

МІНІСТЕРСТВО ОСВІТИ І НАУКИ УКРАЇНИ

Запорізький національний технічний університет

**МЕТОДИЧНІ ВКАЗІВКИ
ДО ЛАБОРАТОРНИХ РОБІТ**

з дисципліни
«Основи теорії планування експерименту»

для студентів спеціальностей 8.17010201 – «Системи технічного захисту інформації, автоматизація її обробки», 8.17010101 – «Безпека інформаційних та комунікаційних систем» усіх форм навчання

2014

Методичні вказівки до лабораторних робіт з дисципліни «Основи теорії планування експерименту» для студентів спеціальностей 8.17010201 – «Системи технічного захисту інформації, автоматизація її обробки», 8.17010101 – «Безпека інформаційних та комунікаційних систем» усіх форм навчання. /Укл.: Л.М. Карпуков, Д.К. Цинько - Запоріжжя: ЗНТУ, 2014. - 49 с.

Укладачі: Л.М. Карпуков, професор, д.т.н.,
Д.К. Цинько, асист.

Рецензент: С.М. Романенко , доцент, к.т.н.

Відповідальний
за випуск: Д.К. Цинько, асист.

Затверджено:
на засіданні кафедри
захисту інформації
протокол № 4 від 25.11.2014 р.

ЗМІСТ

ЛАБОРАТОРНА РОБОТА № 1

Основні параметри та характеристики випадкових величин..... 4

ЛАБОРАТОРНА РОБОТА № 2

Інтервальні оцінки для математичного сподівання і дисперсії
випадкової величини 10

ЛАБОРАТОРНА РОБОТА № 3

Парний регресійний аналіз 18

ЛАБОРАТОРНА РОБОТА № 4

Багатофакторний регресійний аналіз 24

ЛАБОРАТОРНА РОБОТА № 5

Однофакторний експеримент 31

ЛАБОРАТОРНА РОБОТА № 6

Повний факторний експеримент 37

ЛАБОРАТОРНА РОБОТА № 7

Дробовий факторний експеримент 45

ЛАБОРАТОРНА РОБОТА № 1

ОСНОВНІ ПАРАМЕТРИ ТА ХАРАКТЕРИСТИКИ ВИПАДКОВИХ ВЕЛИЧИН

Мета роботи - ознайомлення з основними параметрами та характеристиками випадкової величини, з методами їх статистичної оцінки, з алгоритмами генерації випадкових чисел з рівномірним і нормальним законами розподілу.

1.1 Теоретичні відомості

Повний набір усіх подумки можливих значень випадкової величини x називається генеральною сукупністю. Найважливішою характеристикою генеральної сукупності є функція $f(x)$ розподілу щільності ймовірностей випадкової величини, а найважливішими параметрами - математичне сподівання μ_x і дисперсія σ_x^2 випадкової величини, пов'язані з $f(x)$ наступними формулами:

$$\mu_x = \int_{-\infty}^{\infty} x f(x) dx, \quad (1.1)$$

$$\sigma_x^2 = \int_{-\infty}^{\infty} (x - \mu_x)^2 f(x) dx. \quad (1.2)$$

Математичне сподівання визначає центр розподілу, а дисперсія - розкид значень випадкової величини відносно центру розподілу.

Якщо функція $f(x)$ щільності розподілення ймовірностей випадкової величини x відома, то ймовірність p знаходження x в середині інтервалу, обмеженого значеннями x_1, x_2 , визначається формулою:

$$p = P\{x_1 \leq x \leq x_2\} = \int_{x_1}^{x_2} f(x) dx. \quad (1.3)$$

Вибірка - це кінцеве число незалежно отриманих в процесі експерименту значень величини x . Вибірка обсягу n складається з n значень: $x_1, x_2, \dots, x_n$. Чисельна оцінка статистичних характеристик генеральної сукупності по вибірці називається точковою. Точкова оцінка називається спробою, якщо її значення прагне до істинного значення оцінюваної характеристики при необмеженому збільшенні обсягу вибірки.

За вибіркою оцінка математичного сподівання μ_x виконується за допомогою вибіркового середнього:

$$\bar{x} = \frac{1}{n} \cdot \sum_{i=1}^n x_i \quad (1.4)$$

Оцінка дисперсії σ_x^2 виконується за допомогою вибіркової дисперсії:

$$S^2 = \frac{1}{n-1} \cdot \sum_{i=1}^n (x_i - \bar{x})^2 \quad (1.5)$$

Квадратичні коріння з σ_x^2 і S^2 називаються середньоквадратичними відхиленнями.

Оцінка виду функції $f(x)$ щільності розподілу досліджуваної випадкової величини проводиться шляхом побудови гістограми $f_n(x)$. Гістограма являє собою графік відносної частоти потрапляння елементів вибірки в кожен з підінтервалів, на які розбивається весь діапазон зміни випадкової величини. Відносна частота для кожного підінтервала розраховується по:

$$\omega_k = \frac{n_k}{n}, \quad (1.6)$$

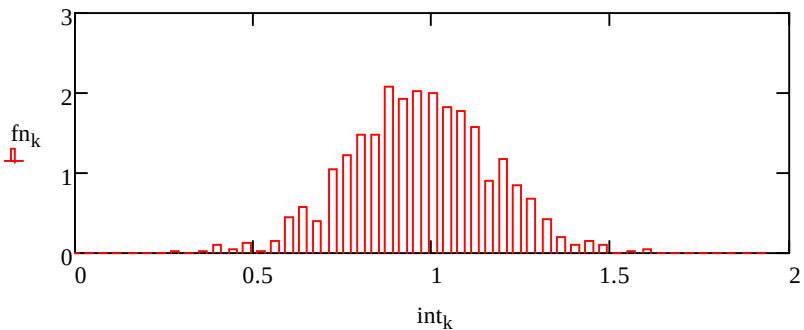
де n - обсяг вибірки, n_k - кількість елементів вибірки, що потрапили в k -й підінтервал.

Графічно гістограма зображується у вигляді сукупності прямокутників, основою яких є ширина підінтервалів, а висотою - відносні частоти, помножені на ширину відповідних підінтервалів. При такій побудові сума площ всіх прямокутників дорівнює одиниці.

При необмеженому збільшенні обсягу вибірки n і зменшенні ширини підінтервалів гістограма $f_n(x)$ буде прагнути до істинного вигляду функції $f(x)$ щільності розподілу ймовірностей випадкової величини.

Приклад Mathcad-програми, що реалізує побудову гістограми, має наступний вигляд:

$$\begin{aligned} M &:= 50 & x_{\min} &:= 0 & x_{\max} &:= 2 & h &:= \frac{x_{\max} - x_{\min}}{M} \\ k &:= 1..M \\ \text{int}_k &:= x_{\min} + (k - 1) \cdot h \\ f_n &:= \frac{\text{hist}(\text{int}, x)}{N \cdot h} \end{aligned}$$



Тут M - число підінтервалів гістограми; $x_{\min}$, $x_{\max}$ - інтервал зміни випадкової величини x ; h - довжина підінтервалів; int - вектор з координатами початків підінтервалів; $\text{hist}(\text{int}, x)$ - функція побудови гістограми; тип графіка - **bar**.

При моделюванні випадкових величин з різними видами розподілів щільності ймовірностей широко використовуються програмно реалізовані датчики випадкових чисел, які, як правило,

будуються на основі датчика з рівномірним розподілом випадкових чисел на інтервалі $[0,1]$.

Одним з методів побудови датчиків випадкових чисел з рівномірним законом $f(x)=1$ розподілу щільності ймовірностей є метод Фібоначчі. Метод заснований на рекуррентному використанні співвідношення

$$x_i = (a x_{i-1} + c)_{\text{mod}(m)}, \quad (1.7)$$

де i - порядковий номер числа; a, c - великі цілі числа; $\text{mod}(m)$ - операція по модулю m .

Рекомендовані значення: $a = 1664525$, $c = 1013904223$, $m = 2^{30}$, $x_0 = 67890$. Для отримання випадкових чисел з інтервалу $[0,1]$ застосовується нормування: x_i/m .

В системі Mathcad є вбудована функція **rnd(N)**, при $N=1$ генеруюча випадкові числа з рівномірним розподілом на інтервалі $[0,1]$.

Помилки даних експерименту носять випадковий характер і являють собою результат підсумовування великої кількості різних взаємно незалежних випадкових впливів малої інтенсивності. Тому на підставі теореми Ляпунова (центральної граничної теореми) з високою вірогідністю можна стверджувати, що випадкові величини, відповідні даними експерименту, характеризуються нормальним законом розподілу ймовірностей:

$$N(x) = \frac{1}{\sqrt{2\pi}\sigma_x} e^{-\frac{1}{2}\left(\frac{x-\mu_x}{\sigma_x}\right)^2}. \quad (1.8)$$

де μ_x - математичне сподівання, σ_x^2 - середньоквадратичне відхилення випадкової величини.

