

МІНІСТЕРСТВО ОСВІТИ І НАУКИ УКРАЇНИ
НАЦІОНАЛЬНИЙ ТЕХНІЧНИЙ УНІВЕРСИТЕТ УКРАЇНИ
“КИЇВСЬКИЙ ПОЛІТЕХНІЧНИЙ ІНСТИТУТ
імені Ігоря Сікорського”



НАУКОВІ ДОСЛІДЖЕННЯ СКЛАДНИХ ТЕХНІЧНИХ СИСТЕМ

Навчальний посібник: ч.1

*Рекомендовано Методичною радою КПІ ім. Ігоря Сікорського
як навчальний посібник для здобувачів ступеня бакалавра
за спеціальностями 131 «Прикладна механіка»*

Київ
КПІ ім. Ігоря Сікорського
2023

Рецензенти: Володимир ДРАГОБЕЦЬКИЙ – док-р техн. наук,
Кременчуцький національний університет Імені Михайла
Остроградського

Відповідальний редактор: Олександр КРАВЕЦЬ – канд. тех. наук, Національний
технічний університет України «КПІ ім. Ігоря
Сікорського»

*Гриф надано Методичною радою КПІ ім. Ігоря Сікорського
(протокол № 7 від 27.04.2023 р.) за поданням Вченої ради
Механіко-машинобудівного інституту (протокол № 5 від 26.12.2022 р.)*

Електронне навчальне видання

*Саленко Олександр Федорович, професор, докт. техн. наук
Вакуленко Сергій Валентинович, старший викладач*

НАУКОВІ ДОСЛІДЖЕННЯ СКЛАДНИХ ТЕХНІЧНИХ СИСТЕМ НАВЧАЛЬНИЙ ПОСІБНИК ч.1.

Наукові дослідження складних технічних систем: навч.посібник. ч.1. для студентів спеціальності 131 «Прикладна механіка» / О.Ф. Саленко, С.В. Вакуленко; КПІ ім. Ігоря Сікорського. – Електронні текстові дані (1 файл: 3,7 Мбайт). – Київ : КПІ ім. Ігоря Сікорського, 2023. – 102 с.

Навчальний посібник розкриває окремі теми при вивченні дисципліни "Наукові дослідження складних технічних систем" для студентів рівня підготовки "Магістр" і має на меті поліпшити та систематизувати знання щодо математичного апарату, що використовується при дослідженнях

Матеріал посібника направлений на практичне закріплення і використання отриманих знань і надбання навичок з основ сучасних методів, методики і підходів щодо дослідження складних технічних систем, прогнозування їх поведінки у різних несталіх умовах зовнішнього середовища, оцінки надійності функціонування складних технічних систем та використання набутих знань у професійній діяльності. Посібник буде корисним при виконанні лабораторних робіт а також при роботі над курсовою роботою із дисципліни. Такі роботи виконуються за індивідуальними завданням відповідно до наукових напрямків, за якими ведеться робота науковцями кафедри.

© О.Ф. Саленко, С.В. Вакуленко 2023

ЗМІСТ

ВСТУП	6
1 ЗАГАЛЬНІ ПОЛОЖЕННЯ	7
2 ВИПАДКОВІ ВЕЛИЧИНИ ТА ЗАКОНИ ЇХ РОЗПОДІЛУ	8
2.1 Закони розподілу дискретних випадкових величин.....	8
2.2 Числові характеристики та властивості дискретних випадкових величин	11
2.3 Закони розподілу безперервних випадкових величин	14
2.4 Числові характеристики безперервних випадкових величин	15
2.5 Вибіркові аналоги інтегральної та диференціальної функцій розподілу	18
2.6 Методи описової статистики у STATGRAPHICS	22
2.7 Нормальний розподіл та його числові характеристики	34
3 ДИСПЕРСІЙНИЙ АНАЛІЗ	
3.1 Сутність задачі дисперсійного аналізу	37
3.2 Дисперсійни аналіз	38
3.3 Ранговий однофакторний аналіз	46
3.4 Критерій Краскела - Уолліса (H-критерій)	48
3.5 Однофакторний ранговий та дисперсійний аналіз у статистичному	

пакеті STATGRAPHICS	51
4 ПЛАНУВАННЯ ЕКСПЕРИМЕНТУ	69
4.1 Повний факторний експеримент (ПФЕ).....	70
4.2 Побудова експерименту в пакеті STATGRAPHICS	75
4.3 Аналіз математичної моделі та результатів експерименту	77
4.4 Алгоритм реалізації ПФЕ.....	79
5 ДОСЛІДЖЕННЯ ПРОЦЕСУ ЗВАРЮВАННЯ ТА РЕАЛІЗАЦІЯ РЕГРЕСІЙНОЇ МОДЕЛІ.....	81
5.1 Аналіз існуючих регресійних моделей параметрів зварного шва	81
5.2 Проведення досліджень та розробка математичної моделі.....	82
5.3 Аналіз результатів дослідження.....	87
СПИСОК РЕКОМЕНДОВАНОЇ ЛІТЕРАТУРИ.....	95
Основна рекомендована література.....	95
Додаткова рекомендована література	96
ДОДАТОК А Приклад оформлення титульного аркуша.....	97
ДОДАТОК Б Приклад проведення регресійного аналізу процесу зварювання.....	99

ВСТУП

На сучасному етапі розвитку сучасної технології науково технічний прогрес знаходить своє відображення в якісних змінах знарядь праці, продуктах праці, у впровадженні нових автоматизованих технологічних процесів як у системах матеріального виробництва так і у системах управління різних рівнів. В результаті з'являються нові вироби й технології, змінюється номенклатура послуг, що надаються підприємствам, установам та фізичним особам, виникають принципово нові економічні об'єкти. Відбувається подальше зростання складності технічних об'єктів і систем, розширюється коло їх інформаційної взаємодії, збільшується невизначеність поведінки технічних систем в умовах швидкозмінного зовнішнього середовища. Застосування нейронних мереж, елементів нечіткої логіки в алгоритмах і засобах автоматизації технічних об'єктів потребує зміни інструментальних засобів, використовуваних для дослідження таких систем, створення адекватного опису їх функціонування, визначення параметрів надійності тощо.

Як показує інженерна практика, спеціалістам, задіяним у створенні та експлуатації складних технічних систем, конче необхідні знання у галузі методів і прийомів проведення наукових досліджень складних технічних систем, внаслідок чого можуть бути виявлені порушення сталості в роботі таких систем, оцінена надійність функціонування в тих чи інших умовах, визначені напрямки удосконалення елементів, вузлів, компонентів та підсистем складних технічних систем.

Тож вивчення основ проведення наукових досліджень складних технічних систем, методів прямого та непрямого експериментів, направлених на встановлення закономірностей функціонування та забезпечення надійності складних технічних систем поліпшить конкурентні переваги фахівця спеціальності 131- Прикладна механіка, зробить його більш затребуваним на сучасному ринку.

1 ЗАГАЛЬНІ ПОЛОЖЕННЯ

Метою вивчення навчальної дисципліни є засвоєння основ сучасних методів, методик і підходів щодо дослідження складних технічних систем, прогнозування їх поведінки у різних несталих умовах зовнішнього середовища, оцінки надійності функціонування складних технічних систем та використання набутих знань у професійній діяльності.

Предметна область дисципліни - теоретичне, комп'ютерне та експериментальне дослідження сучасних науково-технічних проблем прикладної механіки і вирішення задач динаміки, міцності, оптимізації, ресурсу, надійності та безпеки конструкцій і машин, композитних структур, споруд, пристроїв і агрегатів.

Вивчення даного курсу формує у студентів, відповідно до стандарту вищої освіти магістрів спеціальності 131 – Прикладна механіка, такі компетенції:

1. Експериментальні навички. Вміти проводити експерименти під наглядом наставника, а також описувати, аналізувати і критично оцінювати експериментальні дані і бути знайомим з найбільш важливими експериментальними методами.
2. Здатність планувати і проводити експериментальні дослідження та аналізувати дані, отримані від них. Здатність скоротити число дослідів, знайти оптимум, отримати кількісні оцінки впливу чинників і визначити помилки.
3. Дослідницькі навички. Здатність формулювати (роблячи презентації, або представляючи звіти) нові гіпотези та наукові задачі в області прикладної механіки, вибирати належні напрями і відповідні методи для їх розв'язку, беручи до уваги наявні ресурси.
4. Здатність формулювати складні проблеми оптимізації і прийняття рішень, а також інтерпретувати рішення у вихідних контекстах проблем. Здатність досліджувати моделі технічних систем для отримання оптимальних параметрів конструкції, технологічних режимів.
5. Здатність розв'язувати проблеми. Здатність формулювати, аналізувати та синтезувати рішення наукових проблем і задач прикладної механіки на абстрактному рівні шляхом декомпозиції їх на складові, які можна дослідити окремо в їх більш та менш важливих аспектах.

2. ВИПАДКОВІ ВЕЛИЧИНИ ТА ЗАКОНИ ЇХ РОЗПОДІЛУ

2.1. Закони розподілу дискретних випадкових величин

Випадковою величиною X називається числова функція $X=X(\omega)$ від елементарної події, яка визначена на множині елементарних результатів Ω , та така, що при довільному x множині тих ω , для яких $X(\omega) < x$, належить алгебрі подій.

Дискретною випадковою величиною називається випадкова величина з кінцевою або кінцевою множиною можливих значень.

Законом розподілу випадкової величини називається будь-яке правило, що дозволяє знаходити ймовірності всіляких подій, пов'язаних із цією випадковою величиною. Для дискретних випадкових величин найпростішою формою закону розподілу є ряд розподілу у вигляді таблиці, в одному рядку якої перераховані всі значення випадкової величини, а в другому рядку - ймовірності, що їм відповідають.

X	x_1	x_2	x_3	...	x_n
P	p_1	p_2	p_3	...	p_n

Отже, дискретна випадкова величина X в результаті досвіду прийме одне зі своїх можливих значень, тобто відбудеться одне з повної групи подій $\omega_1 = (X = x_1), \omega_2 = (X = x_2), \dots, \omega_n = (X = x_n), \quad \Omega = \{\omega_1, \omega_2, \dots, \omega_n\}$. Ймовірності, що відповідають цим подіям, такі $p_1 = P(X = x_1), p_2 = P(X = x_2), \dots, p_n = P(X = x_n)$. Очевидно, $\sum_{i=1}^n p_i = 1$, тому що $x_i, i = \overline{1, n}$ утворюють повну групу подій.

Графічне зображення ряду розподілу називається багатокутником розподілу дискретної випадкової величини.

Найбільш загальною формою закону розподілу є функція розподілу. Функцією розподілу випадкової величини X називається ймовірність нерівності $X < x$, що розглядається як функція параметра x ,

$$F(x) = P(X < x) \quad (2.1.1)$$

Найчастіше визначену таким чином функцію розподілу називають інтегральною функцією розподілу або інтегральним законом розподілу.

Функція розподілу - найуніверсальніша характеристика, вона повністю визначає випадкову величину. Функція розподілу будь-якої випадкової величини має наступні властивості:

1. $0 \leq F(x) \leq 1$ для всіх x ;
2. $F(x_1) \leq F(x_2)$, якщо $x_1 < x_2$;
3. $F(-\infty) = 0, F(\infty) = 1$;

у всіх точках області визначення функція безперервна зліва, а саме $F(x-0) = F(x)$ або $\lim_{x \rightarrow x_0-0} F(x) = F(x_0)$.

Можна представити, що будь-яка функція $F(x)$, яка має ці властивості, може бути функцією розподілу деякої випадкової величини. Графік $F(x)$ у загальному випадку є графіком не спадної функції, значення якої починаються від нуля і досягають одиниці, причому в окремих точках функція може мати розриви першого роду. Якщо відомий ряд розподілу дискретної випадкової величини, можна легко побудувати функцію розподілу

$$F(x) = P(X < x) = \sum_{x_i < x} P(X = x_i), \quad (2.1.2)$$

де підсумовування поширюється на всі значення x_i , які менше x .

Приклад. На шляху руху автомобіля шість світлофорів, кожен з них дозволяє або забороняє подальший рух автомобіля з ймовірністю 0.5. Скласти ряд розподілу та побудувати функцію розподілу числа світлофорів, пройдених автомобілем до першої зупинки. Рух автомобіля або закінчується на k -му світлофорі, якщо до цього він проходить $k-1$ світлофор без затримки, а на k -му буде зупинений, або автомобіль пройде всі світлофори і зупиненим не буде. Нехай випадкова величина X – число світлофорів, пройдених автомобілем. Очевидно, що X може набувати наступні значення 0,1,2,3,4,5,6. Величина X підпорядкована геометричному закону розподілу з додатковою умовою, що випробування буде закінчено при умові, якщо X набуде значення шість. Отже,

$$\begin{aligned} P(X = k) &= qp^k, k = 0,1,2,3,4 \\ P(X = 6) &= 1 - \sum_{i=0}^5 P(X = i) \end{aligned} \quad (2.1.3)$$

причому, очевидно, що $p=1/2, q=1-p=1/2$. Тоді ряд розподілу випадкової величини X має наступний вигляд.

X	0	1	2	3	4	5	6
P	$\frac{1}{2}$	$\frac{1}{4}$	$\frac{1}{8}$	$\frac{1}{16}$	$\frac{1}{32}$	$\frac{1}{64}$	$\frac{1}{64}$

Дійсно, $p_1 = P(X = 0) = qp^0 = q = 1/2$, $p_2 = P(X = 1) = qp = 1/4$. Знаючи ряд розподілу, легко поставити багатокутник розподілу та функцію розподілу, користуючись формулою $F(x) = P(X < x) = \sum_{x_i < x} P(X = x_i)$ (рис. 2.1).

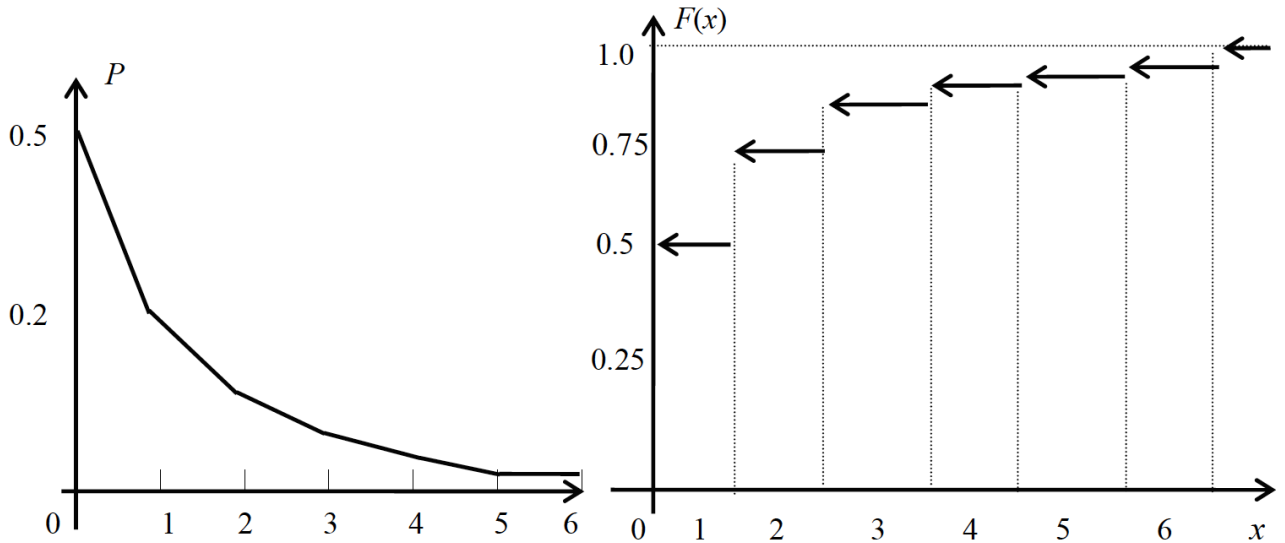


Рис. 2.1. Багатокутник розподілу на функцію розподілу дискретної випадкової величини

Дійсно, $F(0) = P(X < 0) = 0$, $F(1) = P(X < 1) = \sum_{x_i < 1} P(X = x_i) = p_1 = \frac{1}{2}$. Тоді функція розподілу може бути виражена у такому вигляді:

$$F(x) = \begin{cases} 0, & x \leq 0 \\ \frac{1}{2} = 0.5, & 0 < x \leq 1 \\ \frac{3}{4} = 0.75, & 1 < x \leq 2 \\ \frac{7}{8} = 0.875, & 2 < x \leq 3 \\ \frac{15}{16} = 0.9375, & 3 < x \leq 4 \\ \frac{31}{32} = 0.96875, & 4 < x \leq 5 \\ \frac{63}{64} = 0.984375, & 5 < x \leq 6 \\ 1, & x > 6. \end{cases} \quad (2.1.4)$$

2.2. Числові характеристики та властивості дискретних випадкових величин

Ряд розподілу або функція розподілу дискретної випадкової величини є її вичерпними характеристиками, проте вони є досить громіздкими, через що виникає потреба в менш «об'ємних» характеристиках. Такими є показники положення і розсіювання. Характеристики положення дають деяке середнє положення випадкової величини, навколо якого вона групується, а характеристики розсіювання вказують міру розсіювання випадкової величини навколо її середнього положення.

Найбільш уживана характеристика середнього положення це математичне очікування - середнє зважене серед значень x_i , причому кожне x_i за умови опосередкування має враховуватися з вагою p_i . Таким чином, математичне очікування дискретної випадкової величини дорівнює

$$m_X = M(X) = \frac{x_1 p_1 + x_2 p_2 + \dots + x_n p_n}{p_1 + p_2 + \dots + p_n} = \frac{\sum_{i=1}^n x_i p_i}{\sum_{i=1}^n p_i} = \sum_{i=1}^n x_i p_i \quad (2.2.1)$$

Якщо правої частини формули (2.2.1) стоїть ряд, то $M(X) = \sum_{i=1}^{\infty} x_i p_i = \lim_{n \rightarrow \infty} \sum_{i=1}^n x_i p_i$ причому ряд повинен збігатися абсолютно. Математичне очікування цього конкретного розподілу може й не існувати.

Математичне очікування випадкової величини X пов'язане своєрідною залежністю із середнім арифметичним значенням спостереженої випадкової величини X при великій кількості дослідів. Ця залежність того ж типу, що між частотою і ймовірністю, а саме, при великій кількості дослідів середнє арифметичне значень X сходиться ймовірно до свого математичного очікування.

Властивості математичного очікування:

1. $M(C) = C$, $C = \text{const}$. Постійну величину можна розглядати як випадкову, яка приймає лише одне значення з ймовірністю рівної одиниці, тобто. $M(C) = \sum_{i=1}^1 C \cdot 1 = C$.

2. Константу можна виносити за знак математичного очікування, тобто. $M(C \cdot X) = C \cdot M(X)$. Дійсно, $M(CX) = \sum_{i=1}^n C x_i p_i = C \sum_{i=1}^n x_i p_i = CM(X)$

3. Властивість адитивності: $M(X + Y) = M(X) + M(Y)$, так як $M(X + Y) = \sum_{i=1}^n (x_i + y_i) p_i = \sum_{i=1}^n x_i p_i + \sum_{i=1}^n y_i p_i = M(X) + M(Y)$

Сукупність другої та третьої властивостей називається властивістю

"лінійності" та виражається наступною рівністю $M(C_1X_1 + C_2X_2 + \dots + C_nX_n) = C_1M(X_1) + C_2M(X_2) + \dots + C_nM(X_n)$. Зокрема, якщо $Y = kX + b$, тоді $M(Y) = M(kX + b) = kM(X) + b$.

4. Властивість монотонності: якщо $X \geq Y$, тоді $M(X) \geq M(Y)$.

5. Мультиплікативна властивість: для незалежних випадкових величин X та Y справедливо $M(X \cdot Y) = M(X) \cdot M(Y)$.

Крім математичного очікування як характеристики положення випадкової величини часто використовуються мода і медіана.

Моду дискретної випадкової величини X називається таке значення x_k , $k = \overline{1, n}$, для якої

$$P(X = d_x) = \max_k P(X = x_k), \quad (2.2.2)$$

тобто мода є найбільш ймовірним значенням дискретної випадкової величини, якщо це значення єдине. Мода може бути і не єдиною, тобто, розподіл може мати декілька мод (мультимодальний розподіл). Медіаною дискретної випадкової величини X називається число h_x , що задовольняє умову

$$P(X < h_x) = P(X \geq h_x) = 1/2. \quad (2.2.3)$$

Оскільки це рівняння в загальному випадку може мати кілька коренів, то значення медіани може бути єдиним.

