$F_P = \frac{S_{ад}^2}{S_{\{y\}}^2} = \frac{SS_{ад} / f_{ад}}{SS_{\{y\}} / f_{\{y\}}},$$

Де дисперсія відновлення $S_{\{y\}}^2$ приймається за дисперсію помилки S_{ε}^2 (для $h=1$ приймають $S_{\{y\}}^2 = S_{(\bar{y})}^2$ - дисперсія середніх).

При порівнянні статистики та табличного значення критерію Фішера F_T можливі такі висновки відносно H_0

а) Якщо $F_P \leq F_T \{ \alpha; f_{ад}; f_{\{y\}} \}$, то H_0 приймається: модель адекватна

де α -рівень значущості прийняття рішення $\alpha=1-p$ – ймовірність правильності прийнятого рішення про адекватність моделі;

$f_{ад}$ – число ступенів вільностей при перевірці моделі на адекватність (знаменник)

$f_{\{y\}}$ – число ступенів вільностей у виразах при розрахунку дисперсії відновлення.

Ступінь адекватності моделі (для лінійної моделі – ступінь її лінійності):

$$\xi_2(F) = \frac{F_P}{F_T} \geq 1,$$

при цьому залишки ступеню неадекватності (при виборі лінійної моделі – ступінь її нелінійності):

$$\xi_2(F) = \frac{F_P}{F_T} < 1,$$

б) Якщо $F_P > F_T \{ \alpha; f_{ад}; f_{\{y\}} \}$,

то H_0 відкидається: стверджується з рівнем значущості α (ймовірності ризику прийняти невірне рівняння), що вибрана модель адекватна (статистично не відповідає отриманим експериментальним даним) зі ступенем неадекватності (для лінійної моделі ступінь її нелінійності):

$$\xi_2(F) = \frac{F_P}{F_T} > 1,$$

при цьому залишки ступеня адекватності (при виборі лінійної моделі – ступінь її лінійності):

$$\xi_2(F) = \frac{F_P}{F_T} \leq 1,$$

При цьому приймають рішення перейти до планування другого чи більш високого порядку та отримати модель другого або більш високого порядку.

Контрольні питання

1. Опишіть план знаходження відрядкової дисперсії вихідної величини.
2. Що таке критерій Стюдента і де він використовується?
3. Чим визначається F-критерій Фішера і як його застосовують?
4. Опишіть методику розрахунку дисперсії адекватності.
5. Опишіть методику розрахунку дисперсії відновлення.
6. Опишіть методику перевірки адекватності моделі.

ТЕМА 8. МЕТРИКИ ЕФЕКТИВНОСТІ WEB СЕРВІСІВ. МЕТРИКИ ЕФЕКТИВНОСТІ БАЗ ДАНИХ. ЕКСПЕРИМЕНТИ ТА ОПТИМІЗАЦІЯ WEB СЕРВІСІВ. ОСОБЛИВОСТІ ВИКОНАННЯ ЕКСПЕРИМЕНТІВ ТА ОПТИМІЗАЦІЇ (Т'ЮНІНГУ) БАЗ ДАНИХ.

План.

1. Метрики ефективності web сервісів.
2. Метрики ефективності баз даних.
3. Експерименти та оптимізація web сервісів.
4. Особливості виконання експериментів та оптимізації (т'юнінгу) баз даних

викладач: Любименко Олена Миколаївна
olena.liubymenko@donntu.edu.ua

8.1. Метрики ефективності web сервісів

- **Веб-сервіс (англ. web service)** — програмна система, що ідентифікується веб-адресою, зі стандартизованими інтерфейсами. Веб-служби можуть взаємодіяти один з одним і з іншими програмами за допомогою повідомлень, заснованих на певних протоколах. В побуті веб-сервісами називають послуги, що надаються в Інтернеті.
- **Метрика** — це міра, яка дозволяє отримати числове значення деяких властивостей ПЗ.
- **Метрика** — іншими словами це показник поточного стану досягнення цілі. Тому що метрики без цілі, заради того аби щось міряти — не мають змісту. Хоча керівництву, клієнтам є до вподоби гарна аналітика, красиві графіки, показники тощо [<https://qagroup.com.ua/publications/what-is-metrics/>].

Які головні цілі метрик?

- Зменшити кількість дефектів.
- Пришвидшити строки релізу.
- Придумати метрик можна безкінечну кількість, так само їх групувати ми можемо за різними ознаками. Найперше метрики поділяють на **Базові (Базові показники/Base Metrics)** і **Обчислені показники (Calculated metrics)**.
- **Базові показники** їх ще називають **абсолютними** — це необроблені дані, зібрані тестувальниками.
- **Обчислені показники або похідні** — виводяться з співвідношення даних, зібраних у базових показниках.