Приклад генератора випадкових чисел з нормальним розподілом:

$$x_i = \mu_x + \sigma_x \sqrt{2|\ln[\text{rnd}(1)]|} \sin[2\pi \text{rnd}(1)] \quad (1.9)$$

Інший варіант генератора випадкових чисел з нормальним розподілом:

$$x_i = \mu_x + \sigma_x \left[\left(\sum_{j=1}^k \text{rnd}(1) \right) - \frac{k}{2} \right], \quad k = 12. \quad (1.10)$$

В системі Mathcad є вбудована функція **rnorm(n,μ,σ)**, генеруюча випадкові числа з нормальним розподілом.

1.2 Лабораторне завдання

1. Визначити вхідні дані:

- математичне сподівання $\mu_x = c$,
- дисперсія $\sigma_x^2 = 0.3 + (a + b + c)/100$,

де a, b, c - три останні цифри номера студентського квитка.

2. Для рівномірного розподілу випадкових чисел на інтервалі [0,1] визначити по (1.1), (1.2) математичне сподівання і дисперсію.

3. По (1.8) побудувати графік нормального розподілу і по (1.1), (1.2) обчислити математичне сподівання і дисперсію. По (1.3) визначити ймовірності попадання випадкової величини в інтервали. $\pm \sigma_x, \pm 2\sigma_x, \pm 3\sigma_x$. Відзначити на графіку ці інтервали.

4. За співвідношенням (1.7), (1.9), (1.10) скласти програми для генерації випадкових чисел.

5. Для обсягу вибірки $n = 1000$ по (1.7) і за допомогою функції **rnd(1)**, згенерувати випадкові числа. Для кожної з двох вибірок побудувати графіки x_i від i , точками відзначити випадкові числа. За вибірками побудувати гістограми для 50 підінтервалів. По (1.4), (1.5) оцінити математичне сподівання і дисперсію. Порівняти з теоретичними значеннями.

6. Для обсягу вибірки $n = 1000$ по (1.9), (1.10) і за допомогою функції **rnorm(n, μ, σ)**, згенерувати випадкові числа. Для кожної з

трьох вибірок побудувати графіки χ_i від i , точками відзначити випадкові числа, на графіках відзначити рівні μ_x , $\mu_x + 3\sigma_x$, $\mu_x - 3\sigma_x$. За вибірками побудувати гістограми для 50 підінтервалів. По (1.4), (1.5) оцінити математичне сподівання і дисперсію. Порівняти з теоретичними значеннями.

7. На одному і тому ж графіку побудувати функцію (1.8) і гістограму по вибірці від функції **rnorm(n, μ , σ)**, для гістограми використовувати тип графіка **step**. Порівняти гістограму з істинним розподілом (1.8).

1.3 Контрольні питання

1. Що таке генеральна сукупність випадкової величини?
2. Визначення математичного сподівання.
3. Визначення дисперсії, середньоквадратичного відхилення.
4. Як визначити ймовірність попадання випадкової величини в заданий інтервал?
5. Що називається вибіркою?
6. Що таке точкова оцінка? Коли точкова оцінка спроможна?
7. Запишіть формулу оцінки математичного сподівання.
8. Запишіть формулу оцінки дисперсії.
9. Що таке гістограма? Як графічно зображується гістограма?
10. Чому рівні математичне сподівання і дисперсія для рівномірного розподілу на інтервалі $[0, 1]$?
11. Запишіть формулу для нормального розподілу і намалюйте відповідний графік.
12. Які функції виконують **rnd(1)**, **rnorm(n, μ , σ)**?

ЛАБОРАТОРНА РОБОТА № 2

ІНТЕРВАЛЬНІ ОЦІНКИ ДЛЯ МАТЕМАТИЧНОГО СПОДІВАННЯ І ДИСПЕРСІЇ ВИПАДКОВОЇ ВЕЛИЧИНИ

Мета роботи - ознайомлення з процедурами розрахунку довірчих інтервалів для математичного сподівання і дисперсії випадкової величини.

2.1 Теоретичні відомості

Інтервал $[\epsilon_1, \epsilon_2]$, в якому із заданою ймовірністю p знаходиться істинне значення оцінюваного параметра випадкової величини, називається довірчим інтервалом, а ймовірність p - довірчою ймовірністю. Залишкову ймовірність $\alpha = 1 - p$ прийнято називати рівнем значущості.

При обробки даних експерименту типовими завданнями є розрахунки інтервальних оцінок для математичного сподівання μ_x і дисперсії σ_x^2 випадкової величини x . Ці завдання включають в себе:

- розрахунок довірчого інтервалу для μ_x при відомому σ_x^2 ;
- розрахунок довірчого інтервалу для μ_x при невідомому σ_x^2 ;
- розрахунок довірчого інтервалу для σ_x^2 при невідомому μ_x .

Рішення даних завдань засноване на таких положеннях математичної статистики:

- кожен можливий результат вимірювання є випадкова величина, що підкоряється нормальному закону розподілу ймовірностей з одним і тим же математичним очікуванням μ_x ;
- дисперсії σ_x^2 всіх випадкових величин при вимірах з однаковим ступенем точності (рівноточні виміри) повинні бути однаковими.

Тому випадкова величина, відповідна вибіркового середньому (1.4),

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

також матиме нормальний розподіл з тим же математичним очікуванням μ_x , і середнім квадратичним відхиленням $\sigma_x / \sqrt{n}$. Цей розподіл має такий вигляд:

$$N(x) = \frac{1}{\sqrt{2\pi} \sigma_x / \sqrt{n}} e^{-\frac{1}{2} \left(\frac{x - \mu_x}{\sigma_x / \sqrt{n}} \right)^2} . \quad (2.1)$$

2.1.1 Процедура розрахунку довірчого інтервалу для математичного сподівання при відомій дисперсії.

Для розрахунку довірчого інтервалу для математичного сподівання μ_x при відомому σ_x^2 зазвичай використовується нормальний нормований розподіл з $\mu_x = 0$, $\sigma_x = 1$:

$$N_n(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2} x^2} . \quad (2.2)$$

Розрахунок довірчого інтервалу виконується наступним чином:

1. За вибіркою $x_1, x_2, \dots, x_n$ об'єму n за допомогою (1.4) обчислюється вибіркове середнє $\bar{x}$.
2. Складається статистика:

$$m = \frac{\bar{x} - \mu_x}{\sigma_x / \sqrt{n}} , \quad (2.3)$$

яка має розподіл (2.2).

3. Здається довірна ймовірність p і по (1.3) складається рівняння для кордонів довірчого інтервалу $[-\epsilon, \epsilon]$:

$$p = P\{|m| \leq \varepsilon\} = 2 \int_0^{\varepsilon} N_n(x) dx \quad (2.4)$$

З рішення цього рівняння визначається ε , відповідна заданому значенню p .

4. Від знайденого довірчого інтервалу $-\varepsilon \leq m \leq \varepsilon$ для випадкової величини m переходять до шуканого інтервалу для m_x :

$$\bar{x} - \frac{\varepsilon \sigma_x}{\sqrt{n}} \leq \mu_x \leq \bar{x} + \frac{\varepsilon \sigma_x}{\sqrt{n}}. \quad (2.5)$$

2.1.2 Процедура розрахунку довірчого інтервалу для математичного сподівання при невідомій дисперсії.

Якщо дисперсія невідома, що зазвичай має місце на практиці, то для її оцінки використовується вибіркова дисперсія (1.5),

$$S^2 = \frac{1}{n-1} \cdot \sum_{i=1}^n (x_i - \bar{x})^2,$$

і застосовуються такі положення математичної статистики:

- якщо $x_1, x_2, \dots, x_n$ випадкові величини з нормальним розподілом, то сума квадратів цих величин підпорядковується χ^2 -розподіленню (хі-квадрат розподіл);

- якщо χ_m^2 - випадкова величина з хі-квадрат розподілом і з m ступенями свободи, а y - випадкова величина з нормальним розподілом, то їх відношення $y / \sqrt{\chi_m^2 / m}$ підпорядковується розподілу Стюдента (t-розподілу).

Розподіл Стюдента симетричний відносно початку координат і має наступний вигляд:

$$C_m(x) = \frac{1}{\sqrt{m\pi}} \frac{\Gamma\left(\frac{m+1}{2}\right)}{\Gamma\left(\frac{m}{2}\right) \left(1 + \frac{x^2}{m}\right)^{\frac{m+1}{2}}}, \quad (2.6)$$

де m - число свободи, Γ - гамма-функція.

При великих значеннях обсягів вибірки розподіл Стюдента наближається до нормального розподілу.

Розрахунок довірчого інтервалу для математичного сподівання випадкової величини при невідомій дисперсії виконується наступним чином:

1. За вибіркою $x_1, x_2, \dots, x_n$ об'єму n за допомогою (1.4), (1.5) обчислюються вибіркове середнє $\bar{x}$ і вибіркова дисперсія S^2 .

2. Складається статистика:

$$t = \frac{\bar{x} - \mu_x}{S/\sqrt{m}}, \quad (2.7)$$

яка має розподіл Стюдента з ступенем свободи $m = n - 1$.

3. Задаються значенням довірчої ймовірності p і складається рівняння для довірчого інтервалу $[-\epsilon, \epsilon]$:

$$p = P\{|m| \leq \epsilon\} = 2 \int_0^{\epsilon} C_m(x) dx, \quad (2.8)$$

З рішення цього рівняння визначається ϵ , відповідна заданому значенню p .

