Міністерство аграрної політики та продовольства України
Миколаївський національний аграрний університет

Кафедра тракторів та сільськогосподарських машин

**Завірюха М. В.**

# КОНСПЕКТ ЛЕКЦІЙ

з дисципліни: «Технологія наукових досліджень»

Миколаїв, 2014

Вступна лекція

## ТЕМА: ПОНЯТТЯ ЕКСПЕРИМЕНТУ ТА ІСТОРІЯ РОЗВИТКУ СТАТИСТИЧНОГО АНАЛІЗУ

Експеримент як певний вид людської діяльності використовується в різних галузях науки і промисловості. Правильна організація експерименту, застосування методів математичної статистики й прийомів планування дозволяють суттєво скоротити час експерименту, роблять дії експериментатора цілеспрямованими та організованими, сприяють підвищенню продуктивності праці і надійності отриманих результатів.

Методи математичного планування експеримента виникли в результаті розвитку математичної статистики. Розробка цих методів була викликана необхідністю різкого підвищення економічної ефективності наукових досліджень, які останнім часом вимагають більших витрат з боку суспільства.

Експериментальні дослідження різноманітні, але здебільшого вони зводяться до трьох математичних задач:

1. Статистична обробка результатів експеримента, відбір значимих факторів (дисперсійний та регресійний аналізи).
2. Створення математичної моделі процесу і визначення її параметрів (факторний експеримент).
3. Оптимізація процесу, тобто визначення умов, що забезпечують оптимальний режим будь-якого процесу або оптимальні властивості будь-якого продукту (методи оптимізації: метод крутого сходження, симплексне планування).

При використанні методів планування та організації експерименту розв'язок цих задач забезпечується при мінімумі дослідів і з найбільшою продуктивністю.

Джерела планування експерименту йдуть у глибоку стародавність, коли люди неусвідомлено прагнули використовувати свої знання для підвищення ефективності роботи. Однак як самостійна дисципліна, планування та організація експерименту виникла на початку XX століття. Це пов'язано з іменем великого вченого-статиста Р.А. Фишера, який в 1935 році видав першу книгу «Планування експерименту». Він застосував методи математичної статистики для вивчення факторів, що впливають на врожайність сільськогосподарських культур.

Надалі методи планування та організації експерименту почали бурхливо розвиватися. Великий внесок у розвиток дисципліни внесли вчені Бокс і Уилсон. В 50-х роках XX століття ними був запропонований метод крутого сходження для більш ефективного пошуку оптимуму функції відгуку. Даний метод являє собою багатокрокову процедуру руху до оптимуму, причому на кожному кроці експериментальні дані використовуються для локальної апроксимації поверхні регресійної моделі, яка лінійно або полиномиально залежить від факторів і, отже, залежить від параметрів процесу. Даний метод широко використовують і в наш час.

В Україні планування експерименту бурхливо розвивалося в 70-х роках XX століття і пов'язане з іменами таких учених, як Ю.П. Адлер, В.В. Минів, Ю. В. Грановський та ін. Потім планування експерименту ввійшло в моду, його стали застосовувати в тих випадках, коли без нього можна було легко обійтися. Однак до 80-их років стало ясно, що раціональне планування експерименту - це важлива, але далеко не єдина умова досягнення успіху в науковій праці. Ажіотаж навколо планування експерименту влігся, а відповідні методи та алгоритми (повний, дробний факторний експерименти, круте сходження та ін.) стали звичними і загальновживаними, особливо при широком використанні комп'ютера. Сьогодні будь-який фахівець, що працює в області метрології, фізики та інших галузях науки, повинен уміти в необхідних випадках застосовувати методи планування експерименту у своїй роботі.

У наш час планування експерименту використовується не тільки в хімії, фізиці, інших природно-наукових дисциплінах, але і в економічних, соціологічних та технічних науках.

Планування експерименту не можна цілком уважати частиною математичної теорії статистики, хоча дослідження в рамках цього предмета продовжують поставляти нові класи планів і нові узагальнення. Статистична теорія і практика життєво необхідні для професійного планування експерименту. Однак центральним моментом предмета є зв'язок між експериментатором і статистом при визначенні питань, відповіді на які слід одержати з експерименту.

**Модуль 1**

# ОСНОВИ СТАТИСТИКИ ПРИ ПЛАНУВАННІ ЕКСПЕРИМЕНТУ

Лекція №2

## ТЕМА: ОСНОВНІ ПОЛОЖЕННЯ ПЛАНУВАННЯ ЕКСПЕРИМЕНТУ

### 1. Мета і задачі планування експерименту

Основною метою застосування планування експерименту в дослідженнях є одержання найкращих (оптимальних) результатів при найменшому числі дослідів. У зв'язку із цим задачами планування експерименту є:

1. Установити мету дослідження і виробити наукову гіпотезу.
2. З апріорної інформації або попередньо проведеного дисперсійного аналізу обрати значимі фактори, що впливають на процес.
3. Створити план експерименту, що дозволяє з найменшим числом дослідів одержати математичну модель процесу.
4. Використовуючи методи оптимізації експерименту, здійснити пошук оптимальних умов проведення процесу.

### 2. Активний і пасивний експеримент

Існують кілька класифікацій експерименту. Розглянемо одну з найпоширеніших класифікацій, по якій експерименти підрозділяються на активний і пасивний експеримент.

У пасивному експерименті експериментатор не змінює фактори, згідно із планом експерименту, а тільки реєструє їх значення та одержані результати, збирає статистичний матеріал і обробляє дослідні дані для одержання математичної моделі процесу, використовуючи регресійний, кореляційний або дисперсійний аналізи.

В активному експерименті досліди проводяться по попередньо розробленому плану. При цьому змінюється не один, а відразу декілька факторів, і оцінюється їх вплив на процес. Схема будь-якого активного експерименту припускає наступні стадії:

1. Формулювання мети дослідження, вибір параметрів і факторів.

2. Уточнення області зміни факторів на основі апріорної інформації.

3. Побудова плану експерименту, що включає схему експерименту число незалежних дослідів, порядок їх проведення.

4. Проведення дослідів за планом.

5. Статистична обробка результатів експерименту.

6. Інтерпретація результатів експерименту.

7. Оптимізація експерименту, використовуючи відповідні методи.

8. Створення остаточної математичної моделі процесу в області оптимуму.

## 3. Короткі відомості з теорії ймовірностей і математичної статистики

*Теорія ймовірностей* – наука, що вивчає закономірності, притаманні випадковим явищам, тобто явищам, які при відтворенні одних і тих же дослідів протікають по-різному. Математична статистика займається розробкою методів реєстрації, опису і аналізу даних, одержаних у результаті спостережень над випадковими явищами.

*Подія* – результат дослідів чи спостережень.

*Випадкові події* – появу яких неможливо передбачити.

*Невипадкові* – можна передбачити.

*Випадкова величина* – показник, за яким оцінюють випадкові події.

*Детерміновані події* – ними є як випадкові, так і невипадкові, але випадкова подія зумовлена багатьма факторами, які врахувати наперед неможливо.

Появу події. Яка нас цікавить, оцінюють ймовірністю

$$P(A) = \frac{m}{n}, \tag{1}$$

де $m$ – число дослідів, де має місце подія.

$n$ – загальне число подій.

$$0 \le P(A) \le 1 . \tag{2}$$

При експериментальних дослідженнях визначаємо частоту події з серії дослідів.

$$P^*(A) = \frac{m}{n}, \tag{3}$$

де $m$ – число дослідів, де має місце подія.

$n$ – загальне число дослідів.

*P*(A)* – називають також статистичною ймовірністю.

При великій кількості дослідів має місце залежність

$$\left| P(A) - P^*(A) < \varepsilon \right|, \tag{4}$$

де ε – як завгодно мале число, яке не дорівнює 0.

*Закон розподілу* – співвідношення, яке зв'язує можливі значення випадкової величини й відповідні їм ймовірності. Він може бути заданий у вигляді таблиці – *ряду розподілу*, а у вигляді графіка – багатокутники розподілу.

| № досліду | 1 | 2 | … | n |
|---|---|---|---|---|
| $x_i$ | $x_1$ | $x_2$ | … | $x_n$ |

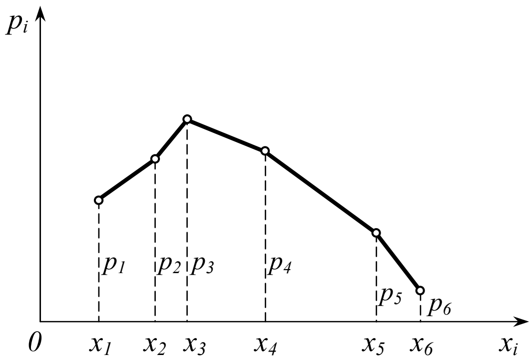


Рис. 1. Схема многокутника розподілу

*Проста статистична сукупність* – одержана сукупність значень випадкової величини.

Найчастіше це таблиці наступного виду:

| № досліду | 1 | 2 | ... | n |
|---|---|---|---|---|
| $x_i$ | $x_1$ | $x_2$ | … | $x_n$ |

Випадкові величини і статистичні розподіли оцінюють рядом числових характеристик: математичне сподівання, дисперсія, середнє квадратичне відхилення.

*Математичне сподівання* випадкової величини – сума добутків усіх можливих значень випадкової величини на їх ймовірність.

$$M[x] = \sum P_i x_i \ . \qquad (5)$$

Для статистичного розподілу аналогією математичного сподівання є середнє арифметичне, або середнє статистичне випадкової величини

$$x_{сер} = M^*[x] = \frac{1}{n}\sum_{i=1}^{n} x_i \ , \qquad (6)$$

де $n$ – число вимірювань

$x_i$ – $i$-тий результат вимірювання.

Дисперсія перервної випадкової величини:

$$D[x] = \sum_{i=1}^{n} \left(x_i - x_{ср}\right)^2 p_i \ . \qquad (7)$$

Але частіше використовують середнє квадратичне відхилення:

$$S[x] = \sqrt{D[x]} \ . \qquad (8)$$

Дисперсію статистичного розподілу називають *статистичною вибіркою* або *емпіричною дисперсією*:

$$D^*[x] = \frac{1}{n-1}\sum_{i=1}^{n} \left(x_i - x_{ср}\right)^2 \ . \qquad (9)$$

*Статистичне середнє квадратичне відхилення* окремого вимірювання статистичного розподілу

$$S_c = \sqrt{\frac{\sum_{i=1}^{n}\left(x_i - x_{cp}\right)^2}{n-1}} \; . \qquad (10)$$

Якщо провести класифікацію статистичного матеріалу на розряди і вважати наближено значення випадкової величини у кожному розряді постійним і рівним середньому значенню розряду, то після визначення частот середнє статистичне, дисперсія і середнє квадратичне відхилення становитимуть

$$x_{cp} = \sum_{j=1}^{k} \tilde{x}_j P_j^* \; , \qquad (11)$$

$$D_x^* = \frac{n}{n-1}\sum_{j=1}^{k}\left(\tilde{x}_j - x_{cp}\right)^2 P_j^* \; , \qquad (12)$$

де $\tilde{x}_j$ – середнє значення *j*-го розряду;

$k$ – число розрядів.

Оскільки $P_j = \frac{m_j}{n}$, то

$$x_{cp} = \frac{1}{n}\sum_{j=1}^{k} \tilde{x}_j \cdot m_j \; , \qquad (13)$$

$$D_x^* = \frac{n}{n-1}\sum_{j=1}^{k}\left(\tilde{x}_j - x_{cp}\right)^2 m_j \; , \qquad (14)$$

$$S_c = \sqrt{\frac{1}{n-1}\sum_{j=1}^{k}\left(\tilde{x}_j - x_{cp}\right)^2 m_j} \; , \qquad (15)$$

де $m_j$ – кількість значень, що потрапляють на *j*-й розряд.

*Коефіцієнт варіації* – відношення середнього квадратичного відхилення до статистичного середнього

$$v = \frac{S_c}{x_c} \cdot 100\% \; . \qquad (16)$$

## 4. Приклад виконання статистичного аналізу

На ряді зразків був виміряний момент у кілограмометрах, необхідний для запирання коробки радіопередавача. Результати наведені нижче:

| | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| 0,62 | 0,40 | 0,60 | 0,53 | 0,45 | 0,65 | 0,59 | 0,70 | 0,60 | 0,60 |
| 0,64 | 0,58 | 0,80 | 0,60 | 0,57 | 0,60 | 0,70 | 0,60 | 0,60 | 0,60 |
| 0,46 | 0,60 | 0,75 | 0,51 | 0,70 | 0,75 | 0,55 | 0,60 | 0,55 | 0,42 |
| 0,48 | 0,66 | 0,52 | 0,58 | 0,73 | 0,73 | 0,57 | 0,55 | 0,65 | 0,60 |
| 0,66 | 0,67 | 0,67 | 0,70 | 0,58 | 0,60 | 0,50 | 0,50 | 0,80 | 0,50 |

Обчислити середнє арифметичне, вибіркову дисперсію і побудувати гистограмму частот.

**Розв'язок.** Визначаємо кількість інтервалів по формулі Старджесса:

$$r = 1 + 3{,}31 \ln n = 1 + 3{,}31 \ln 50 = 1 + 3{,}31 \cdot 1{,}70 = 6{,}61 \approx 7 \,. \qquad (17)$$

Потім обчислюємо ширину інтервалу $h$:

$$h = \frac{x_{\max} + x_{\min}}{r} = \frac{0{,}80 + 0{,}40}{6{,}61} = 0{,}0605 \approx 0{,}06 \qquad (18)$$

Визначимо границі інтервалів, частоту потрапляння в інтервали і середини інтервалів (табл. 1).

Таблиця 1

| Границі інтервалів $x_i$ - $x_{i-1}$ | Середини інтервалів $\bar{x}$ | Частота потрапляння в інтервали $m_i$ | Статистична ймовірність (частота) $p_i$ | Різниця між середніми інтервалами і середніми $K_i$ | $m_i K_i$ | $m_i K_i^2$ |
|---|---|---|---|---|---|---|
| 0,37-0,43 | 0,40 | 2 | 0,04 | -3 | -6 | 18 |
| 0,43-0,49 | 0,46 | 3 | 0,06 | -2 | -6 | 12 |
| 0,49-0,55 | 0,52 | 7 | 0,14 | -1 | -7 | 7 |
| 0,55-0,61 | 0,58 | 21 | 0,42 | 0 | 0 | 0 |
| 0,61-0,67 | 0,64 | 6 | 0,12 | 1 | 6 | 6 |
| 0,67-0,73 | 0,70 | 6 | 0,12 | 2 | 12 | 24 |
| 0,73-0,79 | 0,76 | 3 | 0,06 | 3 | 9 | 27 |
| 0,79-0,85 | 0,82 | 2 | 0,04 | 4 | 8 | 32 |
| | | 50 | 1,00 | | 12 | 126 |

$$\bar{x} = 0{,}58 + \frac{12}{50} 0{,}06 = 0{,}58 + 0{,}01 = 0{,}59.$$

У дійсності ця рівність є наближеною, але для визначення параметрів розподілу є цілком прийнятною.

Приведемо обчислення середнього арифметичного по більш жорсткій формулі:

$$\bar{x} = \sum_{1}^{n} x_i p_i = 0{,}40 \cdot 0{,}04 + 0{,}46 \cdot 0{,}04 + 0{,}52 \cdot 0{,}14 + 0{,}58 \cdot 0{,}42 + \\ + 0{,}64 \cdot 0{,}12 + 0{,}76 \cdot 0{,}06 + 0{,}82 \cdot 0{,}04 = 0{,}59 \quad . \qquad (19)$$

Значення середнього квадратнческого відхилення визначаємо по наступній формулі:

$$\sigma = h\sqrt{\frac{\sum_{1}^{n} m_i K_i^2}{n} - \left(\frac{\sum_{1}^{n} m_i K_i}{n}\right)^2} = 0{,}06\sqrt{\frac{126}{50} - \left(\frac{12}{50}\right)^2} = 0{,}0941. \qquad (20)$$

Для досягнення наочності будують різні графіки статистичного розподілу, з яких найчастіше використовують *полігон, гістограму і кумулятивну криву*.

Полігон і гистограмма є графічними зображеннями статистичного ряду, а кумулятивна крива — це графік статистичної функції розподілу.

Графічним представленням теоретичних законів розподілу є багатокутник розподілу, крива розподілу, графіки функції розподілу.

Полігон служить частіше всього для зображення дискретного статистичного ряду, у той час як гистограмма будується тільки для інтервальних рядів. Випадкові ж величини, для яких отримані ті або інші статистичні ряди, можуть бути при цьому - як дискретними, так і безперервними.

Полігон являє собою ламану лінію, відрізки якої з’єднують точки з координатами ($x_i$; $m_i$). Для інтервального ряду будують полігон, з'єднуючи відрізками точки з координатами ($x_{i0}$; $m_i$) або ($x_{i0}$; $p_i$).

Гистограмма являє собою ступінчату фігуру, що складається із прямокутників, основою яких служать відрізки, що зображують інтервали варіаційного ряду, а різниці дорівнюють частотам або частоті відповідних інтервалів, що поділені на ширину інтервалу. У першому випадку площа гістограми дорівнює об'єму спостережень, у другому — одиниці.

Кумулятивна крива — це крива накопичених частот або накопичених частин. Якщо варіаційний ряд дискретний, то крива являє собою ламану лінію, відрізки якої з’єднують точки з координатами *($x_i$, $m_i^{max}$ )* або *($x_i$, $F_n(x)$)*. Для интервального варіаційного ряду будують ступінчату криву. Ширина кожної ступені дорівнює величині інтервалу, а її висота — відповідному до даного інтервалу значень накопиченої частоти або частості.

За даними прикладу побудуємо гістограму і статистичну функцію розподілу. Для цього по осі абсцис відкладемо значення інтервалів, а по осі ординат — значення частот або статистичних імовірностей (рис. 2, 3).

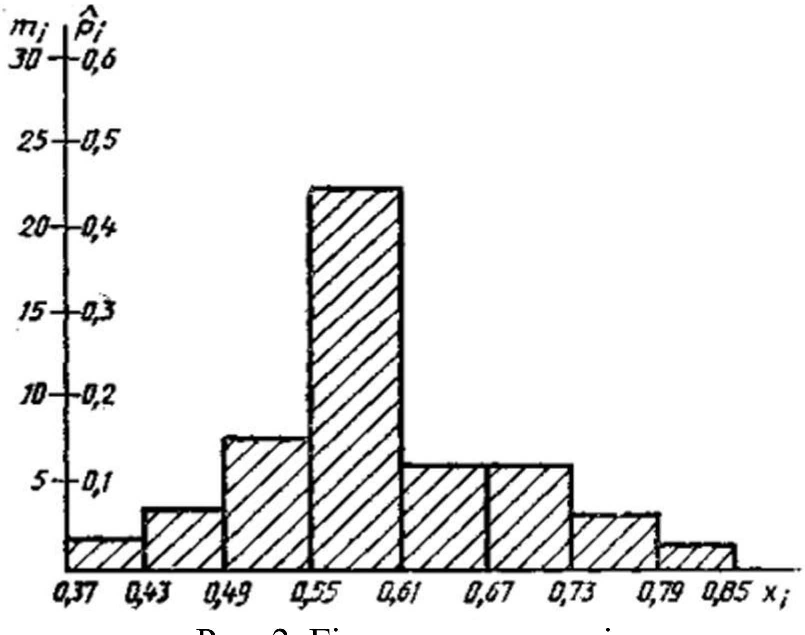


Рис. 2. Гістограма розподілу

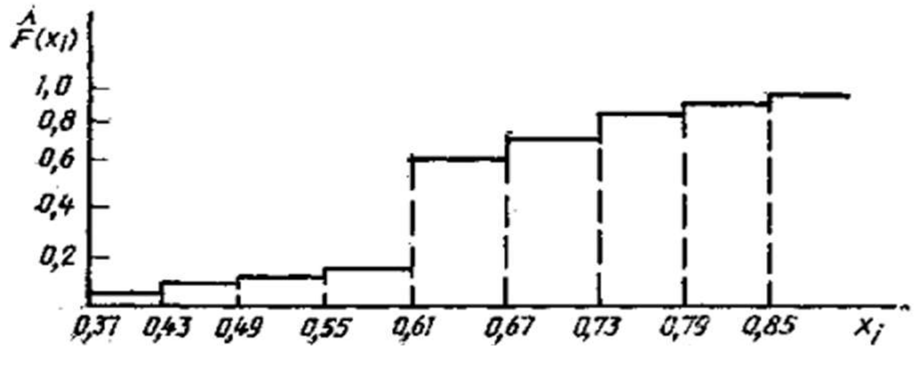


Рис. 3. Статистична функція розподілу

Для побудови статистичної функції розподілу використаємо наступну формулу:

$$F_{i+1}(x_i) = p_i + \overline{F_i}(x_i), \tag{21}$$

і проведемо додаткові обчислення:

$$F_1(x_0) = 0, \quad F_2(x_1) = p_1 + \overline{F_1}(x_0) = 0{,}04 + 0 = 0{,}04,$$

по аналогії:

$$F_3(x_2) = 0{,}06 + 0{,}04 = 0{,}10, \qquad F_4(x_3) = 0{,}14 + 0{,}10 = 0{,}24,$$

$$F_5(x_4) = 0{,}42 + 0{,}24 = 0{,}66, \qquad F_6(x_5) = 0{,}12 + 0{,}66 = 0{,}78,$$

$$F_7(x_6) = 0{,}12 + 0{,}78 = 0{,}90, \qquad F_8(x_7) = 0{,}06 + 0{,}90 = 0{,}96,$$

$$F_9(x_8) = 0{,}96 + 0{,}04 = 1{,}00.$$

Отримані значення відкладемо по осі ординат (рис. 3).