Перейдемо тепер до визначення характеристики розсіювання випадкової величини біля свого математичного очікування. Початковим моментом k -го порядку дискретної випадкової величини X називається математичне очікування k -го ступеня випадкової величини

$$\alpha_k = M(X^k) = \sum_{i=1}^n x_i^k p_i \quad (2.2.4).$$

Це визначення збігається з визначенням початкового моменту в механіці, якщо вірогідність p_i інтерпретувати як масу точок x_i . Зокрема, з формули (2.2.4) випливає, що перший початковий момент є математичним очікуванням, тобто $\alpha_1 = m_x$.

Центральним моментом k -го порядку дискретної випадкової величини X називається математичне очікування k -го ступеня відповідної центрованої випадкової величини

$$\mu_k = M[(X - m_x)^k] = \sum_{i=1}^n (x_i - m_x)^k p_i \quad (2.2.5).$$

Дисперсією випадкової величини X називається математичне очікування квадрата відповідної центрованої величини, а саме її другий центральний момент,

$$D(X) = D_X = \mu_2 = M[(X - m_X)^2] = \sum_{i=1}^n (x_i - m_X)^2 p_i. \quad (2.2.6).$$

Середнім квадратичним відхиленням або стандартним відхиленням (стандартом) випадкової величини X називається величина

$$\sigma_X = \sqrt{D_X} \quad (2.2.7)$$

Для дисперсії з формули (2.2.6) легко виводиться наступна формула, найбільш часто вживана:

$$D_X = \sum_{i=1}^n x_i^2 p_i - m_X^2 \quad (2.2.8)$$

Властивості дисперсії:

1. Дисперсія будь-якої випадкової величини X невід'ємна, причому $D_X = 0$ і тоді, коли X - постійна, тобто $D(X) \geq 0, D(C) = 0$
2. Якщо $Y = X + C$, де $C = const$, то $D(Y) = D(X + C) = D(X)$.
3. Якщо $C = const$, тоді $D(C \cdot X) = C^2 D(X)$.
4. Якщо випадкові величини X та Y незалежні, то $D(X+Y) = D(X) + D(Y)$.

Коефіцієнтом асиметрії називається число A , що визначається формулою

$$A = \frac{\mu_3}{\sigma_X^3} = \frac{\sum_{i=1}^n (x_i - m_X)^3 p_i}{\sigma_X^3} \quad (2.2.9)$$

Коефіцієнт асиметрії служить для характеристики асиметрії багатокутника розподілу. У разі негативного коефіцієнта асиметрії присутній більш пологий нахил багатокутника розподілу спостерігається зліва, інакше - праворуч. У першому випадку асиметрію називають лівосторонньою, у другому – правосторонньою.

Ексцесом або коефіцієнтом крутості називається число

$$E = \frac{\mu_4}{\sigma_x^4} - 3. \quad (2.2.10)$$

Ця характеристика служить для порівняння на «крутість» заданого та нормального розподілу. Ексцес для випадкової величини, розподіленої нормально, дорівнює нулю. Якщо розподілу відповідає негативний ексцес, відповідний багатокутник розподілу має більш пологі вершину в порівнянні з нормальною кривою. У разі позитивного ексцесу багатокутник більш крутий, порівняно з нормальною кривою.

2.3. Закони розподілу безперервних випадкових величин

Безперервною випадковою величиною називається така випадкова величина, ймовірність попадання якої в будь-яку нескінченно малу область є нескінченно малою і для якої при кожному x існує скінченна або нескінченна межа

$$f(x) = \lim_{\Delta x \rightarrow 0} \frac{P(x \leq X < x + \Delta x)}{\Delta x} \quad (2.3.1)$$

Усі основні визначення щодо закону розподілу тут залишаються чинними. Для безперервної випадкової величини неможливо встановити ряд розподілу. Функція ж розподілу для неї існує і є безперервною кривою.

Функцією розподілу безперервної випадкової величини X називається ймовірність наступної нерівності:

$$F(x) = P(X < x) = \int_{-\infty}^x f(t) dt \quad (2.3.2)$$

за умови, що існує така невід'ємна функція $f(x)$ (рис. 2.2), що інтегрується у нескінченних межах. Ця функція називається щільністю розподілу ймовірностей. Справедливі такі співвідношення:

$$\begin{cases} F(x) = \int_{-\infty}^x f(t) dt \\ f(x) = \frac{dF(x)}{dx} \end{cases} \quad (2.3.3)$$

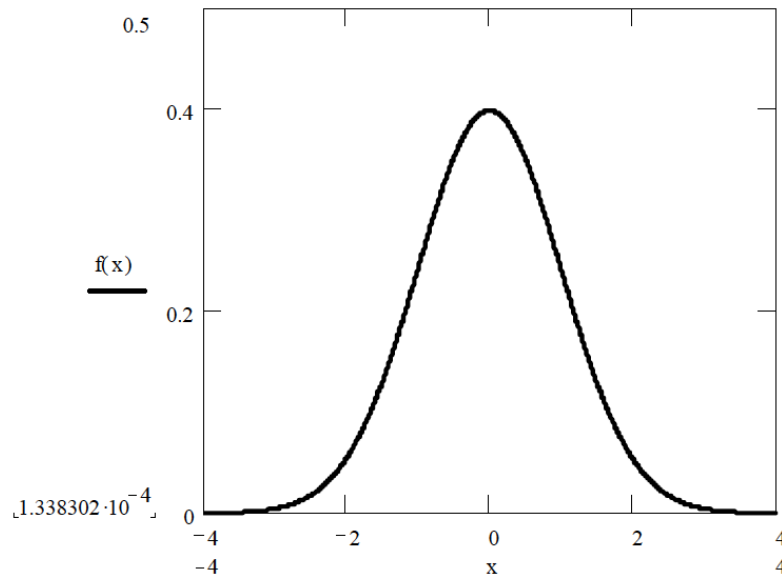


Рис. 2.2. Функція щільності ймовірності безперервної випадкової величини

Функції щільності ймовірностей відповідає крива щільності розподілу, або крива щільності ймовірності. Вона є однією з форм закону розподілу, але не універсальною, бо існує тільки для безперервної випадкової величності. Її деякою аналогією для дискретних випадкових величин є багатокутник розподілу.

Властивості функції щільності розподілу:

1. $f(x) \geq 0$,
2. $\int_{-\infty}^{\infty} f(x) dx = 1$

2.4 Числові характеристики безперервних випадкових величин

Математичне очікування безперервної випадкової величини X із щільністю ймовірності $f(x)$ вважається вираз:

$$M(X) = m_X = \int_{-\infty}^{\infty} xf(x)dx. \quad (2.4.1)$$

Характеристики становища, саме, мода і медіана залишаються у силі для безперервної випадкової величини і навіть визначаються в цьому випадку найбільш природним чином, якщо користуватися поняттям функції щільності розподілу.

Модю безперервної випадкової величини X називається число d_X , що

визначається як точка максимуму функції щільності ймовірності $f(x)$. Отже,

$$f'(d_x) = 0, f'(x < d_x) > 0 \text{ та } f'(x > d_x) < 0. \quad (2.4.2)$$

Медіаною безперервної випадкової величини X називається число d_x , що задовольняє умову

$$\int_{-\infty}^{h_x} f(x)dx = \int_{h_x}^{\infty} f(x)dx = \frac{1}{2}. \quad (2.4.3)$$

Всі визначення для початкових і центральних моментів залишаються чинними, тільки суми замінюються інтегралами.

Дисперсією безперервної випадкової величини називається її другий центральний момент, тобто.

$$D_x = D(X) = \int_{-\infty}^{\infty} (x - m_x)^2 f(x)dx \quad (2.4.4)$$

Квантилем, порядку p розподілу безперервної випадкової величини X називається число t_p , що задовольняє умові

$$P(X < t_p) = p, \text{ або } \int_{-\infty}^{t_p} f(x)dx = p. \quad (2.4.5)$$

Очевидно, що, наприклад, $h_x = t_{0.5}$.

Критичною точкою порядку p розподілу безперервної випадкової величини X називається число k_p , що задовольняє рівнянню:

$$P(X \geq k_p) = p, \text{ або } \int_{k_p}^{\infty} f(x)dx = p. \quad (2.4.6)$$

Квантилі та критичні точки одного і того ж розподілу пов'язані між собою простим співвідношенням $k_p = t_{1-p}$.

Асиметрія та ексцес для безперервних випадкових величин визначаються аналогічно формулам (2.2.9) та (2.2.10).

Приклад. Випадкова величина X підпорядкована закону арксинусу (рис. 2.3) із щільністю розподілу ймовірностей $f(x) = \begin{cases} 0, & |x| \geq a, \\ \frac{1}{\pi\sqrt{a^2-x^2}}, & |x| < a. \end{cases}$ Знайти функцію розподілу $F(x)$ та обчисліть $m_x, D_x, d_x, h_x, k_{0.75}$.

Знайдемо спочатку $F(x)$. За визначенням $F(x) = \int_{-\infty}^x f(t)dt =$

$$= \int_{-a}^x \frac{dt}{\sqrt{a^2-t^2}} = \frac{1}{\pi} \left(\arcsin \frac{t}{a} \Big|_{-a}^x \right) = \frac{1}{\pi} \left(\arcsin \frac{x}{a} + \arcsin 1 \right) = \frac{1}{2} + \frac{1}{\pi} \arcsin \frac{x}{a}.$$

Графіки функції щільності ймовірності та функції розподілу наведені нижче.

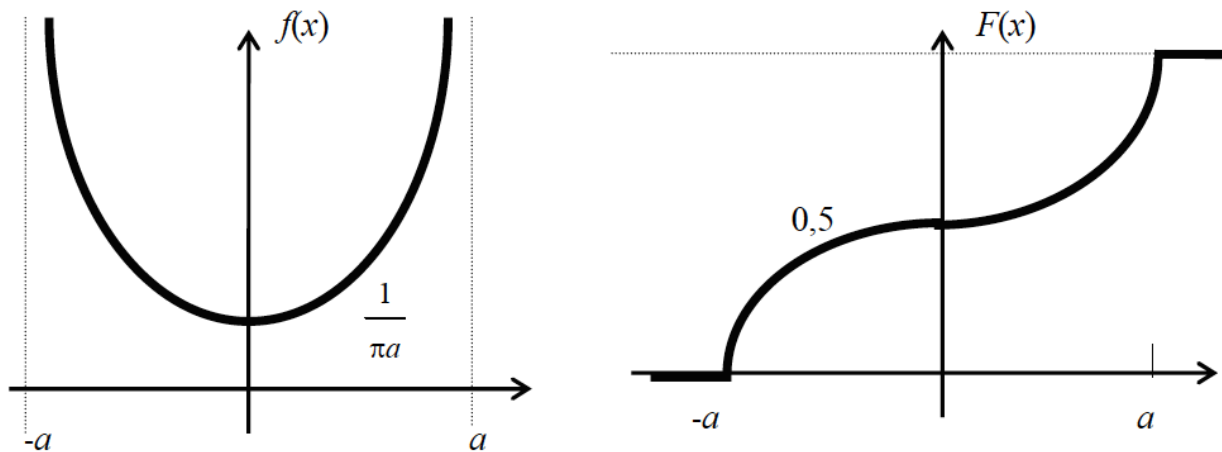


Рис. 2.3. Графіки функцій щільності ймовірності н розподілу закону арксинусу

Визначимо тепер усі числові характеристики, необхідні за умовою завдання.

$$\begin{aligned}
 m_X = \alpha_1 &= \int_{-a}^a x f(x) dx = \int_{-a}^a x \frac{dx}{\pi \sqrt{a^2 - x^2}} = -\frac{1}{2\pi} \int_{-a}^a \frac{d(a^2 - x^2)}{\sqrt{a^2 - x^2}} = \\
 &= -\frac{1}{2\pi} 2\sqrt{a^2 - x^2} \Big|_{-a}^a = 0.
 \end{aligned}$$

Цей результат очевидний і з малюнка функції ймовірності. Знайдемо моду. $f'(x) = \frac{x}{\pi(a^2 - x^2)^{3/2}} = 0$, $x=0$, но $x=0$ - це точка мінімуму, а не максимуму. Отже, моди цей розподіл не має. Медіану також знайдемо за визначенням $\int_{-a}^{h_X} f(x) dx = \int_{-a}^{h_X} \frac{dx}{\pi \sqrt{a^2 - x^2}} = \frac{1}{\pi} \arcsin \frac{x}{a} \Big|_{-a}^{h_X} = \frac{1}{2}$. Звідки $\arcsin \frac{h_X}{a} = 0$, $h_X = 0$. З огляду на симетричності кривої функції щільності ймовірності цей результат теж очевидний з малюнка $f(x)$.

$$\begin{aligned}
 D_X = \mu_2 &= \int_{-a}^a (x - 0)^2 f(x) dx = \int_{-a}^a \frac{x^2 dx}{\pi \sqrt{a^2 - x^2}} = \\
 &= \left(\begin{array}{l} x = a \sin t, dx = a \cos t dt, \\ x^2 = a^2 \sin^2 t, \sqrt{a^2 - x^2} = a \cos t, \\ x = \pm a, t = \pm \pi/2 \end{array} \right) = \frac{1}{\pi} \int_{-\pi/2}^{\pi/2} a^2 \sin^2 t dt = \\
 &= \frac{a^2}{\pi} \int_{-\pi/2}^{\pi/2} \frac{1 - \cos 2t}{2} dt = \frac{a^2}{\pi} \left[\frac{t}{2} \Big|_{-\pi/2}^{\pi/2} - \frac{1}{4} \sin 2t \Big|_{-\pi/2}^{\pi/2} \right] = \frac{a^2}{2}.
 \end{aligned}$$

Знайдемо необхідну критичну точку.

$$\int_{k_{0.75}}^a \frac{dx}{\pi\sqrt{a^2-x^2}} = 0.75 = \frac{1}{\pi} \int_{\arcsin(k/a)}^{\pi/a} \frac{a\cos t}{a\cos t} dt = \frac{1}{\pi} t \Big|_{\arcsin(k/a)}^{\pi/a}$$

$$= \frac{1}{\pi} \left(\frac{\pi}{2} - \arcsin \frac{k}{a} \right) = 0.75.$$

Звідки $\frac{1}{\pi} \arcsin \frac{k}{a} = -0.25, \frac{k}{a} = \sin \left(-\frac{\pi}{4} \right) = -\frac{\sqrt{2}}{2}, k_{0.75} = -\frac{\sqrt{2}a}{2}.$

2.5. Вибіркові аналоги інтегральної та диференціальної функцій розподілу

Предметом математичної статистики вивчення випадкових величин за результатами спостережень. У ній розвиваються методи обробки результатів дослідів, аналізу отриманої із дослідів статистичної інформації, отримання числових оцінок параметрів розподілів.

Центральне поняття математичної статистики – поняття вибірки. Вибірka розуміється так. Нехай проводиться деякий експеримент, пов'язаний із випадковою величиною X функцією розподілу $F(x)$.

Вибіркою обсягу n з генеральної сукупності з функцією розподілу $F(x)$ називається послідовність x_1, x_2, \dots, x_n спостережених значень випадкової величини X , що відповідає n незалежним повторенням даного експерименту. Таким чином, вибірка або вибіркова сукупність - це множина випадково відібраних об'єктів або спостережень над деякою випадковою величиною, а генеральна сукупність - це сукупність всіх об'єктів або всіх можливих значень випадкової величини, з яких проводиться вибірка. Кожен елемент вибірки є конкретна реалізація однієї й тієї ж випадкової величини з функцією розподілу $F(x)$.

Вибіркою об'єму n називається n незалежних випадкових величин X_1, X_2, \dots, X_n , кожна з яких розподілена так само, як деяка випадкова величина X з функцією розподілу $P(X \leq x) = F(x)$. Вибірka називається репрезентативною чи представницькою, якщо вона досить добре представляє кількісні співвідношення генеральної сукупності. Репрезентативність вибірки забезпечується випадковість відбору. Це означає, що будь-який об'єкт вибірки відібраний випадково, при цьому всі об'єкти мають однакову можливість потрапити у вибірку.

Як відомо, існують чотири схеми вибору елементів множин (вибірки): схеми з поверненням елемента, або без повернення та з наступним упорядкуванням, або без упорядкування обраних елементів. Усі ці схеми реалізуються у конкретних інженерних завданнях. Вибірka, впорядкована за зростанням спостережених значень випадкової величини, називається

варіаційним рядом.

Нехай тепер є вибірка x_1, x_2, \dots, x_k обсягу n . Серед елементів x_i можуть бути однаковими. Нехай у вибірці елемент x_i зустрічається n_i разів. Число n_i називається частотою. Очевидно, що $\sum_{i=1}^k n_i = n$. Відношення частоти n_i до обсягу вибірки n називається відносною частотою значення x_i і позначається $w_i = n_i/n$. $\sum_{i=1}^k w_i = \sum_{i=1}^k \frac{n_i}{n} = \frac{1}{n} \cdot n = 1$.

Сукупність пар (x_i, n_i) називається статистичним рядом або статистичним розподілом і зазвичай записується у вигляді таблиці:

X	x_1	x_2	\dots	x_k
n_i	n_1	n_2	\dots	n_k

Якщо X - дискретна випадкова величина, то статистичний ряд записаний у вигляді

X	x_1	x_2	\dots	x_k
w_i	w_1	w_2	\dots	w_k

є аналогом ряду розподілу. Якщо ж X - неперервна величина, то статистичний ряд записується у вигляді

X	$[x_0, x_1]$	$[x_1, x_2]$	\dots	$[x_{k-1}, x_k]$
w_i	w_1	w_2	\dots	w_k

де w_i - відносні частоти потрапляння випадкової величини X в інтервал $[x_{(i-1)}, x_i]$, $i = 1, 2, \dots, k$.

При великому обсязі вибірки n її елементи об'єднуються в групи і отримуємо групований статистичний ряд. Для цього всі інтервали вибірки поділяються на l розрядів (від 6 до 20). Слід пам'ятати, що угруповання завжди вносить певну похибку до обчислень. Ця похибка зростає із зменшенням числа розрядів. Графічним поданням вибірки є полігон частот та гістограма. Полігон частот будується для дискретної випадкової величини. Це графік, точки якого мають координати (x_i, n_i) або (x_i, w_i) . Таким чином, полігон частот для вибірки є податком багатокутника розподілу випадкової

дискретної величини. Для ілюстрації розподілу безперервної випадкової величини будують гістограму (рис. 2.4).

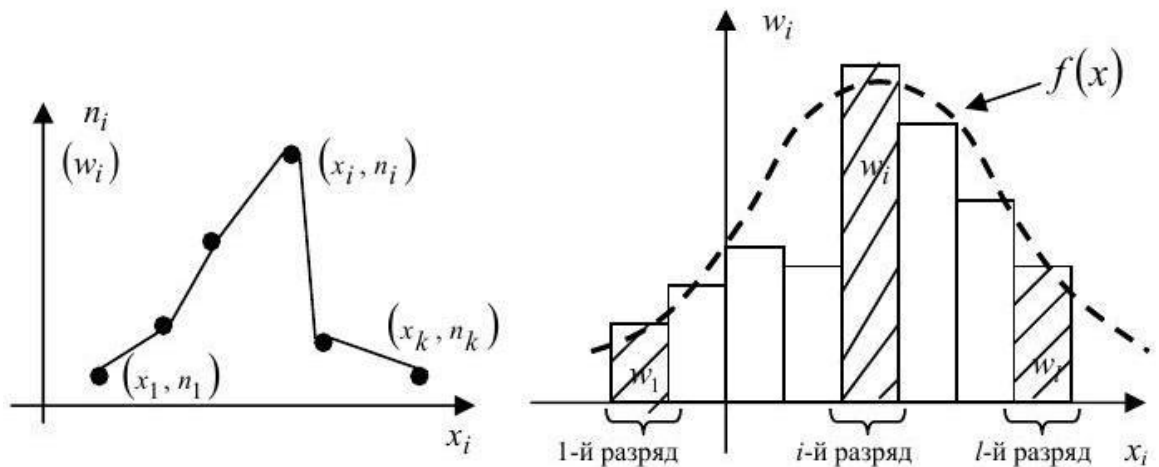


Рис. 2.4. Полігон частот гістограма вибірки

Гістограмою частот групованої вибірки називається ступінчаста фігура, складена з прямокутників, побудованих на інтервалах угруповання (розрядах) так, що площа кожного прямокутника дорівнює чи пропорційна частоті n_i або відносній частоті w_i . Очевидно, що при збільшенні числа дослідів довжину розряду можна необмежено зменшувати, і тоді гістограма все більше наближається деякою кривою, що обмежує одиничну площу. Зрозуміло, що ця крива - графік функції щільності ймовірності безперервної випадкової величини X . Таким чином, гістограма - аналог кривої щільності ймовірності.