4. Від знайденого інтервалу $-\epsilon \leq t \leq \epsilon$ для випадкової величини t переходять до шуканого довірчого інтервалу для m_x :

$$\bar{x} - \frac{\epsilon S}{\sqrt{m}} \leq \mu_x \leq \bar{x} + \frac{\epsilon S}{\sqrt{m}}. \quad (2.9)$$

2.1.3 Процедура розрахунку довірчого інтервалу для дисперсії випадкової величини при невідомому математичному очікуванні.

Розрахунок заснований на наступному положенні математичної статистики:

- якщо $x_1, x_2, \dots, x_n$ - випадкові величини з нормованим нормальним розподілом, то сума квадратів цих величин підпорядковується χ^2 -розподіленого (хі-квадрат розподіл):

$$G_m(x) = \begin{cases} 0, & \text{для } x < 0, \\ \frac{x^{\frac{m}{2}-1}}{2^{\frac{m}{2}} \Gamma\left(\frac{m}{2}\right)} e^{-\frac{x}{2}}, & \text{для } x \geq 0, \end{cases} \quad (2.10)$$

де m - число свободи, Γ - гамма-функція.

Даний розподіл є несиметричним.

Розрахунок довірчого інтервалу для дисперсії випадкової величини при невідомому математичному очікуванні виконується наступним чином:

1. За вибіркою $x_1, x_2, \dots, x_n$ об'єму n обчислюються вибіркове середнє $\bar{x}$ і вибіркова дисперсія S^2 .

2. Складається статистика:

$$g^2 = \frac{(n-1)S^2}{\sigma_x^2}, \quad (2.11)$$

яка має хі-квадрат розподіл зі ступенем свободи $m = n - 1$.

3. Задаються значенням довірчої ймовірності p , знаходять рівень значущості $\alpha = 1 - p$. Оскільки χ^2 -розподіл несиметричний, то складаються два рівняння щодо меж e_1 і e_2 допустимої області

$[e_1, e_2]$ за умови, що ймовірність попадання випадкової величини в критичні області $[0, e_1]$ і $[e_2, \infty]$ дорівнює $\alpha / 2$:

$$\begin{cases} \frac{\alpha}{2} = P\{g^2 \leq \varepsilon_1\} = \int_0^{\varepsilon_1} G_m(x) dx, \\ 1 - \frac{\alpha}{2} = P\{g^2 \leq \varepsilon_2\} = \int_0^{\varepsilon_2} G_m(x) dx. \end{cases}, \quad (2.12)$$

З рішення цього рівняння визначаються межі допустимої області, відповідні заданому значенню p :

$$p = P(\varepsilon_1 \leq g^2 \leq \varepsilon_2) \quad (2.13)$$

4. Від знайденого інтервалу для випадкової величини g^2 переходять до шуканого довірчого інтервалу для σ_x^2 :

$$\frac{(n-1)S^2}{\varepsilon_1} \leq \sigma_x^2 \leq \frac{(n-1)S^2}{\varepsilon_2}. \quad (2.14)$$

2.2 Лабораторне завдання

1. Визначити вхідні дані:

- математичне сподівання $\mu_x = c$,
- дисперсія $\sigma_x^2 = 0.3 + (a + b + c)/100$,
- довірна ймовірність $p = 0.9 + c/100$,

де a, b, c - три останні цифри номера студентського квитка.

2. За допомогою Mathcad-функції **rnorm(n, μ , σ)** отримати вибірку, об'ємом $n = 10$.

3. Виконати розрахунки довірчих інтервалів:

- для μ_x при відомому σ_x^2 ;

- для μ_x при невідомому σ_x^2 ;
- для σ_x^2 при невідомому μ_x .

При розрахунку побудувати двома способами графіки відповідних розподілів: за формулами і за допомогою Mathcad-функцій **dnorm(x, μ , σ)**, **dt(x, m)**, **dchisq(x, m)**, що забезпечують, відповідно, нормальний розподіл, розподіл Стюдента і хі-квадрат розподіл, m - ступінь свободи.

На графіках відзначити - межі довірчих інтервалів. Кордони довірчих інтервалів обчислити двома способами:

- шляхом вирішення рівнянь для кордонів за допомогою Mathcad-функції **root(F(e_0), e_0)**, де $F(e_0)=0$ - рівняння, e_0 - початкове наближення до рішення, його зручно визначати за графіком відповідного розподілу;

- за допомогою Mathcad-функцій **qnorm(p, μ , σ)**, **qt(p, m)**, **qchisq(p, m)**, що забезпечують для нормального розподілу, розподілу Стюдента і хі-квадрат розподілу знаходження по заданій ймовірності p квантилів, тобто кордонів e областей, ймовірність попадання випадкової величини x в які не перевищує p .

2.3 Контрольні питання

1. Запишіть формулу для нормального розподілу та зобразите зовнішній вигляд кривої.

2. Який зв'язок між нормованим та ненормованим розподілом?

3. Що таке довірчий інтервал?

4. Як визначається довірчий інтервал через інтегральну функцію розподілу?

5. Що таке довірка ймовірність?

6. Опишіть алгоритм розрахунку довірчого інтервалу для математичного сподівання при відомій дисперсії.

7. . Як змінюється величина довірчого інтервалу при зміні обсягу вибірки і чому?

8. Зобразіть зовнішній вигляд інтегральної функції розподілу Стюдента та назвіть її основні властивості.

9. Запишіть рівняння для знаходження довірчого інтервалу при використанні розподілу Стюдента.

10.Опишіть алгоритм обчислення довірчого інтервалу для дисперсії при невідомому математичному очікуванні.

11.Зобразіть зовнішній вигляд χ^2 -розподілення та назвіть його основні властивості.

12.Запишіть рівняння для знаходження довірчого інтервалу при використанні χ^2 -розподілення.

13.Перелічіть Mathcad-функції, використані в розрахунках, та вкажіть їх призначення.

ЛАБОРАТОРНА РОБОТА № 3 ПАРНИЙ РЕГРЕСІЙНИЙ АНАЛІЗ

Мета роботи - ознайомлення з процедурою парного регресійного аналізу.

3.1 Теоретичні відомості

Процедура парного регресійного аналізу призначена для побудови регресійних моделей досліджуваних об'єктів або процесів у вигляді найпростішої парної залежності:

$$y = b_0 + b_1 x + e, \quad (3.1)$$

де y - відгук, x - фактор, b_i - коефіцієнти регресії, e - помилка експерименту (випадкова величина).

Процедура аналізу включає в себе наступні кроки:

1. За експериментальними даними y_i , x_i (число дослідів $n > 2$) методом найменших квадратів (МНК) будується залежність

$$Y(x) = \beta_0 + \beta_1 x, \quad (3.2)$$

де b_i - оцінки коефіцієнтів регресії.

За МНК система рівнянь для знаходження цих оцінок має наступний вигляд:

$$\begin{bmatrix} n & \sum_{i=1}^n x_i \\ \sum_{i=1}^n x_i & \sum_{i=1}^n x_i^2 \end{bmatrix} \cdot \begin{bmatrix} \beta_0 \\ \beta_1 \end{bmatrix} = \begin{bmatrix} \sum_{i=1}^n y_i \\ \sum_{i=1}^n y_i x_i \end{bmatrix}. \quad (3.3)$$

Звідси за методом Крамера слідують формули для оцінок:

$$\beta_0 = \left(\sum_{i=1}^n y_i \cdot \sum_{i=1}^n x_i^2 - \sum_{i=1}^n y_i x_i \cdot \sum_{i=1}^n x_i \right) / \Delta, \quad (3.4)$$

$$\beta_1 = \left(n \sum_{i=1}^n y_i x_i - \sum_{i=1}^n x_i \cdot \sum_{i=1}^n y_i \right) / \Delta, \quad (3.5)$$

де $\Delta = n \sum_{i=1}^n x_i^2 - \left(\sum_{i=1}^n x_i \right)^2$ - визначник матриці.