Таблица Б.1 – Массив данных $X_1$

| $X_1$ | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 |
|---|---|---|---|---|---|---|---|---|---|---|
| 1 | 2 | 16 | 6 | 4 | 15 | 14 | 5 | 18 | 2 | 3 |
| 2 | 7 | 13 | 4 | 6 | 12 | 11 | 7 | 9 | 6 | 4 |
| 3 | 7 | 14 | 4 | 3 | 12 | 16 | 7 | 9 | 4 | 7 |
| 4 | 9 | 12 | 8 | 6 | 13 | 10 | 14 | 15 | 7 | 9 |
| 5 | 9 | 11 | 8 | 4 | 14 | 10 | 13 | 11 | 7 | 9 |
| 6 | 4 | 11 | 11 | 8 | 13 | 6 | 12 | 11 | 9 | 5 |
| 7 | 5 | 10 | 12 | 8 | 11 | 6 | 12 | 13 | 9 | 8 |
| 8 | 8 | 10 | 9 | 11 | 4 | 13 | 11 | 12 | 7 | 6 |
| 9 | 6 | 5 | 7 | 11 | 5 | 15 | 11 | 14 | 7 | 6 |
| 10 | 8 | 6 | 13 | 10 | 2 | 12 | 5 | 7 | 8 | 10 |

Таблица Б.2 – Массив данных $X_2$

| $X_2$ | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 |
|---|---|---|---|---|---|---|---|---|---|---|
| 1 | 10,10 | 10,45 | 11,90 | 10,65 | 10,75 | 10,45 | 10,00 | 12,10 | 10,45 | 10,15 |
| 2 | 10,15 | 10,35 | 12,35 | 10,20 | 10,85 | 10,50 | 10,70 | 11,05 | 11,15 | 10,05 |
| 3 | 10,25 | 10,95 | 12,20 | 11,10 | 10,65 | 10,75 | 10,10 | 10,50 | 11,40 | 11,20 |
| 4 | 10,00 | 11,20 | 11,85 | 11,18 | 10,85 | 10,50 | 10,30 | 10,05 | 11,00 | 10,40 |
| 5 | 10,15 | 10,75 | 11,95 | 11,50 | 10,55 | 11,55 | 10,05 | 10,40 | 12,00 | 10,65 |
| 6 | 10,10 | 10,75 | 11,75 | 11,55 | 12,60 | 10,60 | 10,00 | 10,00 | 11,30 | 10,95 |
| 7 | 10,10 | 11,00 | 11,80 | 10,15 | 11,25 | 10,65 | 10,20 | 10,50 | 10,70 | 10,80 |
| 8 | 10,15 | 10,85 | 11,95 | 11,80 | 10,65 | 10,30 | 10,45 | 10,35 | 11,20 | 10,45 |
| 9 | 10,10 | 11,00 | 11,75 | 12,00 | 10,85 | 10,50 | 10,20 | 10,90 | 10,75 | 10,75 |
| 10 | 10,20 | 11,10 | 12,55 | 11,50 | 10,75 | 10,35 | 11,00 | 10,25 | 10,20 | 10,40 |

Таблица Б.3 – Массив данных $X_3$

| $X_3$ | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 |
|---|---|---|---|---|---|---|---|---|---|---|
| 1 | 5,3 | 14,8 | 5,2 | 11,2 | 7,5 | 9,3 | 7,8 | 13,7 | 6,1 | 8,8 |
| 2 | 5,7 | 14,2 | 6,8 | 11,5 | 7,4 | 0,1 | 10,4 | 13,4 | 8,2 | 9,9 |
| 3 | 6,4 | 7,3 | 6,7 | 13,6 | 10,8 | 8,5 | 5,4 | 14,8 | 9,2 | 11,9 |
| 4 | 6,6 | 6,8 | 8,3 | 14,8 | 12,4 | 8,6 | 5,4 | 11,3 | 9,4 | 13,9 |
| 5 | 7,2 | 12,1 | 8,5 | 14,7 | 5,6 | 10,4 | 6,3 | 11,2 | 13,5 | 9,3 |
| 6 | 7,8 | 13,5 | 10,8 | 6,7 | 8,0 | 10,2 | 9,7 | 12,5 | 13,7 | 5,7 |
| 7 | 10,4 | 13,5 | 12,2 | 7,5 | 9,0 | 5,7 | 6,3 | 11,1 | 8,8 | 6,9 |
| 8 | 11,3 | 11,8 | 7,1 | 5,1 | 3,4 | 5,7 | 10,2 | 12,6 | 5,4 | 14,5 |
| 9 | 12,0 | 10,4 | 7,9 | 5,2 | 10,9 | 14,2 | 11,7 | 8,5 | 12,2 | 15,0 |
| 10 | 8,1 | 10,6 | 5,5 | 9,3 | 6,7 | 14,3 | 12,2 | 7,6 | 10,4 | 5,7 |

Таблица Б.4 – Массив данных $X_4$

| $X_4$ | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 |
|---|---|---|---|---|---|---|---|---|---|---|
| 1 | 5,000 | 5,130 | 5,200 | 5,420 | 5,300 | 5,050 | 6,200 | 5,200 | 5,090 | 7,250 |
| 2 | 5,300 | 5,620 | 5,180 | 5,570 | 5,420 | 5,220 | 6,380 | 5,280 | 5,120 | 6,900 |
| 3 | 5,375 | 5,550 | 5,400 | 5,160 | 5,125 | 5,380 | 5,740 | 5,370 | 5,125 | 5,700 |
| 4 | 5,625 | 5,370 | 5,470 | 5,120 | 5,570 | 5,380 | 5,800 | 5,420 | 5,125 | 6,400 |
| 5 | 5,700 | 5,290 | 5,540 | 5,140 | 5,100 | 5,300 | 6,050 | 5,700 | 5,130 | 6,200 |
| 6 | 6,125 | 5,650 | 5,720 | 5,620 | 5,120 | 5,400 | 6,680 | 5,850 | 5,270 | 6,100 |
| 7 | 6,620 | 5,120 | 5,280 | 5,690 | 5,200 | 5,130 | 5,400 | 5,760 | 5,380 | 5,960 |
| 8 | 7,120 | 5,880 | 5,100 | 5,390 | 5,740 | 5,150 | 5,350 | 5,550 | 5,40 | 5,770 |
| 9 | 5,125 | 5,900 | 5,150 | 5,260 | 5,630 | 5,120 | 5,600 | 6,110 | 5,490 | 5,720 |
| 10 | 5,100 | 6,400 | 5,790 | 6,450 | 5,860 | 5,600 | 5,520 | 6,160 | 5,580 | 5,620 |

Таблица Б.5 – Массив данных $X_5$

| $X_5$ | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 |
|---|---|---|---|---|---|---|---|---|---|---|
| 1 | -4,5 | 0,2 | 3,6 | -3,4 | -0,7 | 4,4 | -4,6 | 2,3 | -3,2 | 3,1 |
| 2 | -2,5 | 0,6 | 2,8 | -2,4 | -0,7 | 3,9 | -3,5 | 2,7 | -2,8 | 4,2 |
| 3 | -2,4 | 0,7 | 2,4 | -2,6 | 0,2 | -2,2 | -2,7 | 3,2 | -2,4 | 2,2 |
| 4 | -1,3 | 0,9 | -0,4 | 0,1 | 0,3 | -1,9 | -1,2 | 4,9 | -1,2 | 2,9 |
| 5 | -1,7 | 1,4 | 0,2 | 0,4 | -2,8 | -0,6 | -1,5 | -1,8 | -1,4 | 1,3 |
| 6 | -0,5 | 1,6 | 0,3 | 0,7 | -2,3 | 0,2 | -1,7 | -1,4 | -4,2 | 1,3 |
| 7 | 0,0 | 1,9 | 0,8 | -1,4 | -3,2 | -1,7 | -1,0 | -1,6 | -0,8 | 1,9 |
| 8 | -0,6 | 2,3 | -0,9 | -1,5 | -4,2 | -1,1 | -2,4 | -2,5 | -0,6 | 0,1 |
| 9 | -0,7 | 2,7 | -0,6 | -1,6 | 0,8 | -2,9 | -0,5 | -2,6 | -0,6 | 0,3 |
| 10 | -1,5 | 3,4 | -0,5 | -0,8 | 1,3 | -3,7 | -0,8 | -0,7 | -0,2 | 0,8 |

Таблица Б.6 – Массив данных $X_6$

| $X_6$ | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 |
|---|---|---|---|---|---|---|---|---|---|---|
| 1 | -0,80 | 2,80 | -1,00 | -0,15 | -0,55 | 0,30 | 2,85 | -0,75 | 0,44 | -0,30 |
| 2 | -0,40 | 2,50 | -0,10 | 0,15 | 0,60 | 0,30 | 2,48 | 1,86 | 0,48 | -0,10 |
| 3 | 0,00 | 1,50 | 0,10 | 0,35 | 0,80 | 0,40 | -0,40 | 2,08 | 0,52 | 0,00 |
| 4 | 0,10 | 1,60 | 0,30 | 0,45 | 0,80 | 0,70 | 0,00 | -0,55 | 0,64 | -1,00 |
| 5 | 0,40 | 0,70 | 0,50 | -0,45 | 0,90 | 0,70 | 0,05 | -0,10 | 0,80 | 0,36 |
| 6 | 0,40 | 0,80 | 1,95 | 0,65 | 0,30 | -0,18 | 0,45 | -0,15 | 0,80 | 0,24 |
| 7 | 0,80 | -0,50 | 2,05 | 0,70 | 0,00 | 0,16 | 0,45 | 0,68 | 0,94 | 0,62 |
| 8 | 0,80 | -0,10 | 2,30 | 0,80 | 1,10 | -0,48 | 0,85 | 0,92 | 1,10 | 0,72 |
| 9 | 0,90 | 0,00 | 1,65 | 0,85 | 1,20 | 1,70 | 0,85 | 2,44 | 1,20 | 0,76 |
| 10 | 0,90 | 0,30 | 1,75 | -0,70 | 1,20 | 1,78 | 0,90 | 1,44 | 1,20 | 0,87 |

Таблица Б.7 – Массив данных $X_7$

| $X_7$ | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 |
|---|---|---|---|---|---|---|---|---|---|---|
| 1 | -0,40 | -0,10 | -0,10 | 0,00 | -0,44 | 2,98 | -0,42 | -0,49 | 2,40 | -0,50 |
| 2 | 0,00 | 0,15 | -0,30 | 0,10 | -0,15 | 2,28 | -0,18 | 0,02 | 2,50 | -0,10 |
| 3 | 0,10 | 0,30 | -0,15 | 0,36 | -0,05 | 2,56 | 0,09 | -0,10 | 2,00 | 0,12 |
| 4 | 0,40 | 0,65 | 0,05 | 0,45 | 0,38 | 1,86 | 0,27 | 0,24 | 2,10 | 0,32 |
| 5 | 0,50 | 0,70 | 0,35 | 0,48 | 0,45 | 2,00 | 0,48 | 0,41 | 1,60 | 0,40 |
| 6 | 0,80 | 0,90 | 0,45 | 0,90 | 0,72 | 2,18 | 0,68 | 0,59 | 1,70 | 0,52 |
| 7 | 0,85 | 0,55 | 0,70 | 0,40 | 0,82 | 1,48 | 0,78 | 0,65 | 1,50 | 0,64 |
| 8 | 0,75 | 1,25 | 0,80 | 1,18 | 0,94 | 1,76 | 0,98 | 0,82 | 2,65 | 0,70 |
| 9 | 1,10 | 1,35 | 0,80 | 1,28 | 1,05 | 1,08 | 1,07 | 1,11 | 2,80 | 0,90 |
| 10 | 1,20 | -0,50 | 1,18 | 0,70 | 1,20 | 1,16 | 1,22 | 1,21 | 1,10 | 1,15 |

Таблица Б.8 – Массив данных $X_8$

| $X_8$ | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 |
|---|---|---|---|---|---|---|---|---|---|---|
| 1 | 20,10 | 20,25 | 20,66 | 20,83 | 20,48 | 21,65 | 20,62 | 20,27 | 20,44 | 21,89 |
| 2 | 20,30 | 20,35 | 20,78 | 20,92 | 20,52 | 21,75 | 20,78 | 20,33 | 20,56 | 21,96 |
| 3 | 20,48 | 20,50 | 20,86 | 20,32 | 20,96 | 21,84 | 20,86 | 20,46 | 20,71 | 21,69 |
| 4 | 20,52 | 20,56 | 20,93 | 20,34 | 20,71 | 21,96 | 20,93 | 20,54 | 20,69 | 21,71 |
| 5 | 20,70 | 20,65 | 20,96 | 20,46 | 20,88 | 21,12 | 20,97 | 20,64 | 20,88 | 21,46 |
| 6 | 20,72 | 20,71 | 21,08 | 20,54 | 20,90 | 21,04 | 20,10 | 20,72 | 20,90 | 21,54 |
| 7 | 20,89 | 20,75 | 21,17 | 20,64 | 20,92 | 21,06 | 20,24 | 20,76 | 20,92 | 21,10 |
| 8 | 20,90 | 21,10 | 20,45 | 20,74 | 21,10 | 21,24 | 20,45 | 20,88 | 21,09 | 20,04 |
| 9 | 20,91 | 20,15 | 20,58 | 20,76 | 21,14 | 21,37 | 20,58 | 20,92 | 21,11 | 21,16 |
| 10 | 21,11 | 21,05 | 20,08 | 21,06 | 20,30 | 21,48 | 21,08 | 21,04 | 20,05 | 21,28 |

Таблица Б.9 – Массив данных $X_9$

| $X_9$ | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 |
|---|---|---|---|---|---|---|---|---|---|---|
| 1 | 0,05 | 0,73 | 0,53 | 0,14 | 0,85 | 0,13 | 0,04 | 0,11 | 1,00 | 0,12 |
| 2 | 0,15 | 0,74 | 0,56 | 0,16 | 0,85 | 0,17 | 0,16 | 0,19 | 0,58 | 0,18 |
| 3 | 0,25 | 0,78 | 0,56 | 0,23 | 0,75 | 0,17 | 0,16 | 0,19 | 0,85 | 0,18 |
| 4 | 0,25 | 0,63 | 0,33 | 0,27 | 0,75 | 0,25 | 0,94 | 0,27 | 0,75 | 0,26 |
| 5 | 0,35 | 0,66 | 0,35 | 0,74 | 0,95 | 0,26 | 0,22 | 0,34 | 0,75 | 0,27 |
| 6 | 0,34 | 0,67 | 0,37 | 0,76 | 0,65 | 0,34 | 0,28 | 0,34 | 0,65 | 0,32 |
| 7 | 0,36 | 0,52 | 0,22 | 0,84 | 0,65 | 0,36 | 0,33 | 0,37 | 0,65 | 0,38 |
| 8 | 0,45 | 0,54 | 0,28 | 0,86 | 0,55 | 0,41 | 0,35 | 0,43 | 0,55 | 0,44 |
| 9 | 0,45 | 0,59 | 0,62 | 0,54 | 0,55 | 0,45 | 0,37 | 0,45 | 0,55 | 0,45 |
| 10 | 0,45 | 0,12 | 0,64 | 0,55 | 0,55 | 0,54 | 0,42 | 0,47 | 0,55 | 0,46 |

Таблица Б.10 – Массив данных $X_{10}$

| $X_{10}$ | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 |
|---|---|---|---|---|---|---|---|---|---|---|
| 1 | 93,05 | 94,45 | 93,24 | 93,44 | 94,90 | 93,80 | 93,20 | 94,92 | 93,06 | 93,13 |
| 2 | 93,10 | 94,55 | 93,30 | 93,50 | 94,70 | 93,10 | 93,09 | 94,75 | 93,10 | 93,14 |
| 3 | 93,15 | 94,30 | 93,36 | 93,56 | 94,50 | 93,12 | 93,11 | 94,48 | 93,10 | 93,08 |
| 4 | 93,10 | 94,65 | 93,46 | 93,26 | 94,10 | 93,14 | 93,18 | 94,32 | 93,14 | 93,10 |
| 5 | 93,25 | 94,95 | 93,54 | 93,31 | 94,25 | 93,28 | 93,24 | 94,02 | 93,25 | 93,28 |
| 6 | 93,30 | 94,05 | 93,50 | 93,36 | 94,35 | 93,30 | 93,30 | 94,18 | 93,30 | 93,31 |
| 7 | 93,35 | 94,15 | 93,64 | 93,00 | 93,90 | 93,32 | 93,36 | 93,87 | 93,35 | 93,32 |
| 8 | 93,50 | 93,08 | 93,76 | 93,10 | 93,90 | 93,45 | 93,46 | 93,91 | 93,65 | 93,45 |
| 9 | 93,50 | 93,10 | 93,92 | 93,12 | 93,65 | 93,55 | 93,54 | 93,62 | 93,70 | 93,55 |
| 10 | 93,65 | 93,12 | 94,08 | 93,18 | 93,70 | 93,66 | 93,50 | 93,78 | 93,75 | 93,66 |

Лекція №2

## ТЕМА: ПОХИБКИ ВИМІРЮВАНЬ І ЗАКОНИ РОЗПОДІЛУ

### 1. Похибки вимірювань

При проведенні досліджень поряд із визначенням вимірюваної величини необхідно оцінити і похибку вимірювань.

Похибка не може бути меншою від похибки вимірювального приладу.

Розрізняють абсолютну і відносну похибки:

– абсолютна

$$\Delta = a - x\ , \tag{1}$$

де $a$ – дійсне значення;

$x$ – виміряне;

– відносна (більш зручна)

$$\delta = \frac{\Delta}{a} \cdot 100\%\ . \tag{2}$$

Похибки поділяють на випадкові, систематичні і промахи. Промахи спричинені неправильним записом даних.

Якщо випадкові похибки спричинені різними факторами, дія яких неоднакова у кожному досліді, то вони носять випадковий характер і не можуть бути враховані.

Систематичні – фактори, що їх спричиняють, діють однаково при багаторазових вимірюваннях. Вони бувають чотирьох груп:

1. Їх природа може достатньо точно визначатись і вони називаються поправними.

2. Їх природа відома, але величина невідома. Наприклад, на приладі вказано максимальну похибку, а дійсна може бути меншою.

3. Про їх наявність нам не відомо.

4. Зумовлена властивостями матеріалу (наприклад, при вимірюванні циліндричних тіл – еліпсність).

Якщо експериментальні похибки виявлено та усунуто, то присутні ще випадкові похибки.

Випадкові похибки при проведенні вимірювань у одному і тому ж досліді змінюються як за величиною, так і за знаком. Абсолютна похибка вимірювальних приладів зі шкалами не перевищує ½ найменшої поділки шкали.

Отже, гранична випадкова похибка $\Delta_{ВП}$ є найбільш можливою похибкою, яка має місце при правильному користуванні справним приладом за відсутності систематичних похибок.

Для характеристики точності вимірювання користуються відносним значенням граничної похибки

$$\delta = \frac{\left|\left(a_{зр}\right)_{\max} - x_{\max}\right|}{\left(a_{зр}\right)_{\max}} \cdot 100\% , \tag{3}$$

де $\left(a_{зр}\right)_{\max}$ – результат вимірювання найбільшої можливої величини;

$x_{\max}$ – результат вимірювання цієї ж величини приладом, що використовується.

У довідниках прийнято наводити відносну граничну похибку $\Delta_{гр}$, поділену на найбільш можливе значення величини, яку може виміряти прилад.

У вимірювальних засобах розрізняють статичні і динамічні похибки.

*Статичні* – при вимірюванні постійних величин.

*Динамічні* – при вимірюванні змінних величин (наслідок дії сил інерції).

Залежно від значень допустимої похибки вимірювальні прилади поділяються на еталонні, зразкові (1...4-го розрядів) та робочі (I або II класу).

Вимірювання, за якими одержують дисперсію або середнє квадратичне відхилення однакової величини називають рівно точними.

Нерівноточні вимірювання поділяють на розряди, у яких вимірювання є рівно точними і загальна дисперсія визначається за формулою:

$$S_C^2 = \frac{n_1 S_{C1}^2 + n_2 S_{C2}^2 + \ldots + n_m S_{Cm}^2}{n} =$$

$$= \frac{n_1\left(x_{cp1} - x_{cp}\right) + n_2\left(x_{cp2} - x_{cp}\right) + \ldots + n_m\left(x_{cp\,m} - x_{cp}\right)}{n}, \tag{4}$$

де $n_1, \ldots, n_m$ – число вимірювань у кожному з $m$ розрядів;

$S_{C1}, \ldots, S_{Cm}$ – середнє квадратичне відхилення кожного розряду;

$m$ – число розрядів;

$x_{cp1}, \ldots, x_{cp\,m}$ – середні арифметичні у кожному розряді;

$x_{cp}$ – середнє арифметичне всіх вимірювань

Середня квадратична похибка середнього арифметичного:

$$S_{cpap} = \frac{S_c}{\sqrt{n}}. \tag{5}$$

## 2. Методи виключення результатів із грубими похибками

Грубі похибки (промахи) належать до похибок, які змінюються випадковим чином, при повторних спостереженнях. Вони явно перевищують за своїм значенням похибки, які викликані умовами проведення експерименту. Під промахом розуміється значення похибки, відхилення якого від центру розподілу суттєво перевищує значення, виправдане об'єктивними умовами вимірювань. Тому з погляду теорії імовірності поява промаху малоймовірна.

Причинами грубих похибок можуть бути неконтрольовані зміни умов вимірювань, несправність, помилки оператора та ін.

Для виключення грубих похибок застосовують апарат перевірки статистичних гіпотез.

Оскільки на практиці частіше зустрічаються вимірювання при невідомому СКО (обмежене число спостережень), далі розглянуті наступні критерії перевірки підозрілих (з погляду похибок) результатів спостережень: Ірвіна, Романовского, варіаційного розмаху, Діксона, Смірнова, Шовене.

Оскільки критериальні вимоги (коефіцієнти), що визначають межу, за якою перебувають "грубі" (у вигляді похибок) результати спостережень у різних авторів різні, то перевірку слід виконувати відразу по декільком критеріям. Остаточний висновок про приналежність "підозрілих" результатів розглянутої сукупності спостережень слід робити по більшості критеріїв. Крім цього вибір критерію для визначення грубих похибок повинен виконуватися після побудови гістограми результатів спостережень. По виду гистограммы виконується попередня ідентифікація виду закону розподілу (нормальний, близький до нормальному або відмінний від нього).

**<u>Критерій Ірвіна.</u>** Для отриманих експериментальних даних визначають коефіцієнт по формулі:

$$\lambda = \frac{x_{n+1} - x_n}{S}, \tag{6}$$

де $x_{n+1}$, $x_n$ - найбільші значення випадкової величини;

$S$ - середнє квадратичне відхилення, яке обчислене за всіма значеннями вибірки.

Потім цей коефіцієнт порівнюється з табличним значенням $\lambda_q$, можливі значення якого наведено в таблиці 1.

Таблиця 1

Критерій Ірвіна $\lambda_q$

| Число вимірів *n* | Рівень значимості | |
|---|---|---|
| | *q=0,05* | *q=0,01* |
| 1 | 2 | 3 |
| 2 | 2,8 | 3,7 |
| 3 | 2,2 | 2,9 |
| 10 | 1,5 | 2,0 |
| 20 | 1,3 | 1,8 |
| 30 | 1,2 | 1,7 |
| 50 | 1,1 | 1,6 |

| | | |
|---|---|---|
| 100 | 1,0 | 1,5 |
| 400 | 0,9 | 1,3 |
| 1000 | 0,8 | 1,2 |

Якщо $\lambda > \lambda_q$, то нульова гіпотеза не підтверджується, тобто результат - помилковий, і він повинен бути виключений при подальшій обробці результатів спостережень.

**Критерій Романовского.** Конкуруюча гіпотеза про наявність грубих похибок у підозрілих результатах підтверджується, якщо виконується нерівність:

$$\left| x_{i\ під} - \overline{X}_{ц.р.} \right| \geq t_p S, \tag{7}$$

де $t_p$ - квантиль розподілу Стьюдента при заданій довірчій імовірності із числом ступенів свободи $k = n - k_n$ ( $k_n$ - число підозрілих результатів спостережень). Фрагмент квантилів для розподілу Стьюдента представлено в таблиці 2.

Точкові оцінки розподілу $\overline{X}_{ц.р.}$ і СКВ $S$ результатів спостережень обчислюються без врахування $k_n$ підозрілих результатів спостережень.

Таблица 2

Критерий Стьюдента $t_{КР}$ (квантилі Стьюдента)

| Довір-ча імовір-ність p | Число степенів свободи | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | 3 | 4 | 5 | 6 | 8 | 10 | 12 | 18 | 22 | 30 | 40 | 60 | 120 | ∞ |
| 0,90 | 2,35 | 2,13 | 2,01 | 1,94 | 1,86 | 1,81 | 1,78 | 1,73 | 1,72 | 1,70 | 1,68 | 1,67 | 1,66 | 1,64 |
| 0,95 | 3,18 | 2,78 | 2,57 | 2,45 | 2,31 | 2,23 | 2,18 | 2,10 | 2,07 | 2,04 | 2,02 | 2,00 | 1,98 | 1,96 |
| 0,99 | 5,84 | 4,60 | 4,03 | 3,71 | 3.36 | 3,17 | 3,06 | 2,98 | 2,82 | 2,75 | 2,70 | 2,86 | 2,62 | 2,58 |

**Критерій варіаційного розмаху.** Є одним із простих методів виключення грубих похибок вимірювань (промахів). Для його використання визначають розмах варіаційного ряду впорядкованої сукупності спостережень $(x_1 < x_2 < \ldots < x_к < \ldots < x_n)$.

$$R_n = x_n - x_1. \qquad (8)$$

Якщо який-небудь член варіаційного ряду, наприклад $x_к$, сильно відрізняється від усіх інших, то роблять перевірку, використовуючи наступну нерівність:

$$\overline{X} - z \cdot R_n < x_k < \overline{X} + z \cdot R_n, \qquad (9)$$

де $\overline{X}$ - вибіркове середнє арифметичне значення, обчислене після виключення передбачуваного промаху;

$z$ - критеріальне значення.

Нульову гіпотезу (про відсутність грубої похибки) ухвалюють, якщо зазначена нерівність виконується. Якщо $x_к$ не задовольняє умові (9), то цей результат виключають із варіаційного ряду.

Коефіцієнт $z$ залежить від числа членів варіаційного ряду $n$, що представлено в таблиці 3.

Таблиця 3

Критерій варіаційного розмаху

| $n$ | 5 | 6 | 7 | 8-9 | 10-11 | 12-15 | 16-22 | 23-25 | 26-63 | 64-150 |
|---|---|---|---|---|---|---|---|---|---|---|
| $z$ | 1,7 | 1,6 | 1,5 | 1,4 | 1,3 | 1,2 | 1,1 | 1,0 | 0,9 | 0,8 |

**Критерій Діксона.** Критерій заснований на припущенні, що похибка вимірювань підлягає нормальному закону (попередньо необхідна побудова гістограми результатів спостережень) і перевірка гіпотези про приналежність нормальному закону розподілу. При використанні критерію обчислюють коефіцієнт Діксона (досліджуване значення критерію) для перевірки найбільшого або найменшого екстремального значення залежно від числа вимірювань. У таблиці 4 наведені формули для обчислення коефіцієнтів. Коефіцієнти $r_{10}$, $r_{11}$ застосовують, коли є одне викидання, а $r_{21}$ і $r_{22}$ - коли два викидання. Потрібне первісне впорядкування результатів вимірювань (об'єму вибірки). Критерій застосовується, коли вибірка може містити більше однієї грубої похибки.