Метрики ефективності і якість роботи web сервісів.

Завдання цього набору метрик оцінити наскільки якісно web сервіси виконують свої завдання, визначити рівень компетенцій. Володіючи таким набором показників можна порівнювати роботу web сервісів у різні моменти часу або з іншими, зовнішніми web сервісами.

- **Швидкість роботи (velocity)**— розраховується як відношення реалізованих story points (або вимог, або user stories) за кілька, наприклад, 4-5 ітерацій (Sprint) до кількості обраних ітерацій.
- **Velocity = Кількість story points за N ітерацій/N**
- Призначення метрики виявити можливості роботи web сервісів для подальшого планування обсягу робіт і аналізу трендів її розвитку.
- **Ефективність тестів і тестових наборів** — цікава метрика тим що вона є спірною. Відображає як багато помилок в середньому дозволяють виявити тест-кейси. Ця метрика каже про якість тест дизайну.
- Найкраще розраховувати дану метрику для всіх наборів тестів, проте не заборонено і по окремому.
- **Кількість виявлених помилок/Кількість тест-кейсів у тестовому наборі**

8.2 Метрики ефективності баз даних.

- База даних - організована відповідно до певних правил і підтримувана в пам'яті комп'ютера сукупність даних , що характеризує актуальний стан деякої предметної області і використовується для задоволення потреб користувачів

Показники бази даних(Database Metric)- це

- метрики, що характеризують стан конкретної бази даних та її об'єктів: таблиць, індексів та ін., у тому числі:
 - Загальні метрики БД: кількість таблиць, список таблиць, кількість індексів, список індексів, кількість екстентів, кількість об'єктів, середній розмір об'єкта, обсяг даних у базі, обсяг даних на диску;
 - Метрики таблиць та індексів: кількість записів у таблиці, обсяг записів у таблиці, список полів таблиці, первинний ключ, список зовнішніх ключів, кількість зовнішніх ключів. Обсяг та тип індексу.

Класифікація баз даних

За характером інформації:

- фактографічні (одиниця зберігання - факт, тобто про один об'єкт зберігається певний факт, наприклад банківський рахунок, код товару);
- документальні (одиниця зберігання – документ, наприклад, БД «Звід законів» зберігається як сукупність документів).

За структурою даних:

- ієрархічні (побудовані структурою «дерево», т. е. складаються з різних рівнів);
- мережеві (є різновидом ієрархічної структури; об'єкти можуть мати кілька зв'язків з вищим або нижчим рівнем);
- реляційні (БД як таблиці).

За способом зберігання даних:

- централізовані (основна БД на центральному ПК, до якої користувачі звертаються зі своїх комп'ютерів);
- розподілені (елементи БД зберігаються різних ПК і пов'язані між собою).

8.3 Експерименти та оптимізація web сервісів.

Оптимізація web сервісів (Search Engine Optimization (SEO)) –це комплекс дій щодо внутрішньої та зовнішньої оптимізації, спрямованих на підвищення позицій сайту в пошуковій видачі.

Цілі SEO-оптимізації :

- збільшення трафіку ресурсу;
- виведення сайту у топ певним пошуковим запитам;
- підвищення конверсії;
- покращення позицій сайту у пошуку щодо проектів конкурентів.

Методи оптимізації

- **Біла оптимізація** — оптимізаторська робота над ресурсом без застосування офіційно заборонених кожною пошуковою системою методів розкручування ресурсу — без впливу на пошукові алгоритми сайтів.

Методи білої зовнішньої пошукової оптимізації:

- Реєстрація в самостійних каталогах (вручну або за допомогою спеціальних ресурсів);
- Реєстрація в каталогах пошукових систем, таких як Top-100, [DMOZ](#) (AOL), Апорт, [Yahoo](#) тощо;
- Соціальні мережі;
- Прес-релізи;
- Активна діяльність на тематичних блогах;
- Розміщення новин на сайтах-новинниках, таких як [BuzzFeed](#), [Reddit](#) (<https://uk.wikipedia.org/wiki>)

сіра оптимізація

- можна віднести додавання великої кількості ключових слів в текст сторінки, оптимізація полягає спочатку в підборі ключових запитів для конкретної вебсторінки, визначенні розміру цільового «SEO-тексту» і необхідної частоти ключових слів у ньому, а потім у формулюванні пропозицій і фраз, які містять кількаразове повторення ключових запитів певну кількість разів у різних відмінках, однині та множині, а також різних формах дієслів. Ці параметри можуть потім коригуватися за результатами видачі пошукових систем.
- Сіра оптимізація відрізняється від чорної тим, що вона офіційно не заборонена, але її використання все одно може бути розцінене як неприродне завищення популярності сайту.