2. Обчислюються вибіркові середні відгуку та фактора:

$$\bar{y} = \frac{1}{n} \sum_{i=1}^n y_i, \quad \bar{x} = \frac{1}{n} \sum_{i=1}^n x_i. \quad (3.6)$$

3. Обчислюється дисперсія відгуку:

$$S_y^2 = \frac{1}{n-1} \sum_{i=1}^n (y_i - \bar{y})^2. \quad (3.7)$$

4. Обчислюється залишкова дисперсія:

$$S_o^2 = \frac{1}{n-1} \sum_{i=1}^n (y_i - Y(x_i))^2. \quad (3.8)$$

5. Визначається коваріаційна матриця та оцінки дисперсій коефіцієнтів регресії:

$$\begin{bmatrix} C_{00} & C_{01} \\ C_{10} & C_{11} \end{bmatrix} = \begin{bmatrix} n & \sum_{i=1}^n x_i \\ \sum_{i=1}^n x_i & \sum_{i=1}^n x_i^2 \end{bmatrix}^{-1}, \quad (3.9)$$

$$S_{b_0}^2 = S_o^2 C_{00}, S_{b_1}^2 = S_o^2 C_{11}. \quad (3.10)$$

6. Визначається значущість коефіцієнтів регресії.
Висувається для кожного коефіцієнта гіпотеза

$$H_0: \beta_i = 0.$$

Складаються статистики, які характеризуються розподілом Стьюдента,

$$\tau_i = \frac{\beta_i}{\sqrt{S_{b_i}^2}}. \quad (3.11)$$

За розподілом Стьюдента

$$C_m(x) = \frac{1}{\sqrt{m\pi}} \frac{\Gamma\left(\frac{m+1}{2}\right)}{\Gamma\left(\frac{m}{2}\right) \left(1 + \frac{x^2}{m}\right)^{\frac{m+1}{2}}}, \quad (3.12)$$

де m - число свободи, Γ - гамма-функція, для $m=n-2$ і для заданого рівня значущості α обчислюються кордони $\pm \varepsilon$ критичної області. При $\tau_i \geq \varepsilon$ коефіцієнти регресії статистично значимо відрізняються від нуля для заданого рівня значущості

7. Перевіряється адекватність моделі.

Складається статистика, що характеризується розподілом Фішера,

$$f = \frac{S_y^2}{S_o^2}. \quad (3.13)$$

За розподілом Фішера

$$F_{m,r}(x) = \frac{\Gamma\left(\frac{m+1}{2}\right) \left(\frac{m}{2}\right)^{\frac{m}{2}} x^{\frac{m}{2}-1}}{\Gamma\left(\frac{m}{2}\right) \Gamma\left(\frac{r}{2}\right) \left(1 + \frac{m}{r} x\right)^{\frac{m+r}{2}}} \quad (3.14)$$

з числами свободи $m=n-1$ і $r=n-2$ для заданого рівня значущості α обчислюються межа ϵ_1, ϵ_2 критичної області. При $f \geq \epsilon_2$ модель вважається статистично адекватною для заданого рівня значущості.

3.2 Лабораторне завдання

Побудувати статистичну модель, яка характеризує помісячну залежність між збитком і реалізованими загрозами. Дослідити статистичну значущість коефіцієнтів моделі і виконати статистичну оцінку її адекватності. Визначити ймовірності, для яких коефіцієнти моделі статистично значущі, а модель статистично адекватна

Дані для розрахунку визначаються за програмою:

```

ORIGIN := 1
~~~~~
Grivni := runif(12,1,4)      u := runif(12,1,6)