Таблиця 4

Формули коефіцієнтів Діксона

| Число вимірювань $n$ (об'єм вибірки) | Коефіцієнт Діксона | Для найменшого екстремального значення параметра | Для найбільшого експериментального параметра |
|---|---|---|---|
| 3-7 | $r_{10}$ | $\frac{x_2 - x_1}{x_n - x_1}$ | $\frac{x_n - x_{n-1}}{x_n - x_1}$ |
| 8-10 | $r_{11}$ | $\frac{x_2 - x_1}{x_{n-1} - x_1}$ | $\frac{x_n - x_{n-1}}{x_n - x_2}$ |
| 11-13 | $r_{21}$ | $\frac{x_3 - x_1}{x_{n-1} - x_1}$ | $\frac{x_n - x_{n-2}}{x_n - x_2}$ |
| 14-25 | $r_{22}$ | $\frac{x_3 - x_1}{x_{n-2} - x_1}$ | $\frac{x_n - x_{n-2}}{x_n - x_3}$ |

Обчислені для вибірки по формулах значення коефіцієнтів Діксона $r$ порівнюють із прийнятим (табличним) значенням критерію Діксона $r_q$ (таблиця 5).

Нульова гіпотеза про відсутність грубої похибки виконується, якщо виконується нерівність $r < r_q$.

Якщо $r > r_q$, то результат є грубою похибкою і виключається з подальшої обробки.

Таблиця 5

Критеріальні значення коефіцієнтів Діксона ( при прийнятому рівні значимості $q$)

| Статистика | Число вимірювань | $r_q$ при рівні значимості $q$ | | | |
|---|---|---|---|---|---|
| | | 0,1 | 0,05 | 0,02 | 0,01 |
| 1 | 2 | 3 | 4 | 5 | 6 |
| $r_{10}$ | 3 | 0,886 | 0,941 | 0,976 | 0,988 |
| | 4 | 0,679 | 0,765 | 0,846 | 0,899 |
| | 5 | 0,557 | 0,642 | 0,729 | 0,780 |
| | 6 | 0,482 | 0,560 | 0,644 | 0,698 |
| | 7 | 0,434 | 0,507 | 0,586 | 0,637 |

Продовження табл. 5

| 1 | 2 | 3 | 4 | 5 | 6 |
|---|---|---|---|---|---|
| $r_{11}$ | 8 | 0,479 | 0,554 | 0,631 | 0,683 |
| | 9 | 0,441 | 0,512 | 0,587 | 0,636 |
| | 10 | 0,409 | 0,477 | 0,551 | 0,597 |
| $r_{21}$ | 11 | 0,517 | 0,576 | 0,538 | 0,679 |
| | 12 | 0,490 | 0,546 | 0,605 | 0,642 |
| | 13 | 0,467 | 0,521 | 0,578 | 0,615 |
| $r_{22}$ | 14 | 0,462 | 0,546 | 0,602 | 0,641 |
| | 15 | 0,472 | 0,525 | 0,579 | 0,616 |
| | 16 | 0,452 | 0,507 | 0,559 | 0,595 |
| | 17 | 0,438 | 0,490 | 0,542 | 0,577 |
| | 18 | 0,424 | 0,475 | 0,527 | 0,561 |
| | 19 | 0,412 | 0,462 | 0,514 | 0,547 |
| | 20 | 0,401 | 0,450 | 0,502 | 0,535 |
| | 21 | 0,391 | 0,440 | 0,491 | 0,524 |
| | 22 | 0,382 | 0,430 | 0,481 | 0,514 |
| | 23 | 0,374 | 0,421 | 0,472 | 0,505 |
| | 24 | 0,367 | 0,413 | 0,464 | 0,497 |
| | 25 | 0,360 | 0,406 | 0,457 | 0,489 |

**Критерії "3σ", Райта.** Критерій "правило трьох сигм" є одним з найпростіших для перевірки результатів, що підлягають нормальному закону розподілу. Суть правила трьох сигм: якщо випадкова величина розподілена нормально, то абсолютна величина її відхилення від математичного очікування не перевершує потроєного середнього квадратичного відхилення.

На практиці правило трьох сигм застосовують так: якщо розподіл досліджуваної випадкової величини невідомий, але умова, що зазначена в наведеному правилі, виконується, тобто підстави допускати, що досліджувана величина розподілена нормально; а якщо ні, то вона не розподілена нормально. Із цією метою для вибірки (включаючи підозрілий результат) обчислюється центр розподілу і оцінка СКВ результата спостережень. Результат, який задовольняє умову

$$\left|x_{i\ nid} - \overline{X}_{ц.р.}\right| \geq 3S, \qquad (10)$$

вважається, що має грубу похибку і віддаляється, а раніше обчислені характеристики розподілу уточнюються.

Цьому критерію аналогічний критерій Райта, заснований на тому, що якщо залишкова похибка більше чотирьох сігм, то цей результат вимірювань є грубою похибкою і повинен бути виключений при подальшому обробітку. Обидва критері надійні при числі вимірювань більше 20...50. Їх доцільно застосовувати, коли відома величина генерального середньоквадратичного відхилення ($S$).

Може виявитися, що при нових значеннях $\overline{X}_{ц.р.}$ і $S$ інші результати потраплять у категорію аномальних.

**Критерій Смирнова** використовується при об'ємах вибірки $n$>25 або при відомих значеннях генерального середнього і СКВ. Він установлює лояльніші межі грубої похибки. Для реалізації цього критерію обчислюються дійсні значення квантилів розподілу (досліджуване значення критерію) по формулі:

$$\beta = \frac{\max\left|x_{i\ nid} - \overline{X}\right|}{S}. \qquad (11)$$

Знайдене значення рівняється із критериальным, наведеним у таблиці 6.

Таблиця 6

Квантилі розподілу $\beta_q$

| Об'єм вибірки $n$ | Граничне значення $\beta_q$ при рівні значимості $q$ | | | | |
|---|---|---|---|---|---|
| | 0,100 | 0,050 | 0,0010 | 0,005 | 0,001 |
| 1 | 2 | 3 | 4 | 5 | 6 |
| 1 | 1,282 | 1,645 | 2,326 | 2,576 | 3,090 |
| 2 | 1,632 | 1,955 | 2,575 | 2,807 | 3,290 |
| 3 | 1,818 | 2,121 | 2,712 | 2,935 | 3,403 |
| 4 | 1,943 | 2,234 | 2,806 | 3,023 | 3,481 |
| 5 | 2,036 | 2,319 | 2,877 | 3,090 | 3,540 |
| 6 | 2,111 | 2,386 | 2,934 | 3,143 | 3,588 |

Продовження табл. 6

| 1 | 2 | 3 | 4 | 5 | 6 |
|---|---|---|---|---|---|
| 7 | 2,172 | 2,442 | 2,981 | 3,188 | 3,628 |
| 8 | 2,224 | 2,490 | 3,022 | 3,227 | 3,662 |
| 9 | 2,269 | 2,531 | 3,057 | 3,260 | 3,692 |
| 10 | 2,309 | 2,568 | 3,089 | 3,290 | 3,719 |
| 15 | 2,457 | 2,705 | 3,207 | 3,402 | 3,820 |
| 20 | 2,559 | 2,799 | 3,289 | 3,480 | 3,890 |
| 25 | 2,635 | 2,870 | 3,351 | 3,539 | 3,944 |
| 30 | 2,696 | 2,928 | 3,402 | 3,587 | 3,988 |
| 40 | 2,792 | 3,015 | 3,480 | 3,662 | 4,054 |
| 50 | 2,860 | 3,082 | 3,541 | 3,716 | 4,108 |
| 100 | 3,076 | 3,285 | 3,723 | 3,892 | 4,263 |
| 250 | 3,339 | 3,534 | 3,946 | 4,108 | 4,465 |
| 500 | 3,528 | 3,703 | 4,108 | 4,263 | 4,607 |

**Критерій Шовене** застосовується для законів, що не суперечать нормальному, і будується на визначенні числа очікуваних результатів спостережень $n_{оч}$, які мають настільки ж великі похибки, як і підозрілий. Гіпотеза про наявність грубої попохибки приймається, якщо виконується умова:

$$n_{оч} < 0{,}5.$$

Порядок перевірки гіпотези наступний:

1) обчислюються середнє арифметичне $\overline{X}$ і СКВ $S$ результатів спостережень для всієї вибірки;

2) з таблиці нормованого нормального розподілу (Додаток В - інтегральна функція нормованого нормального розподілу) по величині $z = \dfrac{\left| x_{i\ під} - \overline{X}_{ц.р.} \right|}{S}$ визначається ймовірність появи підозрілого результату в генеральній сукупності чисел $n$:

$$P\left( zS < \left| x_{i\ під} - \overline{X}_{ц.р.} \right| \right). \qquad (12)$$

3) число очікуваних результатів $n_{оч}$ визначається по формулі:

$$n_{оч} = n \cdot P. \qquad (13)$$

Зазначені вище критерії в багатьох випадках виявляються достатньо "жорсткими". Тоді рекомендується користуватися критерієм грубої похибки *k*, що залежать від об'єму вибірки *n* і прийнятої довірчої ймовірності *P*.

Таблиця 7

Залежність критерію грубої похибки від об'єму вибірки *n* і довірчої ймовірності *P*

| *n* | *P* = 95,00 | *P* = 99,00 | *P* = 99,73 | *n* | *P* = 95,00 | *P* = 99,00 | *P* = 99,73 |
|---|---|---|---|---|---|---|---|
| 9 | 4,42 | 7,10 | 11,49 | 25 | 3,84 | 5,14 | 6,25 |
| 10 | 4,31 | 6,99 | 10,26 | 30 | 3,80 | 5,00 | 5,95 |
| 12 | 4,16 | 6,38 | 8,80 | 40 | 3,75 | 4,82 | 5,56 |
| 15 | 4,03 | 5,88 | 7,66 | 50 | 3,73 | 4,70 | 5,34 |
| 20 | 3,90 | 5,41 | 6,73 | | | | |

Для розподілів, відмінних від нормального, таких класів, як двох модальні кругловершинних композицій нормального і дискретного розподілу з ексцесом $\varepsilon$=1,5-3,0; гостровершинних двумодальних; композицій дискретного двозначного розподілу і розподілу Лапласа з ексцесом $\varepsilon$ =1,5-6,0; композицій рівномірного розподілу з експонентним розподілом ексцесу $\varepsilon$=1,8-6,0 і класом експонентних розподілів у межах зміни ексцесу $\varepsilon$=1,8-6,0 границя грубої похибки визначається величиною $\pm\left(t_{гр} \cdot \sigma\right)$ або $\pm\left(t_{гр} \cdot S\right)$, де:

$$t_{гр} = 1{,}2 + 3{,}6 \cdot \left(1 - \gamma\right) \cdot \lg\frac{n}{10}, \qquad (14)$$

де $\gamma$ - контрэксцесс;

$$t_{гр} = 1{,}55 + 0{,}8 \cdot \sqrt{\varepsilon - 1} \cdot \lg\frac{n}{10}, \qquad (14)$$

Похибка у визначенні оцінок *S* СКВ і $t_{гр}$ є негативно коррелированными, тобто зростання СКВ *S* супроводжується зменшенням $t_{гр}$. Тому визначення меж грубої похибки для законів, відмінних від нормального, з ексцесом $\varepsilon < 6$ за допомогою критерію $t_{гр}$ є досить точним і може широко використовуватися на практиці.

Оцінки $\overline{X}$, $S$ і $\varepsilon$ повинні обчислюватися після виключення підозрілих результатів з вибірки. Після розрахунків границь грубої похибки результати спостережень, які опинилися усередині границь, вертаються, а раніше знайдені характеристики розподілу уточнюються.

Для рівномірного розподілу за межі грубої похибки можна прийняти величину ±1,8S.

## 3. Приклад виключення результатів із грубими похибками

Розглянемо приклад застосування критеріїв для виключення грубих похибок при вимірюванні швидкості ударної хвилі. Отримані результати, представлені в таблиці 8.

Таблиця 8

Результати спостережень

| Швидкість $V$, км/с | 3,42 | 3,43 | 3,44 | 3,45 | 3,46 | 3,47 | 3,48 | 3,49 | 3,50 |
|---|---|---|---|---|---|---|---|---|---|
| Частота $m_i$ | 1 | 2 | 4 | 5 | 4 | 2 | 1 | 0 | 1 |

Потрібно визначити, не чи містить результат спостереження $V=3,50$ км/с грубу похибку.

Для графічного визначення виду закону розподілу побудуємо гістограму. При побудові розбивка на інтервали здійснюємо таким чином, щоб виміряні значення виявилися серединами інтервалів, що показано на рис. 1.

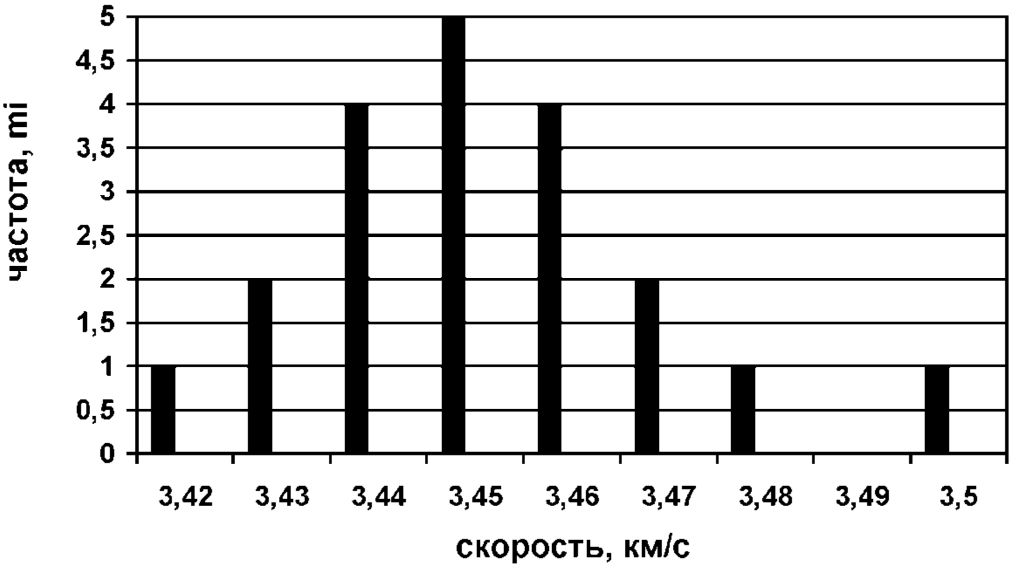


Рис. 1. Загальний вид гістограми

По виду гістограми приблизно ідентифікуємо дослідний розподіл нормальним. Обчислюємо оцінки $\overline{X}$, $S$:

$$\overline{X} = \frac{1}{n}\sum_{i=1}^{n} x_i = 3{,}452,$$

$$\overline{S} = M_k \cdot \sqrt{\frac{1}{n-1}\sum_{i=1}^{n}\left(x_i - \overline{X}\right)^2} = 0{,}019 км/с.$$

Вирішимо задачу за допомогою критеріїв, встановлених для нормального розподілу.

**<u>Перевірка за критерієм Ірвіна.</u>** Для отриманих експериментальних даних визначаємо коефіцієнт по формулі:

$$\lambda = \frac{x_{n+1} - x_n}{S} = \frac{3{,}50 - 3{,}42}{0{,}057} = 1{,}40.$$

Порівняємо цей коефіцієнт з табличним значенням $\lambda_q$. Оскільки $\lambda = 1{,}4 > \lambda_{0,05} = 1{,}3$, то нульова гіпотеза не підтверджується, тобто результат 3,50 (результат, який знаходиться далі від центру розподілу) - помилковий, і він повинен бути виключений при подальшій обробці результатів спостережень.

**Перевірка за критерієм Романовского.** Конкуруюча гіпотеза про наявність грубих похибок у підозрілих результатах підтверджується, якщо виконується нерівність:

$$\left|x_{i\ nід} - \overline{X}_{ц.р.}\right| \geq t_{0,05} S,$$

Обчислимо дальність підозрілого результату від центру розподілу:

$$\left|x_{i\ nід} - \overline{X}\right| = 3,50 - 3,452 = 0,048 км/с.$$

$$\left|x_{i\ nід} - \overline{X}\right| = \left|3,42 - 3,452\right| = 0,032 км/с.$$

Визначимо межу похибки: $t_{0,05} S = 2,1 \cdot 0,019 = 0,0399$.

Оскільки $0,048 км/с >$ 0,0399 *км/с*, то можна зробити висновок, що результат *V=3,5 км/с* є грубою похибкою і повинен бути виключений з подальшого обробітку. А $0,032 км/с <$ 0,0399 *км/с*, і тому результат *V=3,42 км/с* не є промахом і може використовуватись у подальшому обробітку даних.

**Перевірка за критерієм варіаційного розмаху**

Визначаємо розмах ряду за формулою:

$$R_n = x_n - x_1 = 3,50 - 3,42 = 0,08 км/с.$$

Обираємо члени варіаційного ряду 3,50 і 3,42, які сильно відрізняється від усіх інших, і робимо перевірку, використовуючи наступну нерівність:

$$\overline{X} - z \cdot R_n < x_k < \overline{X} + z \cdot R_n,$$

$$3,452 - 1,1 \cdot 0,08 < 3,50 < 3,452 + 1,1 \cdot 0,08,$$

Отже, обидва значення входять в даний діапазон і можуть бути визначені, як нормальні.

**Перевірка за критерієм Діксона**

Знаходимо коефіцієнти Діксона для об’єму вибірки в 20 одиниць.

$$r_{22}^{\min} = \frac{x_3 - x_1}{x_{n-2} - x_1} = \frac{3,44 - 3,42}{3,47 - 3,42} = 0,4\,.$$

$$r_{22}^{\min} = \frac{x_n - x_{n-2}}{x_n - x_3} = \frac{3{,}50 - 3{,}47}{3{,}50 - 3{,}44} = 0{,}5.$$

Нульова гіпотеза про відсутність грубої похибки виконується, якщо виконується нерівність $r < r_{0,05}$. У нашому випадку $r_{0,05} = 0{,}45$, тоді $r_{22}^{\min} < r_{0,05}$, тобто значення 3,42 не є промахом і може використовуватись у подальших дослідженнях. І $r_{22}^{\min} > r_{0,05}$, тобто значення 3,50 є промахом і може бути вилучене з подальшого дослідження.

**Перевірка за критерієм "3σ", Райта.** Обчислимо дальність підозрілого результату від центру розподілу:

$$x_{i\ під} - \overline{X} = 3{,}50 - 3{,}452 = 0{,}048 км/с.$$

Визначимо межу похибки: $3S = 3 \cdot 0{,}019 = 0{,}057$ *км/с*.

Оскільки $x_{i\ під} - \overline{X} = 0{,}048 км/с < 3S = 0{,}057$ *км/с*, то можна зробити висновок, що результат $V = 3{,}5$ *км/с* не містить грубої похибки.

**Перевірка за критерієм Смірнова.** Для реалізації цього критерію обчислюємо дійсні значення квантилів розподілу (досліджуване значення критерію) по формулі:

$$\beta = \frac{\max\left|x_{i\ під} - \overline{X}\right|}{S} = \frac{\left|3{,}5 - 3{,}452\right|}{0{,}019} = 2{,}52.$$

Так як $\beta = 2{,}52 < \beta_{0,05} = 2{,}799$, то дане значення можна вважати нормальним і не виключати його з подальшого дослідження.

**Перевірка за критерієм Шовене.** При знаходженні характеристик розподілу беруть участь усі спостереження і тому $X$=3,452 км/с; $S$=0,019 км/с.

Обчислюємо квантиль $z$ по формулі:

$$z = \frac{\left|x_{i\ під} - \overline{X}_{ц.р.}\right|}{S} = \frac{3{,}5 - 3{,}452}{0{,}019} = 2{,}526.$$

По таблиці В. 1 Додатка У визначаємо ймовірність виходу результатів за квантиль ± z:

$$P\left(2{,}526 \cdot S < \left| x_{i\ \text{під}} - \overline{X}_{ц.р.} \right|\right) = (1 - 0{,}9943) \cdot 2 = 0{,}0114.$$

Тоді очікуване число спостережень із результатом $|x_i| > 3{,}50 км / с$ .

$$n_{оч} = n \cdot P\left(2{,}526 \cdot S < \left| x_{i\ \text{під}} - \overline{X}_{ц.р.} \right|\right) = 20 \cdot 0{,}0114 = 0{,}228.$$

Тому що $n_{оч}$ < 0,5, то доходимо висновку про наявність грубої похибки в результаті спостереження $x_i$ =3,50 км/с.

Оскільки більшість критеріїв (3 з 4-х розглянутих) показали наявність грубої похибки, то результат спостереження необхідно виключити з вибірки.

## Модуль 2

## ПАСИВНІ МЕТОДИ МАТЕМАТИЧНОЇ СТАТИСТИКИ

Лекція №3

## ТЕМА: ОДНОФАКТОРНИЙ ДИСПЕРСІЙНИЙ АНАЛІЗ

### 1. Сутність і задачі дисперсійного аналізу

Дисперсійний аналіз - це дослідження неоднородною числового матеріалу шляхом розкладання загальної дисперсії на складові. Дисперсійний аналіз припускає нормальний розподіл вимірюваної величини і дозволяє визначити значимість того або іншого факгора на функцію відгуку. Експеримент проводиться, задаючи кожний фактор на $k$ рівнях, а на кожному рівні параметр виміряється в $n$ паралельних доскідах. Дисперсійний аналіз особливо ефективний при вивченні декількох факторів.

Отже, задачами дисперсійного аналізу є:

1. Оцінка значимості впливу факторів на функцію відгуку.
2. Встановлення обмежень значимості фактору (тобто в якому діапазоні змін даний фактор значимий).

Дисперсійний аналіз широко використовується в різних видах досліджень. Приведемо два приклади, що наочно описують застосування дисперсійного аналізу.

Відомо, що Земля обертається навколо Сонця і навколо своєї осі. Однак останнім часом з'ясувалося (зі спостереження по зірках), що Земля піддається ще невеликим зрушенням і переміщенням (орбіти, осі і т.д.). Щоб оцінити, на скільки значимі ці відхилення по відношенню до похибки приладів, використовувався дисперсійний аналіз.

На будь-якому виробництві продукція, що випускається, проходить через кілька стадій обробітку. Щоб поліпшити якість продукції, немає необхідності вдосконалювати всі стадії. Можна оцінити за допомогою дисперсійного аналізу, яка зі стадій впливає на якість продукції, і її вдосконалювати в першу чергу.

Дисперсійний аналіз підрозділяється на однофакторний і багатофакторний. Розглянемо однофакторний дисперсійний аналіз.

## 2. Однофакторний дисперсійний аналіз

### 2.1. Основні положення і допущення

В однофакторному дисперсійному аналізі оцінюється значимість одного фактору при сталості всіх інших, тобто оцінюється ступінь впливу цього фактору на функцію відгуку. При цьому даний фактор варіюється на $k$ рівнях. Для кожного рівня проводиться рівне число дослідів - $n$. Загальне число досвідів: $N = kn$.

Наприклад, якщо розглядається значимість впливу товщина леза (фактор) на зусилля різання стебел (функція відгуку), то товщину леза встановлюють на декількох рівнях, припустимо 1, 2 і 3 мм ($k$=3, тобто розглядаэться три рівні). Для кожної товщини леза проводять три паралельні досліди ($n$=3). При цьому загальне число дослідів $N=3{\cdot}3=9$ дослідів. Одержуємо 9 значень функції відгуку (зусилля різання стебел) при різних температурах.

Для зручності обчислень будується таблиця дисперсійного аналізу, де по горизонталі розташовуються рівні фактору, а по вертикалі - кількість паралельних дослідів для кожного рівня.

Таблиця 1

Таблиця однофакторного дисперсійного аналізу

| j\i | 1 | 2 | .... | $k$ |
|---|---|---|---|---|
| 1 | $Y_{11}$ | $Y_{21}$ | .... | $Y_{k1}$ |
| 2 | $Y_{12}$ | $Y_{22}$ | .... | $Y_{k2}$ |
| ... | .... | .... | .... | .... |
| $n$ | $Y_{1n}$ | $Y_{2n}$ | .... | $Y_{kn}$ |

де $Y_{ij}$ - значення функції відгуку для *i*-того рівня, *j*-того номера паралельного досліду.

Значення функції відгуку будемо розглядати як суму трьох складових:

$$Y_{ij} = \mu + \gamma_i + \varepsilon_{ij}\,, \tag{1}$$

де $\mu = M\{Y_{ij}\}$ - математичне очікування функції відгуку;

$\gamma_i$ - вплив фактору на *i*-тому рівні;

$\varepsilon_{ij}$ - похибка вимірювання на *i*-тому рівні.