Введемо, нарешті, поняття вибіркової функції розподілу. Нехай ϵ вибірка обсягу n , x - деяке дійсне число, n_x - число вибірових значень випадкової величини X , менших за x . Тоді число n_x/n є відносною частотою спостережених значень у вибірці значень X , менших за x , тобто відносною частотою появи події $X < x$. Зрозуміло, що із зміною x змінюватиметься і величина n_x/n . Це означає, що відносна частота n_x/n - функція аргументу x . А оскільки ця функція знаходиться за вибіровими дослідними даними, то її називають вибірковою, статистичною або емпіричною.

Статистичною або емпіричною функцією розподілу називають функцію $F^*(x)$, визначальну для кожного значення x відносну частоту події $X < x$, тобто:

$$F^*(x) = \begin{cases} 0, & x \leq x_1 \\ \sum_{i=1}^k w_i, & x_k < x \leq x_{k+1}, k = 1, 2, \dots, n-1, \\ 1, & x > x_n. \end{cases} \quad (2.5.1)$$

Формально емпірична функція розподілу має всі властивості інтегральної функції розподілу. Маючи статистичний ряд, легко отримати статистичну функцію розподілу.

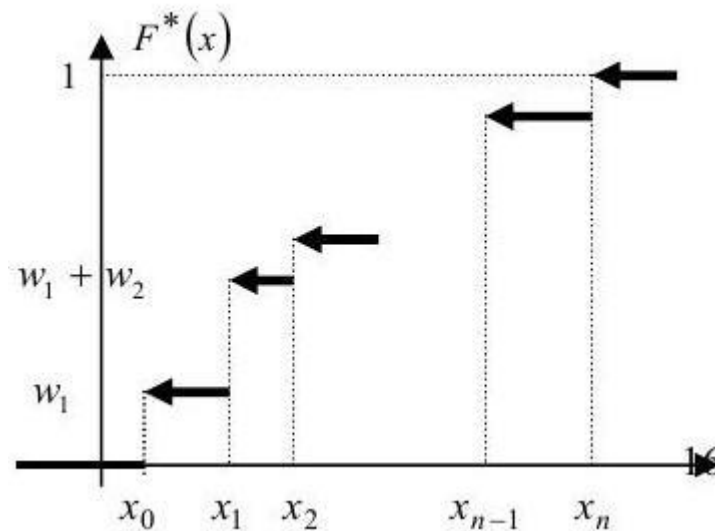


Рис. 2.5. Графік емпіричної функції розподілу

Дійсно:

$$F^*(x_1) = 0, F^*(x_2) = w_1,$$

$$F^*(x_3) = w_1 + w_2, \dots$$

$$F^*(x_k) = \sum_{i=1}^{k-1} w_i$$

$$F^*(x_{n+1}) = \sum_{i=1}^n w_i = 1$$

На графіку цієї функції (рис. 2.5) видно всі основні особливості емпіричної функції розподілу. Вона не зменшується, а її значення знаходяться в інтервалі $[0,1]$. Різкі стрибки графіка функції $F^*(x)$, що надають їй ступінчастого вигляду, мають місце в тих точках, яким відповідають кінці розрядів, а величина стрибка дорівнює відносній частоті розряду. Часто графік $F^*(x)$ будують у вигляді безперервної кривої, з'єднуючи точки графіка, що відповідають кінцям або серединам розрядів, відрізками прямої (рис. 2.6).

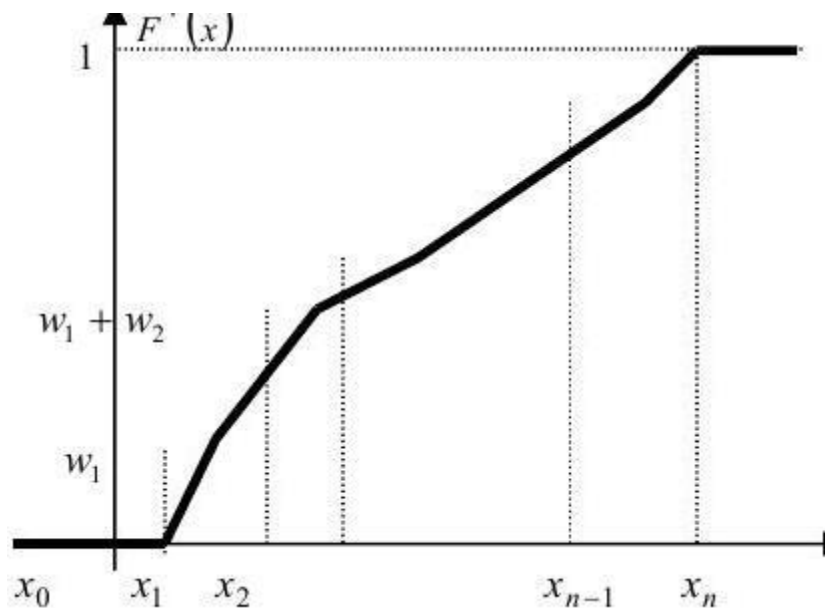


Рис. 2.6. Кумулятивна крива

Зазначимо, що подібний графік емпіричної функції розподілу, що дає наближене уявлення про графік теоретичної функції $F(x)$ часто називають кумулятивною кривою.

Так як за теоремою Бернуллі (Яків Бернуллі (1654-1705) - швейцарський математик) відносні частоти w_i при $n \rightarrow \infty$ сходяться за ймовірністю до відповідних вірогідностей подій, то при $n \rightarrow \infty$ $F^*(x)$ наближається до інтегральної функції розподілу. Про збіжність $F^*(x)$ до $F(x)$ доведено теорему, яка носить ім'я авторів - *теорема Глівенко-Кантеллі* (Валерій Іванович Глівенко (1896-1940) – радянський математик, Франческо Паоло Кантеллі (1875-1966) – італійський математик). За цією теоремою емпірична функція розподілу $F^*(x)$ рівномірно за x з ймовірністю $1 - \epsilon$ сходиться при $n \rightarrow \infty$ до теоретичного розподілу $F(x)$, тобто.

$$P \left\{ \lim_{n \rightarrow \infty} \sup_{-\infty < x < +\infty} |F^*(x) - F(x)| = 0 \right\} = 1 \quad (2.5.2)$$

Сенс цієї теореми у цьому, що зі збільшенням обсягу вибірки n у емпіричної функції розподілу зникають властивості випадковості і вона наближається до теоретичної функції розподілу. Аналогічно, якщо n велике, то значення гістограми $w_n(x)$ у точці x приблизно дорівнює $[F(x_{i+1}) - F(x_i)] / \Delta_i$, де $\Delta_i = x_{i+1} - x_i$, а x_i, x_{i+1} - кінці інтервалу, в якому знаходиться x . Якщо Δ_i - мало, то гістограма $w_n(x)$ досить добре відтворює функцію густини $f(x)$.

2.6 Методи описової статистики у пакеті STATGRAPHICS

Існують різноманітні методи обробки даних, що мають різну

складність і нерідко потребують великих обчислювальних ресурсів. Це величезний світ, створений багаторічними працями професійних математиків та прикладних наукових фахівців.

Разом з тим слід зазначити, що значна частина методів і статистичних пакетів швидко застаріває. Це пов'язано з стрімкими темпами розвитку галузі. У таких умовах виграє програмне забезпечення, що має дуже високі споживчі якості. Пакет STATGRAPHICS (STATistical GRAPHICs System) вигідно відрізняється від інших статистичних пакетів зручністю інтерфейсу користувача і обсягом використовуваних методів обробки, належить до класу універсальних пакетів. Цей пакет, створений американською корпорацією Manugistics, одна із найефективніших систем статистичного аналізу даних.

З моменту виходу першої версії STATGRAPHICS еволюцію зазнали майже всі основні складові пакету. Зараз STATGRAPHICS Plus for Windows включає понад 250 статистичних та системних процедур, що застосовуються у наукових задачах дослідження складних системи, бізнесі, економіці, маркетингу, біології, соціології, на виробництві та інших галузях. Весь пакет має модульну структуру. Кожна статистична процедура STATGRAPHICS супроводжується інтегрованою в систему відмінною графікою. Усі елементи графічних відображень можуть бути піддані корекції та перетворення. Для цього потрібно вибрати потрібний елемент, клацнувши на ньому лівою кнопкою миші, потім клацнути правою кнопкою. Тоді на екрані з'явиться діалогове вікно, в якому можна виконати необхідні зміни.

Власний варіант програми розрахунку можна зберегти як файл StatFolio. Якщо виникне потреба у обробці іншої множини даних за складеною схемою аналізу, необхідно в цей варіант просто завантажити новий файл даних. Результати розрахунків, таблиці та графіки будуть видані автоматично.

У пакеті реалізовано засіб допомоги користувачеві – статконсультант (StatAdvisor), який представляє інтерпретацію результатів. Щоб викликати статконсультанта, потрібно клацнути лівою кнопкою миші на графічному або табличному вікні пакета, а потім на піктограмі StatAdvisor. З'явиться консультаційне вікно, яке містить вичерпні поради, роз'яснення та рекомендації. Після запуску STATGRAPHICS на екрані монітора з'являється наступна заставка (рис. 2.7).

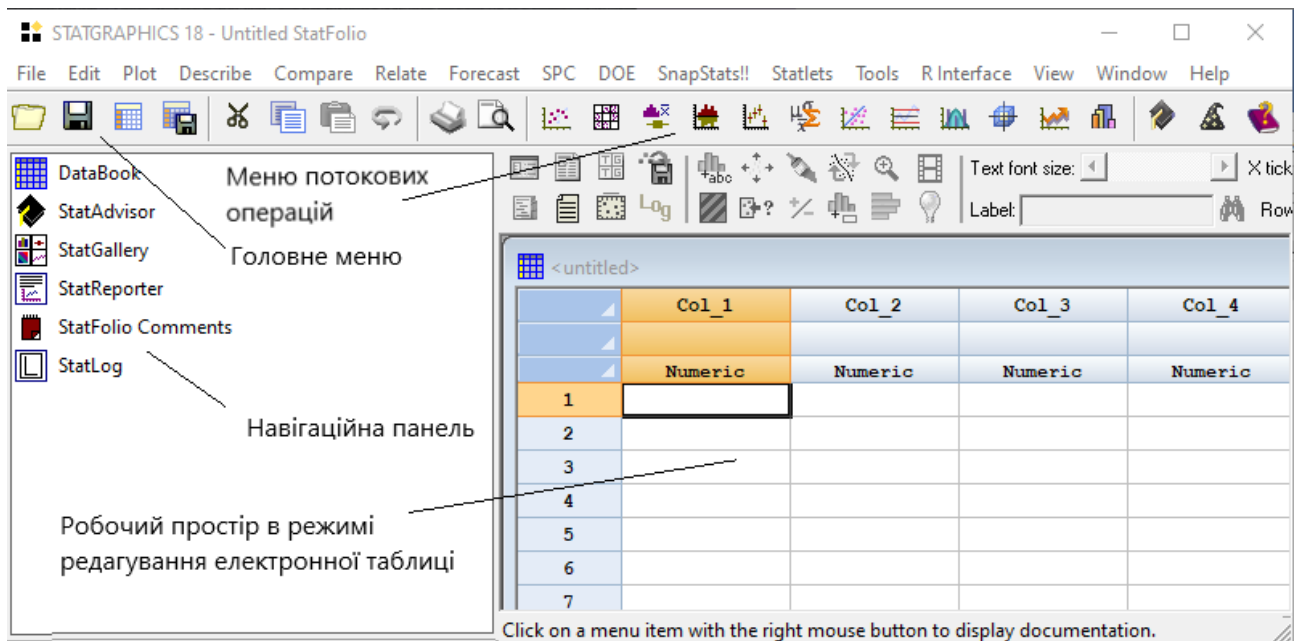


Рис. 2.7. Вікно STATGRAPHICS

Багато пунктів головного меню виконують ті ж дії, що й у більшості інших Windows – програм. Більшість статистичних процедур згруповані у пунктах Describe, Compare та Relate. Меню Describe містить статистичні методи аналізу за однією та багатьох змінних, процедури підбору розподілів, засоби табуляції даних. Меню Compare включає методи порівняння двох та більше вибірок даних, процедури одно- та багатофакторного дисперсійного аналізу.

Меню Relate містить процедури простого, поліноміального та множинного регресійного аналізу. У цій та наступних лабораторних роботах буде детально розібрано вміст цих пунктів меню. Вирішимо найпростіше завдання описової статистики. Знайдемо розмах вибірки, число і довжину інтервалів, складемо таблицю частот, побудуємо гістограму частот, і навіть обчислимо всі числові характеристики наступної вибірки.

Числа вибірки представлені як тривалість роботи електронних ламп одного типу в годинах:

13.4	14.7	15.2	15.1	13.0	8.8	14.0	17.9	15.1	16.5
16.6	14.2	16.3	14.6	11.7	16.4	15.1	17.6	14.1	18.8
11.6	13.9	18.0	12.4	17.2	14.5	16.3	13.7	15.5	16.2
8.4	14.7	15.4	11.3	10.7	16.9	15.8	16.1	12.3	14.0
17.7	14.7	16.2	17.1	10.1	15.8	18.3	17.5	12.7	20.7
13.5	14.0	15.7	21.9	14.3	17.7	15.4	10.9	18.2	17.3
15.2	16.7	17.3	12.1	19.2					

Скористаємося пунктом меню Describe (опис даних). Нам необхідно виконати аналіз однієї змінної. Цей аналіз у пакеті STATGRAPHICS містить процедури обчислення таких характеристик:

1. Сумарні статистики: середнє, медіана, мода, середнє геометричне, дисперсія, стандартне відхилення, мінімум, максимум, розмах, нижній та верхній кватиль, міжкватильний розмах, коефіцієнти асиметрії та ексцесу.
2. Відсотки та табуляції частот.
3. Гістограми та графік щільності.
4. Довірчі інтервали.
5. Перевірка гіпотез про середній та медіан, знаковий та ранговий тест.
6. Графіки "дерево з листям", "ящик з вусами", квантильний графік, графік нормального розподілу, симетричний графік та діаграма розсіювання.

Введемо вихідні дані в нову електронну таблицю, для чого натиснемо на піктограму Untitled в лівому нижньому куті робочого поля (рис. 2.8). Потрібно назвати змінну, яку ми будемо вводити в перший стовпець. Клацніть правою кнопкою миші по заголовку Col_1, з'явиться контекстне меню, в якому вибрати команду Modify Column (Визначити стовпець).

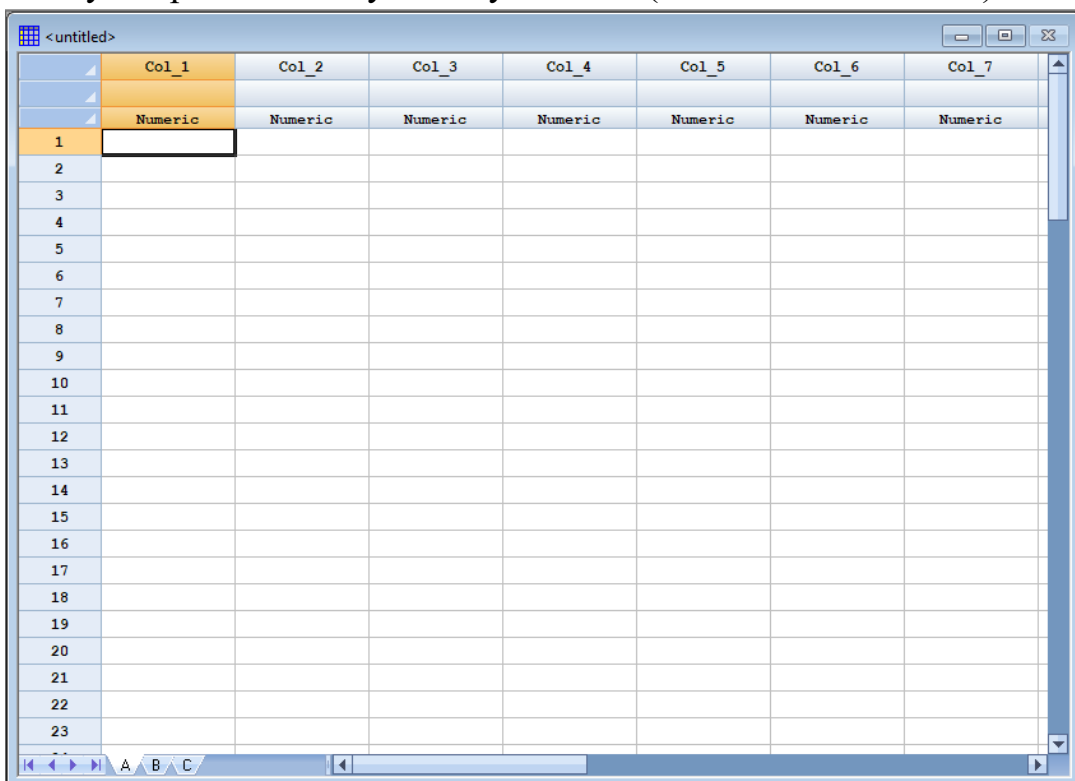


Рис. 2.8. Електронна таблиця STATGRAPHICS

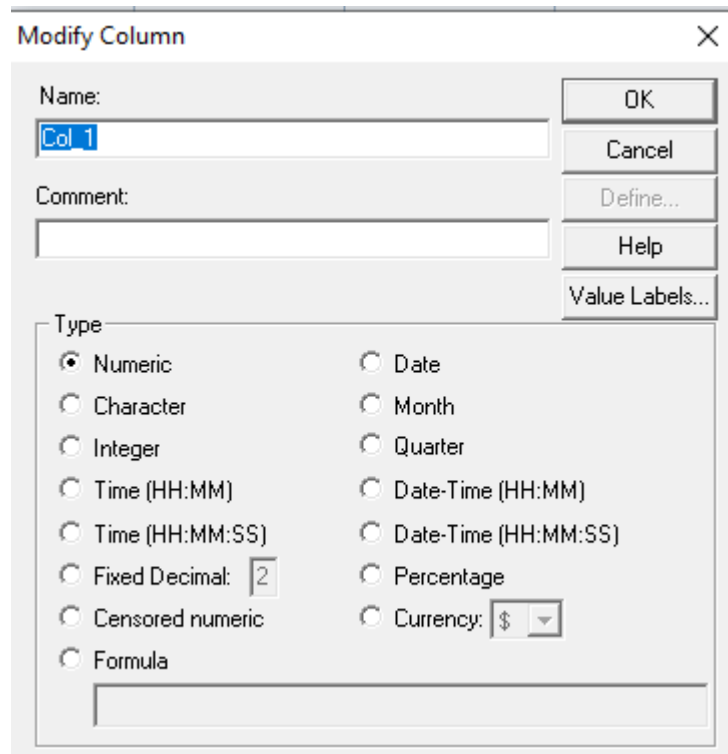


Рис. 2.9. Панель модифікації стовпця

З'явиться нове вікно діалогу (рис. 2.9), в якому потрібно вибрати відповідний тип даних, що вводяться (Numeric) і ввести ім'я змінної (Lamp).

Слід мати на увазі, що роздільником цілої та дробової частини чисел у пакеті STATGRAPHICS є кома, а всі імена вводяться латинськими літерами. Після введення даних у першу колонку їх можна перетворювати, викликавши з контекстного меню пункт Generate Date (Генерувати дані). Допускається більше сотні маніпуляцій зі змінними за допомогою операторів, що надаються. Зрештою, потрібно зберегти файл даних командою File→Save Data File As, ввести ім'я файлу та натиснути ОК. У заголовку таблиці замість <untitled> з'явиться вказане ім'я.