i := 1..12                  Ugrozii := floor{ui}

```

Тут вектор Grivni містить помісячний збиток в т. гр., вектор Ugrozi - кількість загроз.

Лабораторне завдання включає в себе:

1. Побудова за даними щомісячного збитку парної регресійної моделі (3.2). Отриману залежність зобразити на графіку, точками відзначити вихідні дані.
2. Розрахунок по (3.6) - (3.10) точкових оцінок параметрів моделі.

3. Перевірку по (3.11) - (3.14) гіпотез про значущість коефіцієнтів регресії і адекватності моделі з знаходженням мінімальних значень рівня значущості, для яких гіпотези відкидаються.

При перевірці гіпотез побудувати графіки відповідних розподілів, на графіках відзначити кордони допустимих областей. Розрахувати ймовірності попадання статистик в допустимі області і порівняти їх зі значеннями, які були використані в розрахунках кордонів.

В розрахунках можна використовувати статистичні функції системи Mathcad.

Для розподілу Стюдента:

- $dt(x, m)$ – функція щільності ймовірності,
- $pt(\epsilon, m)$ – значення функції розподілу в точці ϵ ,
- $qt(p, m)$ – значення квантиля для ймовірності p (визначення аргументу функції розподілу, для якого ймовірність дорівнює або менше заданого значення p).

Аналогічно визначаються функції для розподілу Фішера: $dF(x, m1, m2)$, $pF(\epsilon, m1, m2)$, $qF(p, m1, m2)$. Тут m , $m1$, $m2$ - ступені свободи.

3.3 Контрольні питання

1. Запишіть і поясніть формулу для парної регресійної моделі.
2. Як за методом найменших квадратів обчислюються оцінки коефіцієнтів регресії?
3. Як обчислюються вибіркові середні відгуку і фактора?
4. Як обчислюється дисперсія відгуку і залишкова дисперсія?
5. Як визначаються коваріаційна матриця та оцінки дисперсій коефіцієнтів регресії?
6. 6. Запишіть і поясніть формулу статистики для перевірки значущості коефіцієнтів регресії.
7. Намалюйте графік розподілу Стюдента і поясніть як визначаються кордони критичної області.
8. Запишіть і поясніть формулу статистики для перевірки адекватності моделі регресії.

9. Намалуйте графік розподілу Фішера і поясніть як визначаються кордони критичної області.

10. Чому рівні ступені свободи для розподілів, використаних в роботі?

ЛАБОРАТОРНА РОБОТА № 4 БАГАТОФАКТОРНИЙ РЕГРЕСІЙНИЙ АНАЛІЗ

Мета роботи - ознайомлення з процедурою проведення експерименту і обробки експериментальних даних при багатofакторному регресійному аналізі.

4.1 Теоретичні відомості

Метою проведення експерименту і обробки експериментальних даних при багатofакторному регресійному аналізі є отримання регресійної моделі досліджуваного об'єкта у вигляді

$$y = b_0 + b_1x_1 + b_2x_2 + \dots + b_nx_n + e \quad (4.1)$$

де y - функція відгуку (вихідна змінна);

x_i - Фактори (вхідні змінні);

b_i - Коефіцієнти регресії;

e - випадкова величина (похибка вимірювань)

Основні умови регресійного аналізу:

– похибка вимірювань e - випадкова величина з нормальним законом розподілу з нульовим математичним очікуванням і дисперсією σ_e^2 , яка називається дисперсією відтворюваності

– значення похибок e в різних дослідах некорреліровані між собою.

Регресійний аналіз включає в себе оцінку коефіцієнтів досліджуваної моделі і статистичний аналіз отриманих результатів.

Процедура регресійного аналізу складається з наступних етапів:

1. Вибір регресійної моделі і проведення $m > (n+1)$ експерименту для складання системи рівнянь щодо оцінок B_i коефіцієнтів регресії b_i :

$$Y(X) = B_0 + B_1X_1 + B_2X_2 + \dots + B_nX_n \quad (4.5)$$

4. Оцінка залишкової дисперсії у вигляді суми квадратів відхилень вимірюваних значень від розрахованих за моделлю:

$$S_{\text{ост}}^2 = \frac{1}{m - (n + 1)} \sum_{i=1}^m [Y_i - Y_i(X)]^2, \quad (4.6)$$

де Y_i – відгук у i -му досліді,

$Y_i(X)$ – значення відгуку, отримане за моделлю (4.5).

Зауваження. Якщо модель виявиться адекватною об'єкту, то єдиною причиною розходження між даними досвіду і моделлю є похибка e вимірювань з дисперсією відтворюваності σ_e^2 . Отже, залишкова дисперсія може служити оцінкою дисперсії σ_e^2 . На практиці заздалегідь невідомо, що модель адекватна, тому оцінку дисперсії відтворюваності виконують по серії повторних дослідів.

5. Кожен i -й досвід повторюємо r раз при незмінних умовах. Отримуємо набори значень відгуків:

$$Y_{ik}, i = 1, 2, \dots, m, k = 1, 2, \dots, r.$$

Обчислюємо вибіркове середнє для відгуку в i -м досвіді, усереднене по серії з r його значень:

$$\bar{Y}_i = \frac{1}{r} \sum_{k=1}^r Y_{ik} - i = 1, 2, \dots, m. \quad (4.7)$$

Проводимо оцінювання дисперсії відтворюваності

$$S_e^2 = \frac{1}{m} \sum_{i=1}^m S_i^2, \quad (4.8)$$

$$\text{де } S_i^2 = \frac{1}{(r-1)} \sum_{k=1}^r (Y_{ik} - \bar{Y}_i)^2.$$

6. Виконуємо точкове оцінювання параметрів коефіцієнтів регресії.

Оцінка математичного очікування μ_{b_i} коефіцієнта регресії b_i збігається з його оцінкою B_i .

Оцінки дисперсії коефіцієнта регресії $S_{b_i}^2 = \sigma_e^2 C_{ii}$.

Оцінка коефіцієнта коваріації для i -го і k -го коефіцієнтів регресії $K_{ik} = \sigma_e^2 C_{ik}$.

Для оцінки дисперсії відтворюваності σ_e^2 використовується S_e^2 .

7. Перевірка значущості коефіцієнтів регресії. Перевірка проводиться для кожного коефіцієнта і зводиться до перевірки основної гіпотези $H_0 : b_i = 0$ та альтернативної $H_1 : b_i \neq 0$, $i=1,2,\dots,n$. Для перевірки складаються статистики, що мають розподіл Стьюдента:

$$t_i = \frac{|B_i|}{S_{b_i}}, \quad (4.9)$$

де $S_{b_i} = \sqrt{S_e^2 C_{ii}}$ – оцінка середньоквадратичного відхилення для коефіцієнта B_i .

За розподілом Стьюдента при числі ступенів свободи $[m(r-1)]$, рівній числу ступенів свободи оцінки S_e^2 в (4.8) і заданому рівні значимості α визначається квантиль $t_\alpha [m(r-1)]$. Якщо $t_i > t_\alpha [m(r-1)]$, то нульова гіпотеза повинна бути відкинута і, отже, коефіцієнт B_i - збережений в моделі.

8. Перевірка адекватності моделі. Перевірка проводиться шляхом порівняння залишкової дисперсії $S_{\text{ост}}^2$ з дисперсією відтворюваності S_e^2 , отриманої при повторних дослідях. При

неадекватній моделі залишкова дисперсія буде значимо переважати над S_e^2 . Для перевірки обчислюється статистика

$$F = \frac{S_{\text{ост}}^2}{S_e^2} \quad (4.10)$$

з розподілом Фішера зі ступенями свободи $[m-(n+1)]$ в чисельнику і $[m(r-1)]$ в знаменнику. Здається рівень значимості α і визначається квантиль $F_\alpha [m-(n+1), m(r-1)]$, якщо $F \leq F_\alpha [m-(n+1), m(r-1)]$, то приймається гіпотеза про адекватність моделі. В іншому випадку гіпотеза відкидається.

9. Перевірка працездатності моделі. Для перевірки обчислюється коефіцієнт детермінації:

$$R^2 = 1 - \frac{r[m-(r-1)]S_{\text{ост}}^2 + m(r-1)S_e^2}{r \sum_{i=1}^m [Y_i - \bar{Y}]^2 + m(r-1)S_e^2}, \quad (4.11)$$

$$\text{де } \bar{Y} = \frac{1}{m} \sum_{i=1}^m \bar{Y}_i.$$

Коефіцієнт детермінації показує, яка частина із загального розсіювання вимірних значень відгуку щодо його середнього значення визначається обраною регресійною моделлю. Межі зміни коефіцієнта: $0 \leq R^2 \leq 1$. Якщо $R^2=0$, то вплив факторів не проявляється, а дані вимірів визначаються випадковими впливами. Чим ближче R^2 до одиниці, тим краще якість пророкувань дослідних даних обраною моделлю. На практиці приймають $R_{\min}^2 = 0.75$, при $R^2 \geq R_{\min}^2$ вважають, що модель цілком працездатна.

4.2 Лабораторне завдання

Досліджуваний об'єкт має три входи і один вихід. Реальна характеристика об'єкта являє собою лінійну залежність між вихідний реакцією (відгуком) і трьома входними впливами (факторами):

$$y = b_0 + b_1x_1 + b_2x_2 + b_3x_3 \quad (4.12)$$

де $b_0 = 1 + c$;

$$b_1 = 0.5 + \frac{b + c}{5};$$

$$b_2 = 0.1 + \frac{a + b + c}{5};$$

$$b_3 = 0.2 + \frac{(1 + c)(1 + b)}{10}$$

Тут а, b, с - три останні цифри номера студентського квитка.

В процесі вимірювань додається нормально розподілена випадкова функція ϵ з нульовим математичним очікуванням і середньоквадратичним відхиленням

$$s_e = 0.02 \left(1 + \frac{2 + a}{10 + b + c} \right).$$

Межі зміни значень факторів при вимірах:

$$x_{1\min} = \frac{a + b + 2c}{2}, x_{2\min} = \frac{(a + 1)(b + 2)}{c + 1}, x_{3\min} = a + 3b + c;$$

$$x_{1\max} = 1.3x_{1\min}, x_{2\max} = 1.4x_{2\min}, x_{3\max} = 1.5x_{3\min}$$

Тут а, b, с - три останні цифри номера студентського квитка.

Потрібно:

1. Скласти імітаційну модель об'єкта, додавши до (4.12) Mathcad- функцію, генеруючу випадкову величину ϵ ;

2. Обрати регресійну модель досліджуваного об'єкта, провести експерименти, скориставшись імітаційної моделлю досліджуваного об'єкта.