Якщо середні значення функції відгуку рівні на всіх *k* рівнях: $Y_1=Y_2=...=Y_k$, тоді фактор незначимо впливає на функцію відгуку ($\gamma_i \approx 0$).

При виконанні дисперсійного аналізу приймаються наступні допущення:

1. Випадкові похибки вимірювань мають нормальний розподіл.

2. Фактори впливають лише на зміну середніх значень, а дисперсія вимірювань залишається постійної, експерименти є равноточними. Дисперсії паралельних дослідів повинні бути однорідні.

3. Число паралельних дослідів однакове для всіх рівнів фактора.

**2.2. Порядок виконання однофакторного дисперсійного аналізу**

1. Для обраного фактору А визначити число і значення рівнів, на якому буде варіюватися фактор.

2. Провести *n* паралельних визначень функції відгуку для кожного рівня фактору А.

3. Побудувати таблицю однофакторного дисперсійного аналізу, увівши в ячейки таблиці отримані значення функції відгуку.

4. Якщо отримані значення функції відгуку незручні для розрахунків, можна перейти до їх перетворених значень.

5. Знайти середні значення вимірів на кожному *i*-ому рівні:

$$\overline{Y_i} = \frac{\sum_{j=1}^{n} Y_{ij}}{n}. \tag{2}$$

6. Знайти загальне середнє значення функції відгуку для всієї вибірки з $N$ вимірів:

$$\overline{Y} = \frac{1}{k}\sum_{i=1}^{k}\overline{Y_i} = \frac{1}{N}\sum_{i=1}^{k}\sum_{j=1}^{n} Y_{ij}. \tag{3}$$

7. Знайти загальну вибіркову дисперсію $S^2$:

$$S^2 = \frac{\sum_{i=1}^{k}\sum_{j=1}^{n}\left(Y_{ij} - \overline{Y}\right)^2}{N-1}. \tag{4}$$

8. Загальна вибіркова дисперсія розкладається на складові, які характеризують вплив фактору А і фактору випадковості. Фактор випадковості при цьому легко оцінити завдяки наявності повторних дослідів на кожному рівні. Для цього визначимо вибіркову дисперсію на кожному рівні:

$$S_i^2 = \frac{1}{n-1}\sum_{j=1}^{n}\left(Y_{ij} - \overline{Y_i}\right)^2. \tag{5}$$

9. Визначити вибіркову дисперсію, що характеризує фактор випадковості:

$$S_{вип}^2 = \frac{1}{k}\sum_{i=1}^{k} S_i^2. \tag{6}$$

10. Якщо врахувати, що дисперсія фактору А не пов'язана з якою-небудь випадковою величиною, то визначимо її але формулі:

$$S_A^2 = \frac{n}{k-1}\sum_{i=1}^{k}\left(Y_i - \overline{Y}\right)^2. \tag{7}$$

11. Якщо дисперсія $S_A^2$ значимо відрізняється від $S_{вип}^2$, вплив фактору А вважається істотним. Перевіряється дана гіпотеза по критерию Фішера. Вплив фактору А вважається значимим, якщо

$$\frac{S_A^2}{S_{вип}^2} > F_{табл(1-p)}\left(f_1;\ f_2\right), \tag{8}$$

де $F_{табл}$ - табличне значення критерію Фишера для довірчої імовірності - р і чисел ступенів свободи: $f_1=k-1$; $f_2=k(n-1)=N-k$.

### 2.3. Алгоритм розрахунку для однофакторного дисперсійного аналізу

1. Використовуючи таблицю однофакторного дисперсійного аналізу (табл. 1.), знайти суми по стовпцям:

$$A_i = \sum_{j=1}^{n} Y_{ij}. \tag{9}$$

2. Знайти суму квадратів всіх спостережень:

$$SS_1 = \sum_{i=1}^{k}\sum_{j=1}^{n} Y_{ij}^2. \tag{10}$$

3. Знайти суму квадратів сум по стовпцях, ділену на число вимірювань у стовпці:

$$SS_2 = \frac{1}{n}\sum_{i=1}^{k} A_i^2. \tag{11}$$

4. Знайти квадрат загальної суми, ділений на число всіх вимірювань (коригувальний член):

$$SS_3 = \frac{1}{N}\sum_{i=1}^{k}\left(A_i\right)^2. \tag{12}$$

5. Знайти суму квадратів для стовпця:

$$SS_A = SS_2 - SS_3. \tag{13}$$

6. Знайти загальну суму квадратів:

$$SS_{заг} = SS_1 - SS_3. \tag{14}$$

7. Знайти залишкову суму квадратів для оцінки похибки експерименту:

$$SS_{зал} = SS_1 - SS_2. \tag{15}$$

8. Знайти дисперсію фактора А:

$$S_A^2 = \frac{SS_A}{k-1}. \tag{16}$$

9. Знайти дисперсію фактору випадковості:

$$S_{вип}^2 = \frac{SS_{зал}}{k(n-1)}. \tag{17}$$

10. Резуль таты розрахунків можна представити у вигляді табл. 2.

Таблиця 2

Розрахунки однофакторного дисперсійного аналізу

| Джерело дисперсії | Число ступенів свободи | Сума квадратів | Дисперсії |
|---|---|---|---|
| Фактор А | *k-1* | $SS_A = SS_2 - SS_3.$ | $S_A^2 = \frac{SS_A}{k-1}.$ |
| Залишок | *k(n-1)* | $SS_{зал} = SS_1 - SS_2.$ | $S_{вип}^2 = \frac{SS_{зал}}{k(n-1)}.$ |
| Загальна сума | *kn- 1* | $SS_{заг} = SS_1 - SS_3.$ | $S^2 = \frac{SS_{заг}}{kn-1}.$ |

Якщо відношення дисперсії фактору А до дисперсії похибки менше або дорівнює табличному значенню критерію Фішера:

$$\frac{S_A^2}{S_{вип}^2} \le F_{табл(1-p)}\left(f_1;\ f_2\right), \tag{18}$$

то вплив фактору А слід уважати незначним. При цьому загальна дисперсія пов'язана лише з фактором випадковості і може служити оцінкою для дисперсії відтворюваності.

При інтерпретації результатів дисперсійного аналізу необхідно мати на увазі, що дуже низьке значення дисперсійного відношення (18) може бути пов'язане з тим, що вплив якогось важливого не контрольованого фактору не було рандомізовано. Це може збільшити дисперсію всередині рівнів, а дисперсію між рівнями залишити незмінною, що зменшує дисперсійне відношення. Результати експерименту при цьому не підкоряються моделі (1).

## 2.4. Оцінка обмежень значимості фактора

Щоб визначити, з якого рівня даний фактор стає значимий, скористаємося критерієм Стьюдента. Дня цього визначимо дисперсію відхилень між двома середніми значеннями функції відгуку для двох різних рівнів фактору, наприклад, *1*-того і *m*-того рівнів факторів ($1 \neq m$):

$$S^2_{(\overline{Y_1}-\overline{Y_m})} = \frac{2}{n} S^2_{вип}\,, \tag{19}$$

Далі знайдемо табличне значення критерію Стьюдента ($t_{p,\ f}$) для довірчої ймовірності $p$ (звичайно $p=0,95$) і числа ступенів свободи $f=2n-2$. Знайдемо критерій *K*, відносно якого будемо порівнювати різниці середніх значень функції відгуку для кожних двох рівнів:

За реперний (основний) рівень можна прийняти той, де спостерігається найбільше (або найменше) середнє значення функції відгуку, і выдносно нього порівнювати інші рівні фактору.

Якщо $\left|\overline{Y_1} - \overline{Y_m}\right| > K$ , то вплив фактору значимий на цьому рівні ы навпаки.

## 2.5. Приклад використання однофакторного дисперсійного аналізу в сільському господарстві

Для визначення впливу гостроти леза ножа подрібнювача на зусилля різання стебел кукурудзи, проведено 5 серій дослідів, що відрізняються гостротою леза ножа: $\varepsilon_1$=0,05 мм, $\varepsilon_2$=0,10 мм, $\varepsilon_3$=0,20 мм, $\varepsilon_4$=0,40 мм, $\varepsilon_5$=0,80 мм ( при сталості інших умов експерименту). При кожній гостроті леза проведено шість паралельних дослідів для визначення зусилля різання стебел кукурудзи в динамічних умовах на маятниковому копрі. Чи впливає гострота леза ножа подрібнювача на результат вимірів? І якщо впливає, то з якого значення? Результати вимірювань приведені в табл. 3.

Таблиця 3

Зусилля різання стебел кукурудзи в динамічних умовах при п'яти рівнях гостроти леза ножа

| № | $\varepsilon_1$ | $\varepsilon_2$ | $\varepsilon_3$ | $\varepsilon_4$ | $\varepsilon_5$ |
|---|---|---|---|---|---|
| 1 | 0,360 | 0,390 | 0,420 | 0,400 | 0,470 |
| 2 | 0,410 | 0,360 | 0,410 | 0,440 | 0,440 |
| 3 | 0,370 | 0,400 | 0,410 | 0,420 | 0,420 |
| 4 | 0,390 | 0,410 | 0,380 | 0,450 | 0,430 |
| 5 | 0,370 | 0,400 | 0,430 | 0,390 | 0,450 |
| 6 | 0,380 | 0,380 | 0,400 | 0,410 | 0,430 |
| $\overline{A_i}$ | 0,380 | 0,390 | 0,410 | 0,420 | 0,440 |

Розв'язок

Перетворимо значення функції відгуку (зусилля різання стебел кукурудзи) до більш зручних значень, користуючись формулою:

$$Y_{ij} = (A_{ij} - 0,42) \cdot 100. \tag{2.21}$$

Тоді таблиця перетворених значень буде мати наступний вигляд (табл. 4).

Таблиця 4

Перетворені значення функції відгуку

| № | $Y_1$ | $Y_1^2$ | $Y_2$ | $Y_2^2$ | $Y_3$ | $Y_3^2$ | $Y_4$ | $Y_4^2$ | $Y_5$ | $Y_5^2$ |
|---|---|---|---|---|---|---|---|---|---|---|
| 1 | -6 | 36 | -3 | 9 | 0 | 0 | -2 | 4 | 5 | 25 |
| 2 | -1 | 1 | -6 | 36 | -1 | 1 | 2 | 4 | 2 | 4 |
| 3 | -5 | 25 | _2 | 4 | -1 | 1 | 0 | 0 | 0 | 0 |
| 4 | -3 | 9 | -1 | 1 | -4 | 16 | 3 | 9 | 1 | 1 |
| 5 | 5 | 25 | _2 | 4 | 1 | 1 | -3 | 9 | 3 | 9 |
| 6 | -4 | 16 | -4 | 16 | -2 | 4 | -1 | 1 | 1 | 1 |
| $\sum_{j=1}^{n} Y_{ij}$ | -24 | 112 | -18 | 70 | -7 | 23 | -1 | 27 | 12 | 40 |

1. Знайдемо суму квадратів усіх спостереженні.

$$SS_1 = \sum_{i=1}^{k} \sum_{j=1}^{n} Y_{ij}^2 = 272. \tag{22}$$

2. Знайдемо суму квадратів сум по стовпцям, ділену на число вимірювань у стовпці:

$$SS_2 = \frac{1}{n}\sum_{i=1}^{k} A_i^2 = \frac{1}{6}\left(24^2 + 18^2 + 7^2 + 1^2 + 12^2\right) = 182{,}3. \qquad (23)$$

3. Знайдемо квадрат загальної суми, який поділений на число всіх вимірювань (коригувальний член):

$$SS_3 = \frac{1}{N}\sum_{i=1}^{k}\left(A_i\right)^2 = \frac{1}{30}(-38)^2 = 48{,}13. \qquad (24)$$

4. Знайдемо суму квадратів для стовпця:

$$SS_A = SS_2 - SS_3 = 182{,}3 - 48{,}13 = 134{,}17. \qquad (25)$$

5. Знайдемо загальну суму квадратів:

$$SS_{заг} = SS_1 - SS_3 = 272 - 48{,}13 = 223{,}87. \qquad (26)$$

6. Знайдемо залишкову суму квадратів для оцінки похибки експерименту:

$$SS_{зал} = SS_1 - SS_2 = 272 - 182{,}3 = 89{,}7. \qquad (27)$$

7. Знайдемо дисперсію фактору А:

$$S_A^2 = \frac{SS_A}{k-1} = \frac{134{,}17}{5-1} = 33{,}54. \qquad (28)$$

8. Знайдемо дисперсію фактору випадковості:

$$S_{вип}^2 = \frac{SS_{зал}}{k(n-1)} = \frac{89{,}7}{5(6-1)} = 3{,}59. \qquad (29)$$

9. Результати розрахунків можна представити у вигляді табл. 5

Таблиця 5

Розрахунок однофакторного дисперсійного аналізу

| Джерело дисперсії | Число степенем свободи | Сума квадратів | Дисперсії |
|---|---|---|---|
| Фактор А | 4 | 134,17 | 33,54 |
| Залишок | 25 | 89,7 | 3,59 |
| Загальна сума | 29 | 223,87 | 7,71 |

$F_{табл} = 2{,}8$ для *(1-p)=0,05 і* $f_1=4$, $f_2=25$.

$$\frac{S_A^2}{S_{вип}^2} = 9{,}34 \le F_{табл 0,05}(4;\ 25) = 2{,}8\,. \tag{30}$$

Отже, гострота леза ножа подрібнювача значимо впливає на зусилля різання стебел кукурудзи в динамічних умовах.

Визначимо, з якого рівня гострота леза починає значимо впливати на зусилля різання стебел кукурудзи.

Знайдемо значення

$$S^2_{(\overline{Y_1}-\overline{Y_m})} = \frac{2}{n} S^2_{вип} = \frac{2}{6} 3{,}59 = 1{,}19\,. \tag{31}$$

Далі знайдемо табличне значення критерію Стьюдента ($t_{p,\ f}$) для довірчої ймовірності $p = 0{,}95$ і числа ступенів свободи $f=2n-2=10$. Тоді $t_{0,95;10}=2{,}23$.

Знайдемо критерій $K$, відносно якого будемо порівнювати різниці середніх значень функції відгуку для кожних двох рівнів:

$$K = t_{p,f} \cdot S^2_{(\overline{Y_1}-\overline{Y_m})} = 2{,}23 \cdot 1{,}19 = 2{,}676. \tag{32}$$

За реперний (основний) рівень приймемо п'ятий рівень, де середнє значення зусилля різання найбільше.

$\left|\overline{Y_5} - \overline{Y_1}\right| = \left|2 - \left(\frac{-24}{6}\right)\right| = 6 > K = 2{,}676.$- вплив гостроти леза значний.

$\left|\overline{Y_5} - \overline{Y_2}\right| = \left|2 - \left(\frac{-18}{6}\right)\right| = 5 > K = 2{,}676.$- вплив гостроти леза значний.

$\left|\overline{Y_5} - \overline{Y_3}\right| = \left|2 - \left(\frac{-7}{6}\right)\right| = 3{,}17 > K = 2{,}676.$- вплив гостроти леза значний.

$\left|\overline{Y_5} - \overline{Y_4}\right| = \left|2 - \left(\frac{-1}{6}\right)\right| = 2{,}17 < K = 2{,}676.$- вплив гостроти леза незначний.

По результатам розрахунків можна зробити висновок, що гострота леза ножа подрібнювача починає значимо впливати на зусилля різання стебел кукурудзи при динамічному навантаженні, починаючи з $\varepsilon_4$=0,40 мм.

Лекція №4

# ТЕМА: ДВУХФАКТОРНЫЙ ДИСПЕРСІЙНИЙ АНАЛІЗ

## 1. Основні понятті і припущення

У цьому випадку розглядається вплив відразу двох факторів А і В на функцію відгуку. Причому А змінюється на $a_1, a_2, \ldots, a_k$ рівнях ($k$ - число рівнів фактору А), В змінюється на $b_1, b_2, \ldots, b_m$ рівнях ($m$ - число рівнів фактору В). Слід враховувати, що $k$ незалежно від $m$. Загальна таблиця двухфакторного дисперсійного аналізу для $n$ числа паралельних визначень має наступний вигляд (табл. 1).

Таблиця 1

Таблиця двухфакторного дисперсійного аналізу

| B | A | | | | Сума |
|---|---|---|---|---|---|
| | $a_1$ | $a_2$ | $a_i$ | $a_k$ | |
| $b_1$ | $Y_{111}, Y_{112}, Y_{11n}$ | $Y_{211}, Y_{212}, Y_{21n}$ | $Y_{i11}, Y_{i12}, Y_{i1n}$ | $Y_{k11}, Y_{k12}, Y_{k1n}$ | $B_1$ |
| $b_2$ | $Y_{121}, Y_{122}, Y_{12n}$ | $Y_{221}, Y_{222}, Y_{22n}$ | $Y_{i21}, Y_{i22}, Y_{i2n}$ | $Y_{k21}, Y_{k22}, Y_{k2n}$ | $B_2$ |
| $b_j$ | $Y_{1j1}, Y_{1j2}, Y_{1jn}$ | $Y_{2j1}, Y_{2j2}, Y_{2jn}$ | $Y_{ij1}, Y_{ij2}, Y_{ijn}$ | $Y_{kj1}, Y_{kj2}, Y_{kjn}$ | $B_j$ |
| $b_m$ | $Y_{1m1}, Y_{1m2}, Y_{1mn}$ | $Y_{2m1}, Y_{2m2}, Y_{2mn}$ | $Y_{im1}, Y_{im2}, Y_{imn}$ | $Y_{km1}, Y_{km2}, Y_{kmn}$ | $B_m$ |
| Сума | $A_1$ | $A_2$ | $A_i$ | $A_k$ | |

Загальне число дослідів:

$$N = nkm. \quad (1)$$

Функцію відгуку можна представити у вигляді:

$$Y_{ijn} = \mu + \gamma_A + \gamma_B + \gamma_A \cdot \gamma_B + \varepsilon_{ijn}, \quad (2)$$

де $\mu = M\{Y_{ijn}\}$ - математичне очікування функції відгуку;

$\gamma_A$ - вплив фактору А на *i*-тому рівні;

$\gamma_B$ - вплив фактору В на *j*-тому рівні;

$\gamma_A \cdot \gamma_B$ - вплив взаємодії факторів А та В;

$\varepsilon_{ijn}$ - похибка вимірювання всередині серії дослідів.

Усі припущення, прийняті в однофакторному дисперсійному аналізі, діють і у двухфакторному аналізі.

## 2. Двухфакторний дисперсійний аналіз у відсутності паралельних дослідів

Для полегшення розрахунків досить часто модель спрощується:

1. Не враховується взаємодія факторів.
2. Відсутність паралельних дослідів.
3. Розглядається модель із фіксованими рівнями факторів.

Тоді функція відгуку буде описуватися наступним рівнянням:

$$Y_{ijn} = \mu + \gamma_A + \gamma_B + \varepsilon_{ijn} . \qquad (3)$$

Таблиця двох факторного дисперсійного аналізу у відсутності паралельних дослідів має наступний вигляд (табл. 2)

Таблиця 2

Таблиця двухфакторного дисперсійного аналізу у відсутності паралельних дослідів

| б В | A | | | | Сума |
|---|---|---|---|---|---|
| | $a_1$ | $a_2$ | $a_i$ | $a_k$ | |
| $b_1$ | $Y_{11}$ | $Y_{21}$ | $Y_{i1}$ | $Y_{k1}$ | $B_1$ |
| $b_2$ | $Y_{12}$ | $Y_{22}$ | $Y_{i2}$ | $Y_{k2}$ | $B_2$ |
| $b_j$ | $Y_{1j}$ | $Y_{2j}$ | $Y_{ij}$ | $Y_{kj}$ | $B_j$ |
| $b_m$ | $Y_{1m}$ | $Y_{2m}$ | $Y_{im}$ | $Y_{km}$ | $B_m$ |
| Сума | $A_1$ | $A_2$ | $A_i$ | $A_k$ | |

$Y_{ij}$ - значення функції відгуку на *i*-тому рівні фактору А і *j*-тому рівні фактору В.

Порядок розрахунків у двухфакторном дисперсійному аналізі у відсутності паралельних досліджень аналогічний однофакторному дисперсійному аналізу. У даному пункті приведемо лише алгоритм розрахунків двухфакторного дисперсійного аналізу.

**3. Алгоритм розрахунків двухфакторного дисперсійного аналізу у відсутності паралельних досліджень**

1. Використовуючи таблицю двухфакторного дисперсійного аналізу (табл. 2), знайти суми по стовпцям:

$$A_i = \sum_{j=1}^{m} Y_{ij}. \tag{4}$$

2. Знайти суми по рядкам:

$$B_j = \sum_{i=1}^{k} Y_{ij}. \tag{5}$$

3. Знайти суму квадратів всіх спостережень:

$$SS_1 = \sum_{i=1}^{k} \sum_{j=1}^{m} Y_{ij}^2. \tag{6}$$

4. Знайти суму квадратів сум по стовпцях, ділену на число вимірювань у стовпці:

$$SS_2 = \frac{1}{m} \sum_{i=1}^{k} A_i^2. \tag{7}$$

5. Знайти суму квадратів сум по рядкам, ділений на число вимірювань в рядку:

$$SS_3 = \frac{1}{k} \sum_{j=1}^{m} B_j^2. \tag{8}$$

6. Знайти квадрат загальної суми, ділений на число всіх вимірювань (коригувальний член):

$$SS_4 = \frac{1}{N}\sum_{i=1}^{k}\left(A_i\right)^2 = \frac{1}{N}\sum_{j=1}^{m}\left(B_j\right)^2. \qquad (9)$$

7. Знайти суму квадратів для стовпця (для фактора А):

$$SS_A = SS_2 - SS_4. \qquad (10)$$

8. Знайти суму квадратів для рядка (для фактора В):

$$SS_B = SS_3 - SS_4. \qquad (11)$$

9. Знайти загальну суму квадратів:

$$SS_{заг} = SS_1 - SS_4. \qquad (12)$$

10. Знайти залишкову суму квадратів для оцінки похибки експерименту:

$$SS_{зал} = SS_{заг} - SS_A - SS_B. \qquad (13)$$

11. Знайти дисперсію фактора А:

$$S_A^2 = \frac{SS_A}{k-1}. \qquad (14)$$

12. Знайти дисперсію фактора А:

$$S_B^2 = \frac{SS_B}{m-1}. \qquad (15)$$

13. Знайти дисперсію фактору випадковості:

$$S_{вип}^2 = \frac{SS_{зал}}{(k-1)(m-1)}. \qquad (16)$$

14. Резуль таты розрахунків можна представити у вигляді табл. 3.

Таблиця 3

Розрахунки двухфакторного дисперсійного аналізу

| Джерело дисперсії | Число ступенів свободи | Сума квадратів | Дисперсії |
|---|---|---|---|
| Фактор А | *k-1* | $SS_A = SS_2 - SS_4.$ | $S_A^2 = \frac{SS_A}{k-1}.$ |

| | | | |
|---|---|---|---|
| Фактор В | *m-1* | $SS_B = SS_3 - SS_4.$ | $S_B^2 = \dfrac{SS_B}{m-1}.$ |
| Залишок | *(k-1)(m-1)* | $SS_{зал} = SS_{заг} - SS_A - SS_B.$ | $S_{вип}^2 = \dfrac{SS_{зал}}{(k-1)(m-1)}$ |
| Загальна сума | *km-1* | $SS_{заг} = SS_1 - SS_4.$ | $S^2 = \dfrac{SS_{заг}}{km-1}.$ |

Якщо відношення дисперсії фактору А до дисперсії похибки менше або дорівнює табличному значенню критерію Фішера:

$$\frac{S_A^2}{S_{вип}^2} \le F_{табл(1-p)}\left(f_1;\ f_2\right), \tag{17}$$

де $f_1 = k-1,\ f_2 = (k-1)(m-1),$ то вплив фактору А слід вважати незначним. Аналогічного і для фактору В:

$$\frac{S_B^2}{S_{вип}^2} \le F_{табл(1-p)}\left(f_1;\ f_2\right), \tag{18}$$

де $f_1 = m-1,\ f_2 = (k-1)(m-1),$ то вплив фактору В слід вважати незначним. При цьому загальна дисперсія пов'язана лише з фактором випадковості і може служити оцінкою для дисперсії відтворюваності. Обмеження значимості фаторів визначаються аналогічно одно факторному експерименту.