Проаналізуємо тепер статистичні дані. Виберемо Describe→Numeric Data→One Variable Analysis (Аналіз однієї змінної). З'явиться вікно для завдання змінної, що аналізується. У нашому випадку це Lamp. Після натискання на кнопку ОК виникне поле аналізу однієї змінної з початковим зведенням про введені дані. У цьому повідомленні вказано ім'я Lamp, число спостережень 65 та його межі від 8.4 до 21.9. У верхній частині робочого поля розташовані кнопки меню поточних операцій, за допомогою яких можна змінювати вхідні дані, вибирати табличні та графічні опції та зберігати результати аналізу у файлі даних.

Вікна, в яких відображаються табличні та графічні результати, розкриваються на все робоче поле двома клацаннями лівої кнопки миші.

Клацання правої кнопки миші відкриває доступ до спеціального меню, яке задає параметри графічних зображень або зміни в поточному аналізі даних. Наприклад, при натисканні правої кнопки миші на вікні загальних статистик на екрані виникне наступне діалогове вікно (рис. 2.10). Для обчислення необхідних статистик слід поставити галочку навпроти відповідних назв. Задаємо обчислення всіх сумарних статистик, а також гістограму та графік густини, діаграму розсіювання квантильний графік. В результаті на робоче поле будуть видані такі табличні та графічні зображення (рис. 2.11).

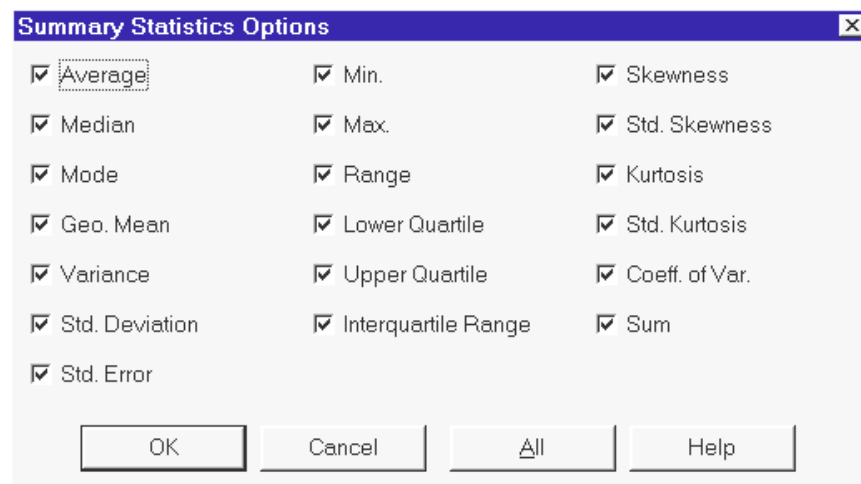


Рис. 2.10. Діалогове вікно завдання загальних статистик

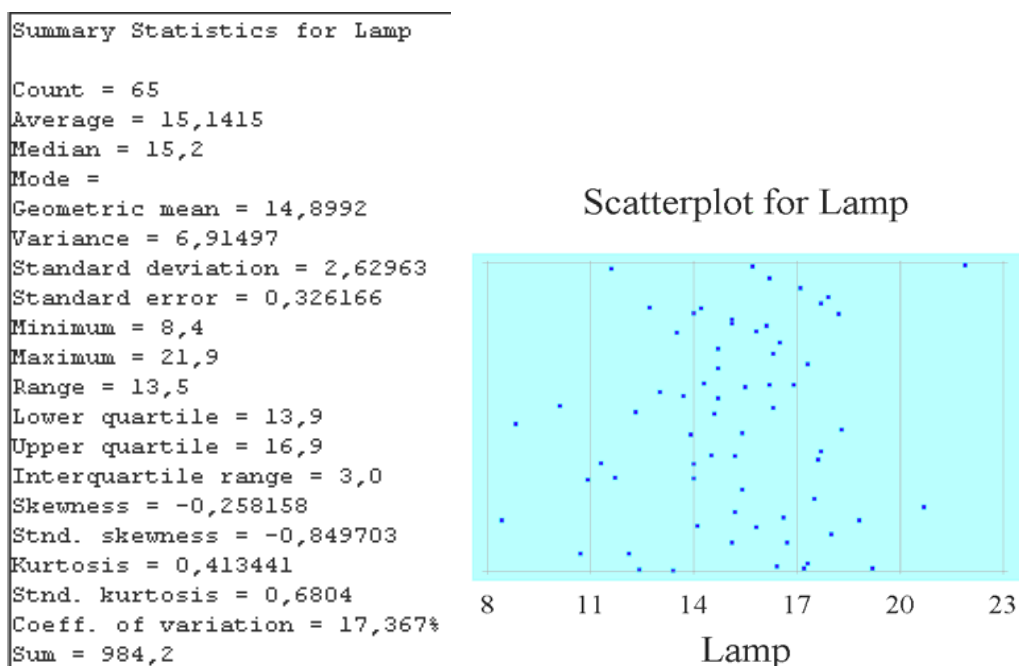


Рис. 2.11. Діаграма розсіювання змінної Lamp

На наведених графіках та таблицях за порядком першою розташована таблиця зі значеннями всіх обчислених статистик вибірки. Праворуч від

таблиці зображено діаграму розсіювання елементів вибірки (див. рис. 2.11). Далі йде таблиця частот, що включає значення верхньої (Lower Limit) і нижньої (Upper Limit) межі інтервалу вибірки, його середину (Midpoint), число (Frequency) і відносну частоту (Relative Frequency) потраплянь в інтервал вибірки, а також їх показники накопичення. Нижче розташовано саму гістограму, побудовану за даними попередньої таблиці (рис. 2.12). Нарешті, останніми наведено таблицю початкових відомостей про елементи вибірки та таблицю процентилей (рис. 2.13). Процентилей рівнів 0.5, 0.25 та 0.75 що відповідають медіані, нижній та верхній квартилі вибірки. Праворуч від таблиць наведено процентильний графік. Меню Describe дозволяє аналізувати безліч змінних, підбирати розподіли та проводити табуляцію даних. Ці можливості, не використані в завданні та будуть частково задіяні в наступних лабораторних роботах.

Class	Lower Limit	Upper Limit	Midpoint	Frequency	Relative Frequency	Cumulative Frequency	Cum. Rel. Frequency
at or below		7,0		0	0,0000	0	0,0000
1	7,0	9,0	8,0	2	0,0308	2	0,0308
2	9,0	11,0	10,0	3	0,0462	5	0,0769
3	11,0	13,0	12,0	8	0,1231	13	0,2000
4	13,0	15,0	14,0	15	0,2308	28	0,4308
5	15,0	17,0	16,0	21	0,3231	49	0,7538
6	17,0	19,0	18,0	13	0,2000	62	0,9538
7	19,0	21,0	20,0	2	0,0308	64	0,9846
8	21,0	23,0	22,0	1	0,0154	65	1,0000
above	23,0			0	0,0000	65	1,0000

Mean = 15,1415 Standard deviation = 2,62963

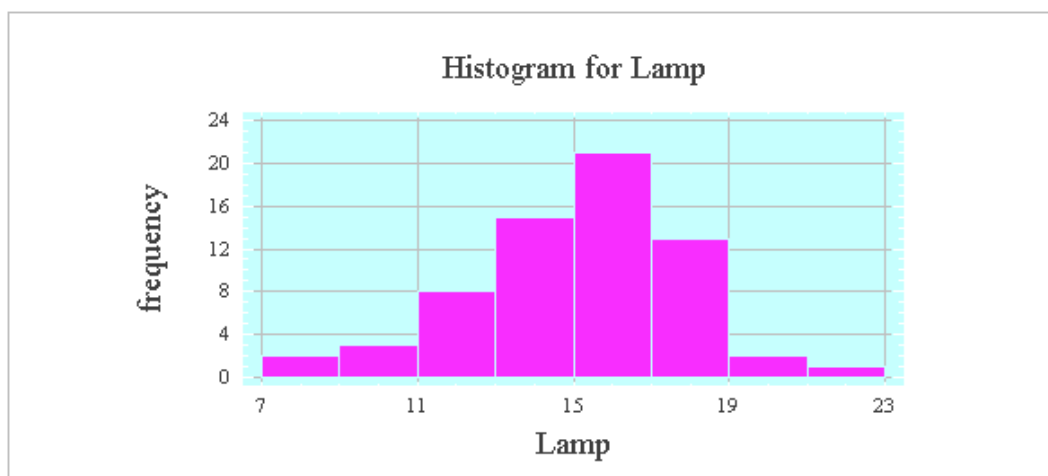


Рис. 2.12. Результати аналізу та гістограма змінної Lamp

Analysis Summary	
Data variable:	Lamp
	65 values ranging from 8,4 to 21,9
Percentiles for Lamp	
1,0%	= 8,4
5,0%	= 10,7
10,0%	= 11,6
25,0%	= 13,9
50,0%	= 15,2
75,0%	= 16,9
90,0%	= 18,0
95,0%	= 18,8
99,0%	= 21,9

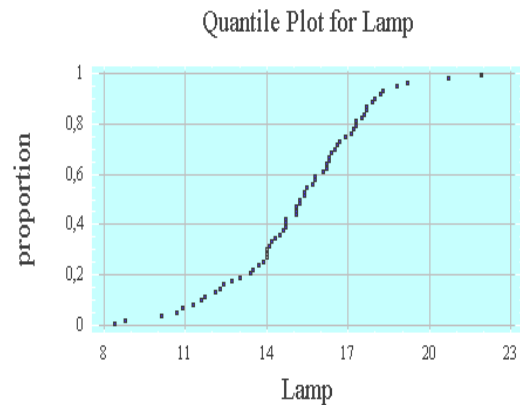


Рис. 2.13. Квантильний графік змінної Lamp

Завдання. За даними вибірками за допомогою пакету STATGRAPHICS обчислити всі сумарні статистики, побудувати гістограму, квантильний графік та діаграму розсіювання. Варіанти прикладу наступні.

1. Урожайність зернових культур в Україні із 1945 по 1989 рр. в центнерах з гектара:

5.6, 4.6, 7.3, 6.7, 6.9, 7.9, 7.4, 8.6, 7.8, 7.7, 8.4, 9.9, 8.4, 11.1, 10.4, 10.9, 10.7, 10.9, 8.3, 11.4, 9.5, 13.7, 12.1, 14.0, 13.2, 15.6, 15.4, 14.0, 17.6, 15.4, 10.9, 17.5, 15.0, 18.5, 14.2, 14.9, 12.6, 15.2, 15.9, 14.4, 16.2, 18.0, 18.3, 17.0, 18.8.

2. Час вирішення контрольної задачі учнями четвертого класу за секунди:

38, 60, 41, 51, 33, 42, 45, 21, 53, 60, 68, 52, 47, 46, 49, 49, 14, 57, 54, 59, 77, 47, 28, 48, 58, 32, 42, 58, 61, 30, 61, 35, 47, 72, 41, 45, 44, 55, 30, 40, 67, 65, 39, 48, 43, 60, 54, 42, 59, 50.

3. Вимірювання ємності затвора-стоку у 80 польових транзисторів дали такі результати:

1.9, 3.1, 1.3, 0.7, 3.2, 1.1, 2.9, 2.7, 2.7, 4.0, 1.7, 3.2, 0.9, 0.8, 3.1, 1.2, 2.6, 1.9, 2.3, 3.2, 4.1, 1.3, 2.4, 4.5, 2.5, 0.9, 1.4, 1.6, 2.2, 3.1, 1.5, 1.1, 2.3, 4.3, 2.1, 0.7, 1.2, 1.5, 1.8, 2.9, 0.8, 0.9, 1.7, 4.1, 4.3, 2.6, 0.9, 0.8, 1.2, 2.1, 3.2, 2.9, 1.1, 3.2, 4.5, 2.1, 3.1, 5.1, 1.1, 1.9, 0.9, 3.1, 0.9, 3.1, 3.3, 2.8, 2.5, 4.0, 4.3, 1.1, 2.1, 3.8, 4.6, 3.8, 2.3, 3.9, 2.4, 4.1, 4.2, 0.9.

4. Додатні відхилення від номінального розміру у партії деталей у міліметрах:

17, 21, 8, 20, 23, 18, 22, 20, 17, 12, 20, 11, 9, 19, 20, 9, 19, 17, 21, 13, 17, 22, 22,

10, 20, 20, 15, 19, 20, 20, 13, 21, 21, 9, 14, 11, 19, 18, 23, 19.

5. Час відновлення діодів з однієї партії у наносекундах:

69, 73, 70, 68, 61, 73, 70, 72, 67, 70, 66, 70, 76, 68, 71, 71, 68, 70, 64, 65, 72, 70, 70, 69, 66, 70, 77, 69, 71, 74, 72, 72, 72, 68, 70, 67, 71, 67, 72, 69, 66, 75, 76, 69, 71, 67, 70, 73, 71, 74.

6. Час реакції в секундах:

8.5, 7.1, 6.7, 6.2, 2.9, 4.4, 6.0, 5.8, 5.4, 8.2, 6.9, 6.5, 6.1, 3.8, 6.0, 6.0, 5.6, 5.3, 7.7, 6.8, 6.5, 6.1, 4.2, 4.7, 5.6, 5.4, 5.3, 7.4, 6.7, 6.4, 6.1, 4.5, 6.0, 5.8, 5.6, 5.1.

7. Діаметри головок заклепок у міліметрах:

13.39, 13.42, 13.38, 13.53, 13.51, 13.20, 13.40, 13.40, 13.28, 13.43, 13.46, 13.53, 13.55, 13.29, 13.24, 13.34, 13.54, 13.66, 13.43, 13.42, 13.38, 13.34, 13.57, 13.26, 13.33, 13.43, 13.50, 13.44, 13.53, 13.48, 13.48, 13.34, 13.36, 13.59, 13.36, 13.44, 13.34, 13.33, 13.25, 13.28, 13.49, 13.33, 13.26, 13.26, 13.55, 13.54, 13.37, 13.31, 13.37, 13.33.

8. Максимальна ємність двадцяти підстроювальних конденсаторів у пікофарадах:

4.45, 4.40, 4.42, 4.45, 4.38, 4.42, 4.36, 4.35, 4.40, 4.45, 4.42, 4.44, 4.36, 4.42, 4.44, 4.38, 4.39, 4.40, 4.42, 4.45.

9. Максимальні витрати води річки Тиса біля міста Хуст за 1950-2003 рр. в кубометрах за секунду:

2.46, 1.69, 1.34, 2.22, 2.18, 1.22, 1.22, 0.75, 1.26, 1.73, 1.74, 3.09, 1.57, 1.97, 2.23, 2.03, 1.58, 0.90, 2.40, 1.65, 1.96, 2.30, 1.79, 1.48, 3.44, 1.91, 3.06, 2.08, 1.06, 1.56, 1.88, 2.10, 2.02, 1.74, 1.18, 2.12, 1.38, 0.90, 1.45, 1.78, 1.97, 2.27, 2.34, 2.44.

10. Відхилення довжини валиків від номінального розміру в міліметрах, відібраних із поточної продукції прецизійного токарного автомата:

1.0, 1.5, -2.5, 0.0, -1.5, 1.0, 1.0, 15.0, -1.0, 2.0, 2.0, 3.0, 11.0, -1.0, 5.0, 4.5, 0.5, 3.5, 8.0, 5.0, 4.5, 3.5, 9.5, 12.5, 7.5, 7.5, 10.0, 8.5, 10.0, -3.0, 5.0, 3.5, -3.0, -14.0, 17.0, -9.0, -13.0, -12.5, 8.5, 12.5, 6.0, 8.5, 0.0, 7.0, -1.0, -3.0, 0.5, 0.0, -2.0, -4.5, 2.0, -10.0, -8.5, -3.5, -11.5, -11.5, -7.5, -11.5, -6.5, 2.0.

11. Проби заліза мали наступні точки плавлення (у градусах Цельсія):

1493, 1519, 1518, 1512, 1512, 1514, 1489, 1508, 1508, 1494, 1509, 1506, 1512, 1483, 1507, 1491, 1490, 1501, 1516, 1492, 1503, 1511, 1515, 1499, 1505.

11. Наведено дані щодо вмісту хрому (у вагових відсотках) у зразках нержавіючої сталі:

17.4, 17.9, 17.6, 18.1, 18.0, 17.6, 18.9, 18.2, 16.9, 17.5, 18.4, 17.8, 17.4, 18.5, 24.6, 20.8, 18.1, 26.0, 21.8, 17.7, 16.7, 18.8, 21.4, 19.5, 18.8.

12. Дано вміст залізистої сироватки (у мікрограмах на 100 мл) у 40 зразках:

111, 107, 100, 99, 102, 106, 109, 108, 104, 99, 107, 108, 106, 98, 105, 103, 110, 105, 104, 100, 101, 96, 97, 102, 107, 113, 116, 113, 110, 98, 96, 108, 103, 104, 114, 114, 113, 108, 106, 99.

13. На телефонній станції проводилися спостереження за кількістю неправильних з'єднань за хвилину. Спостереження протягом години дали наступні результати:

3, 1, 3, 1, 4, 2, 2, 4, 0, 3, 0, 2, 2, 0, 2, 1, 4, 3, 3, 1, 4, 2, 2, 1, 1, 2, 1, 0, 3, 4, 1, 3, 2, 7, 2, 0, 0, 1, 3, 3, 1, 2, 4, 2, 0, 2, 3, 1, 2, 5, 1, 1, 0, 1, 1, 2, 2, 1, 1, 5.

15. При вимірі діаметрів валиків після шліфування отримано наступні результати (у міліметрах):

6.75, 6.77, 6.77, 6.73, 6.76, 6.74, 6.70, 6.75, 6.71, 6.77, 6.79, 6.73, 6.70, 6.74, 6.75, 6.71, 6.70, 6.78, 6.81, 6.69, 6.80, 6.68, 6.74, 6.83, 6.76, 6.82, 6.71, 6.77, 6.75, 6.82, 6.80, 6.72, 6.69, 6.81, 6.74, 6.80, 6.76, 6.77, 6.81, 6.82, 6.73, 6.72, 6.77, 6.78, 6.75, 6.68, 6.72, 6.69, 6.76, 6.70.

16. Октанове число бензину:

84.0, 83.5, 84.0, 85.0, 83.1, 83.5, 81.7, 85.4, 84.1, 83.0, 85.8, 82.4, 82.4, 83.4, 83.3, 83.1, 83.3, 82.4, 83.3, 82.6, 82.0, 83.2, 84.0, 84.2, 82.2, 83.6, 84.9, 83.2, 82.8, 83.4, 80.2, 82.7, 83.0, 85.0, 83.0, 85.0, 83.7, 83.6, 83.1, 82.5.

17. Вимірювання кількості опадів (у сантиметрах), що випали під час кількох ураганів:

1.05, 1.40, 0.69, 1.41, 0.51, 1.49, 1.38, 2.00, 0.96, 1.31, 2.07, 1.02, 0.89, 1.51, 0.66, 1.16, 0.64, 1.07, 0.33, 1.59, 1.11, 1.33, 0.96, 1.40, 1.71, 0.75, 0.75, 0.92, 1.03, 0.78.

18. Результати вимірів деякої фізичної характеристики пластикового

матеріалу, отриманого з кількох партій:

55, 42, 45, 41, 43, 53, 41, 43, 34, 50, 42, 41, 43, 46, 42, 44, 43, 45, 34, 48, 47, 46, 48, 41, 38, 49, 41, 44, 40, 48, 52, 50, 45, 30, 35, 52, 35, 46, 40, 48.

19. Час приготування кави, виражений з точністю до сотих часток хвилини, для кількох типів електричних кавоварок:

1.38, 9.69, 0.39, 1.42, 0.54, 5.94, 0.59, 1.42, 0.39, 1.46, 0.55, 6.15, 0.61, 2.63, 2.44, 0.56, 0.69, 0.71, 0.95, 0.50, 2.69, 2.68, 0.53, 0.72, 0.74, 0.93, 0.53, 5.37, 2.18, 0.97.

20. Втрати металу в сотнях тон за період від встановлення обладнання до моменту руйнування деякої його частини:

84, 60, 40, 47, 34, 46, 67, 92, 95, 40, 98, 60, 59, 108, 86, 117, 46, 93, 100, 92, 93, 79, 66, 82, 68.