3. Виконати регресійний аналіз отриманих експериментальних даних.

4. Записати остаточний вид регресійної моделі, виключивши з неї не значущі коефіцієнти.

Рівень значущості в розрахунках $\alpha=0.05$.

4.3 Контрольні питання

1. Запишіть і поясніть формулу лінійної регресійної моделі.
2. Як за методом найменших квадратів обчислюються оцінки коефіцієнтів регресії?
3. Основні умови регресійного аналізу.
4. Запишіть і поясніть інформаційну матрицю Фішера і ковариаційну матрицю.
5. Формула для оцінки залишкової дисперсії.
6. Оцінка дисперсії відтворюваності для адекватної моделі.
7. Формули точкової оцінки параметрів коефіцієнтів регресії.
8. Методика перевірки значимості коефіцієнтів регресії.
9. Методика перевірки адекватності моделі.
10. Методика перевірки працездатності моделі.
11. Коефіцієнт детермінації, його властивості.

ЛАБОРАТОРНА РОБОТА № 5 ОДНОФАКТОРНИЙ ЕКСПЕРИМЕНТ

Мета роботи - ознайомлення з процедурою планування однофакторного експерименту і статистичної обробки даних експерименту.

5.1 Теоретичні відомості

Однофакторний експеримент (ОФЕ) відноситься до регресійних планів першого порядку. Плани першого порядку призначені для експериментального отримання лінійних регресійних моделей виду

$$y = b_0 + b_1x_1 + b_2x_2 + \dots + b_nx_n + e \quad (5.1)$$

де b_i - коефіцієнти регресії, e - випадкова величина (похибка вимірювань) з нормальним законом розподілу, нульовим математичним очікуванням і дисперсією s_e^2 (дисперсією відтворюваності).

Однофакторний (класичний) експеримент передбачає по чергову зміну кожного з факторів при незмінних значеннях інших. При цьому кожен фактор фіксується на двох рівнях: верхньому (x_{i0}) і нижньому (x_{in}). Отже, число експериментів за цим планом становить $2n$.

Для формалізації та спрощення обробки даних експерименту виконується перехід від реальних значень x_i факторів до нормованих X_i так, щоб верхній рівень x_{i0} відповідав одиниці, нижній - мінус одиниці, а середнє значення фактора - нулю. Нормировка виконується шляхом заміни змінних за формулою

$$X_i = \frac{x_i - x_{i0}}{\Delta x_i} \quad (5.2)$$

де $x_{i0} = \frac{x_{i\text{в}} + x_{i\text{н}}}{2}$ – середнє значення i -го фактора, $\Delta x_i = \frac{x_{i\text{в}} - x_{i\text{н}}}{2}$ – крок варіації i -го фактора.

З урахуванням нормування модель (5.1) запишеться у вигляді

$$y = B_0 + B_1 X_1 + B_2 X_2 + \dots + B_n X_n + e \quad (5.3)$$

Зв'язок між реальними і нормованими коефіцієнтами регресії визначається формулами:

$$b_i = \frac{B_i}{\Delta x_i}, i = 1, 2, \dots, n, \quad (5.4)$$

$$b_0 = B_0 - \sum_{i=1}^n \frac{B_i x_{i0}}{\Delta x_i}. \quad (5.5)$$

Для нормованих факторів складається прямокутна таблиця розміром $m \times n$, число n стовпців якої дорівнює числу факторів, а число m рядків – числу дослідів. Для однофакторного експерименту $m=2n$. Така таблиця зі значеннями нормованих факторів називається матрицею спектра плану експерименту, а набір значень факторів в експерименту – точкою плану експерименту. Наприклад, для отримання двухфакторной моделі

$$y = B_0 + B_1 X_1 + B_2 X_2 + e \quad (5.6)$$

складається матриця спектра плану у вигляді таблиці 5.1.

Матриця спектра плану, доповнена стовпцем зі значеннями відгуків y_k , називається матрицею планування експерименту. Для зручності обчислень матриця плану доповнюється також стовпцем для фіктивної змінної X_0 , яка відповідає одиниці при коефіцієнті регресії B_0 (див. табл. 5.2).

Таблиця 5.1 – Матриця спектра плану однофакторного експерименту при двох факторах

Номер спроби	X ₁	X ₂
1	-1	0
2	+1	0
3	0	-1
4	0	+1

Таблиця 5.2 – Матриця планування однофакторного експерименту

Номер спроби	X ₀	X ₁	X ₂	y _k
1	+1	-1	0	y ₁
2	+1	+1	0	y ₂
3	+1	0	-1	y ₃
4	+1	0	+1	y ₄

Матриця значень факторів, відповідна матриці планування однофакторного експерименту, в загальному випадку має наступний вигляд:

$$F = \begin{bmatrix} +1 & -1 & 0 & \mathbf{L} & 0 \\ +1 & +1 & 0 & \mathbf{L} & 0 \\ +1 & 0 & -1 & \mathbf{L} & 0 \\ +1 & 0 & +1 & \mathbf{L} & 0 \\ \mathbf{L} & \mathbf{L} & \mathbf{L} & \mathbf{L} & \mathbf{L} \\ +1 & 0 & 0 & \mathbf{L} & -1 \\ +1 & 0 & 0 & \mathbf{L} & +1 \end{bmatrix}. \quad (5.7)$$

Матриця F однофакторного експерименту симетрична щодо центру експерименту: $\sum_{k=1}^m X_{ik} = 0$. Стовпці матриці ортогональні один одному: $\sum_{k=1}^m X_{ik} X_{jk} = 0$, $i \neq j$. Тут i - номер фактора; k - номер досвіду; $m = 2n$ - число дослідів.

Плани, у яких стовпці матриці планування експерименту задовольняють умові ортогональності, називаються ортогональними. У ортогональних планів інформаційна матриця Фішера Φ і коваріаційна матриця C перетворюються в діагональні матриці. Для ОФЕ ці матриці мають такий вигляд:

$$\Phi = F^T F = \text{diag}(2n, 2, 2, \dots, 2), \quad (5.8)$$

$$C = (F^T F)^{-1} = \text{diag}\left(\frac{1}{2n}, \frac{1}{2}, \frac{1}{2}, \dots, \frac{1}{2}\right), \quad (5.9)$$

На підставі (5.9) дисперсії оцінок коефіцієнтів регресії:

$$S_{b_0}^2 = \frac{\sigma_e^2}{2n}, \quad (5.10)$$

$$S_{b_i}^2 = \frac{\sigma_e^2}{2}. \quad (5.11)$$

Коваріація відсутня, так як $C_{ik} = 0$ при $i \neq k$.

На підставі (5.10), (5.11) дисперсія передбачення по регресійній моделі для однофакторного експерименту

$$\sigma_y^2 = \sigma_e^2 \left[\frac{1}{2n} + \frac{1}{2} \rho^2 \right]. \quad (5.12)$$

Отже, план ОФЕ є ротатабельним, тому що дисперсія залежить тільки від відстані $\rho = \sqrt{\sum_{k=1}^m X_k^2}$ між точкою плану і центром плану.

Процедура проведення експерименту:

1. Обирається регресійна модель, складається план однофакторного експерименту, проводяться $m = 2n$ експериментів відповідно до плану.

2. Виконується статистичний аналіз відповідно до процедури регресійного аналізу:

- по (4.4) обчислюються оцінки коефіцієнтів регресії, по (4.5) складається регресійна модель і по (4.6) оцінюється залишкова дисперсія;

- проводиться g повторних дослідів і по (4.8) оцінюється дисперсія відтворюваності;

- по (5.10), (5.11) виконуються оцінки дисперсій коефіцієнтів регресії;

- по (4.9) перевіряється значимість коефіцієнтів регресії;

- по (4.10) перевіряється адекватність моделі;

- по (4.11) перевіряється працездатність моделі.

3. Здійснюється перехід по (5.3), (5.4) від нормованих до реальних коефіцієнтів регресії, при цьому незначущі коефіцієнти регресії в модель не включаються. Записується остаточний вид регресійної моделі.

5.2 Лабораторне завдання

Опис досліджуваного об'єкта і порядок вибору його параметрів приведено в лабораторній роботі №4.

Потрібно вибрати регресійну модель, скласти план однофакторного експерименту, здійснити експеримент і виконати регресійний аналіз експериментальних даних.

Порівняти отриману регресійну модель з моделлю, складеною в лабораторній роботі №4.

5.3 Контрольні питання

1. Запишіть і поясніть формулу лінійної регресійної моделі.
2. Принцип проведення однофакторного експерименту і необхідне число дослідів.
3. Запишіть формули нормування факторів.
4. Запишіть формули переходу від нормованих коефіцієнтів регресії до реальних значень.
5. Поясніть на прикладі порядок складання матриці спектра плану експерименту.
6. Поясніть на прикладі порядок складання матриці планування експерименту.
7. Основні властивості матриці F, складеної із значень факторів.
8. Поясніть на прикладі вид інформаційної матриці Фішера і коваріаційної матриці.
9. Формули для оцінок дисперсії і ковариаций коефіцієнтів регресії.
10. Запишіть і поясніть формулу для дисперсії передбачення.
11. Сформулюйте умову ротатбельності плану ОФЕ.
12. Запишіть етапи процедури регресійного аналізу.

ЛАБОРАТОРНА РОБОТА № 6 ПОВНИЙ ФАКТОРНИЙ ЕКСПЕРИМЕНТ

Мета роботи - ознайомлення з процедурою планування повного факторного експерименту та статистичної обробки даних експерименту.

6.1 Теоретичні відомості

Повний факторний експеримент (ПФЕ) відноситься до регресійних планів першого порядку. ПФЕ використовується для отримання лінійних регресійних моделей без взаємодії і з взаємодією. Наприклад, для трьох факторів ($n = 3$) лінійна модель без взаємодії записується як

$$y = b_0 + b_1x_1 + b_2x_2 + b_3x_3 + e, \quad (6.1)$$

а з взаємодією у вигляді

$$y = b_0 + b_1x_1 + b_2x_2 + b_3x_3 + b_{12}x_1x_2 + b_{13}x_1x_3 + b_{23}x_2x_3 + b_{123}x_1x_2x_3 + e \quad (6.2)$$

де b_i, b_{ik}, b_{ikj} - коефіцієнти регресії, e – випадкова величина (похибка вимірювань) з нормальним законом розподілу, нульовим математичним очікуванням і дисперсією s_e^2 (дисперсією відтворюваності).