## 4. Алгоритм розрахунків двухфакторного дисперсійного аналізу з паралельними дослідами

Якщо проводити *n* паралельних дослідів на кожному рівні кожного фактору, то дисперсія фактору випадковості знаходиться для кожного рівня факторів (у кожній ячейкі табл. 1). При цьому слід ураховувати, що дані дисперсії повинні бути однорідні. Однорідність дисперсії можна перевірити по критерію Кохрена у випадку однакового числа паралельних дослідів. Тоді дисперсію фактору випадковості можна розрахувати по формулах:

$$S_{ij}^2 = \frac{1}{n-1}\sum_{u=1}^{n}(Y_{iju} - \overline{Y_{ij}})^2. \qquad (19)$$

$$S_{вип}^2 = \frac{1}{mk}\sum_{i=1}^{k}\sum_{j=1}^{m} S_{ij}^2. \qquad (20)$$

Тоді алгоритм розрахунків дещо відрізняється від описаного вище.

1. Знайти суму визначень у кожній ячейці таблиці 1 двухфакторного дисперсійного аналізу з паралельними дослідами:

$$Y_{ij} = \sum_{u=1}^{n} Y_{iju}\ . \qquad (21)$$

2. Знайти квадрат цієї суми:

$$Y_{ij}^2 = \left(\sum_{u=1}^{n} Y_{iju}\right)^2. \qquad (22)$$

3. Знайти суми по стовпцях (табл. 1):

$$A_i = \sum_{j=1}^{m}\sum_{u=1}^{n} Y_{iju}. \qquad (23)$$

4. Знайти суми по рядках:

$$B_j = \sum_{i=1}^{k}\sum_{u=1}^{n} Y_{iju}. \qquad (24)$$

5. Знайти суму всіх дослідів (загальна сума):

$$SS_0 = \sum_{i=1}^{k}\sum_{j=1}^{m}\sum_{u=1}^{n} Y_{iju}. \qquad (25)$$

6. Знайти суму квадратів усіх спостережень:

$$SS_1 = \sum_{i=1}^{k}\sum_{j=1}^{m}\sum_{u=1}^{n} Y_{iju}^2. \qquad (26)$$

7. Знайти суму квадратів сум по стовпцях, яка поділена на число вимірювань у стовпці:

$$SS_2 = \frac{1}{nm}\sum_{i=1}^{k} A_i^2\ . \qquad (27)$$

8. Знайти суму квадратів підсумків по рядках, ділену на число вимірів у рядку:

$$SS_3 = \frac{1}{nk}\sum_{j=1}^{m} B_j^2 . \tag{28}$$

9. Знайти квадрат загальної суми, який поділений на число всіх вимірювань (коригувальний член):

$$SS_4 = \frac{1}{nmk} SS_0^2 . \tag{29}$$

10. Знаходимо суму квадратів для стовпця (для фактору А):

$$SS_A = SS_2 - SS_4. \tag{30}$$

11. Знайти суму квадратів для рядка ( для фактору В):

$$SS_B = SS_3 - SS_4. \tag{31}$$

12. Знайти загальну суму квадратів:

$$SS_{заг} = SS_1 - SS_4. \tag{32}$$

13. Знайти залишкову суму квадратів для оцінки похибки експерименту:

$$SS_{зал} = SS_1 - \frac{\sum_{i=1}^{k}\sum_{j=1}^{m} Y_{ij}^2}{n}. \tag{33}$$

14. Знайти суму квадратів для оцінки ефекту взаємодії факторів:

$$SS_{AB} = SS_{заг} - SS_A - SS_B - SS_{зал}. \tag{34}$$

15. Знайти дисперсію фактору А:

$$S_A^2 = \frac{SS_A}{k-1}. \tag{35}$$

16. Знайти дисперсію фактору В:

$$S_B^2 = \frac{SS_B}{m-1}. \tag{36}$$

17. Знайти дисперсію фактору випадковості:

$$S_{вип}^2 = \frac{SS_{зал}}{km(n-1)}. \tag{37}$$

18. Знайти дисперсію взаємодії факторів:

$$S_{AB}^2 = \frac{SS_{AB}}{(k-1)(m-1)}. \tag{38}$$

19. Перевірку значимості факторів А і В, а також ефекту взаємодії факторів здійснюють по критерію Фішера. Якщо відношення дисперсії фактору А до дисперсії похибки менше табличного значення критерію Фішера:

$$\frac{S_A^2}{S_{вип}^2} \le F_{табл(1-p)}\left(f_1;\ f_2\right), \tag{39}$$

де $f_1 = k-1,\ f_2 = km(n-1),$ то вплив фактору А слід вважати незначним.

Аналогічно для фактору В. Якщо відношення дисперсії фактору В до дисперсії похибки менше табличного значення критерію Фішера:

$$\frac{S_B^2}{S_{вип}^2} \le F_{табл(1-p)}\left(f_1;\ f_2\right), \tag{40}$$

де $f_1 = m-1,\ f_2 = km(n-1),$ то вплив фактору В слід вважати незначним. Для ефекту взаємодії факторів А і В, якщо відношення дисперсії $S_{AB}^2$ до дисперсії похибки менше табличного значення критерію Фишера:

$$\frac{S_{AB}^2}{S_{вип}^2} \le F_{табл(1-p)}\left(f_1;\ f_2\right), \tag{41}$$

де $f_1 = (m-1)(k-1),\quad f_2 = km(n-1),$ то вплив взаємодії факторів А і В слід вважати незначним.

**Приклад.**

Передбачається, що на максимальну вихідну напругу акумуляторних батарей деякого виду впливають матеріал пластин і температура приміщення, де вони встановлені. У лабораторії проводяться чотири репліки факторного експерименту для трьох типів матеріалу і трьох значень температури; результати спостережень наведені в табл. 4. Порядок, у якому проводилися 36 спостережень, визначався випадковим чином.

Таблиця 4

Дані по максимальній вихідній напрузі

| Вид матеріалу | Температура, ˚С | | | | | | $A_i$ |
|---|---|---|---|---|---|---|---|
| | 10 | Σ | 18 | Σ | 26 | Σ | |
| 1 | 130 155<br>74 180 | 539 | 34 40<br>80 75 | 229 | 20 70<br>82 58 | 230 | 998 |
| 2 | 150 188<br>159 126 | 623 | 136 122<br>106 115 | 479 | 25 70<br>58 45 | 198 | 1300 |
| 3 | 138 110<br>168 160 | 576 | 174 120<br>150 139 | 573 | 96 104<br>82 60 | 342 | 1501 |
| $B_j$ | 1738 | | 1291 | | 770 | | 3799 |

1. Знайдемо суму квадравтів всіх спостережень:

$$SS_1 = \sum_{i=1}^{k}\sum_{j=1}^{m}\sum_{u=1}^{n} Y_{iju}^2 = 130^2 + 155^2 + 74^2 + \\ +180^2 + ... + 96^2 + 104^2 + 82^2 + 60^2 = 478547. \quad (42)$$

2. Знайдемо суму квадратів сум по стовпцях, ділену на число вимірювань у стовпці:

$$SS_2 = \frac{1}{mn}\sum_{i=1}^{k} A_i^2 = \frac{1}{3\cdot 4}(998^2 + 1300^2 + 1501^2) = 411583{,}75. \quad (43)$$

3. Знайдемо суму квадратів сум по рядкам, ділений на число вимірювань в рядку:

$$SS_3 = \frac{1}{kn}\sum_{j=1}^{m} B_j^2 = \frac{1}{3\cdot 4}(1738^2 + 1291^2 + 770^2) = 440018{,}75. \quad (44)$$

4. Знайдемо квадрат загальної суми, ділений на число всіх вимірювань (коригувальний член):

$$SS_4 = \frac{1}{N}\sum_{i=1}^{k}\left(A_i\right)^2 = \frac{1}{N}\sum_{j=1}^{m}\left(B_j\right)^2 = \frac{1}{36}3799^2 = 400900{,}0278. \quad (45)$$

5. Знайдемо суму квадратів для стовпця (для фактора А):

$$SS_A = SS_2 - SS_4 = 411583{,}75 - 400900{,}0278 = 10683{,}72. \quad (46)$$

6. Знайдемо суму квадратів для рядка (для фактора В):

$$SS_B = SS_3 - SS_4 = 440018{,}75 - 400900{,}0278 = 39118{,}72. \qquad (47)$$

7. Знайти загальну суму квадратів:

$$SS_{заг} = SS_1 - SS_4 = 478547 - 400900{,}0278 = 77646{,}96. \qquad (48)$$

8. Знайти залишкову суму квадратів для оцінки похибки експерименту:

$$SS_{зал} = SS_1 - \frac{\sum_{i=1}^{k}\sum_{j=1}^{m} Y_{ij}^2}{n} = 478547 - \\ -\frac{1}{4}\left(539^2 + +229^2 + ... + 342^2\right) = 18230{,}7622. \qquad (49)$$

9. Знайти суму квадратів для оцінки ефекту взаємодії факторів:

$$SS_{AB} = SS_{заг} - SS_A - SS_B - SS_{зал} = \\ = 77646{,}96 - 10683{,}72 - 39118{,}72 - 18230{,}7622 = 9613{,}7578. \qquad (50)$$

10. Знайти дисперсію фактору А:

$$S_A^2 = \frac{SS_A}{k-1} = \frac{10683{,}72}{3-1} = 5341{,}86. \qquad (51)$$

11. Знайти дисперсію фактору В:

$$S_B^2 = \frac{SS_B}{m-1} = \frac{39118{,}72}{3-1} = 19559{,}36. \qquad (52)$$

12. Знайти дисперсію фактору випадковості:

$$S_{вип}^2 = \frac{SS_{зал}}{km(n-1)} = \frac{18230{,}7622}{3 \cdot 3(4-1)} = 675{,}2134. \qquad (53)$$

13. Знайти дисперсію взаємодії факторів:

$$S_{AB}^2 = \frac{SS_{AB}}{(k-1)(m-1)} = \frac{9613{,}7578}{(3-1)(3-1)} = 2403{,}43945. \qquad (54)$$

14. Перевірку значимості факторів А і В, а також ефекту взаємодії факторів здійснюють по критерію Фішера. Якщо відношення дисперсії фактору А до дисперсії похибки менше табличного значення критерію Фішера:

$$\frac{S_A^2}{S_{вип}^2} \le F_{табл(1-p)}\left(f_1;\ f_2\right), \qquad (55)$$

де $f_1 = k-1,\ f_2 = km(n-1),$ то вплив фактору А слід вважати незначним.

$$\frac{5341{,}86}{675{,}2134} = 7{,}9114 > F_{табл(1-0{,}05)}\left(2;27\right) = 3{,}35.$$

Тобто фактор А (вид матеріалу пластин) в значній мірі впливає на максимальну вихідну напругу.

Аналогічно для фактору В. Якщо відношення дисперсії фактору В до дисперсії похибки менше табличного значення критерію Фішера:

$$\frac{S_B^2}{S_{вип}^2} \le F_{табл(1-p)}\left(f_1;\ f_2\right), \qquad (56)$$

де $f_1 = m-1,\ f_2 = km(n-1),$ то вплив фактору В слід вважати незначним.

$$\frac{19559{,}36}{675{,}2134} = 28{,}9677 > F_{табл(1-0{,}05)}\left(2;27\right) = 3{,}35.$$

Тобто фактор В (температура навколишнього середовища) в значній мірі впливає на вихідну напругу.

Для ефекту взаємодії факторів А і В, якщо відношення дисперсії $S_{AB}^2$ до дисперсії похибки менше табличного значення критерію Фишера:

$$\frac{S_{AB}^2}{S_{вип}^2} \le F_{табл(1-p)}\left(f_1;\ f_2\right), \qquad (57)$$

де $f_1 = (m-1)(k-1),\ f_2 = km(n-1),$ то вплив взаємодії факторів А і В слід вважати незначним.

$$\frac{2403{,}43945}{675{,}2134} = 3{,}5595 > F_{табл(1-0{,}05)}\left(4;27\right) = 2{,}73.$$

І взаємодія факторів А і В, які розглядаються суттєво впливають на результати спостережень.

Таблиця 5

Розрахунки двухфакторного дисперсійного аналізу

| Джерело дисперсії | Число ступенів свободи | Сума квадратів | Дисперсії |
|---|---|---|---|

| | | | |
|---|---|---|---|
| Фактор А | 2 | 10683,72 | 5341,86 |
| Фактор В | 2 | 39118,72 | 19559,36 |
| Взаємодія факторів А і В | 4 | 9613,7578 | 2403,43945 |
| Залишок | 27 | 18230,7622 | 675,2134 |
| Загальна сума | 35 | 77646,96 | 2218,4846 |

Аналізуючи проведені результати можна зробити висновок, про те, що обрані фактори на величину максимальної вихідної напруги впливають зі значним ефектом, а також на досліджуваний показник впливає і взаємодія розглянутих факторів. Найбільш суттєвий вплив чинить фактор В (температура навколишнього середовища) виходячи з відношень 55, 56 та 57.

Для полегшення інтерпретації результатів експерименту корисно побудувати залежності середніх відгуків для кожної комбінації вибірок (рис. 1). Про статистично значиму взаємодію свідчить відсутність паралельності ліній графіка. Загалом кажучи, більш висока вихідна напруга спостерігається при низькій температурі незалежно від виду матеріалу. При зміні температури від низької до проміжної вихідна напруга для матеріалу типу 3 навіть зростає, а для матеріалів 1 і 2 — зменшується. При зміні температури від проміжної до високої вихідна напруга для матеріалів 2 і 3 зменшується, а для матеріалу 1 - залишається практично такою ж.

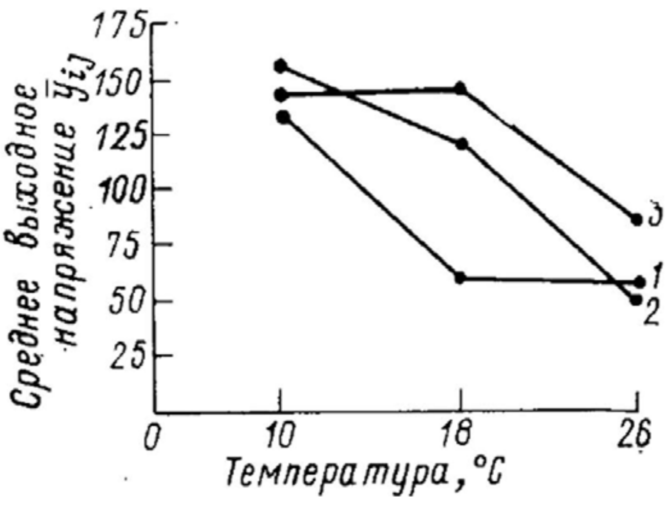

Рис. 1. Залежності середньої вихідної напруги від температури для декількох видів матеріалу

Лекція №5

# ТЕМА: КОРЕЛЯЦІЙНИЙ І РЕГРЕСІЙНИЙ АНАЛІЗИ

## 1. Обробка результатів спільних вимірювань

Спільні вимірювання являють собою проведені одночасно вимірювання двох або декількох, як правило, різних величин для знаходження залежності між ними.

Цей вид вимірювань знаходить широке застосування в наукових, технічних і метрологічних вимірюваннях.

Спільні вимірювання застосовуються в метрологічній практиці при експериментальному визначенні градуювальних характеристик засобів вимірювань, у тому числі різних перетворювачів.

Градуювальна характеристика засобу виміру являє собою залежність між значеннями величин на вході і виході засобів вимірювання. Вона може бути представлена у вигляді таблиці, графіка або формули (тобто в аналітичному вигляді).

Найбільш універсальною формою градуювальної характеристики є її описання у вигляді формули, яку зручно використовувати при автоматизованих випробуваннях із застосуванням ЕОМ.

Кожному вимірювальному приладу або перетворювачу відповідають власна індивідуальна залежність між вхідною величиною $X$ і вихідний $Y$, яка в загальному випадку залежить також і від часу $t$. Функціональна залежність $y=f(x)$, являє собою функцію перетворення вимірювального приладу (або перетворювача) і є градуювальною характеристикою.

При градуюванні виконують спільні вимірювання вхідних і вихідних величин. Якщо число точок вимірювань $n$, то одержують набір результатів вимірювань $(x_i; y_i)$, $i=1...n$, по яких визначають градуювальну характеристику.

У кожній досліджуваній точці вимірювання проводяться багаторазово (при прямому і зворотному напрямку зміни вхідної величини).

Найбільш переважною градуювальною характеристикою є лінійні моделі:

$$Y = \alpha + \beta \cdot X, \tag{1}$$

де $\alpha$ - константа (вільний член);

$\beta$ - коефіцієнти, які визначають за експериментальним даними при градуюванні засобу вимірювання методом регресійного аналізу.

У регресійному аналізі для визначення коефіцієнтів застосовують метод найменших квадратів (МНК), який припускає, що виконано дві основні вимоги:

1) значення вхідних величин $x_i$ відомі точно;

2) результати виміряних вихідних величин $y_i$ містять незалежні випадкові похибки, які розподілені за нормальним законом.

Необхідно спеціально перевірити справедливість виконання умови 2. Грубі похибки, що різко виділяються (промахи) повинні бути виключені. Для цього застосовують розглянуті вище критерії перевірки статичних гіпотез.

Для визначення коефіцієнтів $a$ і $b$ у рівнянні регресії (1) використовують регресійний аналіз:

$$\overline{y} = a + b \cdot x, \tag{2}$$

де $\overline{y}$ - лінія регресії (функція відгуку).

## 2. Методика простого кореляційно-регресійного аналізу

Розглянемо методику регресійного аналізу для пасивного експерименту, коли експеримент заздалегідь не планується.

Рівнянню (1) відповідає парна регресія, коефіцієнти якої визначають по формулах:

$$b=\frac{\sum_{i=1}^{n}x_i\cdot y_i-\frac{1}{n}\cdot\sum_{i=1}^{n}x_i\cdot\sum_{i=1}^{n}y_i}{\sum_{i=1}^{n}x_i^2-\frac{1}{n}\cdot\left(\sum_{i=1}^{n}x_i\right)^2}. \tag{3}$$

$$a=\frac{1}{n}\cdot\left(\sum_{i=1}^{n}y_i-b\cdot\sum_{i=1}^{n}x_i\right). \tag{4}$$

Дисперсія $\overline{y}$ буде складатися із двох компонентів - дисперсії параметра $a$ і дисперсії параметра $b$.

Верхня $y_{гр}^{в}$ і нижня $y_{гр}^{н}$, границі для $\overline{y}$ мають вигляд:

$$y_{гр}^{в}=\overline{y}+t_q\cdot S_{\overline{y_i}}, \qquad y_{гр}^{н}=\overline{y}-t_q\cdot S_{\overline{y_i}}, \tag{5}$$

де $t_q$ - коефіцієнт Стьюдента;

$S_{\overline{y_i}}$ - середнє квадратичне відхилення відгуку $\overline{y}$;

$$S_{\overline{y_i}}=\sqrt{\frac{\sum_{i=1}^{n}\left(y_i-\overline{y_i}\right)^2}{n-2}\cdot\left[\frac{1}{n}+\frac{\left(x_i-\overline{X}\right)^2}{\sum_{i=1}^{n}\left(x_i-\overline{X_i}\right)^2}\right]}, \tag{6}$$

де $\overline{X}=\frac{1}{n}\sum_{i=1}^{n}x_i$ .

Якщо нанести границі на графік (рис. 1), то вони розташуються відповідно вище і нижче лінії регресії у вигляді віток гіперболи, що обмежують довірчу область.

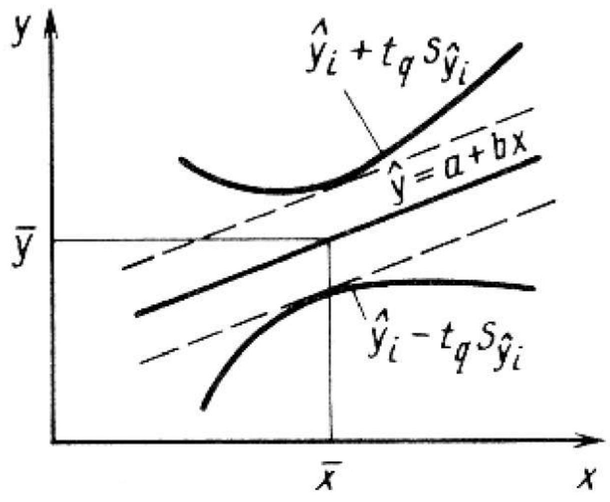


Рис. 1. Межі довірчого інтервалу лінійного рівняння парної регресії

Якщо виникає необхідність перевірки статичної гіпотези про рівність двох рівнянь регресії, то така перевірка включає послідовну перевірку справедливості трьох статичних гіпотез:

1) про залишкові дисперсії;

2) про значення коефіцієнтів регресії *b*;

3) про значення коефіцієнтів (константа) *a*.

У розглянутій парній регресії значення $\bar{y}$ залежать від значень тільки однієї змінної *x*. Однак у загальному випадку $\bar{y}$ може залежати від декількох змінних $x_1,\ x_2,\ldots,x_n$. Це так званий випадок множинної регресії.

Оцінка параметрів регресії звичайно супроводжується розрахунками додаткової характеристики, яка називається коефіцієнтом кореляції.

Вибірковий коефіцієнт кореляції являє собою емпіричну (тобто певну за експериментальними результатами) величину лінійної залежності між *x* і *y*.

У математичній статистиці ступінь кореляції змінних (*n* пар випадкових величин $x_1,\ y_1,\ i=1,\ 2,\ 3,\ \ldots\ ,\ i$), яку оцінюють вибірковим коефіцієнтом кореляції Пірсона:

$$r_b = \frac{n \cdot \sum_{i=1}^{n} x_i \cdot y_i - \sum_{i=1}^{n} x_i \cdot \sum_{i=1}^{n} y_i}{\sqrt{\left[ n \cdot \sum_{i=1}^{n} x_i^2 - \left( \sum_{i=1}^{n} x_i \right)^2 \right] \cdot \left[ n \cdot \sum_{i=1}^{n} y_i^2 - \left( \sum_{i=1}^{n} x_i \right)^2 \right]}}. \qquad (7)$$

Крім зазначених способів оцінки величини коефіцієнта кореляції, можна використовувати також формулу:

$$r_{x,y} = a \frac{\sigma_x}{\sigma_y}, \qquad (7.1)$$

де $a$ — параметр при *x* у рівнянні лінійної регресії;

$\sigma_x$ і $\sigma_y$ — середні квадратичні відхилення для аргументу і функції.

Якщо $r_b > 0$, то при збільшенні *x* зростає *y*, при $r_b < 0$, з ростом *x*, *y* - зменшується. Прийнято вважати, що при виконанні умови $0{,}75 < |r_b| < 0{,}95$ існує сильний зв'язок, а при $0{,}95 < |r_b| < 1$ - функціональна залежність.

Для невеликих значень *n* ( так звана мала вибірка) коефіцієнт кореляції $r_b$ повинен бути скоректований:

$$r_b' = r_b \left[ 1 + \frac{1 - r_b^2}{2 \cdot (n-3)} \right]. \qquad (8)$$

Вибірковий коефіцієнт $r_b$ ($r_b'$) повинен бути перевірений на істотність, тобто чи істотно він відрізняється від нуля.

Для перевірки статичної гіпотези про істотність (значимість) кореляції між досліджуваними величинами *X* і *Y* і побудови довірчих інтервалів для коефіцієнта кореляції використовують перетворення Фішера:

$$U = \frac{1}{2} \cdot ln \frac{1+r}{1-r}, \qquad (9)$$

і апроксимується нормальним законом з дисперсією:

$$\sigma_U = \frac{1}{(n-3)}. \qquad (10)$$

Якщо $|U| > Z_{1-q/2} \cdot \sigma_U$ , то коефіцієнт кореляції суттєво відрізняється від нуля.

Прийнявши, що справедливим є нормальний закон ( для довірчої імовірності $P = 0{,}95$), знаходять верхню $U^B$ і нижню $U^H$ границі довірчого інтервалу:

$$U^B = U + 1{,}96 \cdot \sigma_U, \qquad U^H = U - 1{,}96 \cdot \sigma_U \qquad (11)$$

Якщо число вимірювань $n > 50$, то при довірчій імовірності 0,95 довірчі межі вибіркового значення коефіцієнта кореляції визначалась за формулами:

$$r^B = r_b + 1{,}96 \cdot \frac{1 - r_b^2}{\sqrt{n}}, \quad r^H = r_b - 1{,}96 \cdot \frac{1 - r_b^2}{\sqrt{n}}. \qquad (12)$$

За допомогою перетворення (9) можна встановити рівність між собою двох вибіркових коефіцієнтів кореляції, використовуючи, наприклад, статичний критерій $T$.

При обраному рівні значимості $q$ потрібно перевірити нульову гіпотезу про рівність нулю генерального коефіцієнта кореляції (при конкуруючій гіпотезі: вибірковий коефіцієнт кореляції $r_b$ відмінний від нуля).