21. Результати виміру зросту (у сантиметрах) випадково відібраних 50 студентів:

155.0, 159.1, 167.5, 181.7, 175.0, 164.8, 165.2, 171.6, 180.3, 170.0, 173.9, 168.3, 169.5, 169.0, 162.8, 165.1, 159.0, 161.5, 155.5, 160.8, 161.2, 175.0, 176.1, 167.2, 170.8, 165.2, 168.4, 157.3, 178.0, 182.0, 181.5, 175.0, 177.3, 171.6, 169.0, 165.3, 163.4, 166.0, 172.8, 159.3, 161.2, 157.1, 165.7, 160.4, 174.7, 165.4, 169.3, 173.8, 177.2, 179.6.

22. За даними 40 дослідів отримано наступні експериментальні значення випадкової величини X :

8, 14, 42, 22, -40, 18, -16, 38, -4, 2, -16, 34, 6, -11, 54, 8, 20, 74, -26, 0, 4, -28, 16, -22, 36, 44, 10, -13, 16, 24, -19, 46, 5, -7, 17, 23, 47, -21, 6, 14.

23. Річна кількість опадів (у дюймах) у Львові з 1963 по 2012 рік:

21.59, 16.93, 29.48, 31.60, 26.25, 23.40, 25.42, 21.32, 25.02, 33.86, 22.67, 18.82, 28.44, 26.16, 28.17, 34.08, 33.82, 30.28, 27.92, 27.14, 24.40, 20.35, 26.64, 27.01, 19.21, 27.74, 23.85, 21.23, 28.15, 22.61, 19.80, 27.94, 21.47, 23.52, 22.86, 17.69, 22.54, 23.28, 22.17, 20.84, 38.10, 20.65, 22.97, 24.26, 23.01, 23.67, 26.75, 25.36, 24.79, 27.88.

24. Річна врожайність ячменю (у центнерах на 1 акр) в Україні з 1950 по

1999 рік:

16.7, 16.3, 16.5, 13.3, 16.5, 15.0, 15.9, 15.5, 16.9, 16.4, 14.9, 14.5, 16.6, 15.1, 14.6, 16.0, 16.8, 16.8, 15.5, 17.3, 15.5, 15.5, 14.2, 15.8, 15.7, 14.1, 14.8, 14.4, 15.6, 13.9, 14.7, 14.3, 14.0, 14.5, 15.4, 15.3, 16.0, 16.4, 17.2, 17.8, 14.4, 15.0, 16.0, 16.8, 16.9, 16.6, 16.2, 14.0, 18.1, 17.5.

25. Ряд із 60 рівномірних випадкових чисел, що приймають цілі значення від 0 до 19:

3, 15, 15, 8, 19, 1, 3, 12, 19, 13, 16, 4, 17, 8, 6, 15, 3, 3, 7, 4, 5, 14, 15, 10, 3, 10, 13, 14, 15, 8, 10, 1, 18, 17, 4, 10, 16, 2, 13, 3, 14, 7, 16, 3, 10, 12, 0, 3, 2, 3, 10, 5, 10, 3, 2, 11, 14, 18, 8, 14.

26. Поквартальні індекси роздрібної ціни на антибіотик в Україні у 2011–2018 роках:

295.0, 317.5, 314.9, 321.4, 324.7, 323.7, 322.5, 332.9, 372.9, 380.9, 353.0, 348.9, 354.0, 345.7, 319.5, 317.6, 333.7, 323.9, 312.8, 310.2, 323.2, 342.9, 300.3, 309.8, 304.3, 285.9, 292.3, 298.7, 312.5, 336.1, 295.5, 318.4.

27. В результаті вимірювання контрольованого розміру відібраних виробів отримано такі спостереження:

1.08, 1.10, 1.12, 1.38, 1.18, 1.12, 1.36, 1.25, 1.15, 1.14, 1.40, 1.42, 1.11, 1.22, 1.36, 1.33, 1.35, 1.35, 1.41, 1.21, 1.37, 1.13, 1.15, 1.29, 1.31, 1.17, 1.45, 1.34, 1.17, 1.23, 1.39, 1.06, 1.26, 1.31, 1.37.

28. В результаті зважування 30 проб хімічної речовини отримані такі дані (у міліграмах):

25, 28, 30, 31, 28, 26, 50, 52, 20, 24, 26, 23, 40, 36, 28, 31, 32, 33, 36, 35, 29, 42, 42, 45, 38, 40, 41, 29, 25, 37.

29. Отримано такі результати аналізів на вміст вуглецю (у відсотках) в пробах нелегованої сталі:

0.18, 0.12, 0.12, 0.08, 0.08, 0.12, 0.19, 0.32, 0.27, 0.11, 0.14, 0.23, 0.16, 0.09, 0.08, 0.05, 0.13, 0.17, 0.10, 0.14, 0.30, 0.27, 0.31, 0.24, 0.22, 0.34, 0.14, 0.46, 0.39, 0.24, 0.28, 0.11, 0.42, 0.29, 0.11.

30. Результати лабораторних аналізів 60 зразків сланцевих порід на вміст кремнію (SiO_2) у відсотках:

57.8, 54.6, 54.8, 51.7, 61.1, 62.3, 52.2, 49.2, 53.9, 60.0, 56.2, 55.2, 53.3, 57.9, 54.0, 52.6, 53.8, 53.6, 51.5, 54.0, 50.4, 53.0, 53.3, 51.6, 50.9, 49.6, 52.2, 50.5, 51.1, 52.2, 49.2, 49.3, 48.8, 53.5, 52.8, 52.9, 52.1, 47.3, 49.8, 49.3, 50.1, 54.4, 49.0, 48.9, 51.3, 51.6, 46.2, 50.4, 50.7, 53.1, 52.9, 51.3, 52.7, 46.6, 46.5, 51.3, 51.0, 47.5, 47.7, 44.9.

2.7. Нормальний розподіл та його числові характеристики

В теорії ймовірності нормальний закон займає особливе місце, оскільки є граничним законом для багатьох інших при виконанні деяких дуже нежорстких обмежень. Саме розподіл суми випадкових величин відповідає наближено нормальному закону, якщо серед цих випадкових величин немає таких, що різко виділяються, самі ж випадкові величини окремо можуть бути підпорядкованими будь-якому закону.

Випадкова величина X має нормальний розподіл, якщо її функція щільності ймовірності має вигляд

$$f(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-m)^2}{2\sigma^2}}. \quad (2.6.1)$$

Закон має два параметри m і σ , тобто відноситься до класу двопараметричних законів. Графіки функцій щільності ймовірності та функції розподілу наведено на рис. 2.14.

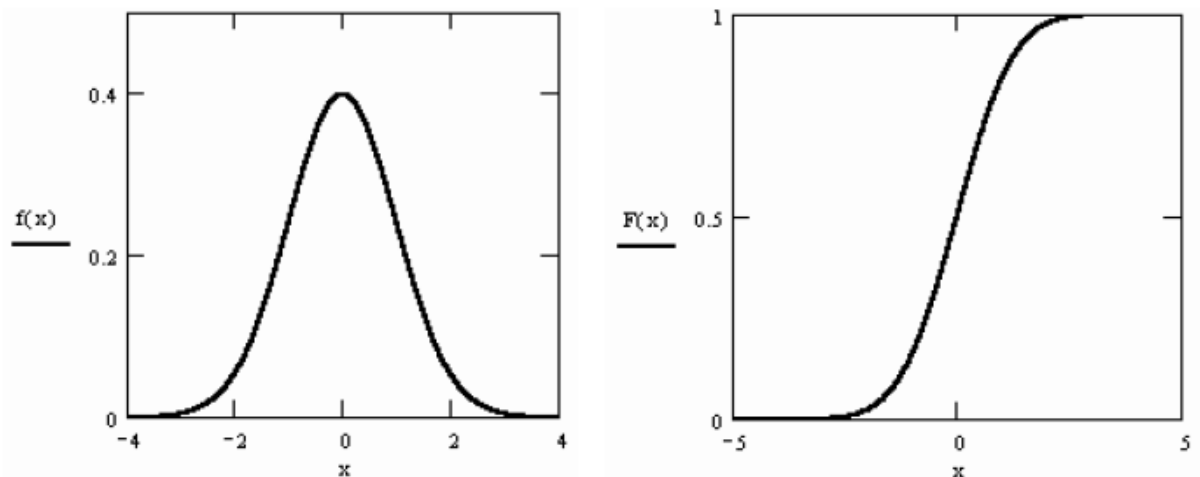


Рис. 2.14. Графіки функції щільності ймовірності та функції розподілу стандартного нормального закону

Знайдемо як завжди спочатку функцію розподілу

$$\begin{aligned}
F(x) &= \int_{-\infty}^x f(t) dt = \int_{-\infty}^x \frac{1}{\sqrt{2\pi\sigma}} e^{-\frac{(t-m)^2}{2\sigma^2}} dt = \\
&= \frac{1}{\sqrt{2\pi\sigma}} \int_{-\infty}^x e^{-\frac{(t-m)^2}{2\sigma^2}} dt = \left\langle \frac{t-m}{\sigma} = u, \right. \\
&\quad \left. dt = \sigma du. \right\rangle = \\
&= \frac{1}{\sqrt{2\pi\sigma}} \int_{-\infty}^{\frac{x-m}{\sigma}} e^{-\frac{u^2}{2}} \sigma du = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\frac{x-m}{\sigma}} e^{-\frac{t^2}{2}} dt.
\end{aligned}$$

Останній інтеграл не виражається через елементарні функції. Він називається функцією Лапласа і позначається:

$$\Phi\left(\frac{x-m}{\sigma}\right) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\frac{x-m}{\sigma}} e^{-\frac{t^2}{2}} dt. \quad (2.6.2)$$

Історично розрізняють кілька різновидів функції Лапласа. Формула (1.6.2) дає нормальну функцію Лапласа, яка й називається функцією Лапласа; функція:

$$\Phi^*\left(\frac{x-m}{\sigma}\right) = \frac{1}{\sqrt{2\pi}} \int_0^{\frac{x-m}{\sigma}} e^{-\frac{t^2}{2}} dt. \quad (2.6.3)$$

називається нормованою функцією Лапласа; нарешті, також широко застосовується формула:

$$\widehat{\Phi}\left(\frac{x-m}{\sigma}\right) = \frac{2}{\sqrt{\pi}} \int_0^{\frac{x-m}{\sigma}} e^{-t^2} dt. \quad (2.6.4)$$

яка називається приведеною функцією Лапласа. Зв'язок між усіма цими функціями легко встановлюється за загальним правилом заміни змінних у певному інтегралі. Наприклад:

$$\Phi(x) = \frac{1}{2} + \Phi^*(x), \quad \widehat{\Phi}(x) = 2\Phi(\sqrt{2}x) - 1$$

Визначимо параметри нормального закону. Обмежимося двома точковими характеристиками: математичним очікуванням та дисперсією

$$m_x = \frac{1}{\sqrt{2\pi\sigma}} \int_{-\infty}^{\infty} x e^{-\frac{(x-m)^2}{2\sigma^2}} dx = \left\langle \frac{x-m}{\sigma\sqrt{2}} = t, \right. \\ \left. dx = \sqrt{2}\sigma dt \right\rangle =$$

$$\begin{aligned}
&= \frac{1}{\sqrt{2\pi}\sigma} \int_{-\infty}^{\infty} (\sqrt{2}\sigma t + m) e^{-t^2} \sqrt{2}\sigma dt = \\
&= \frac{\sigma\sqrt{2}}{\sqrt{\pi}} \int_{-\infty}^{\infty} t e^{-t^2} dt + \frac{m}{\sqrt{\pi}} \int_{-\infty}^{\infty} e^{-t^2} dt = \\
&= \frac{\sigma\sqrt{2}}{2\sqrt{\pi}} (-e^{-t^2}) \Big|_{-\infty}^{\infty} + \frac{m}{\sqrt{\pi}} \int_{-\infty}^{\infty} e^{-t^2} dt = \frac{m}{\sqrt{\pi}} \sqrt{\pi} = m.
\end{aligned}$$

Тут використаний інтеграл Ейлера-Пуассона:

$$\int_{-\infty}^{\infty} e^{-t^2} dt = 2 \int_0^{\infty} e^{-t^2} dt = \sqrt{\pi}.$$

Аналогічно

$$\begin{aligned}
D_X &= \frac{1}{\sqrt{2\pi}\sigma} \int_{-\infty}^{\infty} (x - m_X)^2 e^{-\frac{(x-m_X)^2}{2\sigma^2}} dx = \left\langle \begin{array}{l} \frac{x - m_X}{\sqrt{2}\sigma} = t, \\ dx = \sqrt{2}\sigma dt. \end{array} \right\rangle = \frac{2\sigma^2}{\sqrt{\pi}} \int_{-\infty}^{\infty} t^2 e^{-t^2} dt \\
&= \\
&= \left\langle \begin{array}{l} t = u, dt = du, \\ te^{-t^2} dt = dv, \\ v = -\frac{1}{2} e^{-t^2} \end{array} \right\rangle = \left[-\frac{1}{2} t e^{-t^2} \Big|_{-\infty}^{\infty} + \frac{1}{2} \int_{-\infty}^{\infty} e^{-t^2} dt \right] \frac{2\sigma^2}{\sqrt{\pi}} = \frac{2\sigma^2}{\sqrt{\pi}} \cdot \frac{1}{2} \sqrt{\pi} = \sigma^2.
\end{aligned}$$

Отже, $D_X = \sigma^2$, $\sigma_X = \sigma$. Зрозуміло також значення параметрів m і σ нормального розподілу: m - центр розсіювання – який є центром симетрії розподілу. Це добре видно з графіка функції щільності ймовірності. Розмірність центра розсіювання m дорівнює розмірності випадкової величини X . Дещо складніше ситуація з параметром σ . Цей параметр характеризує форму кривої розподілу. У разі збільшення σ графік дедалі більше «розмазується» по осі OX , тобто. випадкова величина X має більше розсіювання біля центру симетрії. Чим менший σ , тим «гостріше» вершина графіка функції щільності ймовірності. Розмірність D_X збігається із розмірністю X^2 . Легко показати, що для нормального розподілу $d_X = h_X = m_X$. Корисна формула, що виражає будь-який центральний момент нормального розподілу через дисперсію $\mu_k = (k-1)!! D_X$.

3. ДИСПЕРСІЙНИЙ АНАЛІЗ

3.1. Сутність задачі дисперсійного аналізу

Вивчення розсіювання величин, що спостерігаються в експерименті - один з головних предметів прикладної статистики. Дисперсійний аналіз є методом розкладання загальної дисперсії сукупності спостережень на складові. Враховуючи, що розсіювання спостережуваної випадкової змінної X причинно обумовлено впливом безлічі факторів, дисперсійний аналіз можна інтерпретувати як метод поділу ефектів впливу на значення X різних підмножин, що спостерігаються, в загальній множині факторів. Термін "дисперсійний аналіз" вперше ввів Фішер і визначив його як відділення дисперсії, що приписується одній групі причин, від дисперсії, що приписується іншим групам. Модель, що використовується при цьому, узагальнено може бути представлена в наступному вигляді.

$$\begin{array}{l} \text{Значення,} \\ \text{що} \\ \text{спостеріга} \\ \text{ються} \end{array} = \begin{array}{l} \sum \text{ параметрів,} \\ \text{що описують} \\ \text{зумовлені} \\ \text{ефекти} \end{array} + \begin{array}{l} \sum \text{ випадкових} \\ \text{величин, що} \\ \text{описують} \\ \text{невизначені} \\ \text{(залишкові)} \\ \text{ефекти} \end{array} \quad (3.1.1)$$

При дослідженні залежностей однією з найпростіших є ситуація, коли можна вказати лише один фактор, що впливає на кінцевий результат, і цей фактор може набувати лише кінцевої кількості значень (рівнів). Такі завдання, звані завданнями однофакторного аналізу, часто зустрічаються практично. Типовий приклад завдань однофакторного аналізу – порівняння за результатами кількох різних способів дії, спрямованих на досягнення однієї мети.

Для застосування дисперсійного аналізу необхідно спочатку побудувати відповідну статистичну модель та з'ясувати структуру експериментальних даних. Досвід показує, що при зміні способу обробки найбільшої мінливості в першу чергу, як правило, схильне до положення випадкової величини, яке можна охарактеризувати медіаною або середнім значенням. Дотримуючись цього емпіричного правила, в однофакторних

завданнях також зазвичай припускають, що всі спостереження належать деякому сімейному зрушенню розподілів. Часто як таке сімейство розглядається сімейство нормальних розподілів і для обробки даних застосовуються методи дисперсійного аналізу. В інших випадках припущення про нормальність не є правомірним і тоді використовують різні непараметричні методи аналізу, з яких найбільш розроблені рангові методи.

Введемо деякі загальноприйняті терміни, що дозволяють отримати у стислому вигляді опис структури експерименту. Основним є поняття фактора – це якість чи властивість, відповідно до якого класифікуються дані та яке має впливати на кінцевий результат. Кожен фактор має кілька різних рівнів. Рівень - конкретна реалізація фактора - використовується для опису якості, що визначає кожну категорію застосовуваної класифікації.

Структура або схема експерименту, зазвичай звана планом експерименту, описується факторами, що входять до нього, і способом комбінування різних рівнів різних факторів. Нарешті величину результату часто називають відгуком.

Для порівняння впливу факторів на результат потрібний певний статистичний матеріал. Зазвичай його отримують наступним чином: кожен з k способів обробки застосовується кілька разів (не обов'язково одне і те ж число разів) до об'єкта, що досліджується, потім результати реєструються. Дані таких випробувань можуть бути зведені у табл. 3.1.

Таблиця 3.1

Обробки (відповідають рівням факторів)	1	3	...	k
Результати спостережень	x_{11}	x_{12}	...	x_{1k}
	x_{21}	x_{22}	...	x_{2k}

	x_{n1}	x_{n2}	...	x_{nk}

3.2 Дисперсійний аналіз

Для опису даних табл. 3.1 здебільшого виявляється прийнятною адитивна модель. Вона передбачає, що значення відгуку x_{ij} можна як суми вкладу (впливу) чинника і незалежної від вкладу фактора випадкової величини. Зазвичай модель однофакторного дисперсійного аналізу записується як

$$x_{ij} = \mu + T_j + \varepsilon_{ij}, \quad i = 1, 2, \dots, n_j, \quad j = 1, 2, \dots, k \quad (3.2.1)$$

де μ - математичне очікування X загальної сукупності, $T_j = \mu_j - \mu$ - ефект впливу j -го рівня чинника, тобто відхилення від загального середнього рівня при j -й обробці μ_j - математичного очікування X в j -й групі, ε_{ij} - випадкова помилка спостережень.

Зазвичай передбачається лише безперервність закону розподілу величин ε_{ij} та його незалежність. Однак у багатьох випадках про розподіл ε_{ij} можна сказати більше, наприклад, припускають, що величини $\varepsilon_{ij} \in N(0, D)$, тобто мають нормальний розподіл із нульовим середнім та загальною дисперсією, яка невідома. Додаткова інформація про закон розподілу випадкових величин ε_{ij} дозволяє використовувати більш сильні методи в моделі однофакторного аналізу як для перевірки гіпотез, так і для оцінки параметрів. Сукупність цих методів називається однофакторним дисперсійним аналізом.

Найчастіше дисперсійний аналіз заснований на наступних правдоподібних припущення про випадкові величини ε_{ij} .

1. Математичне очікування кожної залишкової випадкової величини дорівнює нулю. Це означає, що вся мінливість у математичних очікуваннях охоплюється параметрами. Його дуже правдоподібне припущення, бо вплив другого члена в моделі (3.1.1) завжди набагато менше першого.
2. Залишкові випадкові величини взаємно незалежні, Це припущення менш очевидно, як перше. Сенс його полягає в тому, що між різними спостереженнями не існує будь-якого зв'язку, який не можна було б пояснити за допомогою членів, що описують обумовлені ефекти.
3. Усі залишкові випадкові величини мають однакове середньоквадратичне відхилення. Його припущення про однорідність дисперсій. У багатьох випадках це припущення не виконується, тому, перш ніж проводити дисперсійний аналіз будь-якого набору даних, важливо розглянути можливі коливання D .
4. Кожна залишкова випадкова величина розподілена за нормальним законом. У загальному випадку справедливість цього припущення найменш ймовірна, ніж трьох інших. Значна частина дисперсійного аналізу може проводитися без прийняття цього припущення, необхідного лише обґрунтування використання деяких формально точних критеріїв для перевірки значущості і формул оцінювання.

Припущення, описані вище, мають форму:

1. $M(\varepsilon_{ij}) = 0$;
2. величини ε_{ij} взаємно незалежні
3. $D(\varepsilon_{ij}) = D = \text{const}$; . (3.2.2)
4. $\varepsilon_{ij} \in N(0, D)$.