чорна оптимізація

- Це методи, які суперечать правилам пошукових систем: використання дорвеїв (сторінок і ресурсів, створених спеціально для роботів пошукових систем, найчастіше з великою кількістю ключових слів на сторінці), прийом під назвою клоакінг (користувачеві віддається одна сторінка, що легко читається, а пошуковому роботу — інша, оптимізована під які-небудь запити), використання прихованого тексту на сторінках сайту, використання «однопіксельних посилань».
- **Чорні методи зовнішньої пошукової оптимізації**
- Обмін посиланнями ((прямий, взаємний, односторонній (купівля посилань));
- Лінкбілдінг;
- Створення та ведення блогів на інших доменах з ціллю залишити посилання
- Дорвеї
- Клоакінг

- **Google PageRank (Google PR)** — один з методів, який Google використовує для визначення релевантності або важливості сторінки. Важливі сторінки отримують більший PageRank і частіше відображаються у верхній частині результатів пошуку. Google PageRank (PR) — це міра від 0 до 10. Google Pagerank базується на зворотних посиланнях. PageRank працює, підраховуючи кількість та якість посилань на сторінку, щоб визначити приблизну оцінку того, наскільки важливий вебсайт.
- **Alexa Traffic Rank** базується на кількості трафіку, зареєстрованого користувачами, які встановили панель інструментів Alexa протягом трьох місяців. Рейтинги на сайті базуються на загальній кількості унікальних відвідувачів та переглядів сторінок. Унікальні відвідувачі визначаються кількістю унікальних користувачів Alexa, які відвідують сайт у певний день. Перегляди сторінок — це загальна кількість запитів URL-адрес користувачів Alexa для сайту.

- **Авторитетність домену (DA)**, метрика вебсайту, розроблена компанією Moz, є прогностичним показником для визначення трафіку вебсайту та рейтингу звичайних пошукових систем. Адміністрація домену базується на різних показниках посилань, таких як кількість пов'язаних кореневих доменів, кількість загальних зворотних зв'язків та відстані зворотних посилань на домашній сторінці вебсайтів.
- **Moz Page Authority**. У порівнянні з авторитетністю домену, який визначає силу рейтингу всього домену або субдомену, Page Authority визначає силу окремої сторінки. Це оцінка, розроблена Мозом у логарифмічній шкалі на 100 пунктів. На відміну від TrustFlow, влада домену не враховує спам.
- **Рейтинг домену (або DR)**. Одна з основних метрик в сервісі Ahrefs.com. Дослівно її можна перевести як «Рейтинг Домену». Багато SEO-фахівці використовують саме це значення, щоб швидко визначити «контрольну вагу» певного сайту і порівняти його з іншими. Самі автори дають цій метриці наступне визначення: «Рейтинг Домену показує загальну силу посилального профілю певного сайту».

8.4 ОСОБЛИВОСТІ ВИКОНАННЯ ЕКСПЕРИМЕНТІВ ТА ОПТИМІЗАЦІЇ (Т'ЮНІНГУ) БАЗ ДАНИХ

Оптимізація - спроба реорганізації фізичного простору, який займає таблиці та індекси бази даних.

Вона може допомогти зменшити займане місце та дещо підвищити продуктивність. Для оптимізації використовується стандартна команда **OPTIMIZE TABLE** із синтаксису MySQL. З докладним описом цієї функції можна ознайомитись у [документації](#) до MySQL.

Оптимізація баз даних - це покращення створення запитів:

- заздалегідь якомога оптимальне фільтрування даних ще до моменту з'єднання їх з іншими таблицями;
- якомога менше за обсягом даних сортувати у результуючий набір;
- по можливості уникати конструкції DISTINCT, LIKE '% ...', OUTER JOIN особливо на великих об'ємах даних;
- якщо у вибірці потрібно лише одне поле від приєднаної таблиці, то не потрібно приєднувати таблицю, а в самій вибірці зробити підзапит;
- при фільтрації, агрегації і вибірці потрібно враховувати наявні індекси, щоб оптимізатор міг ними скористатися;
- повертати тільки ті поля, які дійсно потрібні, а не всі поля з усіх з'єднаних таблиць;
- не перевантажувати умови для з'єднання таблиць, а краще винести частину умови в фільтр.