Якщо нульова гіпотеза відкидається, то це означає, що вибірковий коефіцієнт кореляції значимо відрізняється від нуля, а $X$ і $Y$ корельовані, тобто пов'язані лінійною залежністю.

У якості критерію перевірки нульової гіпотези приймають випадкову величину:

$$T = \frac{r_b \sqrt{n-2}}{\sqrt{1 - r_b}}. \qquad (13)$$

По рівню значимості $q$ і числу степенів свободи $k = n - 2$ знаходимо по табл. 1 (для двухсторонньої критичної області) критичну точку $t_{KP}(q, k)$. Тоді якщо $T > t_{KP}$ - нульова гіпотеза відхиляється, і вибірковий коефіцієнт кореляції значимо відрізняється від нуля, тобто *X* і *У* корельовані.

Таблица 1

Критерий Стьюдента $t_{KP}$ (квантилі Стьюдента)

| Довір-ча імовір-ність р | Число степенів свободи | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | 3 | 4 | 5 | 6 | 8 | 10 | 12 | 18 | 22 | 30 | 40 | 60 | 120 | ∞ |
| 0,90 | 2,35 | 2,13 | 2,01 | 1,94 | 1,86 | 1,81 | 1,78 | 1,73 | 1,72 | 1,70 | 1,68 | 1,67 | 1,66 | 1,64 |
| 0,95 | 3,18 | 2,78 | 2,57 | 2,45 | 2,31 | 2,23 | 2,18 | 2,10 | 2,07 | 2,04 | 2,02 | 2,00 | 1,98 | 1,96 |
| 0,99 | 5,84 | 4,60 | 4,03 | 3,71 | 3.36 | 3,17 | 3,06 | 2,98 | 2,82 | 2,75 | 2,70 | 2,86 | 2,62 | 2,58 |

Стандартна похибка лінійного коефіцієнта кореляції визначається за формулою:

$$m_r = \sqrt{\frac{1 - r^2}{n - 2}}. \tag{14}$$

Оцінка стандартної похибки параметрів лінійної регресії визначалась за наступними формулами:

$$m_b = \sqrt{\frac{\sum_{i=1}^{n}(y - \overline{y_x})^2}{(n-2)\cdot\sum_{i=1}^{n}(x - \overline{x})^2}}, \quad m_a = \sqrt{\frac{\sum_{i=1}^{n}(y - \overline{y_x})^2 \cdot \sum_{i=1}^{n} x_i^2}{(n-2)\cdot n \cdot \sum_{i=1}^{n}(x - \overline{x})^2}}. \tag{15}$$

де $y - \overline{y_x}$ - різниця між фактичними та отриманим в результаті теоретичних досліджень значенням ендогенної змінної.

Максимальне та мінімальне значення для коефіцієнта регресії, тобто довірчий інтервал, визначались за допомогою *t*-критерію Стьюдента наступними формулами:

$$b^B = b + t_{табл} \cdot m_b, \quad b^H = b - t_{табл} \cdot m_b. \tag{16}$$

Аналогічно

$$a^{B} = a + t_{табл} \cdot m_{a}, \quad a^{H} = a - t_{табл} \cdot m_{a}. \qquad (17)$$

Стандартна та відносна похибки рівняння регресії визначались за наступними співвідношеннями:

$$\sigma_{u} = \sqrt{\frac{\sum_{i=1}^{n} \left( y - \hat{y}_{x} \right)^{2}}{n-2}}, \quad \vartheta_{u} = \frac{\sigma_{u}}{\bar{y}} \cdot 100\%. \qquad (18)$$

Оскільки вплив будь-яких з досліджуваних факторів може виявитись несуттєвим, то проводимо перевірку значущості коефіцієнтів регресії за допомогою критерію Стьюдента. Коефіцієнт вважають значущим, якщо виконується нерівність:

$$\left| b_{a} \right| \geq \Delta b_{u} = t\left( 0,05; f_{y} \right) \frac{S_{y}}{\sqrt{n}}, \qquad (19)$$

де $b_a$ – коефіцієнти у рівнянні регресії;

$\Delta b_u$ – довірча границя;

$t\left(0,05; f_u\right)$ – критерій Стьюдента при 5%-му рівні значущості та числі ступенів вільності дисперсії відтворюваності $f_y = n\left(m_0 - 1\right)$

Коефіцієнти, які виявились незначущими виключають із рівняння регресії.

Метою перевірки адекватності математичної (регресійної) моделі є підтвердження того, що дана модель правильно описує досліджуваний процес. Для цього визначаються погрішності математичної моделі й експериментальних даних. Якщо погрішності моделі перевищують погрішності експериментальних даних, то гіпотеза про адекватність математичної моделі відхиляється.

Для встановлення відповідності отриманої моделі експериментальним результатам проводять перевірку її адекватності за допомогою критерію Фішера:

$$F = \frac{S^2_{ад}}{S^2_{відт}}, \qquad (20)$$

де $S^2_{ад}$ - дисперсія адекватності;

$$S^2_{ад} = \frac{1}{f_1} \cdot \sum_{i=1}^{N} (y_i - \overline{y_i})^2, \qquad (21)$$

$y_i$ – результат i-го досліду, який проведений за матрицею планування;

$\overline{y_i}$ - результат i-го значення досліду, передбаченого за допомогою регресійної моделі;

$S^2_{відт}$ - дисперсія відтворюваності.

$$S^2_{відт} = \frac{1}{f_2} \cdot \sum_{i=1}^{n} (y_i - \overline{y_i})^2, \qquad (22)$$

Розподіл випадкової величини F визначається числом ступенів свободи дисперсії адекватності f1 та відтворюваності f2:

$$f_1 = N\text{-}d \quad та \quad f_2 = n\text{-}1, \qquad (23)$$

де $N$ – кількість дослідів за матрицею планування;

$d$ – кількість значимих коефіцієнтів регресії;

$n$ – кількість попередніх повторних дослідів, які проведено для середнього (нульового) рівня факторів

Обравши необхідний рівень значущості і визначивши критичне (табличне) значення розподілу Фішера $F_{табл}(f_1; f_2)$ для заданих чисел ступенів свободи, отримуємо нерівність

$$F = \frac{S^2_{ад}}{S^2_{відт}} > F_{табл}(f_1; f_2), \qquad (24)$$

при виконанні якої дисперсії не можуть бути рівними, тобто модель є неадекватною. У випадку, якщо умова (24) не виконується, то гіпотеза про лінійний характер досліджуваної залежності виключається. Далі досліджувалась можливість описати її рівнянням другої та третьої степені, які мають вигляд:

Далі досліджувалась можливість описати її рівнянням другої та третьої степені, які мають вигляд:

$$y = a \cdot x^2 + b \cdot x + c, \quad \text{або} \quad y = a \cdot x^3 + b \cdot x^2 + c \cdot x + d, \qquad (25)$$

з повторною перевіркою на адекватність.

### 3. Множинна кореляція та регресія

Множинну кореляцію та регресію розглянемо на прикладі залежності між трьома величинами: *z, y* та *x*.

У випадку лінійної залежності для характеристики щільності зв'язку трьох величин використовують часткові та множинні коефіцієнти кореляції.

Частковий коефіцієнт кореляції характеризує зв'язок двох величин при сталому значенні третьої, він позначається літерою $r$ з індексами, які вказують величини, залежність яких розглядається. Наприклад, $r_{xz\ y}$ – частковий коефіцієнт кореляції між величинами $x$ та $z$ при усуненні змін, що вносяться величиною $y$. Часткові коефіцієнти розраховуються через парні коефіцієнти кореляції: $r_{xy}$ , $r_{zy}$ ; $r_{xz}$ за формулою:

$$r_{xz\cdot y} = \frac{r_{xz} - r_{xy} \cdot r_{yz}}{\sqrt{\left(1 - r_{xy}^2\right)\left(1 - r_{yz}^2\right)}};$$

$$r_{xy\cdot z} = \frac{r_{xy} - r_{xz} \cdot r_{yz}}{\sqrt{\left(1 - r_{xy}^2\right)\left(1 - r_{yz}^2\right)}}; \qquad (26)$$

$$r_{yz\cdot x} = \frac{r_{yz} - r_{xz} \cdot r_{xy}}{\sqrt{\left(1 - r_{xy}^2\right)\left(1 - r_{xz}^2\right)}}.$$

Часткові коефіцієнти, так само як і парні коефіцієнти, можуть приймати значення від –1 до +1.

Множинний коефіцієнт кореляції показує щільність лінійної залежності однієї величини від двох інших. Наприклад, коефіцієнт $R_{xy\cdot z}$ показує щільність зв'язку величини $x$ з величиною $y$ і $z$ та розраховується за формулою:

$$R_{x\cdot yz} = \sqrt{\frac{r_{xy}^2 + r_{xz}^2 - 2r_{xy}r_{xz}r_{yz}}{1 - r_{yz}^2}};$$

$$R_{y\cdot xz} = \sqrt{\frac{r_{xy}^2 + r_{yz}^2 - 2r_{xy}r_{xz}r_{yz}}{1 - r_{xz}^2}}; \tag{27}$$

$$R_{z\cdot xy} = \sqrt{\frac{r_{xz}^2 + r_{yz}^2 - 2r_{xy}r_{xz}r_{yz}}{1 - r_{xy}^2}}.$$

Коефіцієнт кореляції $R$ є величина додатна і може приймати значення від 0 до 1. Якщо $R$=0, то між величинами немає лінійної кореляційної залежності. Якщо $R$=1, то існує лінійний функціональний зв'язок між розглянутими величинами.

**Модуль 3**

# МЕТОДИ ВИЯВЛЕННЯ НАЙБІЛЬШ СУТТЄВИХ ФАКТОРІВ

Лекція №6

## ТЕМА: ЕКСПЕРТНИЙ МЕТОД (АПРІОРНЕ РАНЖУВАННЯ ФАКТОРІВ) ВИЯВЛЕННЯ НАЙБІЛЬШ СУТТЄВИХ ФАКТОРІВ

Процеси, які вивчаються в землеробській механіці, визначаються великою кількістю незалежних факторів. Частина факторів роблять істотний вплив на досліджуваній параметр оптимізації, інші фактори роблять незначний вплив. До початку вивчення процесу, звичайно, невідомий ступінь впливу окремих факторів, тому на перших етапах в програму дослідження необхідно включати всі фактори, яке здатні вплинути на параметр оптимізації.

Проте включення в програму дослідження великої кількості факторів вимагає значного збільшення експериментальних робіт. Для скорочення кількості дослідів необхідно заздалегідь виключити фактори, що роблять невеликий вплив і виділити найбільш істотні фактори для їх подальшого детального вивчення.

Інтуїтивний відбір істотних факторів вносить елемент суб'єктивності і може привести до недостовірних результатів. Тому в наукових дослідженнях застосовуються різні експериментально-статистистичні методи аналізу значущості факторів.

### 1. Експертний метод (апріорне ранжування факторів)

Експертний метод припускає використання всієї попередньої (апріорній) інформації про процес, що вивчається, або об’єкт. Процедура проведення експертного методу складається з 4 етапів:

1) опитування і анкетування з даного питання;

2) статистична обробка анкет і складання матриці рангів;

3) оцінка узгодженості думок фахівців;

4) аналіз діаграми рангів і виключення незначущих факторів.

Чисельність експертної групи (7-12 чоловік) визначається з умови досягнення необхідної вірогідності правильного розв'язання.

Як **приклад** розглянемо організацію і проведення експертної оцінки факторів, що впливають на формування валка.

Для вирішення цього завдання були залучені фахівці з проектування, дослідження й експлуатації валкових жниварок. При співбесіді з кожним експертом були заздалегідь розглянуті й обговорені безліч різних показників, які хоча б трохи характеризують якість валків. Було попередньо встановлено 23 таких фактора. Зазвичай експерту пропонується оцінити 7-10 різних факторів, при більшому числі знижується об’єктивність оцінок. Тому при повторній співбесіді з експертами були відсіяні якнайменше значущі фактори. За факторами, що залишилися, експерти встановили коефіцієнти вагомості (табл. 1).

За результатами опитування обчислюється коефіцієнт конкордації $W$ Кендела (узгодження), визначальний ступінь узгодженості думок фахівців за формулою:

$$W = \frac{12S}{m^2\left(k^3 - k\right)}, \tag{1}$$

де $S$ - сума квадратів відхилень;

$m$ - число опитуваних фахівців;

$k$ - число факторів.

Сума квадратів відхилень обчислюється за формулою:

$$S = \sum_{i=1}^{k}\left(\sum_{i=1}^{m} a_{ij} - L\right)^2 , \tag{2}$$

де $a_{ij}$ - ранг (порядковий номер при опитуванні) $i$-го фактора в $j$-го фахівця;

Таблиця 1

Коефіцієнти вагомості факторів

| № фак-тора | Фактор | Експерти | | | | | | | | | | | | $\sum_{i=1}^{m} a_{ij}$ | $\sum_{i=1}^{m} a_{ij} - L$ | $\left(\sum_{i=1}^{m} a_{ij} - L\right)^2$ | $v = \frac{\sum_{i=1}^{m} a_{ij}}{k \cdot L}$ |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 | | | | |
| 1 | Кут нахилу платформи | 1 | 5 | 2 | 5 | 5 | 0 | 0 | 1 | 1 | 3 | 3 | 2 | 28 | -29,9 | 894,01 | 0,0484 |
| 2 | Довжина транспортера | 3 | 2 | 2 | 3 | 2 | 1 | 3 | 2 | 4 | 5 | 2 | 3 | 32 | -25,9 | 670,81 | 0,0553 |
| 3 | Розміщення вивантажувального вікна (двох- і трьох потокові жниварки) | 10 | 10 | 8 | 5 | 5 | 6 | 10 | 10 | 10 | 10 | 7 | 9 | 100 | 42,1 | 1772,4 | 0,1727 |
| 4 | Нерівномірність стерні по висоті | 2 | 4 | 1 | 1 | 1 | 1 | 1 | 2 | 2 | 3 | 1 | 1 | 20 | -37,9 | 1436,4 | 0,0345 |
| 5 | Положення мотовила відносно транспортера і різального апарату | 4 | 5 | 6 | 7 | 2 | 5 | 7 | 1 | 2 | 4 | 7 | 5 | 55 | -2,9 | 8,41 | 0,0950 |
| 6 | Полеглість стебел | 5 | 7 | 6 | 5 | 9 | 3 | 3 | 4 | 4 | 5 | 4 | 7 | 62 | 4,1 | 16,81 | 0,1071 |
| 7 | Частота обертання мотовила | 4 | 5 | 8 | 9 | 10 | 3 | 3 | 3 | 8 | 8 | 7 | 1 | 69 | 11,1 | 123,21 | 0,1192 |
| 8 | Степінь відповідності швидкостей транспортера і жниварки | 4 | 3 | 2 | 4 | 8 | 7 | 4 | 4 | 6 | 7 | 7 | 5 | 61 | 3,1 | 9,61 | 0,1054 |
| 9 | Відстань від різального апарата до транспортера | 6 | 5 | 6 | 3 | 6 | 7 | 8 | 9 | 9 | 5 | 8 | 6 | 78 | 20,1 | 404,01 | 0,1347 |
| 10 | Відповідність форми і розмірів вивантажувального вікна потоку стебел | 10 | 8 | 4 | 10 | 4 | 9 | 6 | 4 | 7 | 5 | 5 | 2 | 74 | 16,1 | 259,21 | 0,1278 |
| Сума | | 49 | 54 | 45 | 52 | 52 | 42 | 45 | 40 | 53 | 55 | 51 | 41 | 579 | - | 5594,9 | - |

$L$ - середнє значення сум рангів із кожного фактору

$$L = \frac{\sum_{i=1}^{k}\sum_{i=1}^{m} a_{ij}}{k}. \tag{3}$$

Якщо фахівцеві важко провести чітку межу між двома факторами, то вводяться «зв'язані» ранги, тобто двом або більше факторам надається одне і те ж місце (номер).

Якщо мають місце "зв'язані" ранги, то коефіцієнт конкордації $W$ визначається за формулою:

$$W = \frac{12 \cdot S}{m^2\left(k^3 - k\right) - m\sum_{i=1}^{m} T_i}, \tag{4}$$

де $T_i = \sum_{j=1}^{m}\left(t_j^3 - t_j\right)$.

Так, для першого фахівця:

| | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| Фактори | $A_1$ | $A_2$ | $A_3$ | $A_4$ | $A_5$ | $A_6$ | $A_7$ | $A_8$ | $A_9$ | $A_{10}$ |
| Ранги | 1 | 3 | 10 | 2 | 4 | 5 | 4 | 4 | 6 | 10 |

Тут $t_j - j$-е - число однакових рангів в $j$-м ранжируванні. Наприклад, якщо який-небудь фахівець призначив ранги, то для цього випадку:

$$T_i = \sum_{1}^{2}\left(3^3 - 3\right) + \left(2^3 - 2\right) = 30.$$

Якщо в окремих фахівців не виявилося "зв'язаних" рангів, то для них $T_i$ буде рівно нулю.

Значення коефіцієнта конкордації змінюється в інтервалі від 0 до 1, і чим більше його значення, тим більша узгодженість думок у фахівців.

Обчислення коефіцієнта конкордації зручніше проводити, склавши матрицю результатів опитування у вигляді алгоритму (останні 4 стовпці табл. 1).

Після обчислення коефіцієнта конкордації визначають його значущість за критерієм $\chi^2$, оскільки величина $m(k-1)W$ має $\chi^2$ - розподіл з числом ступенів вільності $f = k-1$. Значення $\chi^2$ - критерія подання в додатку 1.

Розрахункове значення $\chi^2$ - розподілу визначається за формулою

$$\chi^2 = m(k-1)W. \tag{5}$$

Гіпотеза про наявність згоди між фахівцями приймається, коли $\chi^2_{розр} \geq \chi^2_{табл}$.

Для випадку, що розглядається (див. табл. 1)

$$\sum_{i=1}^{m} T_i = 3(4^3-4) + 6(3^3-3) + 14(2^3-2) = 408,$$

$$W = \frac{12 \cdot 5594,9}{12^2(10^3-10) - 12 \cdot 408} = 0,488,$$

$$\chi^2_{розр} = 12 \cdot (10-1) \cdot 0,488 = 52,67.$$

За додатком 1 при рівні значущості α=0,05 для числа ступенів вільності *f*=10-1=9 $\chi^2_{табл} = 16,9$. Розрахункове значення критерія $\chi^2$ більше за табличне і можна стверджувати, що думки фахівців співпадають.

Після цього будуємо діаграму рангів факторів, яка відображає колективну думку фахівців. Для цього по осі абсцис наносимо фактори в порядку спадання їх рангу, а по осі ординат – суми рангів для відповідного фактора (останній стовпець табл. 1). За допомогою отриманої діаграми проводиться оцінка значущості факторів (рис.1).

Діаграми рангів можуть мати різний вигляд:

а) спадання майже експоненціальне. Цей випадок з великими спадами найбільш сприятливий. Тут можна розділити фактори на групи і за деяким критерієм відсіяти неістотні;

б) спадання підкоряється параболічному закону. І тут можна згрупувати фактори і відсіяти слабовпливаючі;

в) спадання майже лінійне. Цей випадок «поганий», оскільки фахівці, маючи високу згоду в думках, хоча і допускають відмінності в факторах, але невпевнено. Тут краще включати в експеримент усі фактори;

г) розподіл факторів за сумами рангів майже рівномірний. І цей випадок “поганий”. Тут або фахівці не можуть вибрати серед запропонованих найбільш сильновпливаючі фактори, або всі ці фактори дійсно впливають сильно, або низький рівень апріорної інформації. І тут доводиться включати в програму дослідження всі фактори.

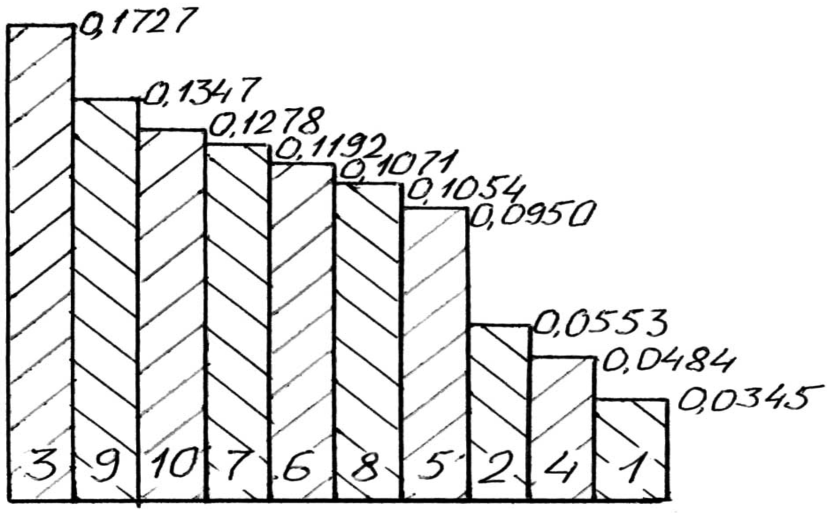


Рис. 1. Діаграма питомих ваг факторів, які впливають на процес формування валка

За аналогією з викладеною методикою можна оцінити перспективність якого-небудь технологічного прийому, способу вирощування сільськогосподарських культур і т.п.

Лекція №7

# ТЕМА: МЕТОД ВИПАДКОВОГО БАЛАНСУ (МЕТОД НАДНАСИЧЕНИХ ПЛАНІВ)

Серед експериментально-статистичних методів виділення суттєвих факторів найбільш ефективні методи, засновані на результатах заздалегідь спланованих експериментів. Найбільш широко використовуються насичені і наднасичені плани.

В першому випадку виходять з передумови, що на вихідний параметр роблять вплив тільки лінійні ефекти і не впливає взаємодія факторів. Застосування методу насичених планів обмежене, оскільки вплив взаємодії факторів на вихідний параметр часто виявляється значним.

Метод наднасичених планів (метод випадкового балансу) дозволяє виділяти найбільш значущі лінійні фактори і їх взаємодії. Цей метод застосовують, якщо число істотних факторів значно менше загального числа факторів, узятих "під підозру".

Число дослідів призначається з умови:

$$n > k + 1\,. \tag{1}$$

де $k$ - число факторів.

Розглянемо конкретний **приклад** використання методу випадкового балансу.

На лабораторній установці вивчався вплив різних факторів на нерівномірність висоти зрізу, яка визначалася за формулою:

$$\overline{y} = \frac{\sum_{i=1}^{m} n_i \cdot y_i}{\sum_{i=1}^{m} n_i}\,, \tag{2}$$

де $y_i = (L_i - H)$ - перевищення дійсної довжини стерні $L_i$ над висотою встановлення $H$ ріжучого апарату; $\sum_{i=1}^{m} n_i$ - число стебел, що зрізуються; $m$ - число повторностей у кожному досліді.

Перелік врахованих факторів поданий в таблиці 3.

Таблиця 3

Фактори, що впливають на нерівномірність висоти стерні в лабораторних умовах

<table>
<tr><th rowspan="2">Фактор і його позначення</th><th colspan="2">Рівні факторів</th></tr>
<tr><th>нижній ( - )</th><th>верхній ( + )</th></tr>
<tr><td>$x_1$ - подача, мм</td><td>60</td><td>100</td></tr>
<tr><td>$x_2$ - довжина стебел від точки закріплення до верхівки колоска, мм</td><td>600</td><td>820</td></tr>
<tr><td>$x_3$ - виніс осі мотовила вперед щодо спинки ножа, мм</td><td>50</td><td>210</td></tr>
<tr><td rowspan="2">$x_4$ - тип ріжучого апарата</td><td colspan="2">двохножовий</td></tr>
<tr><td>с одним рухомим ножем</td><td>с двома рухомими ножами</td></tr>
<tr><td>$x_5$ - установочна висота ріжучого апарата, мм</td><td>70</td><td>150</td></tr>
<tr><td>$x_6$ - кінематичний режим роботи мотовила</td><td>1,3</td><td>1,75</td></tr>
<tr><td>$x_7$ - густота стеблестою пшениці, стебел/м$^2$</td><td>300</td><td>500</td></tr>
<tr><td>$x_8$ - висота осі мотовила над спинкою ножа, мм</td><td>770</td><td>970</td></tr>
</table>

Відзначимо, що з восьми факторів один ( $x_4$ ) має якісний характер: нижній рівень відповідає двохножовому апарату з одним рухомим ножем ($S$=$t$=$t_0$=76,2 мм), верхній – апарату з двома рухомими ножами ($S$=$t$=76,2 мм).

Для планування відсіваючих експериментів скористаємося відомою матрицею (табл. 4), приведеною в роботі [3].