Якщо рівні досліджуваного фактора фіксовані, то ефекти $T_j = \mu_j - \mu$ є фіксованими постійними та їх сума дорівнює нулю, тому що в експерименті вибрано всі можливі значення рівнів.

Звернемося тепер до табл. 3.1. Змінність або варіація значень X_u , що спостерігаються, може бути викликана змінністю рівнів фактора і змінністю значення випадкових величин, що описують невизначені ефекти.

Обчислимо середнє значення кожної групи і загальне середнє всіх спостережень:

$$\begin{aligned}\bar{x}_j &= \frac{1}{n_j} \sum_{i=1}^{n_j} x_{ij} = \frac{1}{n_j} x_{.j}, \quad j = 1, 2, \dots, k, \\ \bar{x} &= \frac{1}{n} \sum_{j=1}^k n_j \bar{x}_j = \frac{1}{n} \sum_{j=1}^k \sum_{i=1}^{n_j} x_{ij}, \quad n = \sum_{j=1}^k n_j, \\ \bar{\varepsilon}_j &= \frac{1}{n_j} \sum_{i=1}^{n_j} \varepsilon_{ij}, \quad \bar{\varepsilon} = \frac{1}{n} \sum_{j=1}^k n_j \bar{\varepsilon}_j = \frac{1}{n} \sum_{j=1}^k \sum_{i=1}^{n_j} \varepsilon_{ij}.\end{aligned}$$

З урахуванням цих формул, перше рівняння моделі однофакторного дисперсійного аналізу (3.2.1) можна спростити. Підсумуємо формулу (3.2.1) по i в межах від одиниці до n_j . Отримаємо $\sum_{i=1}^{n_j} x_{ij} = \sum_{i=1}^{n_j} \mu + \sum_{i=1}^{n_j} T_j + \sum_{i=1}^{n_j} \varepsilon_{ij}$ або $n_j \bar{x}_j = n_j \mu + n_j T_j + n_j \bar{\varepsilon}_j$. Остаточний вираз

$$\bar{x}_j = \mu + T_j + \bar{\varepsilon}_j. \quad (3.2.3)$$

Продовжимо складання по j в межах від одиниці до k . Тоді

$$\begin{aligned}\sum_{j=1}^k \sum_{i=1}^{n_j} x_{ij} &= \sum_{j=1}^k \sum_{i=1}^{n_j} \mu + \sum_{j=1}^k \sum_{i=1}^{n_j} T_j + \sum_{j=1}^k \sum_{i=1}^{n_j} \varepsilon_{ij}, \text{ тобто} \\ n\bar{x} &= \mu \sum_{j=1}^k n_j + \sum_{j=1}^k T_j n_j + \sum_{j=1}^k n_j \bar{\varepsilon}_j.\end{aligned}$$

Оскільки $\mu_j = \mu + T_j$ – відхилення значень μ_j від середнього значення μ , то $\mu = \frac{1}{n} \sum_{j=1}^k n_j \mu_j$, тобто середньозважене значення μ_j .

Тоді $\mu n = \sum_{j=1}^k n_j \mu_j = \sum_{j=1}^k n_j (\mu + T_j) = \sum_{j=1}^k n_j \mu + \sum_{j=1}^k n_j T_j = \mu n + \sum_{j=1}^k n_j T_j$. Звідки $\sum_{j=1}^k n_j T_j = 0$. Остаточне друге рівняння моделі має вид $n\bar{x} = n\mu + n\bar{\varepsilon}$, або

$$\bar{x} = \mu + \bar{\varepsilon}. \quad (3.2.4)$$

Відніmemo з рівняння (3.2.3) рівняння (3.2.4), отримаємо $\bar{x}_j - \bar{x} = T_j + \bar{\varepsilon}_j - \bar{\varepsilon}$. Тоді $M(\bar{x}_j - \bar{x}) = M(T_j + \bar{\varepsilon}_j - \bar{\varepsilon}) = T_j$, оскільки $M(\varepsilon_{ij}) = 0$ через припущення (3.2.2). Так само $D(\bar{x}_j - \bar{x}) = D$.

Розрахуємо загальну тотожність дисперсійного аналізу.

$$\begin{aligned} \sum_{j=1}^k \sum_{i=1}^{n_j} (x_{ij} - \bar{x})^2 &= \sum_{j=1}^k \sum_{i=1}^{n_j} [(x_{ij} - \bar{x}_j) - (\bar{x} - \bar{x}_j)]^2 = \\ &= \sum_{j=1}^k \sum_{i=1}^{n_j} (x_{ij} - \bar{x}_j)^2 + \sum_{j=1}^k \sum_{i=1}^{n_j} (\bar{x} - \bar{x}_j)^2 - \\ &- 2 \sum_{j=1}^k \sum_{i=1}^{n_j} (x_{ij} - \bar{x}_j)(\bar{x} - \bar{x}_j) = \sum_{j=1}^k \sum_{i=1}^{n_j} (x_{ij} - \bar{x}_j)^2 + \\ &+ \sum_{j=1}^k \sum_{i=1}^{n_j} (\bar{x} - \bar{x}_j)^2 - 2 \sum_{j=1}^k (\bar{x} - \bar{x}_j) \sum_{i=1}^{n_j} (x_{ij} - \bar{x}_j). \end{aligned}$$

Проте $\sum_{i=1}^{n_j} (x_{ij} - \bar{x}_j) = \sum_{i=1}^{n_j} x_{ij} - \sum_{i=1}^{n_j} \bar{x}_j = n_j \bar{x}_j - \bar{x}_j \sum_{i=1}^{n_j} 1 = n_j \bar{x}_j - \bar{x}_j n_j = 0$, тоді останній вираз

$$\sum_{j=1}^k \sum_{i=1}^{n_j} (x_{ij} - \bar{x})^2 = \sum_{j=1}^k \sum_{i=1}^{n_j} (\bar{x} - \bar{x}_j)^2 + \sum_{j=1}^k \sum_{i=1}^{n_j} (x_{ij} - \bar{x}_j)^2 \text{ або}$$

$$\sum_{j=1}^k \sum_{i=1}^{n_j} (x_{ij} - \bar{x})^2 = \sum_{j=1}^k n_j (\bar{x}_j - \bar{x})^2 + \sum_{j=1}^k \sum_{i=1}^{n_j} (x_{ij} - \bar{x}_j)^2. \quad (3.2.5)$$

Таким чином, загальна сума квадратів відхилень спостережень від загального середнього \bar{x} розбивається на суму квадратів відхилень вибірових середніх \bar{x}_j від загального середнього \bar{x} та суму квадратів відхилень спостережень x_{ij} від вибірових середніх груп \bar{x}_j (всередині груп), тобто,

$$Q = Q_1 + Q_2, \text{ де } Q = \sum_{j=1}^k \sum_{i=1}^{n_j} (x_{ij} - \bar{x})^2, Q_1 = \sum_{j=1}^k n_j (\bar{x}_j - \bar{x})^2,$$

$$Q_2 = \sum_{j=1}^k \sum_{i=1}^{n_j} (x_{ij} - \bar{x}_j)^2.$$

До формули (6.2.5) входять три члени. Розглянемо їх докладніше. Член $\frac{Q_1}{n} = \frac{1}{n} \sum_{j=1}^k n_j (\bar{x}_j - \bar{x})^2$ дорівнює дисперсії групових середніх, до нього входять k доданків, «свобода» зміни яких обмежена одним співвідношенням

$$\bar{x} = \frac{1}{n} \sum_{j=1}^k \sum_{i=1}^{n_j} x_{ij} = \frac{1}{n} \sum_{i=1}^{n_j} n_j \bar{x}_j = \frac{n_1 \bar{x}_1 + n_2 \bar{x}_2 + \dots + n_k \bar{x}_k}{n_1 + n_2 + \dots + n_k}.$$

З цієї формули випливає єдине рівняння зв'язку $(\bar{x}_1 - \bar{x})n_1 + (\bar{x}_2 - \bar{x})n_2 + \dots + (\bar{x}_k - \bar{x})n_k = 0$. Тому кажуть, що величина Q_1 має $(k - 1)$ ступінь вільності.

Величина Q_2/n дорівнює середній з групових дисперсій. У формулу розрахунку Q_2 входять $n_1 + n_2 + \dots + n_k = n$ доданків. Свобода перших n_1 доданків обмежена одним співвідношенням $\bar{x}_1 = \frac{1}{n_1} \sum_{i=1}^{n_1} x_{i1}$ або $(x_{11} - \bar{x}_1) + (x_{21} - \bar{x}_1) + \dots + (x_{n_11} - \bar{x}_1) = 0$. Таким чином, «свобода» зміни k доданків обмежена k умовами. Це означає, що величина Q_2 має $(n - k)$ ступенів вільності.

Нарешті, у формулу Q/n входять $n_1 + n_2 + \dots + n_k = n$ доданків. На них накладено одне обмеження $\bar{x} = \frac{1}{n} \sum_{j=1}^k \sum_{i=1}^{n_j} x_{ij} = \frac{(x_{11} + x_{12} + \dots + x_{1n_1}) + (x_{21} + x_{22} + \dots + x_{2n_2}) + \dots + (x_{k1} + x_{k2} + \dots + x_{kn})}{n}$ або $(x_{11} - \bar{x}) + (x_{12} - \bar{x}) + \dots + (x_{1n_1} - \bar{x}) + (x_{21} - \bar{x}) + (x_{22} - \bar{x}) + \dots + (x_{2n_2} - \bar{x}) + \dots + (x_{k1} - \bar{x}) + (x_{k2} - \bar{x}) + \dots + (x_{kn_j} - \bar{x}) = 0$. Тому Q має $(n - 1)$ степенів вільності.

За третьою умовою (3.2.2) все генеральні групові дисперсії мають бути рівними, тобто $D_1 = D_2 = \dots = D_k = D$. Знайдемо незміщені оцінки D .

По-перше, переконаємось у тому, що незміщена оцінка дисперсія D дорівнює $Q_2/(n - k)$, т.е. $M\left(\frac{Q_2}{n-k}\right) = D$. Справедливо, що

$$M\left(\frac{Q_2}{n-k}\right) = \frac{1}{n-k} M\left[\sum_{j=1}^k \sum_{i=1}^{n_j} (x_{ij} - \bar{x}_j)^2\right] = \frac{1}{n-k} \sum_{j=1}^k M\left[\sum_{i=1}^{n_j} (x_{ij} - \bar{x}_j)^2\right] \\ = \frac{1}{n-k} \sum_{j=1}^k M(n_j D_j^*),$$

$$\text{де } D_j^* = \frac{1}{n_j} \sum_{i=1}^{n_j} (x_{ij} - \bar{x}_j)^2 = \frac{(x_{1j} - \bar{x}_j)^2 + (x_{2j} - \bar{x}_j)^2 + \dots + (x_{n_j j} - \bar{x}_j)^2}{n_j}$$

- вибіркова групова дисперсія, обчислена за результатами спостережень j -м рівні фактора, $j = 1, 2, \dots, k$.

$$\begin{aligned}
M\left(\frac{Q_2}{n-k}\right) &= \frac{1}{n-k} \sum_{j=1}^k M(n_j D_j^*) = \frac{1}{n-k} \sum_{j=1}^k M[(n_j - 1)\tilde{D}_j] = \\
&= \frac{1}{n-k} \sum_{j=1}^k (n_j - 1)M(\tilde{D}_j) = \frac{1}{n-k} \sum_{j=1}^k (n_j - 1)D_j = \frac{1}{n-k} D_j \sum_{j=1}^k (n_j - 1) = \\
&= \frac{1}{n-k} D_j(n-k) = D_j = D,
\end{aligned}$$

Оскільки $\tilde{D}_j = \frac{n_j}{n_j-1} D_j^*$ - незміщені оцінки групових дисперсій, тобто. $M(\tilde{D}_j) = D_j$. Остання рівність правильна тільки в тому випадку, коли спостереження в j -й групі незалежні і проводяться в однакових умовах. Це справедливо за другою умовою (3.2.2). Отже, $M\left(\frac{Q_2}{n-k}\right) = D$.

Розглянемо тепер питання відмінності обробок (чинників) в табл. 3.1 . Воно зводиться до з'ясування різниці між T_1, T_2, \dots, T_k . Гіпотеза про однорідність даних означає рівності $\mu_1 = \mu_2 = \dots = \mu_k$, т.е. $T_1 = T_2 = \dots = T_k = 0$. Альтернатива про впорядкованість ефектів обробки (про вплив фактора) перетворюється на $T_1 \leq T_2 \leq \dots \leq T_k$, а різниця між i -м і j -м рівнем фактора, природно, характеризується величиною $\mu_i - \mu_j = T_i - T_j$. Отже, нехай $H_0: \mu_1 = \mu_2 = \dots = \mu_k$ та виконуються умови (3.2.2).

Тоді при кожному рівні фактора величина X матиме нормальний розподіл з тим самим математичним очікуванням і однією і тією ж дисперсією, що дорівнює D , тобто, перехід від одного рівня фактора до іншого не вносить жодних змін: існує одна генеральна сукупність, і результати спостережень, наведені в табл. 3.1 - це вибірка обсягу n із цієї генеральної сукупності. Оскільки спостереження незалежні і проведені в однакових умовах, то незмішана оцінка загальної дисперсії D і $\epsilon \tilde{D} = \frac{1}{n-1} \sum_{j=1}^k \sum_{i=1}^{n_j} (x_{ij} - \bar{x})^2$. Таким чином, $M\left[\frac{1}{n-1} \sum_{j=1}^k \sum_{i=1}^{n_j} (x_{ij} - \bar{x})^2\right] = M\left(\frac{1}{n-1} Q\right) = D$. Переконаємось тепер у тому, що при виконанні умов гіпотези H_0 та умов (3.2.2) величина $\frac{1}{k-1} Q_1$ також є незміщеною оцінкою загальної дисперсії. $M\left(\frac{Q_1}{k-1}\right) = \frac{1}{k-1} M(Q_1) = \frac{1}{k-1} M(Q - Q_2) = \frac{1}{k-1} [M(Q) - M(Q_2)] = \frac{1}{k-1} [(n-1)D - -(n-k)D] = D$.

Отже, є три незміщені оцінки однієї і тієї ж дисперсії D , причому оцінки $Q_2/(n-k)$ є незміщеною оцінкою у будь-якому випадку, а оцінки $Q/(n-1)$ та $Q_1/(k-1)$ - тільки при виконанні гіпотези H_0 , тобто, тільки

тоді, коли досліджуваний чинник впливає на результат. Перевірка гіпотези H_0 про рівність груп математичних очікувань ґрунтується на порівнянні дисперсій $s_1^2 = Q_1/(k-1)$ та $s_2^2 = Q_2/(n-k)$. Визначимо спочатку випадкову величину з χ^2 -розподілом. Оскільки вірні припущення (6.2.2), а s_1^2 і s_2^2 є незміщеними оцінками дисперсії D , тоді

$$\frac{(k-1)s_1^2}{D} \in \chi_{k-1}^2 \text{ и } \frac{(n-k)s_2^2}{D} \in \chi_{n-k}^2. \quad (3.2.6)$$

Величини χ_{k-1}^2 і χ_{n-k}^2 незалежні через незалежність s_1^2 і s_2^2 . Тоді відношення

$$\frac{\chi_{k-1}^2/(k-1)}{\chi_{n-k}^2/(n-k)} = \frac{s_1^2}{s_2^2} = \frac{Q_1/(k-1)}{Q_2/(n-k)}$$

мають F -розподіл із числом ступенів свободи $k-1$ і $n-k$,

$$\frac{Q_1/(k-1)}{Q_2/(n-k)} \in F_{k-1, n-k} \quad (3.2.7)$$

Тоді $H_0: T_1 = T_2 = \dots = T_k = 0$

$H_1: T_i \neq T_j, i \neq j, 1 \leq i \leq k, 1 \leq j \leq k$

Гіпотеза H_0 приймається лише на рівні значимості α , якщо вибіркове значення статистики $F_{k-1, n-k}$ менше $F_{1-\alpha/2}(k-1, n-k)$ або більше $F_{\alpha/2}(k-1, n-k)$. В цьому випадку \bar{x} і $s_2^2 = Q_2/(n-k)$ є незміщеними оцінками математичного очікування та дисперсії вибірки (спостережень x_{ij}). Оцінка s_2^2 не залежить від виду нульової гіпотези H_0 . Оцінка s_1^2 суттєво використовує основне припущення гіпотези суттєво використовує основне припущення гіпотези H_0 . Вона дає близький K результат тільки у тому випадку, коли гіпотеза H_0 вірна. При порушенні H_0 оцінка s_1^2 має тенденцію до зростання, тим більше, чим більше відхилення від H_0 . Зіставляючи один з одним ці дві оцінки, ми можемо укладати, що H_0 слід відкинути, якщо вони виявляються значно різні. Практично обчислення Q, Q_1 та Q_2 зручно проводити за формулами: $Q = A - C, Q_1 = B - C, Q_2 = A - B$, де

$$A = \sum_{j=1}^k \sum_{i=1}^{n_j} x_{ij}^2, B = \sum_{j=1}^k \frac{1}{n_j} (\sum_{i=1}^{n_j} x_{ij})^2 = \sum_{j=1}^k \frac{1}{n_j} x_{\cdot j}^{-2},$$

$$C = \frac{1}{n} (\sum_{j=1}^k \sum_{i=1}^{n_j} x_{ij})^2 = \frac{1}{n} \bar{x}^2. \quad (3.2.8)$$

Для контролю правильності розрахунків використовують тотожність $Q = Q_1 + Q_2$. Якщо гіпотеза про рівність середніх відхиляється, потрібно

визначити, які саме групи середніх мають значну різницю. Для цього часто використовуються не самі оцінки величин T_j , деякі лінійні комбінації цих величин. Для визначення вводиться поняття контрасту. Контрастом параметрів T у моделі адитивного впливу фактора на відгук називається величина $L_k = \sum_{j=1}^k c_j T_j$, де $c_j, j = 1, 2, \dots, k$ - константи, що однозначно визначаються з формулювань основної та альтернативної гіпотез, причому $\sum_{j=1}^k c_j = 0$. Зрозуміло, що різниця $T_i - T_j$ є найпростішим прикладом контрасту, коли $c_i = 1, c_j = -1, c_l = 0$ при всіх $l \neq i$ та $l \neq j$. Оцінки контрастів такі:

$$M(L_k) = \hat{L}_k = \sum_{j=1}^k c_j \bar{x}_j, D(L_k) = \hat{D}_k = \frac{Q_2}{n-k} \sum_{j=1}^k \frac{c_j^2}{n_j} \quad (3.2.9)$$

Межа довірчого інтервалу для L_k має вид

$$\hat{L}_k \pm \sqrt{\hat{D}_k \cdot \sqrt{(k-1)F_{1-\alpha}(k-1, n-k)}}. \quad (3.2.10)$$

Приклад. Передбачається, що вибірки отримані з нормально розподілених генеральних сукупностей з рівними дисперсіями. Треба перевірити гіпотезу про рівність середніх. Якщо H_0 приймається, знайти незміщені оцінки середнього та дисперсії. У разі відхилення H_0 провести попарне порівняння середніх, використовуючи метод лінійних контрастів. Прийняти $\alpha = 0.05$.

Номер вибірки	Спостереження				
1	6	5	12	9	10
2	14	11	5	6	—
3	12	4	7	—	—

Найшвидше завдання вирішується за формулами (3.2.8). Для цього продовжимо вихідну таблицю ще кількома стовпцями.