При проведенні ПФЕ зазвичай використовуються всілякі комбінації значень факторів при двох рівнях їх зміни - верхньому (x_{i_v}) і нижньому (x_{i_n}). Такий експеримент позначають ПФЕ 2^n , де 2^n - число можливих комбінацій з n факторів на двох рівнях.

Для формалізації та спрощення обробки даних повного факторного експерименту виконується перехід по формулі (5.2) від реальних значень x_i факторів до нормованих X_i так, щоб верхній

рівень відповідав одиниці, а нижній - мінус одиниці. Зв'язок між реальними b_i і нормованими B_i коефіцієнтами регресії визначається формулами (5.3), (5.4)

Для нормованих факторів складається матриця спектра плану експерименту. Наприклад, для трьохфакторної моделі без взаємодії

$$y = B_0 + B_1 X_1 + B_2 X_2 + B_3 X_3 + e, \quad (6.3)$$

матриця спектра плану ПФЕ 2^3 представляється таблицею 6.1.

Таблиця 6.1 - Матриця спектра плану ПФЕ 2^3 .

Номер спроби	X_1	X_2	X_3
1	-1	-1	-1
2	-1	-1	+1
3	-1	+1	-1
4	-1	+1	+1
5	+1	-1	-1
6	+1	-1	+1
7	+1	+1	-1
8	+1	+1	+1

По матриці спектра плану складається матриця планування експерименту шляхом додавання стовпця з +1 для фіктивної змінної X_0 і стовпців зі значеннями відгуків y_k (див. табл. 6.2).

Таблиця 6.2 - Матриця планування ПФЕ 2^3

Номер спроби	X_0	X_1	X_2	X_3	y
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

Продовження табл. 6.2 - Матриця планування ПФЕ 2^3

Номер спроби	X0	X1	X2	X3	y
5	+1	+1	-1	-1	y_5
6	+1	+1	-1	+1	y_6
7	+1	+1	+1	-1	y_7
8	+1	+1	+1	+1	y_8

Матриця значень факторів, відповідна матриці планування ПФЕ 2^3 , буде мати наступний вигляд:

$$F = \begin{bmatrix} +1 & -1 & -1 & -1 \\ +1 & -1 & -1 & +1 \\ +1 & -1 & +1 & -1 \\ +1 & -1 & +1 & +1 \\ +1 & +1 & -1 & -1 \\ +1 & +1 & -1 & +1 \\ +1 & +1 & +1 & -1 \\ +1 & +1 & +1 & +1 \end{bmatrix}. \quad (6.4)$$

Плани ПФЕ 2^n є центрованими і ортогональними, так як виконуються умови центрування $\sum_{k=1}^m X_{ik} = 0$ і ортогональності

$\sum_{k=1}^m X_{ik} X_{jk} = 0$, $i \neq j$. Тут i – номер фактора; k – номер досвіду; $m=2^n$ – число спроб. Тому інформаційна матриця Φ і ковариационная матриця C ПФЕ 2^n – діагональні:

$$\Phi = F^T F = 2^n I_{n+1}, \quad (6.5)$$

$$C = (F^T F)^{-1} = \frac{1}{2^n} I_{n+1}, \quad (6.6)$$

де I_{n+1} – одинична матриця, розмірністю $n+1$.

Звідси випливають формули для дисперсії оцінок коефіцієнтів регресії:

$$S_{b_0}^2 = S_{b_i}^2 = \frac{\sigma_e^2}{2^n}. \quad (6.7)$$

Ковариация відсутня, так як $C_{ik} = 0$ при $i \neq k$.

Дисперсія передбачення по регресійній моделі

$$\sigma_y^2 = \frac{\sigma_e^2}{2^n} [1 + \rho^2]. \quad (6.8)$$

Дисперсія залежить тільки від відстані $\rho = \sqrt{\sum_{k=1}^m X_k^2}$ між точкою

плану і центром плану, тому план ПФЕ є ротатбельним.

Для регресійних моделей (6.2) з урахуванням взаємодії факторів повне число усіх можливих коефіцієнтів регресії, включаючи b_0 , дорівнює 2^n , тобто дорівнює числу дослідів ПФЕ. Матриця спектра плану для таких моделей і, відповідно, матриця планування доповнюються стовпцями для комбінацій добутоків факторів. Наприклад, для моделі (6.2) ПФЕ 2^3 можливі наступні комбінації: X_1X_2 ; X_1X_3 ; X_2X_3 ; $X_1X_2X_3$. Стовпці для цих комбінацій отримують шляхом перемноження елементів стовпців, відповідних факторів в базовій таблиці (див. табл. 6.2). Отримані таким чином стовпці будуть ортогональними між собою і по відношенню до стовпців базової таблиці. Тому розширена матриця планування (см. Табл. 6.3) буде ортогональною, а інформаційна матриця і ковариационная матриця - діагональні. Вирази для них збігаються з (6.5), (6.6), де одинична матриця береться з розмірністю 2^n . Отже, для оцінки дисперсій

коефіцієнтів при парних і потрійному взаємодіях може бути застосована формула (6.8).

Таблиця 6.3 - Розширена матриця планування ПФЕ 2³

Номер спроби	X ₀	X ₁	X ₂	X ₃	X ₁₂	X ₁₃	X ₂₃	X ₁₂₃	y
1	+1	-1	-1	-1	+1	+1	+1	-1	y ₁
2	+1	-1	-1	+1	+1	-1	-1	+1	y ₂
3	+1	-1	+1	-1	-1	+1	-1	+1	y ₃
4	+1	-1	+1	+1	-1	-1	+1	-1	y ₄
5	+1	+1	-1	-1	-1	-1	+1	+1	y ₅
6	+1	+1	-1	+1	-1	+1	-1	-1	y ₆
7	+1	+1	+1	-1	+1	-1	-1	-1	y ₇
8	+1	+1	+1	+1	+1	+1	+1	+1	y ₈

Число рівнянь, що складаються за табл.6.3 дорівнює 8, тобто дорівнює числу оцінюваних коефіцієнтів регресії в моделі (6.2).

Процедура проведення експерименту:

1. Складається план повного факторного експерименту для моделі (6.2), проводяться $m = 2^n$ експериментів відповідно до плану.

2. Виконується статистичний аналіз відповідно до процедури регресійного аналізу:

- по (4.4) обчислюються оцінки коефіцієнтів регресії, по (4.5) складається регресійна модель і по (4.6) оцінюється залишкова дисперсія;

- проводиться g повторних дослідів і по (4.8) оцінюється дисперсія відтворюваності;

- по (6.8) виконуються оцінки дисперсій коефіцієнтів регресії;

- по (4.9) перевіряється значимість коефіцієнтів регресії;
- по (4.10) перевіряється адекватність моделі;
- по (4.11) перевіряється працездатність моделі.

3. Виробляється перехід по (5.3), (5.4) від нормованих до реальних коефіцієнтів регресії, при цьому незначущі коефіцієнти регресії в модель не включаються. Записується остаточний вид регресійної моделі.

6.2 Лабораторне завдання

Досліджуваний об'єкт має чотири входи і один вихід. Реальна характеристика об'єкта являє собою нелінійну залежність між вихідною реакцією (відгуком) і чотирма вхідними впливами (факторами):

$$y = a_0 + 0.76a_1x_1(1 + a_2x_2) - a_3(x_2 - 1.83)(x_3 + a_4x_1x_3), \quad (6.9)$$

$$\text{де } a_0 = 1 + c; \quad a_1 = 0.5 + \frac{b - c}{5}; \quad a_2 = 0.1 + \frac{a + b + c}{5};$$

$$a_3 = 0.2 + \frac{(1 + c)(1 + b)}{10}; \quad a_4 = 0.3 + \frac{a - b + c}{5}.$$

Тут a, b, c - останні цифри номера студентського квитка.

В процесі вимірювань додається нормально розподілена випадкова функція ε з нульовим математичним очікуванням і

$$\text{дисперсією } s_e^2 = \frac{2 + a}{5 + b + c}.$$

Межі зміни значень факторів при вимірах:

$$x_{1\min} = \frac{a + b + 2c}{2}, \quad x_{2\min} = \frac{(a + 1)(b + 2)}{c + 1}, \quad x_{3\min} = a + 3b + c, \quad x_{4\min} = x_{1\min};$$

$$x_{1\max} = 1.3x_{1\min}, \quad x_{2\max} = 1.4x_{2\min}, \quad x_{3\max} = 1.5x_{3\min}, \quad x_{4\max} = x_{1\max}$$

Тут а, b, с - останні цифри номера студентського квитка

Потрібно:

- скласти імітаційну модель об'єкта, додавши до (6.9) Mathcad-функцію, генеруючу випадкову величину e з $m_x=0$ і з заданим s_e^2 ;
- для регресійних моделей досліджуваного об'єкта, без урахування і з урахуванням взаємодії факторів, провести експерименти, скориставшись імітаційної моделлю досліджуваного об'єкта;
- виконати регресійний аналіз отриманих експериментальних даних;
- записати остаточний вид регресійних моделей, виключивши з них незначущі коефіцієнти. Рівень значущості в розрахунках $\alpha = 0.05$.

6.3 Контрольні питання

1. Запишіть і поясніть формулу лінійної регресійної моделі без взаємодії і з взаємодією.
2. Принцип проведення повного факторного експерименту при двох рівнях зміни факторів і необхідне число дослідів.
3. Запишіть формули нормування факторів.
4. Скільки і які взаємодії факторів можна врахувати в ПФЕ 2^n ?
5. Поясніть на прикладі порядок складання матриці планування ПФЕ 2^n .
6. Якими властивостями володіє матриця планування ПФЕ 2^n ?
7. Поясніть на прикладі вид інформаційної та ковариаційної матриці.
8. Формули для оцінок дисперсії і ковариаций коефіцієнтів регресії.
9. Як розрахувати оцінки коефіцієнтів регресійного рівняння?
10. Запишіть і поясніть формулу для дисперсії передбачення і поясніть умову ротатабельності плану ПФЕ.

11. Запишіть етапи процедури регресійного аналізу.

12. Якого вигляду набуває регресійна модель по закінченню регресійного аналізу?

ЛАБОРАТОРНА РОБОТА № 7 ДРОБОВИЙ ФАКТОРНИЙ ЕКСПЕРИМЕНТ

Мета роботи - ознайомлення з процедурою планування дробового факторного експерименту і статистичної обробки даних експерименту.

7.1 Теоретичні відомості