Таблиця 4

Матриця відсіваючого експерименту

| Номер досліду | Фактори | | | | | | | | Значення критерію оптимізації, $\overline{y}$ мм |
|---|---|---|---|---|---|---|---|---|---|
| | $x_1$ | $x_2$ | $x_3$ | $x_4$ | $x_5$ | $x_6$ | $x_7$ | $x_8$ | |
| 1 | - | - | - | - | + | - | + | + | 16 |
| 2 | + | + | - | - | - | + | + | + | 7 |
| 3 | + | - | + | - | + | - | - | - | 12 |
| 4 | + | - | - | + | + | + | + | - | 18 |
| 5 | - | + | + | - | - | - | - | + | 17 |
| 6 | - | + | - | + | + | + | - | + | 21 |
| 7 | - | - | + | + | - | + | - | - | 18 |
| 8 | + | + | + | + | - | - | + | - | 15 |
| 9 | - | + | - | + | + | - | + | + | 23 |
| 10 | + | + | - | - | - | - | - | - | 9 |

В останньому стовпці цієї таблиці наведені результати усереднених в 3-кратній повторності дослідів (нерівномірність висоти зрізу).

Для аналізу отриманих результатів будується спеціальна діаграма (рис. 2). На вісь абсцис наносять всі фактори і їх рівні, а на вісь ординат - дослідні значення $\overline{y}$ . Кожний з факторів розглядається окремо на верхньому і нижньому рівнях.

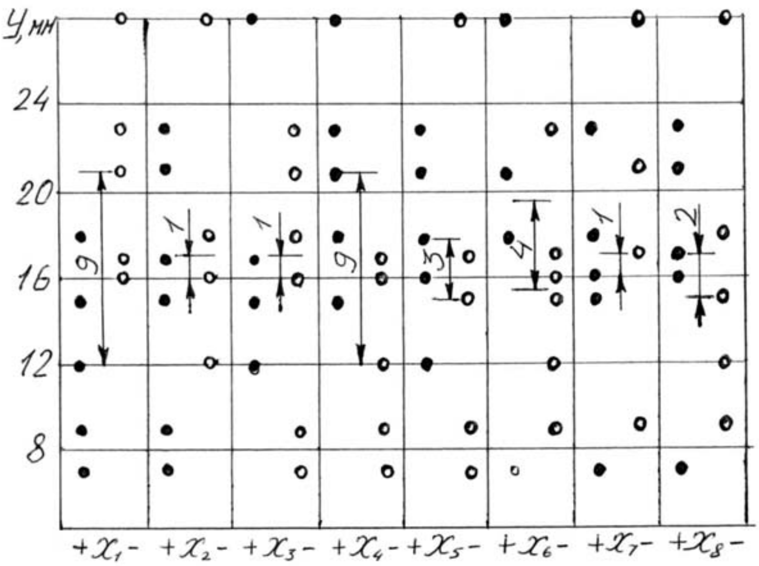


Рис. 2. Виділення значущих факторів за допомогою $\eta$-правила

Після побудови діаграми необхідно для кожного з факторів визначити число виступаючих точок. Наприклад, для фактора $x_1$ число додатніх значень, розташованих нижче від'ємних, дорівнює 4 (15, 12, 9 і 7 мм). Число від'ємних значень цього фактора, розташованих вище додатніх дорівнює 3 (21, 23, і 28 мм). Значить, для фактора $x_1$ число виступаючих точок $q$=4+3.

Для виділення суттєвих факторів доцільно використовувати спосіб, що ґрунтується на числі виступаючих точок і на розбіжності медіан [4]. Медіана завжди припадає на середній випадок варіаційного ряду. Наприклад, по фактору $x_4$ для рівня (+) впорядкований ряд критеріїв оптимізації буде: 26, 23, 21, 18, 15 мм. Медіаною цього ряду стала величина 21 мм. Для фактора $x_6$ із рівнем (+) впорядкований варіаційний ряд: 28, 21, 18, 7 мм, в якому медіана:

$$Me = \frac{(21+18)}{2} = 13,5 \text{ мм.}$$

З факторів, що розглядаються, виділяються значущі, ті, для яких $\eta_1 = Me \cdot q$ має більше значення.

За цим правилом визначаємо величину $\eta$ для факторів, що розглядаються: $\eta_1 = 9 \cdot 7 = 63$ , $\eta_2 = 1 \cdot 3 = 3$ , $\eta_3 = 1 \cdot 3 = 3$ , $\eta_4 = 9 \cdot 7 = 63$ , $\eta_5 = 3 \cdot 3 = 9$ , $\eta_6 = 4 \cdot 2 = 3$ , $\eta_7 = 1 \cdot 2 = 2$ $\eta_8 = 2 \cdot 2 = 4$ .

Як бачимо, $x_1$ (подача) і $x_4$ (тип ріжучого апарату) мають більш високий ступінь впливу на результат експерименту. Перевіримо значущість цих факторів по $t$ -критерію, для цього побудуємо таблицю з двома входами.

Таблиця 5

Обчислення ефектів факторів $x_1$ і $x_4$ (в мм)

| Оцінюваний фактор | $+x_1$ | $-x_1$ |
|---|---|---|
| $+x_4$ | $\sum y_1 = (18 + 15) = 33$<br>$\overline{y}_1 = 16,5$ | $\sum y_2 = (21 + 28 + 23) = 72$<br>$\overline{y}_2 = 24$ |
| $-x_4$ | $\sum y_3 = (7 + 12 + 9) = 28$<br>$\overline{y}_3 = 9,3$ | $\sum y_4 = (16 + 17) = 33$<br>$\overline{y}_4 = 16,5$ |

В клітинках цієї таблиці записані величини критерію оптимізації, отримані в тих дослідах, де фактори $x_1$ і $x_4$ знаходилися на верхніх або на нижніх рівнях. Ефекти виділених факторів:

$$x_1 = \frac{\overline{y}_1 + \overline{y}_3}{2} - \frac{\overline{y}_2 + \overline{y}_4}{2} = \frac{16,5 + 9,3}{2} - \frac{24 + 16,5}{2} = -7,25 \text{ мм.}$$

$$x_4 = \frac{\overline{y}_1 + \overline{y}_2}{2} - \frac{\overline{y}_3 + \overline{y}_4}{2} = \frac{16,5 + 24}{2} - \frac{9,3 + 16,5}{2} = 7,25 \text{ мм.}$$

Середньоквадратична помилка, яка характеризує розсіювання відносно середніх (див. табл. 5), визначається за формулою:

$$S_R = \sqrt{\frac{\sum y_i^2}{n_i - 1} - \frac{\left(\sum y_i\right)^2}{n_i\left(n_i - 1\right)}}, \qquad (3)$$

де $n_i$ - число вимірів по кожній з клітинок.

Обчислимо значення $S_R$ , використовуючи табл. 6.

Таблиця 6

Розрахунок $S_R$

| Номер клітинки | $\sum y_i$ | $\left(\sum y_i\right)^2$ | $\sum y_i^2$ | $n_i$ | $S_R^2$ | $\frac{S_R^2}{n_i}$ |
|---|---|---|---|---|---|---|
| 1 | 33 | 1089 | 549 | 2 | 4,5 | 2,25 |
| 2 | 72 | 5184 | 1754 | 3 | 13 | 4,33 |
| 3 | 28 | 734 | 274 | 3 | 7 | 2,33 |
| 4 | 33 | 1089 | 545 | 2 | 0,5 | 0,25 |

Сума значень в останньому стовпці таблиці:

$$\sum \frac{S_R^2}{n_i} = 2,25 + 4,33 + 2,33 + 0,25 = 9,16 .$$

Розрахункові значення $t$-критерію:

$$t_{x_1} = \frac{\left(\bar{y}_1 + \bar{y}_2\right) - \left(\bar{y}_3 + \bar{y}_4\right)}{\sqrt{\sum \frac{S_R^2}{n_i}}} = \frac{\left(16,5 + 24\right) - \left(9,3 + 16,5\right)}{\sqrt{9,16}} = 4,87 , \tag{4}$$

$$t_{x_4} = \frac{\left(\bar{y}_1 + \bar{y}_3\right) - \left(\bar{y}_2 + \bar{y}_4\right)}{\sqrt{\sum \frac{S_R^2}{n_i}}} = \frac{\left(16,5 + 9,3\right) - \left(24 + 16,5\right)}{\sqrt{9,16}} = 4,87 .$$

Як вже наголошувалося, табличне значення $t_T$ критерію Стьюдента залежить від рівня значущості і числа ступенів вільності. Число ступенів вільності, пов'язане з $S_R$, визначаємо за формулою:

$$f = \sum n_i - K_1 , \tag{5}$$

де $K_1$ - число клітинок в табл. 5.

В нашому випадку $f$=(2+3+3+2)-4=6. При ступені ризику 0,2 отримаємо $t_T = 1,94$ [6, додаток 1]. Оскільки $t_{x_1} = t_{x_4} > t_T$ , то признаємо фактори $x_1$ і $x_4$ значущими з вірогідністю 80 %.

Для визначення наступних за значущістю факторів необхідно виключити вплив ефектів $x_1$ і $x_4$ на нерівномірність висоти зрізу.

Перетворимо результати спостережень (останній стовпець в табл. 4), щоб “зняти” дію ефектів від $x_1$ і $x_4$. Для цього до всіх результатів табл. 4 на рівні ($+x_1$) додаємо величину його ефекту зі зворотним знаком (+7,25), а на рівні ($+x_4$) додаємо (-7,25). З урахуванням цього коректування складаємо нову таблицю і діаграму розсіювання (рис.3).

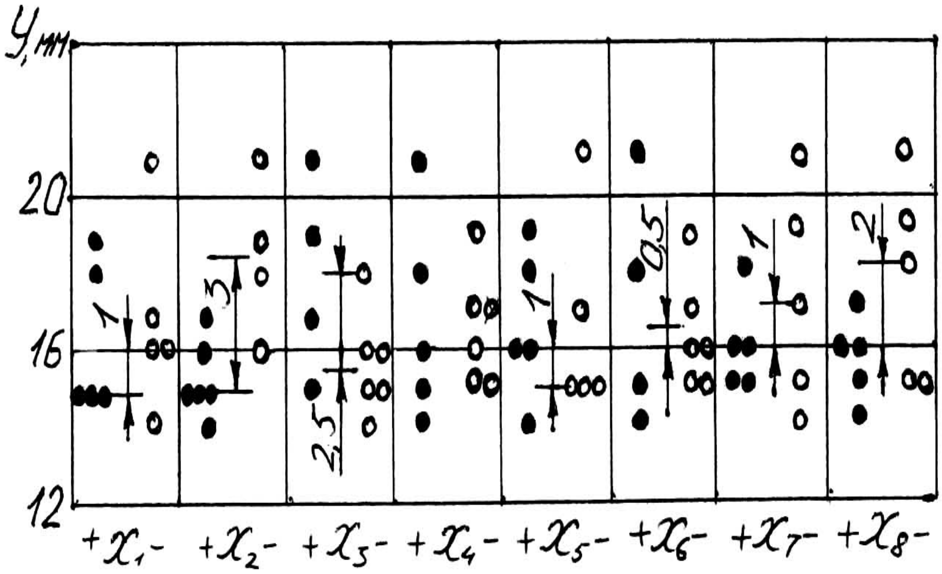


Рис. 3. Скоректована діаграма розсіювання

Таблиця 7

Скоректовані результати експерименту

| Номер досліду | Фактори | | | | | | | | Значення критерію оптимізації, $\bar{y}$ мм |
|---|---|---|---|---|---|---|---|---|---|
| | $x_1$ | $x_2$ | $x_3$ | $x_4$ | $x_5$ | $x_6$ | $x_7$ | $x_8$ | |
| 1 | - | - | - | - | + | - | + | + | 16 |
| 2 | + | + | - | - | - | + | + | + | 14,25 |
| 3 | + | - | + | - | + | - | - | - | 19,25 |
| 4 | + | - | - | + | + | + | + | - | 18 |
| 5 | - | + | + | - | - | - | - | + | 17 |
| 6 | - | + | - | + | + | + | - | + | 13,75 |
| 7 | - | - | + | + | - | + | - | - | 20,75 |
| 8 | + | + | + | + | - | - | + | - | 15 |
| 9 | - | + | - | + | + | - | + | + | 15,75 |
| 10 | + | + | - | - | - | - | - | - | 16,25 |

Використовуючи $\eta$ -плавило, підрахуємо питому вагу факторів: $\eta_1 = 1 \cdot 2 = 2$ , $\eta_2 = 3 \cdot 3 = 9$ , $\eta_3 = 2,5 \cdot 3 = 7,5$ , $\eta_4 = 0$ , $\eta_5 = 1 \cdot 2 = 9$ , $\eta_6 = 0,5 \cdot 2 = 1$ , $\eta_7 = 1 \cdot 3 = 3$ $\eta_8 = 2 \cdot 4 = 8$ . Аналіз скоректованої діаграми показує, що ефекти $x_1, x_4, x_5$ наблизилися до нульової величини; трохи зросли ефекти від факторів $x_6$ і $x_7$. Найбільш значущими виявляються ефекти від факторів $x_2$ (довжина стебел) $x_3$ (винесення мотовила) і $x_8$ (висота мотовила над ріжучим апаратом). Оскільки фактор $x_2$ залежить від умов зростання культури, то ним важко керувати. Надалі оцінюємо тільки ефекти керованих факторів: $x_3$ і $x_8$. Аналогічно повторюємо всі операції першого етапу статистичного аналізу (табл. 8, 9).

Таблиця 8

Обчислення ефектів факторів $x_3$ і $x_8$ (в мм)

| Оцінюваний фактор | $+x_3$ | $-x_3$ |
|---|---|---|
| $+x_8$ | $\sum y_5 = y_5 = 17$ | $\sum y_6 = 16 + 14,25 + 13,75 + 15,75 = 59,75$, $\bar{y}_6 = 14,9$ |
| $-x_8$ | $\sum y_7 = 19,25 + 20,75 + 15 = 55$, $\bar{y}_3 = 9,3$ | $\sum y_8 = 18 + 16,25 = 34,25$ $\bar{y}_8 = 17,10$ |

Ефекти виділених факторів:

$$x_3 = \frac{\bar{y}_5 + \bar{y}_7}{2} - \frac{\bar{y}_6 + \bar{y}_8}{2} = 1,65 \text{ мм};$$

$$x_8 = \frac{\bar{y}_5 + \bar{y}_6}{2} - \frac{\bar{y}_7 + \bar{y}_8}{2} = -1,75 \text{ мм}.$$

Таблиця 9

Розрахунок середньоквадратичної помилки $S_R$

| Номер клітинки | $\sum y_i$ | $\left(\sum y_i\right)^2$ | $\sum y_i^2$ | $n_i$ | $S_R^2$ | $\frac{S_R^2}{n_i}$ |
|---|---|---|---|---|---|---|
| 1 | 17 | 289 | 289 | 1 | 0 | 0 |
| 2 | 59,75 | 3570 | 896 | 4 | 1,1 | 0,275 |
| 3 | 55 | 3025 | 1026 | 3 | 8,83 | 2,943 |
| 4 | 34,25 | 1173 | 588 | 2 | 1,5 | 0,75 |

Сума $\sum \frac{S_R^2}{n_i} = 0,275 + 2,943 + 0,75 = 3,97$ .

Розрахункові значення $t$-критерію:

$$t_{x_3} = \frac{(\bar{y}_5 + \bar{y}_6) - (\bar{y}_7 + \bar{y}_8)}{\sqrt{\sum S_R^2 / n_i}} = \frac{(17 + 14,9) - (18,3 + 17,1)}{\sqrt{3,97}} = 1,757;$$

$$t_{x_8} = \frac{(\bar{y}_5 + \bar{y}_7) - (\bar{y}_7 + \bar{y}_8)}{\sqrt{\sum S_R^2 / n_i}} = \frac{(18,3 + 17,1) - (17 + 14,9)}{\sqrt{3,97}} = 1,686.$$

При ступені ризику $\alpha = 0,2$ і $f$=(1+4+3+2)-4=6 табличне значення критерію Стьюдента буде $t_T = 1,44$ . Оскільки $t_{x_3} > t_{x_8} > t_T$ , то вважаємо фактори $x_3$ і $x_8$ значущими.

Величина ефектів лінійних факторів, що впливають на нерівномірність висоти стерні, представлена на рис. 4.

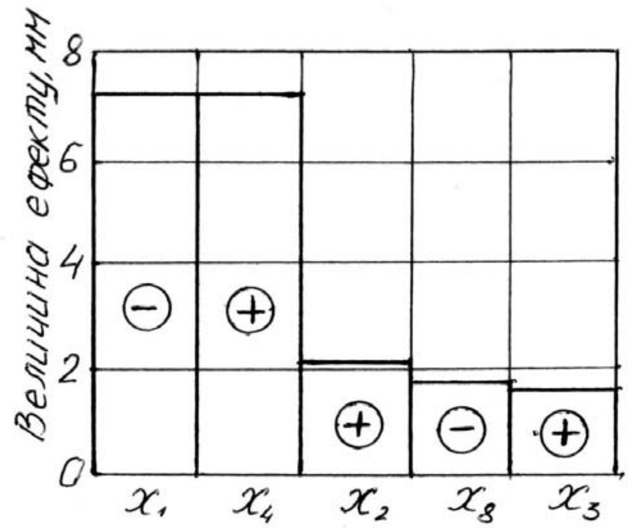


Рис. 4. Ефекти факторів, що впливають на нерівномірність висоти стерні

Знайшовши всі значущі лінійні ефекти, потрібно встановити також ефекти двухфакторних взаємодій [5]. Найбільш суттєві взаємодії повинні мати і найбільше число точок, що виділяються, на верхньому і нижньому рівнях. З аналізу діаграм розсіювання видно (див. рис. 4), що найбільш істотними взаємодіями можуть бути: $x_1 x_4$, $x_1 x_2$, $x_1 x_8$ і т. д.

**Модуль 4**

# ОПТИМІЗАЦІЯ ОБ'ЄКТІВ ДОСЛІДЖЕННЯ

Лекція №8

## ТЕМА: КАНОНІЧНЕ ПЕРЕТВОРЕННЯ ЛІНІЙНИХ МОДЕЛЕЙ

Після одержання адекватної математичної моделі другого порядку, допустимо типу, необхідно визначити координати оптимуму (якщо він існує) і вивчити властивості поверхні відгуку в межах оптимуму.

Найчастіше використовують канонічні перетворення математичної моделі. Для аналізу і систематизації рівняння другого порядку приводиться до типової канонічної форми виду:

$$Y - Y_S = B_{11}X_1^2 + B_{22}X_2^2 + B_{kk}X_k^2, \tag{1}$$

де $Y$ — значення критерію оптимізації;

$Y_S$ — значення критерію оптимізації в оптимальній точці;

$X_1, X_2, X_k$ — нові осі координат, повернені відносно старих $x_1, x_2, x_k$;

$B_{11}, B_{22}, B_{kk}$ — коефіцієнти регресії в канонічній формі.

При канонічному перетворенні рівняння (1) проводиться перенос початку координат у нову точку $S$ і поворот старих осей на деякий кут у факторному просторі, у результаті чого зникають лінійні члени і змінюється значення вільного члена. Щоб здійснити перенос початку координат в особливу точку поверхні відгуку, необхідно продифференціювати функцію відгуку по кожній зміній і, прирівняти до нуля часткові похідні, розв'язати отриману систему рівнянь, тобто знайти значення факторів, які оптимізують величину критерію оптимізації. Розв'язком системи k-лінійних рівнянь є координати оптимуму (центру фігури, якщо поверхня має центр). При числі факторів $k \le 3$ про функцію відгуку можна одержати наочне геометричне представлення. Так, залежність $Y$ від єдиного фактору $X$ зображується на площині у вигляді прямої або кривої лінії. Причому кожний дослід дає пари значень ($Y$; $X$), тобто точку на даній лінії. Залежність $Y$ від двох факторів ($X_1;X_2$) описується площиною або криволінійною поверхнею в тривимірному просторі. У випадку трьох і більш факторів наочне зображення неможливе, але математики без особливої складності описують відповідну функцію як деяку гіперповерхню, і розраховують координати точок, положення екстремумів і навіть маршрути руху по такій поверхні.

Розташування факторів у координатному просторі називається факторним простором. Геометричне зображення функції відгуку у факторному просторі називається поверхнею відгуку.

Існує кілька типів поверхні відгуку. Розглянемо чотири основні типи (рис. 1):

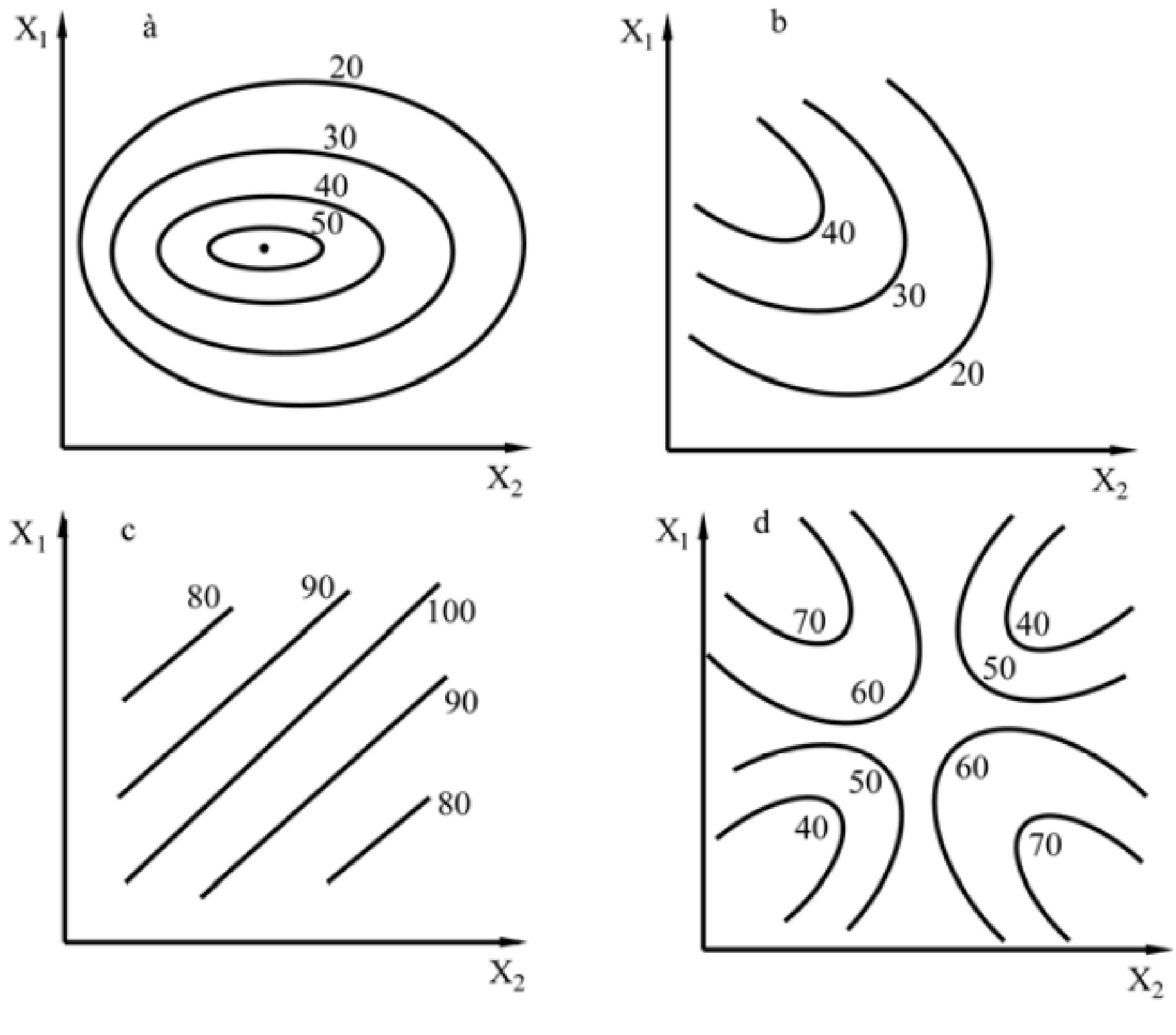


Рис. 1. Двовимірний перетин різних типів поверхні відгуку:

a) «Вершина», b) «Стаціонарне піднесення», c) «Хребет», d) «Сідло».

1. «Вершина». Такий вид поверхні відгуку описується функцією, що має максимум або мінімум в одній точці факторного простору (рис. 1a).

2. «Стаціонарне піднесення» - плавне зростання (убування) функції відгуку при зміні факторів (рис. 1b). При такому типі поверхні відгуку максимум (або мінімум) не спостерігається.