Номер вибірки	Спостереження					n_j	$\sum_{i=1}^{n_j} x_{ij}$	$\sum_{i=1}^{n_j} x_{ij}^2$	$\frac{1}{n_j} \left(\sum_{i=1}^{n_j} x_{ij} \right)^2$
	6	5	12	9	10				
1	6	5	12	9	10	5	42	386	352.8
2	14	11	5	6	-	4	36	378	324
3	12	4	7	-	-	3	23	209	176.3

Тоді $n = 12, k = 3, A = 973, B = 851.133, C = 850.083$. Звідки $Q = A - C = 122.917, Q_1 = B - C = 1.05, Q_2 = A - B = 121.867$. Перевіримо справедливість розрахунків: $Q_1 + Q_2 = 1.05 + 121.867 = 122.917 = Q$.

$$H_0: T_1 = T_2 = T_3 = 0, H_1: T_i \neq T_j, i \neq j, 1 \leq i \leq 3, 1 \leq j \leq 3.$$

$$F_{\text{теор.}}^{(1)} = F_{\alpha/2}(k-1, n-k) = F_{0.025}(3-1, 12-3) = F_{0.025}(2, 9),$$

$$F_{\text{теор.}}^{(2)} = F_{1-\alpha/2}(k-1, n-k) = F_{0.975}(2, 9). F_{\text{выб.}} = \frac{Q_1/2}{Q_2/9} = \frac{0.525}{13.541} = 0.039.$$

Таблиці F - розподілів з квантилями для малих ймовірностей дуже рідкісні, тому квантили $F_{\text{теор.}}^{(1)}$ и $F_{\text{теор.}}^{(2)}$ були обчислені в математичному пакеті Mathcad. $F_{\text{теор.}}^{(1)} = 0.025, F_{\text{теор.}}^{(2)} = 5.715$. Для гіпотези про рівність середніх

$$s_2^2 = Q_2/(n-k) = 13.541.$$

3.3. Ранговий однофакторний аналіз

В останні роки активно розвивалися методи математичної статистики, для яких не потрібні жодні припущення про розподіл, за винятком припущення, що це розподіл безперервний. Ці методи називаються непараметричними чи вільними від розподілу.

Якщо відсутня інформація про розподіл спостережень, то безпосередньо використовувати для перевірки нульової гіпотези кількісні значення спостережень x_{ij} стає важко. У цьому випадку найпростіше спиратися у своїх висновках тільки на ставленні «більше менше» між спостереженнями, оскільки вони не залежать від розподілу спостережень.

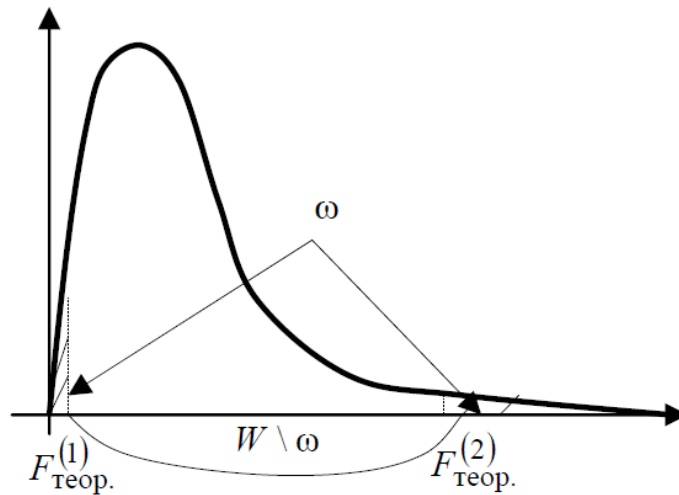


Рис. 3.1 Критична область статистики для гіпотези про рівності середніх

В цьому випадку вся корисна інформація міститься у рангах. Отримаємо вихідної вибірки варіаційний ряд, тобто, розташуємо вибіркові значення у порядку зростання. Кожній величині з цього ряду можна порівняти її ранг, що дорівнює порядковому номеру величини в загальному варіаційному ряду. Зауважимо, що й спостереження однорідні, тобто. вся вибірка взята з однієї і тієї ж генеральної сукупності, будь-який розподіл рангів рівномірний, а загальна кількість способів угруповання рангів, наприклад, при двох підвиборках обсягів n і m дорівнює є рівними числу способів, якими можна витягти m предметів з $N = n + m$, C_{n+m}^m .

Відповідні критерії для перевірки нульової гіпотези називають рангові, вони придатні для будь-яких безперервних розподілів спостережень. Більше того, вони підходять і тоді, коли виміри X_y зроблено в порядковій шкалі, наприклад, є тестовими балами або експертними оцінками. Основні формули рангового однофакторного аналізу виведено у припущенні, що з чисел X_y немає збігів. При наявності збігів використовуються середні ранги, при цьому теоретична схема діє як наближена, і надійність її висновків знижується. Для обліку збігів запроваджуються спеціальні поправки. Припишемо кожному спостереженню X_y у загальному варіаційному ряду його ранг r_{ij} . Тоді табл. 3.1 перетворюється на табл. 3.2

Таблиця 3.2

Обробки (відповідають рівням факторів)	1	3	...	k
Ранги результатів спостережень	r_{11}	r_{12}	...	r_{1k}
	r_{21}	r_{22}	...	r_{2k}

	$r_{n1 1}$	$r_{n2 2}$...	$r_{nk k}$
--	------------	------------	-----	------------

Загальна методика перевірки статистичних гіпотез радить скласти деяку статистику, а саме, функцію від рангів r_{ij} , яка лягла б в основу критерію перевірки гіпотези. Основна вимога до цієї статистики така: її розподіл за гіпотези H_0 має помітно відрізнятись від її розподілу при альтернативах. Наприклад, часто в якості статистики беруть суму рангів однієї підвиборки. Раціональність такої процедури полягає в тому, що якщо один розподіл (однієї підвиборки) зміщений відносно іншого, то це має проявитися в тому, що маленькі ранги повинні в основному відповідати одній підвиборці, а великі - іншій, внаслідок чого відповідні суми рангів повинні бути маленькими або більшими залежно від того, яка альтернатива має місце.

3.4. Критерій Краскела - Уолліса (H-критерій)

Якщо не можна сказати щось визначене про альтернативи H_0 , можна скористатися для її перевірки вільним від розподілу H-критерієм. Він був запропонований Краскелом та Уоллісом і є узагальненням двовибіркового критерію Вілкоксона.

Побудуємо загальний варіаційний ряд, що містить $n_1 + n_2 + \dots + n_k = n$ елементів, де n_j - кількість спостережень у j -й підвиборці (на j -му рівні фактора). $R_{.j}$ - сума рангів кожної обробки, а саме, кожного стовпця табл. 6, а \bar{R}_j - середнє арифметичне цих рангів. Формули для їх знаходження наступні:

$$R_{.j} = \sum_{i=1}^{n_j} r_{ij}, \quad \bar{R}_j = \frac{1}{n_j} \sum_{i=1}^{n_j} r_{ij} = \frac{1}{n_j} R_{.j}$$

$$M(\bar{R}_j) = M\left(\frac{1}{n_j} R_{.j}\right) = \frac{1}{n} \sum_{k=1}^n k = \frac{1}{n} \cdot \frac{n(n+1)}{2} = \frac{n+1}{2}$$

$M(\bar{R}_j)$ як середнє арифметичне всіх рангів від одиниці до n , а $1 + 2 + \dots + n = \frac{n(n+1)}{2}$. Звідки $M(R_{.j}) = \frac{n_j(n+1)}{2}$.

Якщо між стовпцями немає систематичних відмінностей, середні ранги \bar{R}_j не повинні значно відрізнятись від середнього рангу, розрахованого по всій сукупності чисел r_{ij} . Математичне очікування середнього рангу, дорівнює $M(\bar{R}) = (n+1)/2$. Більш складнішим чином розраховується

дисперсія. Для \bar{R}_j вона дорівнює $D(\bar{R}_j) = \frac{(n+1)(n-n_j)}{12n_j}$. Якщо $n \rightarrow \infty$, то дріб $\frac{\bar{R}_j - M(\bar{R}_j)}{\sqrt{D(\bar{R}_j)}}$ має межі як стандартний нормальний розподіл, який

використовували Краскел і Уолліс для побудови статистичного критерію, який вони позначили літерою H і яка має наступний вигляд

$$H = \sum_{j=1}^k \frac{\left[\bar{R}_j - \frac{n+1}{2}\right]^2}{\frac{(n+1)(n-n_j)}{12n_j}} \left(1 - \frac{n_j}{n}\right). \quad (3.4.1)$$

Краскел и Уолліс показали, що асимптотично статистика H має χ^2 -розподіл з $(k-1)$ ступеня свободи, де k - кількість підвибірки (рівні факторів). Часто статистика H записується в одному з наступних двох типів:

$$H = \frac{12}{n(n+1)} \sum_{j=1}^k n_j \left(\bar{R}_j - \frac{n+1}{2}\right)^2, \quad (3.4.2)$$

або

$$H = \frac{12}{n(n+1)} \sum_{j=1}^k \frac{R_{.j}^2}{n_j} - 3(n+1). \quad (3.4.3)$$

Якщо два або більше спостережень збігаються, то найкраща процедура полягає в тому, що спостереження збігу потрібно віднести до того ж рангу, що є рівним середньому арифметичному рангу, ці спостереження повинні бути отриманими, якби вони не збігалися. Ця операція залишає без змін кількість рангів та математичне очікування кількості рангів. Але формула обчислення змін дисперсії, оскільки дисперсія статистики \bar{R}_j залежить від суми квадратів рангів, які зміняться від такої заміни. Вид статистики H також зміниться, тому вона виправляється відповідною корекцією.

Якщо є багато збігів, рекомендується використовувати модифіковану форму статистики H' :

$$H' = \frac{H}{1 - \sum_{j=1}^p \frac{T_j}{n^3 - n}} \quad (3.4.4)$$

де p - число груп спостережень, що співпадають, $T_j = (t_j^3 - t_j)$, t_j - число груп спостережень, що співпадають в j групі.

Приклад. Кислота безперервно концентрується на деякому типі обладнання, в результаті чого частина обладнання іржавіє і з часом руйнується. Втрати металу (в сотнях тонн) за період від установки обладнання до моменту руйнування деякої його частини зафіксовані в таблиці для трьох ливарних майстерень А, В і С. Перевірити нульову гіпотезу, за якою середня тривалість служби металу така сама для всіх трьох майстерень.

Майстерня	Втрати металу									
	А	84	60	40	47	34	46			
В	67	92	95	40	98	60	59	108	86	117
С	46	93	100	92	92					

Жодних правдоподібних припущень про імовірнісний розподіл втрат металу в цьому завданні зробити не можна. Скористаємося ранговим методом Краскела - Уолліса. Слід зазначити, що величини, наведені у вихідній таблиці, мають сенс самі по собі, ніж у порівнянні з іншими величинами. Хоча при переході від величин втрат металу до їх рангів відбувається певна втрата інформації, але така інформація, по-перше, не така значна, по-друге, компенсується тим, що від невідомого закону розподілу величин x_{ij} ми переходимо до величин r_{ij} , розподіл яких при гіпотезі H_0 відомо.

Основна гіпотеза H_0 визначає постійний термін служби металу у всіх трьох майстерень, тобто, постійний рівень втрат, отже, визначає однорідність вихідних вибірок. Позначимо втрати металу в j -й групі через μ_j . Тоді

$$H_0 : \mu_1 = \mu_2 = \mu_3,$$

$$H_1 : \mu_i \neq \mu_j, i \neq j, 1 \leq i \leq 3, 1 \leq j \leq 3.$$

Спочатку отримаємо варіаційний ряд і припишемо кожному спостереженню його ранг. У зв'язку з наявністю в таблиці збігів користуватимемося середніми рангами.

Спостереження	34	40	40	46	46	47	59	60	60	67	84
Номер спостережень у варіаційному рядку	1	2	3	4	5	6	7	8	9	10	11
Ранг	1	2.5	2.5	4.5	4.5	6	7	8.5	8.5	10	11

Спостереження	86	92	92	92	93	95	98	100	108	117	
Номер спостережень у варіаційному рядку	12	13	14	15	16	17	18	19	20	21	
Ранг	12	14	14	14	16	17	18	19	20	21	

Загальна кількість спостережень $n=21$. Складемо тепер із вихідної таблиці таблицю рангів і доповнимо її двома стовпцями, що містять $R_{.j}$ та \bar{R}_j .

Майстерня	Ранги втрат металу										$R_{.j}$	\bar{R}_j
A	11	8.5	2.5	6	1	4.5					33.5	5.583
B	10	14	17	2.5	18	8.5	7	20	12	21	130	13.00
C	4.5	16	19	14	14						67.5	13.50

Для обчислення статистики Фаркела – Уолліса зручніше використовувати формулу (3.4.3). Тоді $H = \frac{12}{21 \cdot 22} \left(\frac{33.5^2}{6} + \frac{130^2}{10} + \frac{67.5^2}{5} \right) - 3 \cdot 22 = 6.423$. Оскільки є збіги, скоригуємо статистику H .

У нашому випадку є чотири групи спостережень що збігаються: 40,40; 46,46; 60,60; 92,92,92. Обчислимо поправки за такою формулою (6.4.4.): $T_1 = (2^3 - 2) = 6, T_2 = 6, T_3 = 6, T_4 = (3^3 - 3) = 24$. Знаменник дробу у виразі для H' дорівнює: $1 - \sum_{j=1}^4 \frac{T_j}{(21^3 - 21)} = 1 - \frac{6+6+6+24}{9240} = 0.995$.

$$\text{Тоді } H' = \frac{H}{0.995} = 6.455.$$

Як було зазначено, величина H асимптотично розподілена згідно із законом χ^2 з числом ступенів свободи $k - 1$, тобто в даному випадку рівним двом. Знайдемо квантиль χ^2 - розподіли: $\chi_{0.95,2}^2 = 5.99$. Отже, з використанням правостороннього критерію $H' > \chi_{0.95,2}^2$, тобто $H' \in \omega$, і гіпотеза H_0 має бути відкинута з рівнем значущості $\alpha = 0.05$.

3.5. Однофакторний ранговий та дисперсійний аналіз у статистичному пакеті STATGRAPHICS

Дисперсійний аналіз застосовується для виявлення впливу виділеного набору факторів на результативну ознаку. Загальна ідея дисперсійного

аналізу полягає у розкладанні загальної дисперсії результативної ознаки на частини, зумовлені впливом контрольованих факторів, та залишкову дисперсію, що викликається випадковими обставинами.

Відомо багато моделей дисперсійного аналізу. Вони класифікуються, з одного боку, за математичною природою факторів (детерміновані, випадкові та змішані), з іншого боку – за кількістю контрольованих факторів (однофакторні та багатофакторні моделі). За способом організації вихідних даних серед моделей дисперсійного аналізу виділяють повні та неповні k -факторні плани, повні та неповні блокові плани та рандомізовані блокові плани. У STATGRAPHICS реалізовані всі перелічені моделі дисперсійного аналізу.

Розв'яжемо в пакеті STATGRAPHICS наступне завдання однофакторного дисперсійного аналізу.

Час хімічної реакції при різному вмісті каталізатора розподілився наступним чином (у секундах):

Зміст каталізатор., %	Номер експерименту												Сума
	1	2	3	4	5	6	7	8	9	10	11	12	
5	5.9	6.0	7.0	6.5	5.5	7.0	8.1	7.5	6.2	6.4	7.1	6.9	80.1
10	4.0	5.1	6.2	5.3	4.5	4.4	5.3	5.4	5.6	5.2	-	-	51.0
15	8.2	6.8	8.0	7.5	7.0	7.2	7.9	8.1	8.5	7.8	8.1	-	85.1

Вважаючи, що вибірки отримані з нормально розподілених генеральних сукупностей з рівними дисперсіями, перевірити нульову гіпотезу H_0 про рівність середніх. Прийняти $\alpha = 0.1$. Розкриємо електронну таблицю SRATGRAPHICS і введемо до неї значення спостережень (величини x_{ij} - значення результативної ознаки) та значення градацій фактора (можна вводити закодовані значення, наприклад, 1, 2, 3), оскільки це показано на рис. 6.2. Доступ до процедур аналізу здійснюється з пункту меню Compare → Analysis of Variance → One-Way ANOVA (однофакторний дисперсійний аналіз). Скорочення ANOVA походить від виразу "Analysis of variance". У вітчизняній літературі замість терміна "аналіз варіацій" використовується термін "дисперсійний аналіз".

	observ	factor	Col_3	Col_4
1	5,9	5		
2	6	5		
...		
11	7,1	5		
12	6,9	5		
13	4	10		
14	5,1	10		
...		
21	5,6	10		
22	5,2	10		
23	8,2	15		
24	6,8	15		
...		
32	7,8	15		
33	8,1	15		

Рис. 3.2. Електронна таблиця STATGRAPHICS

Відразу з'являється вікно однофакторного дисперсійного аналізу (рис. 6.3). У комірці Dependent Variable (Залежна змінна) введемо Observ, а комірці Factor (Фактор) ім'я Factor. Натисніть ОК. На екрані з'явиться виведення однофакторного дисперсійного аналізу, в якому підтверджується, що введено 33 спостереження, для яких зафіксовано три рівні фактора. Внизу під цими включено повідомлення StatAdvisor з рекомендаціями щодо подальшого аналізу.

У додатковому меню, що виникло, відкриємо вікно Tabular Options і відзначимо всі процедури для цього меню. Вкажемо призначення всіх процедур, що входять до цього меню. Analysis Summary (Виведення аналізу). Звіт цього вікна вже відчинений у вигляді заставки. На ній вказані найзагальніші відомості про вибірку.

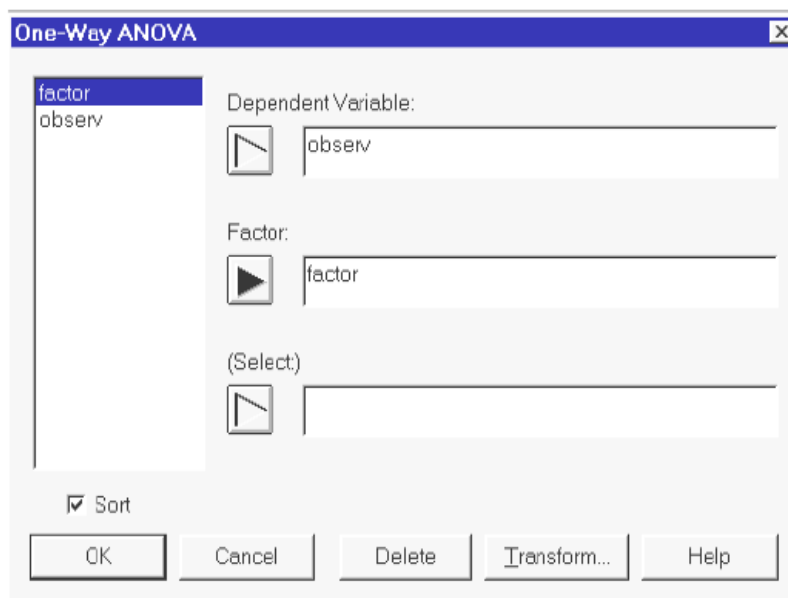


Рис. 3.3. Вікно для завдання змінних

Summary Statistics (Опис даних). Зміст числової інформації, що розміщена в цьому звіті, зрозуміло з контексту. Спочатку аналізується

інформація про фактор. Наводяться значення рівнів фактора, кількість спостережень на кожному рівні, середні, дисперсії та стандартні відхилення на кожному рівні та по всій вибірці. Потім наводяться найменші та найбільші значення членів вибірки, їх стандартні асиметрії та ексцеси, нарешті, в останній таблиці вміщені сумарні значення спостережень за факторами та загалом за вибіркою.

ANOVA Table (Таблиця дисперсійного аналізу). Призначення цієї таблиці – дати відповідь питання наявності значного впливу рівнів чинника досліджуваного відгук, тобто. на наявність ефектів обробки. У першій колонці Source (Джерело варіації) вказано дві частини, на які розкладається загальна дисперсія за формулою (3.2.5) Between groups (Між групами) та Within groups (Всередині груп). Далі наводиться загальна дисперсія Total (corr.) (Разом (скориговане значення)).

Другий стовпець містить суму квадратів між групами, всередині груп та загальну, тобто величини Q_1 , Q_2 і Q , третій – відповідне число ступенів свободи. У четвертому стовпці знаходяться значення дисперсій: між групами величина s_1^2 всередині груп величина s_2^2 . У стовпці F-ratio виводиться значення F -статистики, нарешті, стовпець p-Value містить рівень значущості цієї статистики (рис. 3.4).

Means table (Таблиця середніх). Деякі дані з цієї таблиці вже були наведені в Summary Statistics. Стовпчик Std. error (pooled s) (Об'єднана стандартна помилка) містить s_2 . У двох останніх стовпцях зазначеної таблиці знаходяться межі довірчих інтервалів для середніх із третього стовпця. Звернемо увагу на те, що в таблиці середніх наведено довірчі 95% інтервали, побудовані за методикою LSD (рис. 3.5). Натискання правою кнопкою миші в полі заставки Table of Means відкриє наступне додаткове меню (рис. 3.6), в якому задаються різні способи побудови довірчих інтервалів. Відомості про методи побудови стандартних довірчих інтервалів (Confidence Interval) можна знайти у [19].