Контрольні питання

1. Опишіть метрики ефективності web сервісів?
2. Які головні цілі метрик?
3. Опишіть якість роботи web сервісів?
4. Опишіть метрики ефективності баз даних?
5. Опишіть методи оптимізації?
6. Опишіть метрики для оптимізації пошукової системи?
7. Опишіть особливості оптимізації (т'юнінгу) баз даних?

СПИСОК ВИКОРИСТАНОЇ ЛІТЕРАТУРИ

1. Теорія планування експериментів: Виконання розрахунково-графічної роботи [Електронний ресурс] : навч. посіб. для студ. спеціальності 131 «Прикладна механіка», спеціалізації «Технологія машинобудування» / С.М. Лапач ; КПІ ім. Ігоря Сікорського. – Електронні текстові данні (1 файл: 3,31 Мбайт). – Київ : КПІ ім. Ігоря Сікорського, 2020. – 86 с.
2. Комп'ютерні моделі [Електронний ресурс]. – Режим доступу: <https://storinka.click/297-kompyutern-model.html>. – (дата звернення: 18.03.2023).
3. Конспект лекцій з курсу «Планування і обробка результатів експерименту» : для студ. ден. і заоч. форм навч. спеціальності 141 – Електроенергетика, електротехніка та електромеханіка / Харків. нац. унт міськ. госп-ва ім. О. М. Бекетова ; уклад. Л. А. Назаренко. – Харків : ХНУМГ ім. О. М. Бекетова, 2018. – 163 с.
4. Статюха, Г.О. Вступ до планування оптимального експерименту : навч. посіб. / Г.О. Статюха, Д.М. Складанний, О.С. Бонаренко. – К. : ІВЦ «Політехніка», 2011. – 117 с. – Режим доступу: https://ela.kpi.ua/bitstream/123456789/18103/1/Vstup_planyvanya_SSB.pdf. – (дата звернення: 18.03.2023).
5. Планування і обробка даних наукового експерименту : конспект лекцій / В.В. Полтавець. – Донецьк : ДВНЗ ДонНТУ, 2008. – 52 с.

6. Засименко, В.М. Основи теорії планування експерименту : навч. посіб. / В.М. Засименко. – Львів : вид-во. ДУ «ЛП», 2000. – 205 с.
7. Мотигін, В.В. Планування експерименту в інженерних дослідженнях (лабораторний практикум) : навч. посіб. / В.В. Мотигін, С.М. Павлов. – Вінниця : ВДТУ, 2001. – 82 с.
8. Ляшенко, Б. М. Методи обчислень : навч.–метод. посіб. для студ. фіз.–мат. ф-ту / Б. М. Ляшенко, О. М. Кривонос, Т. А. Вакалюк. – Житомир : вид-во ЖДУ, 2014. – 228 с.
9. Методи обробки експериментальних даних / А.А. Горват, О.О. Молнар, В.В. Мінькович. – Ужгород : вид-во УжНУ «Говерла», 2019. – 182 с. – Режим доступу: http://www.dgma.donetsk.ua/docs/kafedry/avp/metod/ДПЕ_Навчальний_посібник_Горват.Pdf. – (дата звернення: 18.03.2023).
10. Шейко, В.М. Організація та методика науково-дослідницької діяльності : підручник / В.М. Шейко, Н.М. Кушнарєнко. – 6-те вид., перероб. і допов. – К. : Знання, 2008. – 310 с. – Режим доступу: <https://chmnu.edu.ua/wp-content/uploads/2016/07/SHejko-V.M.-ta-inshi-Organizatsiya-tametodika-naukovo-doslidnitskoyi-diyalnosti.pdf>. – (дата звернення: 18.03.2023).
11. Грищук, Ю.С. Основи наукових досліджень : навч. посіб. / Ю.С. Грищук. – Харків : НТУ «ХПІ», 2008. – 232 с. – Режим доступу: <http://web.kpi.kharkov.ua/ea/wp-content/uploads/sites/25/2017/02/OND-Ukr.pdf>. – (дата звернення: 18.03.2023).

Схемоконспект лекцій з дисципліни «Теорія та практика комп'ютерних експериментів» для аспірантів спеціальності 122 «Комп'ютерні науки» всіх форм навчання [Електронний ресурс] / укл. О.М. Любименко. – Луцьк : ДонНТУ, 2023. – 126 с.

Комп'ютерний набір і верстка: Любименко Олена Миколаївна
Укладач: Любименко О.М., доц., к.ф.-м.н., доц.

Донецький національний технічний університет
м. Луцьк Волинської обл., вул. Потебні, 56

Луцьк

2023