3. «Хребет» або «Яр» - функція відгуку має множину значень максимумів або мінімумів, розташованих у вигляді хребта (яру) (рис. 1c).

4. «Сідло» - на двох ділянках факторного простору спостерігається збільшення функції відгуку, а на дві інших ділянках - зменшення (рис. 1d).

Зустрічаються поверхні відгуку і з більш складними конфігураціями. Якщо число факторів більше двох, для наглядного зображения функції відгуку користуються двовимірними перетинами поверхні відгуку (рис. 1). Тут не має ні одного центру з максимальним значенням критерія оптимізації і центром може бути будь-яка точка на осі, яка відповідає нульовому значенню коефіцієнта регресії канонічної форми.

Приведення функції відгуку до канонічної форми необхідно проводити і при числі факторів більше 3.

Тут ми розглянемо метод двовимірних перетинів, але перед цим викладемо процедуру канонічного перетворення цільової функції. Як ми вже відзначали, спочатку здійснюється перенос початку координат у точку *S*. Для цього необхідно знайти часткові похідні по $x_1, x_2, x_k$ у математичній моделі обєкту дослідження другого порядку, яка має вигляд

$$y = b_0 + \sum^{k} b_i x_i + \sum_{i<j}^{k} b_{ij} x_i x_j + \sum^{k} b_{ii} x_i^2 , \qquad (2)$$

і прирівняти їх до нуля. Розв'язком системи отриманих рівнянь будуть знайдені координати нового центру в старих осях ($x_1, x_2, x_k$). Підставляючи знайдені значення в рівняння (2), визначають величину критерію оптимізації в точці *S(Ys)*. Якщо головний визначник системи рівнянь дорівнює нулю, то поверхня не має центру. Тоді точку *S* поміщають у старий початок координат або в будь-яку іншу точку із кращим значенням критерію оптимізації на головних напрямках поверхні.

Після переносу центру в точку *S* рівняння (2) матиме вигляд:

$$Y = Y_S + \sum_{ij}^{k} b_{ij} x_i \, x_j + \sum^{k} b_{ii} x_i^2 \,, \tag{3}$$

Потім у новому початку координат *S* осі координат повертають на кут α до сполучення з головними осями геометричної поверхні. У результаті цих операцій одержують канонічну форму другого порядку типу (1).

Для визначення кута повороту α є справедливим рівняння:

$$tg 2\alpha = \frac{b_{ij}}{b_{ii} - b_{jj}} = \frac{b_{12}}{b_{11} - b_{22}}. \tag{4}$$

Тоді коефіцієнти регресії в канонічній формі при *k=2* визначаються із співвідношень:

$$\begin{cases} B_{11} = b_{11}\cos^2\alpha + b_{12}\cos\alpha\sin\alpha + b_{22}\sin^2\alpha; \\ B_{22} = b_{11}\sin^2\alpha - b_{12}\cos\alpha\sin\alpha + b_{22}\cos^2\alpha; \\ B_{12} = 2(b_{22} - b_{11})\cos\alpha\sin\alpha + b_{12}(\cos^2\alpha + \sin^2\alpha); \\ B_1 = \dfrac{b_{13}\cos\alpha}{2} + \dfrac{b_{23}\sin\alpha}{2}; \\ B_2 = \dfrac{b_{23}\cos\alpha}{2} - \dfrac{b_{13}\sin\alpha}{2}. \end{cases} \tag{5}$$

Або ж для того, щоб отримати значення $B_{11}$ та $B_{22}$ необхідно розв'язати характеристичне рівняння

$$f(B) = \begin{vmatrix} b_{11} - B & \dfrac{b_{12}}{2} \\ \dfrac{b_{12}}{2} & b_{22} - B \end{vmatrix} = B^2 - (b_{11} + b_{22})B + (b_{11}b_{22} - \frac{b_{12}^2}{4}) = 0 \,. \tag{6}$$

Два корені характеристичного рівняння (6) є невідомими значеннями $B_{11}$ та $B_{22}$.

При *k=3* для знаходження коефіцієнтів канонічної форми $B_{11}$, $B_{22}$ та $B_{33}$ необхідно скласти і розв'язати характеристичне рівняння поверхні другого порядку.

Незважаючи на те, що при кількості факторів *k>3* наочне уявлення про геометричний образ функції відгуку стає неможливим, тому що збільшується різноманіття фігур, які можуть бути описані рівняннями другого порядку, і підвищується розмірність факторного простору, тому і у цьому випадку необхідно проводити канонічне перетворення рівняння регресії, оскільки канонічний аналіз рівняння полегшує інтерпретацію результатів дослідження.

Виконаємо канонічне перетворення наступного рівняння регресії, яке отримане в результаті проведення D-оптимального плану Кіфера $Ki_{75}$:

$$\begin{gathered} y = 0{,}487 + 0{,}119x_2 - 0{,}072x_6 + 0{,}02x_8 - 0{,}218x_{11} - \\ - 0{,}122x_{14} - 0{,}092x_2x_6 - 0{,}0014x_2x_8 - 0{,}094x_2x_{11} - \\ - 0{,}004x_2x_{14} - 0{,}063x_6x_8 + 0{,}042x_6x_{11} + 0{,}088x_6x_{14} - \\ - 0{,}051x_8x_{11} + 0{,}059x_8x_{14} + 0{,}109x_{11}x_{14} + 0{,}197x_2^2 + 0{,}207x_6^2 + \\ + 0{,}069x_8^2 + 0{,}167x_{11}^2 + 0{,}13x_{14}^2. \end{gathered} \tag{7}$$

Таблиця 1

Фактори, інтервали і рівні їх варіювання в експериментах

по описанню області оптимуму

| Вид плану | Рівень і інтервал варіювання факторів | Фактор | | | | |
|---|---|---|---|---|---|---|
| | | Колова швидкість молотків $v_м$, м/с | Подача матеріалу *Q*, кг/год | Число молотків $z_м$, шт | Довжина різки $L_{ср}$, мм | Зазор *S*, мм |
| | | $x_2$ | $x_6$ | $x_8$ | $x_{11}$ | $x_{14}$ |
| D-оптимальний план Кіфера $Ki_{75}$ | +1 | 120 | 800 | 48 | 35 | 45 |
| | 0 | 100 | 550 | 32 | 25 | 25 |
| | -1 | 80 | 300 | 16 | 15 | 5 |
| | ε | 20 | 250 | 16 | 10 | 20 |

1. Система диференціальних рівнянь, отримана з рівняння (7),

$$\begin{cases} \dfrac{\partial y}{\partial x_2} = 0,119 - 0,092x_6 - 0,0014x_8 - 0,094x_{11} - 0,004x_{14} + 0,394x_2 = 0; \\ \dfrac{\partial y}{\partial x_6} = -0,072 - 0,092x_2 - 0,063x_8 + 0,042x_{11} + 0,088x_{14} + 0,414x_6 = 0; \\ \dfrac{\partial y}{\partial x_8} = 0,02 - 0,0014x_2 - 0,063x_6 - 0,051x_{11} + 0,059x_{14} + 0,138x_8 = 0; \\ \dfrac{\partial y}{\partial x_{11}} = -0,218 - 0,094x_2 + 0,042x_6 - 0,051x_8 + 0,109x_{14} + 0,334x_{11} = 0; \\ \dfrac{\partial y}{\partial x_{14}} = -0,122 - 0,004x_2 + 0,088x_6 + 0,059x_8 + 0,109x_{11} + 0,26x_{14} = 0. \end{cases} \quad (8)$$

Координати нового центру в старих осях координат отримані розв’язанням системи рівнянь (8), а розв’язком рівняння (1) при підстановці в нього ровних (оптимальних) значень визначається величина параметру оптимізації в точці *S(Y_S)*.

Для визначення коефіцієнтів $B_{2\text{-}2}$, $B_{6\text{-}6}$, … , $B_{14\text{-}14}$ розв’язували характеристичне рівняння:

$$f(B) = \begin{vmatrix} 0,197 - B & -0,046 & -0,0007 & -0,047 & -0,002 \\ -0,046 & 0,207 - B & -0,0315 & 0,021 & 0,044 \\ -0,0007 & -0,0315 & 0,069 - B & -0,0255 & 0,0295 \\ -0,047 & 0,021 & -0,0255 & 0,167 - B & 0,0545 \\ -0,002 & 0,044 & 0,0295 & 0,0545 & 0,13 - B \end{vmatrix} = 0 . \quad (9)$$

Складено по рівнянню (1):

$$f(B)=\begin{vmatrix} b_{2-2}-B & \frac{b_{2-6}}{2} & \frac{b_{2-8}}{2} & \frac{b_{2-11}}{2} & \frac{b_{2-14}}{2} \\ \frac{b_{2-6}}{2} & b_{6-6}-B & \frac{b_{6-8}}{2} & \frac{b_{6-11}}{2} & \frac{b_{6-14}}{2} \\ \frac{b_{2-8}}{2} & \frac{b_{6-8}}{2} & b_{8-8}-B & \frac{b_{8-11}}{2} & \frac{b_{8-14}}{2} \\ \frac{b_{2-11}}{2} & \frac{b_{6-11}}{2} & \frac{b_{8-11}}{2} & b_{11-11}-B & \frac{b_{11-14}}{2} \\ \frac{b_{2-14}}{2} & \frac{b_{6-14}}{2} & \frac{b_{8-14}}{2} & \frac{b_{11-14}}{2} & b_{14-14}-B \end{vmatrix}=0 \quad . \tag{9}$$

В результаті розвязання рівняння (8) були отримані наступні значення коефіцієнтів: $B_{2\text{-}2}$=0,171, $B_{6\text{-}6}$=0,2904, $B_{8\text{-}8}$=0,0261, $B_{11\text{-}11}$=0,1655, $B_{14\text{-}14}$=0,114.

Рівняння регресії (1), яке представлено в канонічній формі має наступний вид:

$$Y-0{,}4037=0{,}171X_2^2+0{,}2904X_6^2+0{,}0261X_8^2+0{,}1655X_{11}^2+0{,}114X_{14}^2. \tag{10}$$

Оскільки всі коефіцієнти при квадратичних членах мають позитивні знаки, то поверхня відгуку, описана рівнянням (1) представляє не що інше як пятивимірний параболоїд з координатами центру поверхні $x_{2S}$ = -0,1704; $x_{6S}$ =0,023; $x_{8S}$ =0,052106; $x_{11S}$ =0,52152; $X_{14S}$=0,25176 (фактори відповідно мають значення: швидкість молотків 96,9 м/с; подача матеріалу в дробарку 555,8 кг/год; кількість молотків на молотковому полі ротора 31,2 шт., середня довжина часток матеріалу, що подрібнюється, 30,2 мм і зазор між кінцями молотків і камерою дроблення 30 мм). Центр фігури знаходиться в області експерименту, а також поблизу апріорно обраного центру експерименту.

## 2. Вивчення поверхні відгуку за допомогою двомірних перетинів

Після канонічного перетворення ы визначення виду поверхні відгуку починається її аналіз, який зручно проводити за допомогою двовимірних перетинів.

Для полегшення розрахунків аналіз звичайно проводиться із закодованими величинами факторів.

Побудова двовимірних перетинів функції відгуку виконується в наступній послідовності. У модель (2) підставляються закодовані значення (оптимальні або близькі до оптимальних) усіх факторів, крім досліджуваних двох. Далі в отриманому рівнянні визначається центр поверхні відгуку шляхом визначення часткових похідних по кожному фактору, які прирівнюються до нуля. По формулах (4) (5) і (6) проводиться канонічне перетворення моделі другого порядку. Після канонічного перетворення визначається тип поверхні відгуку в перетині і починається графоаналітичний аналіз отриманого рівняння. На графіку в координатах незалежних змінних з натуральному масштабі наноситься центр (особлива точка) поверхні відгуку і з нього проводять координатні осі головних напрямків канонічного рівняння. Потім, надаючи різні значення критерію оптимізації в канонічному рівнянні, будується серія кривих рівного виходу (ізоліній), в області допустимих значень зміни незалежних змінних. По кривим перетинів роблять висновки про зміну величини критерію оптимізації залежно від натуральних значень розглянутих факторів. Геометрична інтерпритація усіх можливих двовимірних значень дає уявлення про значення критерію оптимізації, яке він буде приймати при варіюванні рівнів кожної пари факторів.

У випадку, якщо поверхня відгуку має центр, який розташований не в області експерименту, або не має його, двовимірні перетини дають уявлення про деякий умовний оптимум функції, тобто про найбільш сприятливі комбінації факторів досліджуваного процесу, так як розв'язок системи диференціальних рівнянь по кожному з факторів в цьому випадку не дає практичного результату. В цьому випадку дослідження поверхні відгуку за допомогою двомірних перетинів є розв'язком поставленої задачі.

Як приклад приводимо вивчення методом двовимірних перетинів поверхні відгуку, представленої рівнянням, наприклад (2).

З десяти можливих двовимірних перетинів представляється доцільним у першу чергу досліджувати ті перетини, які мають найбільше практичне значення.

Двовимірний перетин поверхні відгуку, що характеризує показник енергоємності подрібнення залежно від окружної швидкості молотків ($x_2$) і зазору між їх кінцями і камерою дроблення ($x_{14}$). Для одержання цього перетину підставляємо значення $x_6=0,\ x_8=0,\ x_{11}=0$ у рівняння (2). У результаті отримали наступне рівняння:

$$y = 0{,}487 + 0{,}119x_2 - 0{,}122x_{14} - 0{,}004x_2x_{14} + 0{,}197x_2^2 + 0{,}13x_{14}^2. \qquad (11)$$

Визначаємо координати центру поверхні диференціюванням рівняння (11) і розв'язком отриманої системи рівнянь:

$$\begin{cases} \dfrac{\partial y}{\partial x_2} = 0{,}119 - 0{,}004x_{14} + 0{,}494x_2 = 0, \\ \dfrac{\partial y}{\partial x_{14}} = -0{,}122 - 0{,}004x_2 + 0{,}26x_{14} = 0, \end{cases} \qquad (12)$$

$$x_{2S} = -0{,}298, \quad x_{14S} = 0{,}463.$$

Підставляючи значення $x_{2S}$ й $x_{14S}$ у рівняння (12), отримаємо значення показника енергоємності в центрі поверхні $Y_S = 0{,}441$.

Проводимо канонічне перетворення рівняння (12), для чого вирішуємо характеристичне рівняння:

$$f(B) = \begin{vmatrix} b_{2-2} - B & \dfrac{b_{2-14}}{2} \\ \dfrac{b_{2-14}}{2} & b_{14-14} - B \end{vmatrix} = B^2 - (b_{2-2} + b_{14-14})B + (b_{2-2}b_{14-14} - \frac{b_{2-14}^2}{4}) = 0 \ .(13)$$

У нашому випадку

$$f(B) = \begin{vmatrix} 0{,}197 - B & -0{,}002 \\ -0{,}002 & 0{,}130 - B \end{vmatrix} = B^2 - (b_{2-2} + b_{14-14})B + (b_{2-2}b_{14-14} - \frac{b_{2-14}^2}{4}) = 0$$

.

(14)

Загальними числами (коренями) даного характеристичного рівняння будуть: $B_{2-2}$=0,199; $B_{14-14}$=0,128, а саме рівняння в канонічній формі запишеться:

$$Y - 0{,}441 = 0{,}199 X_2^2 + 0{,}128 X_{14}^2. \qquad (15)$$

Визначення кута повороту осей координат у точцы $S$ виконуємо по формулі (4).

У нашому випадку

$$tg2\alpha = \frac{b_{12}}{b_{11} - b_{22}} = \frac{-0{,}004}{0{,}197 - 0{,}13} = -0{,}0597.$$

$$2\alpha = arctg(-0{,}0597) = -3°24' \Rightarrow \alpha = -1°42'. \qquad (16)$$

Підставляючи різні значення показника енергоємності з рівняння (15), одержуємо рівняння відповідних контурних кривих — еліпсів, у сукупності, що представляють ціле сімейство сполучених еліпсів (лінії рівного значення показника енергоємності). Результати розрахунків наведені в табл. 2 і представлені на рис. 2. З рис. 2 слідує, що мінімальне значення показника енергоємності ($E$) з розглянутому перетині поверхні відгуку при інших факторах ($x_6$, $x_8$, $x_{11}$), узятих на нульовому рівні, дорівнює 0,441 кВт·год/(т.од.ст.подр) і має місце при окружній швидкості молотків близько 95 м/с і зазорі між їх кінцями і камерою подрібнення, яке приблизно дорівнює 35 мм. На основі цього рисунку можна зробити висновок по те, що оптимальні значення розглянутих факторів можуть перебувати в межах $v_м$=90…98 м/с і зазор $S$=28...40 мм.

Таблиця 2

Допоміжна таблиця для розрахунків координат основних точок при побудові двомірного перетину

| Величина параметра $Y$ | $X_2$ | $X_{14}$ | Величина параметра $Y$ | $X_2$ | $X_{14}$ |
|---|---|---|---|---|---|
| 0,441 | 0 | 0 | 0,570 | 0 | ±0,010 |
| 0,450 | 0 | ±0,265 | 0,570 | ±0,805 | 0 |
| 0,450 | ±0,213 | 0 | 0,610 | 0 | ±1,150 |
| 0,470 | 0 | ±0,470 | 0,610 | ±0,920 | 0 |
| 0,470 | ±0,380 | 0 | 0,690 | 0 | ±1,390 |
| 0,490 | 0 | ±0,625 | 0,690 | ±0,120 | 0 |
| 0,490 | ±0,496 | 0 | 0,790 | 0 | ±1,650 |
| 0,530 | 0 | ±0,834 | 0,790 | ±1,32 | 0 |
| 0,530 | ±0,670 | 0 | | | |

Двумірний перетин поверхні відгуку, яке описано рівнянням (7) по факторах: число молотків на молотковому полі ротора ($x_8$) і зазор ($x_{14}$) представлені на рис. 3. Для одержання цього перетину підставляємо значення: $x_2$=0, $x_6$=0, $x_{11}$=0 у рівняння (7) і знаходимо:

$$y = 0{,}487 + 0{,}02x_8 - 0{,}122x_{14} + 0{,}059x_8x_{14} + 0{,}069x_8^2 + 0{,}13x_{14}^2. \qquad (17)$$

При диференціюванні рівняння (17) були знайдені координати центру: $x_{8S}$=-0,385; $x_{14S}$=0,556. Підставляючи значення $x_{8S}$ і $x_{14S}$ у рівняння (17), одержали значення показника енергоємності в центрі поверхні: $Y_S$=0,449.

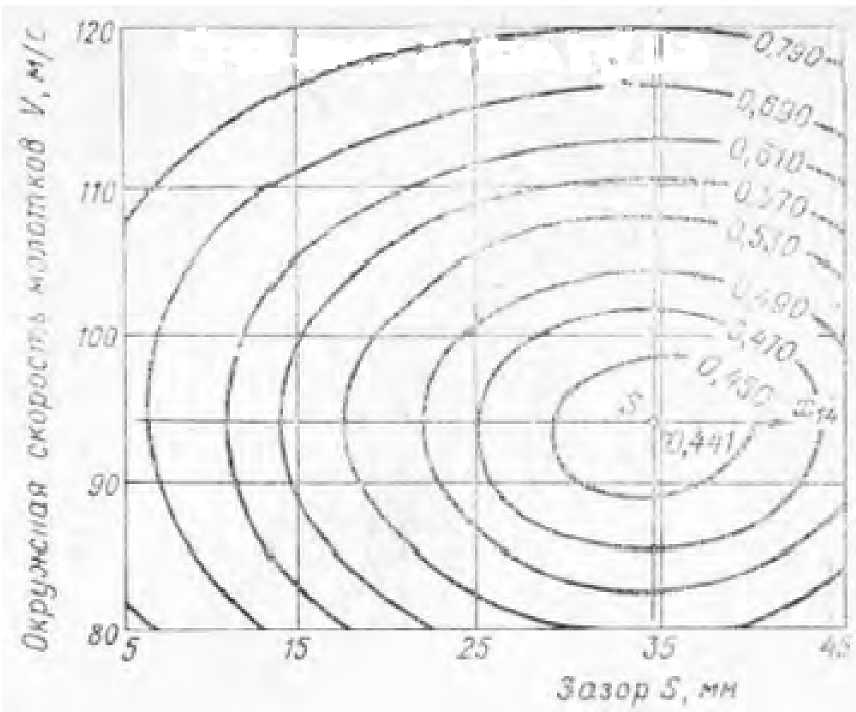

Рис. 2. Двомірний перетин поверхні відгуку, що характеризує показник енергоємності при $x_6=0; x_8=0; x_{11}=0$ (показані лінії рівної енергоємності)

Кут повороту нових осей координат, визначений по формулі (4), дорывнюэ α=-22°.

Після розрахунків коефіцієнтів рівняння регресії в канонічній формі виявилося, що: $B_{8-8}$=0,0565; $B_{14-14}$=0,1425.

Рівняння регресії в канонічній формі запишеться

$$Y - 0{,}449 = 0{,}0565\,X_8^2 + 0{,}1425\,X_{14}^2 .$$

Правильність розрахунків підтверджується перевіркою, тобто порівнянням сум коефіцієнтів при квадратичних членах 0,068785+0,13031=0,199095; 0,0565+0,1425=0,199.

Тут слід зазначити, що при негативному куті α, осі в центрі повертаються по годинниковій стрілці (рис. 3), а при позитивному — проти годинникової стрілки (рис. 4).

На підставі отриманих даних побудований двомірний перетин (рис. 3), де представлена система контурних кривих, що є еліпсами.

Аналізуючи рис. 3, бачимо, що мінімальне значення показника енергоємності в перетині області оптимуму відносно кількості молотків на молотковому полі ротора ($x_8$) і зазору $s$ при інших факторах ($x_2$, $x_6$, $x_{11}$), що зафіксовані на нульовому рівні, дорівнює 0,449 кВт·год/(т.од.ст.подр.), а область оптимуму знаходить ся в межах $z_м$=20...30 шт. і зазор $s$=30...40 мм. При цьому зі зменшенням зазору $s$ між кінцями молотків і камерою дроблення до 5...10 мм показник енергоємності *(Е)* збільшується в 1,5...2 рази.

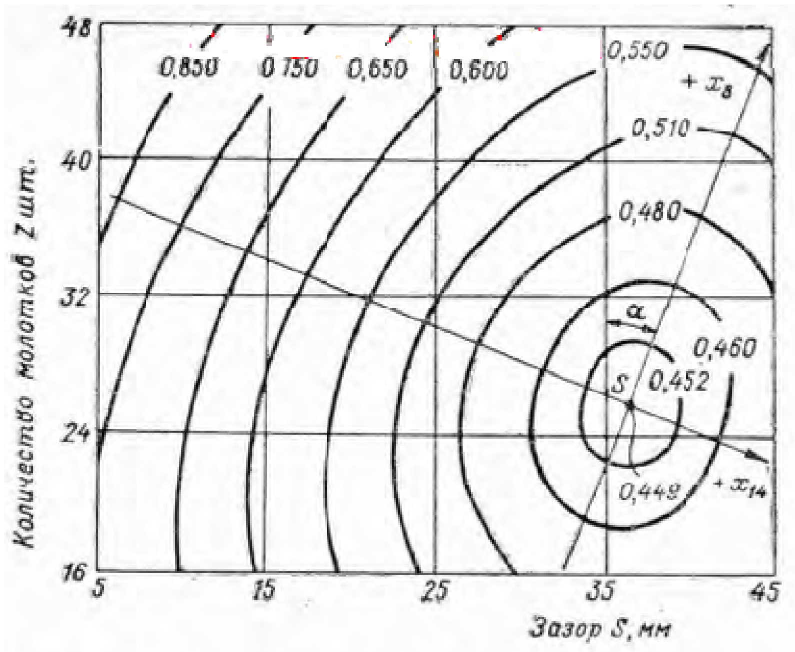


Рис. 4. Двомірний перетин поверхні відгуку, що характеризує показник енергоємності при $x_2=0; x_6=0; x_{11}=0$ (показані лінії рівної енергоємності)

Подібним чином можна побудувати і проаналізувати і всі інші із восьми, що залишилися нерозглянутих двомірних перетинів.

Компромісні задачі можна вирішувати ы за допомогою двовимірних перетинів. Суть методу полягає в тому, що при реалізації матриці плану одночасно розраховуються два критерії оптимізації і складаються два рівняння регресії. Розв'язок задачі із трьома і більш критеріями оптимізації стає надзвичайно громіздким і тому не застосовується.

Аналіз рівнянь регресії й побудова двовимірних перетинів проводиться одночасно і лінії рівного рівня для обох критеріїв оптимізації наносяться на одне креслення. При цьому оптимальне значення одного критерію оптимізації ухвалюється залежно від значення іншого.