Дробовий факторний експеримент (ДФЕ) відноситься до планів першого порядку, призначених для отримання лінійних регресійних моделей. План ДФЕ будується на основі плану повного факторного експерименту ПФЕ 2^n , однак для скорочення кількості дослідів з безлічі n факторів в експерименті беруть участь тільки $(n-p)$ чинників, які називаються **основними**. Позначається дробовий факторний експеримент з $(n-p)$ основними факторами як ДФЕ 2^{n-p} . Частини ПФЕ, використовувані в ДФЕ, називаються дробовими репліками. Наприклад, ДФЕ 2^{3-1} , ДФЕ 2^{4-1} , ДФЕ 2^{5-1} являють собою $\frac{1}{2}$ репліки (напіврепліку) планів ПФЕ 2^3 , ПФЕ 2^4 , ПФЕ 2^5 (кількість дослідів скорочується вдвічі); ДФЕ 2^{4-2} , ДФЕ 2^{5-2} – $\frac{1}{4}$ репліки ПФЕ 2^4 , ПФЕ 2^5 (кількість дослідів скорочується в чотири рази); ДФЕ 2^{5-3} – $\frac{1}{8}$ репліки ПФЕ 2^5 (кількість дослідів скорочується у вісім разів). Слід зазначити, що дробовий факторний експеримент дозволяє скоротити кількість дослідів, але при цьому збільшуються помилки у визначенні коефіцієнтів регресії, а також виключається можливість складання деяких парних взаємодій факторів. Застосування ДФЕ ефективно в тому випадку, якщо у досліджуваного об'єкта вплив на відгук ефектів взаємодії чинників відсутній або малий, тобто при побудові регресійної моделі об'єкта коефіцієнти регресії b_{ijk} при парних добутках факторів можуть бути виключені.

При побудові плану ДФЕ 2^{n-p} спочатку складається план повного факторного експерименту для $(n-p)$ основних факторів, потім цей план доповнюється p стовпцями для неосновних факторів. Кожен з цих p стовпців отримують шляхом поелементного перемноження від двох до $(n-p)$ стовпців основних факторів. Правила перемноження стовпців основних факторів називаються **генераторами плану**.

Для прикладу розглянемо складання плану ДФЕ для регресійної моделі вигляду

$$y = b_0 + b_1x_1 + b_2x_2 + b_3x_3 + b_{12}x_1x_2 + b_{13}x_1x_3 + b_{23}x_2x_3 + b_{123}x_1x_2x_3 + e \quad (7.1)$$

Тут вісім невідомих коефіцієнтів, для яких план ПФЕ 2^3 дає 8 рівнянь. Якщо припустити, що ефекти взаємодії малі або незначущі, то можна використовувати план ДФЭ 2^{3-1} , обравши x_1, x_2 в якості основних факторів, відповідна матриця планування представлена у табл. 7.1.

Таблиця 7.1 – Матриця планування ДФЕ 2^{3-1}

Номер спроби	x_0	x_1	x_2	x_1x_2 (x_3)	y_k
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

Нехай є достатня впевненість, що в обраних інтервалах варіювання факторів x_1 і x_2 коефіцієнт b_{12} буде незначущим. Тоді стовпець x_1x_2 можна використовувати як генератор плану для неосновного фактору x_3 ($x_3 = x_1x_2$). За планом ДФЕ 2^{3-1} може бути складена наступна модель:

$$y = b_0 + b_1x_1 + b_2x_2 + b_3x_3 + e \quad (7.2)$$

При цьому оцінки отриманих коефіцієнтів будуть змішаними, так як знайдені коефіцієнти рівняння регресії будуть спільними оцінками лінійних ефектів і ефектів взаємодії. Наприклад, якщо по табл.6.1 скласти стовпці для добутоків x_1x_3 , x_2x_3 и $x_1x_2x_3$, то виявиться, що елементи стовпця x_1x_2 співпадають з елементами

стовпця x_2 , елементи стовпця x_2x_3 - з елементами стовпця x_1 , а елементи стовпця $x_1x_2x_3$ - з елементами стовпця x_0 . Отже, оцінка коефіцієнтів моделі (7.2) виходить змішаної з оцінками коефіцієнтів моделі (7.1) оцінка b_0 - с $b_0 + b_{123}$, оцінка b_1 - с $b_1 + b_{23}$, оцінка b_2 - с $b_2 + b_{13}$ і оцінка b_3 - с $b_3 + b_{12}$.

При плануванні потрібно заздалегідь знати, які ефекти оцінюються спільно. Для цього користуються поняттям визначального контраст плану, який дозволяє, не вивчаючи план експерименту, визначити ефекти змішання оцінок коефіцієнтів регресії. **Визначальний контраст (ВК)** отримують шляхом множення генератора плану на неосновні фактори, так щоб в результаті отримає одиницю. Наприклад, для ДФЕ 2^{3-1} можливий генератор плану $x_3 = x_1x_2$. Оскільки завжди $x_i^2 = 1$, то, помножуючи генератор ліворуч і праворуч на x_3 , отримуємо визначальний контраст:

$$1 = x_1x_2x_3. \quad (7.3)$$

Відзначимо, що одиниця в ВК відповідає $x_0 = 1$ при коефіцієнті b_0 .

Для визначення ефектів змішання множать ВК на основні фактори. З контрасту (7.3) слідують наступні зв'язки:

$$\begin{aligned} x_1 &= x_2x_3, & b_1 &\rightarrow b_1 + b_2b_3, \\ x_2 &= x_1x_3, & b_2 &\rightarrow b_2 + b_1b_3, \\ x_3 &= x_1x_2, & b_3 &\rightarrow b_3 + b_1b_2, \\ x_0 &= x_1x_2x_3 = 1, & b_0 &\rightarrow b_0 + b_1b_2b_3. \end{aligned}$$

У загальному випадку при плануванні ДФЕ необхідно складати одночасно кілька генераторів плану і відповідних їм визначальних контрастів. На їх основі вводиться **узагальнений визначальний контраст (УВК)**, який будується з вихідних контрастів і з результатів їх почленного перемноження. Наприклад, x_1, x_2, x_3 - основні фактори ДФЕ 2^{5-2} . Нехай для неосновних факторів x_4, x_5 складені генератори у

вигляді: $x_4 = x_1x_2$, $x_5 = x_1x_2x_3$. Визначальні контрасти для цих генераторів: $1 = x_1x_2x_4$, $1 = x_1x_2x_3x_5$. Відповідний узагальнений визначальний контраст складатиметься з цих ВК і додаткового контрасту, отриманого шляхом їх перемноження:

$$1 = x_1x_2x_4 = x_1x_2x_3x_5 = x_3x_4x_5 \quad (7.4)$$

В результаті множення УВК (7.4) на чинники, знаходяться ефекти змішання коефіцієнтів регресії. Наприклад, для x_1 , x_3 і x_1x_3 отримаємо:

$$x_1 = x_2x_4 = x_2x_3x_5 = x_1x_3x_4x_5,$$

$$b_1 \rightarrow b_1 + b_{24} + b_{235} + b_{1345};$$

$$x_3 = x_1x_2x_3x_4 = x_1x_2x_5 = x_4x_5,$$

$$b_3 \rightarrow b_3 + b_{45} + b_{125} + b_{1234};$$

$$x_1x_3 = x_2x_4 = x_2x_5 = x_1x_4x_5,$$

$$b_{13} \rightarrow b_{13} + b_{24} + b_{25} + b_{145}.$$

Порядок контрасту - це кількість елементів в ньому. Максимальний порядок у контрастів плану характеризує роздільну здатність плану. Наприклад, план ДФЕ 2^{4-1} з генератором $x_4 = x_1x_2$ і контрастом $1 = x_1x_2x_4$ має роздільну здатність три і позначається ДФЕ 2_{III}^{4-1} . План ДФЕ 2^{4-1} з $x_4 = x_1x_2x_3$ і $1 = x_1x_2x_3x_4$ має роздільну здатність чотири і позначається ДФЕ 2_{IV}^{4-1} . ДФЕ з найбільшою роздільною здатністю називається **головним**. Головним планам віддається перевага, тому що чим вище порядок змішування, тим менше ймовірність того, що відповідні коефіцієнти регресії будуть значущими.

План ДФЕ ортогональний, для лінійних моделей без взаємодії він є ротатбельним. Інформаційна матриці Φ і коваріаційна матриця C плану ДФЕ 2^{n-p} визначаються формулами, аналогічними (6.5), (6.6):

$$\Phi = F^T F = 2^N I_{N+1}, \quad (7.5)$$

$$C = (F^T F)^{-1} = \frac{1}{2^N} I_{N+1}, \quad (7.6)$$

де $N = n - p$.

Звідси випливають оцінки дисперсії коефіцієнтів регресії:

$$S_{b_0}^2 = S_{b_i}^2 = \frac{\sigma_e^2}{N}. \quad (7.7)$$

Процедура проведення експерименту:

1. Складається план повного факторного експерименту для моделі (6.2), проводяться $m=2^n$ експериментів відповідно до плану.

2. Виконується статистичний аналіз відповідно до процедури регресійного аналізу:

- по (4.4) обчислюються оцінки коефіцієнтів регресії, по (4.5) складається регресійна модель і по (4.6) оцінюється залишкова дисперсія;

- проводиться g повторних дослідів і по (4.8) оцінюється дисперсія відтворюваності;

- по (6.8) виконуються оцінки дисперсій коефіцієнтів регресії;

- по (4.9) перевіряється значимість коефіцієнтів регресії;

- по (4.10) перевіряється адекватність моделі;

- по (4.11) перевіряється працездатність моделі.

3. Здійснюється перехід по (5.3), (5.4) від нормованих до реальних коефіцієнтам регресії, при цьому незначущі коефіцієнти регресії в модель не включаються. Записується остаточний вид регресійної моделі.

7.2 Лабораторне завдання

Опис досліджуваного об'єкта і порядок вибору його параметрів наведені в лабораторній роботі №6.

Потрібно:

- скласти імітаційну модель об'єкта, додавши до (6.9) mathcad-функцію, генеруючу випадкову величину ε з $\mu_\varepsilon=0$ і з заданим s_ε^2 ;
- для регресійної моделі дфє 2^{4-1} досліджуваного об'єкта провести експерименти при двох різних наборах основних факторів, скориставшись імітаційною моделлю досліджуваного об'єкта;
- виконати регресійний аналіз отриманих експериментальних даних;
- записати остаточний вид регресійних моделей, виключивши з них незначущі коефіцієнти. Рівень значущості в розрахунках $\alpha = 0.05$.

7.3 Контрольні питання

1. Дробовий факторний експеримент, позначення, основні фактори.
2. Приклад складання плану ДФЕ 2^{3-1} .
3. Генератори плану, їх призначення, приклад застосування.
4. Визначальний контраст, приклад застосування.
5. Узагальнений визначальний контраст, приклад застосування.
6. Порядок контрасту, роздільна здатність, позначення ДФЕ при заданій роздільній здатності.
7. Як визначаються ефекти змішання факторів?
8. Вид матриці Фішера для ДФЕ 2^{n-p} .
9. Вид коваріаційної матриці для ДФЕ 2^{n-p} .
10. Оцінки дисперсії коефіцієнтів регресії для ДФЕ 2^{n-p} .
11. Процедура проведення ДФЕ.
12. Якого вигляду набуває регресійна модель по закінченню регресійного аналізу?