Analysis Summary

Dependent variable: observ
Factor: factor

Number of observations: 33
Number of levels: 3

Summary Statistics for observ

factor	Count	Average
5	12	6,675
10	10	5,1
15	11	7,73636
Total	33	6,55152

factor	Variance	Standard deviation
5	0,538409	0,733764
10	0,411111	0,641179
15	0,292545	0,540875
Total	1,53883	1,24049

factor	Minimum	Maximum
5	5,5	8,1
10	4,0	6,2
15	6,8	8,5
Total	4,0	8,5

factor	Std. skewness	Std. kurtosis
5	0,423072	-0,0936714
10	-0,265288	0,0096087
15	-0,75687	-0,536064
Total	-0,646217	-1,08188

factor	Sum
5	80,1
10	51,0
15	85,1
Total	216,2

ANOVA Table for observ by factor

Analysis of Variance					
Source	Sum of Squares	Df	Mean Square	F-Ratio	P-Value
Between groups	36,6945	2	18,3472	43,87	0,0000
Within groups	12,548	30	0,418265		
Total (Corr.)	49,2424	32			

Рис. 3.4. Результати однофакторного дисперсійного аналізу

Table of Means for observ by factor with 95,0 percent LSD intervals					
factor	Count	Mean	Std. error (pooled s)	Lower limit	Upper limit
5	12	6,675	0,186696	6,40539	6,94461
10	10	5,1	0,204515	4,80466	5,39534
15	11	7,73636	0,194998	7,45477	8,01796
Total	33	6,55152			

Рис. 3.5. Таблиця середніх

Multiple Range Tests (Множинні порівняння) видає результати аналізу множинних порівнянь середніх (рис. 3.7). У стовпці Homogeneous Groups (Однорідні групи) вертикальними стовпцями із символів зірочок виділено можливі однорідні групи спостережень. У випадку розглянутого приклада таких груп три і кожна з яких відповідає одному з трьох різних рівнів чинника. Отже, всі групи неоднорідні, і об'єднати в одну загальну групу не можна.

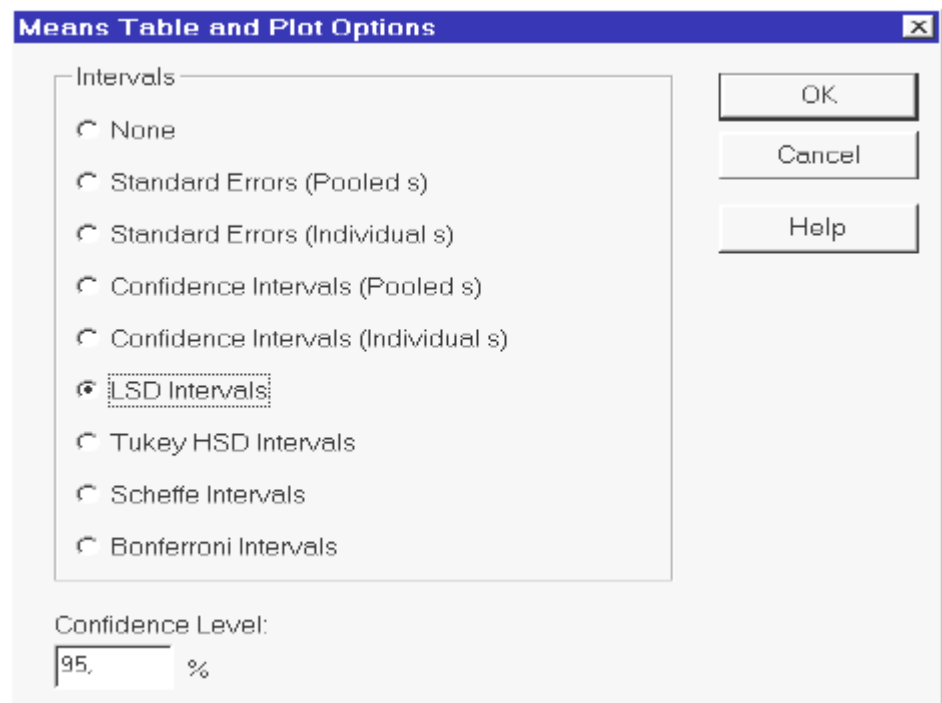


Рис. 3.6. Вікно для завдання різних способів побудови довірчих інтервалів

Multiple Range Tests for observ by factor

Method: 95,0 percent LSD			
factor	Count	Mean	Homogeneous Groups
10	10	5,1	X
5	12	6,675	X
15	11	7,73636	X
Contrast			Difference +/- Limits
5 - 10			*1,575 0,565537
5 - 15			*-1,06136 0,551337
10 - 15			*-2,63636 0,577104

Рис. 3.7. Однорідні групи спостережень та контрасти

Далі у таблиці наводяться значення лінійних контрастів, обчислених за формулою $L_k = \sum_{j=1}^k c_j T_j$. В нашому випадку $H_0: m_1 = m_2 = m_3$, де m_i - середня i -тої підвиборки (рівня обробки). Тоді $L_{k_1} = m_1 - m_2, c_1 = 1, c_2 = -1, c_3 = 0, L_{k_2} = m_1 - m_3, c_1 = 1, c_2 = 0, c_3 = -1, L_{k_3} = m_2 - m_3, c_1 = 0, c_2 = 1, c_3 = -1, \bar{x}_1 = 6.675, \bar{x}_2 = 5.1, \bar{x}_3 = 7.73636$.

Оцінки та дисперсії лінійних контрастів обчислюються за формулою (3.2.9):

$$\begin{aligned} \hat{L}_{k_1} &= c_1 \bar{x}_1 + c_2 \bar{x}_2 = 6.675 - 5.1 = 1.575, \hat{L}_{k_2} = c_1 \bar{x}_1 + c_3 \bar{x}_3 = \\ &= 6.675 - 7.73636 = -1.06136, \bar{L}_{k_3} = c_2 \bar{x}_2 + c_3 \bar{x}_3 = 5.1 - 7.73636 \\ &= -2.63636 \end{aligned}$$

і так далі. Нарешті, у стовпці під заголовком +/- Limits наведені межі довірчого LSD інтервалу для лінійних контрастів, обчислені за формулою, аналогічною формулою (3.2.10).

Після натискання правою кнопкою миші в полі заставки Multiple Range Test з'являється додаткове меню, подібне до меню Table of Means, в якому можна задати різні способи побудови довірчих інтервалів.

Variance Check (Тести дисперсії). Ця процедура включає результати трьох статистичних критеріїв Кокрена (англійського математика Вільяма Геммела Кохрена), Бартлетта і Хартлі для порівняння розкидів спостережень на різних рівнях фактора (рис. 3.8). Критерії Кокрена та Бартлетта перевіряють на однорідність ряд дисперсій, тобто нульову гіпотезу виду $H_0: D_1 = D_2 = \dots = D_k$. В даному випадку D_i - дисперсія відповідної підвиборки на i -му рівні фактора. За цими двома критеріями, крім значень статистик критеріїв, наводяться також значення мінімальних рівнів значущості. Слід зазначити, що критерії Кокрена і Бартлетта дуже чутливі до відхилення моделі спостережень від нормальності, у

інтерпретації результатів цих критеріїв потрібна певна обережність. Інформацію про критерії Кокрена, Бартлетта та Хартлі можна знайти в [1, 2, 4, 8].

Kruskal – Wallis Tests (Ранговий однофакторний аналіз Краскела – Уолліса) досліджує ефект дії одного фактора класифікації для збалансованого чи незбалансованого плану.

У колонці factor стоять мітки відповідних способів обробки (факторів), в колонці Sample Size (Розмір вибірки) – число спостережень на кожному рівні фактора. У колонці Average Rank (Середній ранг) – відповідна величина рангу для кожної групи. Під таблицею наведено значення для асимптотичної апроксимації, скоригованої для випадку збігаються спостережень за формулою (3.4.4), та мінімальний рівень значущості цієї статистики (p-Value).

Перелічені вище процедури лише трохи вирішують питання правомірності застосування дисперсійного аналізу до аналізованих даних. Це питання є визначальним і залежить від достовірності висновків, отриманих у результаті аналізу. Для більш детального розгляду вихідної вибірки в пакеті STATGRAPHICS можуть бути застосовані критерії χ^2 та Колмогорова для перевірки згоди з нормальним розподілом, візуальний (окомірний) метод перевірки нормальності, критерії асиметрії та ексцесу.

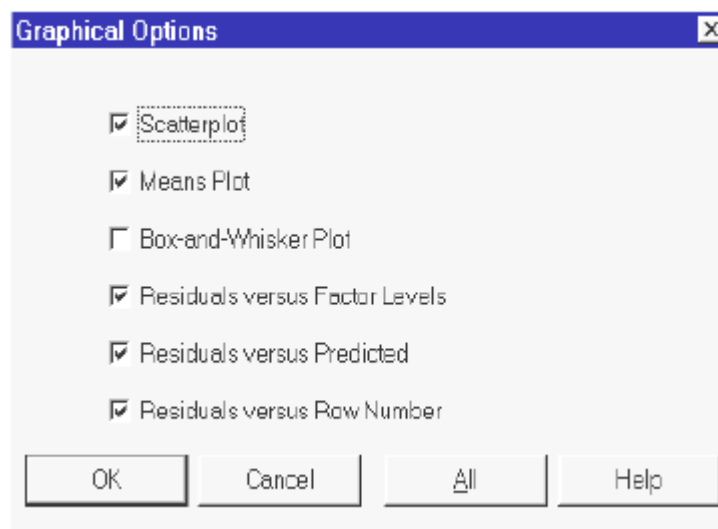


Рис. 3.9. Панель графічних процедур однофакторного дисперсійного аналізу

Розглянемо процедури вікна Graphics Options додаткового меню. У розділі One-Way ANOVA можна будувати різного типу графіки (рис. 6.9). Зазначимо всі пункти крім третього. В результаті отримаємо наступні графіки: Scatterplot – діаграма розсіювання вихідної вибірки. Means Plot реалізує графічне представлення даних таблиці, що видається процедурою

Table of Means (рис. 3.10). Процедури Residuals versus Factor Levels, Residuals versus Predicted та Residuals versus Row Number дають графіки залишків в одній із трьох можливих форм: залежно від рівня фактора, залежно від передбачених значень або в залежності від номера спостереження у векторі введення даних (рис. 3.11). Кожна з цих форм підкреслює свій аспект у можливих причинах порушення однорідності розподілу залишків.

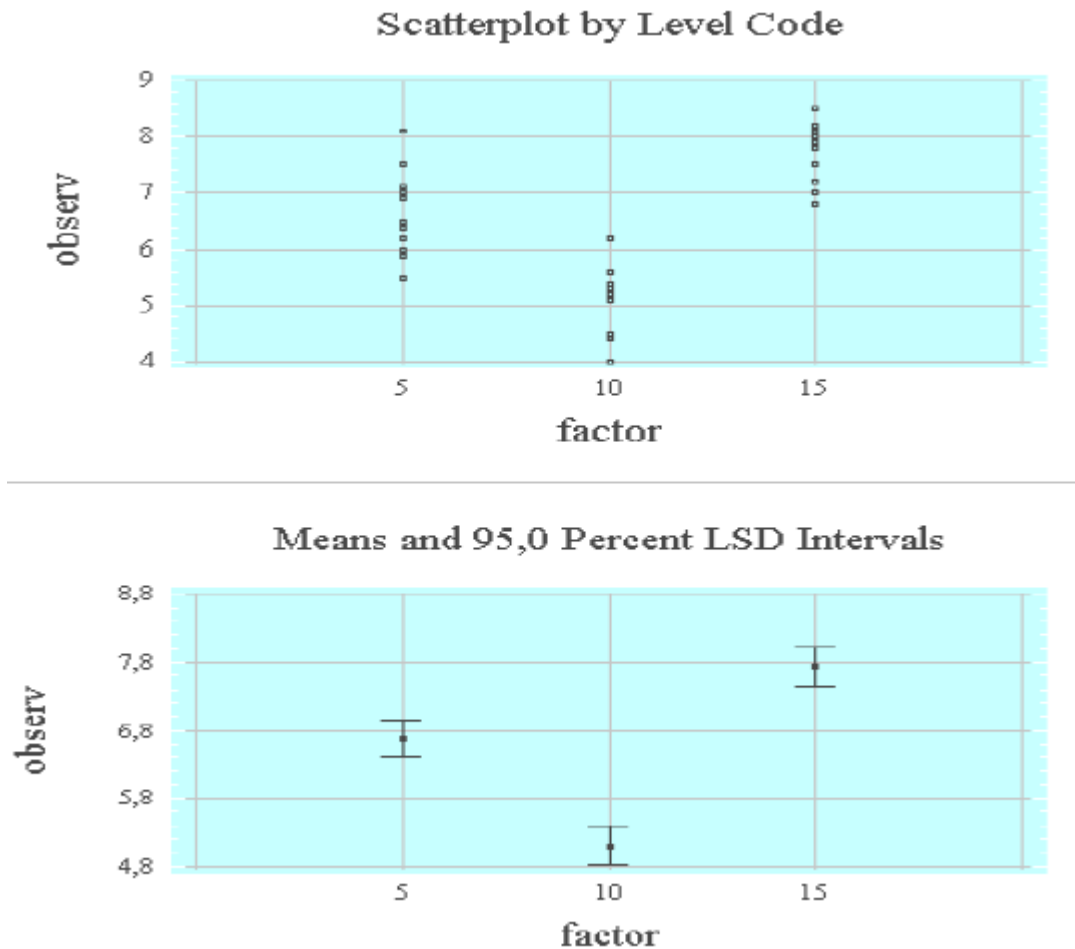


Рис. 3.10. Діаграма розсіювання вибірки та довірчі інтервали для середніх факторів

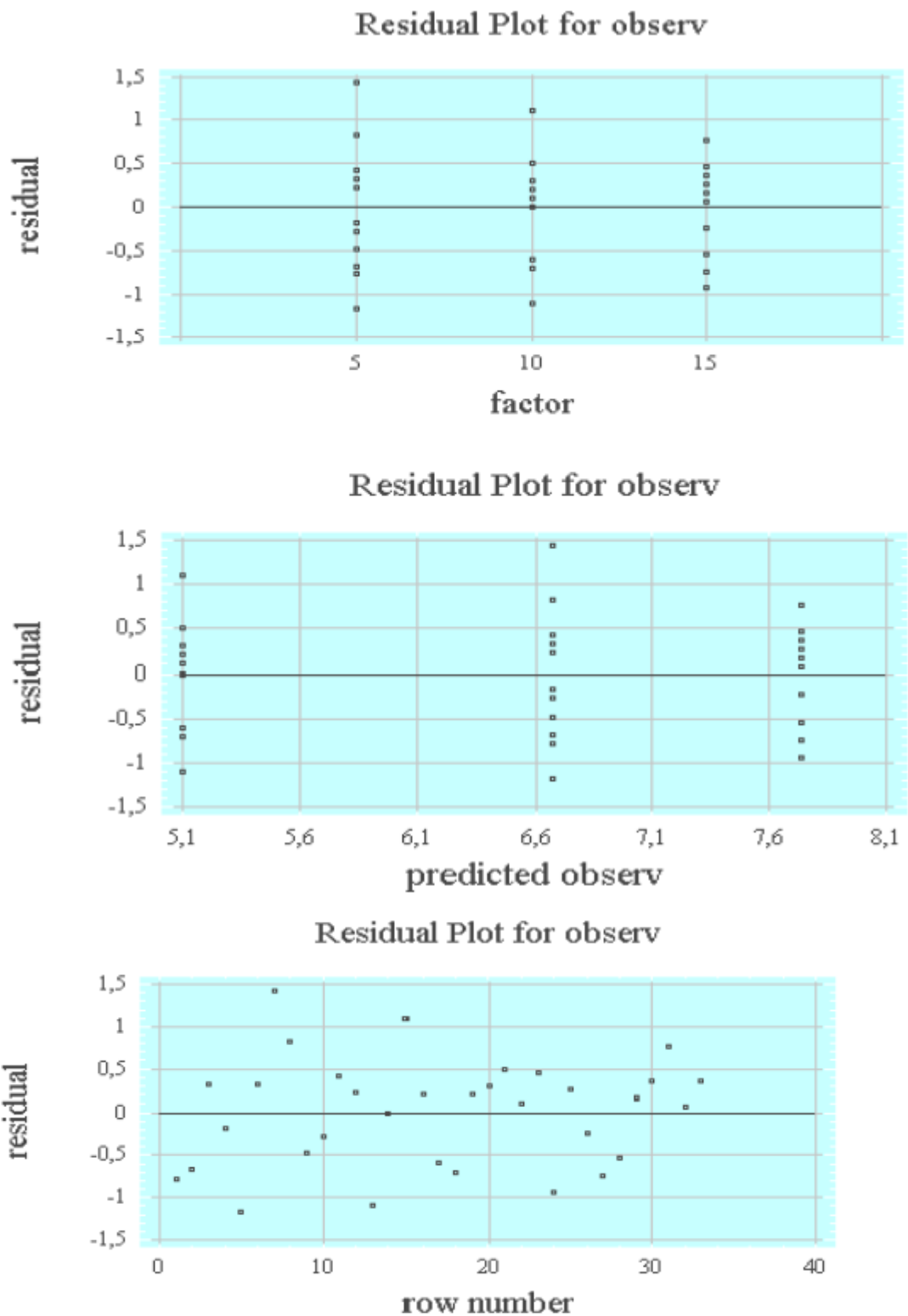


Рис. 3.11. Графіки залишків

Завдання. За допомогою розглянутих процедур пакету STATGRAPHICS вирішити завдання однофакторного дисперсійного аналізу. Скрізь рівень значності прийняти рівним 0.05. В кожному варіанті

завдання перевірити гіпотезу H_0 про рівність середніх. Якщо гіпотеза H_0 приймається, то знайти незміщені оцінки середнього та дисперсії. Якщо H_0 відхиляється, провести попарне порівняння середніх, використовуючи метод лінійних контрастів.

1. В трьох магазинах, що продають товари одного виду, дані товарообігу за вісім місяців роботи (у тис. грн.) склали наступне зведення:

Мага- зин	Місяць							
	1	2	3	4	5	6	7	8
I	19	23	26	18	20	20	18	35
II	20	20	32	27	40	23	22	18
III	16	15	18	26	19	17	19	18

2. У наступній таблиці наведено результати обстеження 60 працівників виробництва, у яких фіксувалося середнє годинникове вироблення в натуральних одиницях продукції. Взяти за фактор – стаж роботи.

Стаж	Вік		
	від 25 до 35 років	від 35 до 45 років	від 45 до 55 років
Від 1 до 4 років	19, 20, 20, 20, 22,	19, 20, 20, 23, 25,	18, 19, 20, 21, 23,
Від 4 до 7 років	30, 31, 32, 32, 34,	20, 29, 30, 31, 31,	19, 25, 25, 26, 26,
Від 7 до 10 років	35, 35, 39, 40, 41,	36, 40, 41, 42, 45,	24, 24, 24, 25, 25,
Більше 10 років	40, 40, 41, 41, 42,	28, 31, 35, 36, 40,	20, 24, 25, 31, 32.

3. Розв'язати завдання № 2 з тими самими даними, взявши за фактор, що впливає на середній годинний виробіток, вік працівника.

4.

Номер вибірки	Спостереження								
	1	2	3	4	5	6	7	8	9
1	12	4	7	8	5	9	6	-	-
2	14	11	5	6	3	-	-	-	-
3	6	5	12	9	10	7	11	4	5

5.

Номер вибірки	Спостереження					
	1	2	3	4	5	6
1	4	2	3	4	5	3

2	6	5	4	7	6	8
3	8	9	10	7	8	6

6.

Номер вибірки	Спостереження							
	1	2	3	4	5	6	7	8
1	9	8	8	7	9	-	-	-
2	8	11	8	9	10	12	-	-
3	9	10	7	11	8	10	12	13

7. Наведено дані про вміст імуноглобуліну IgA у сироватці крові (мг %) у хворих п'яти вікових груп:

Вікова група	Вміст IgA (мг %)											
	1	83	85	-	-	-	-	-	-	-	-	-
2	84	85	85	86	86	87	-	-	-	-	-	-
3	86	87	87	87	88	88	88	88	88	89	90	
4	89	90	90	91	-	-	-	-	-	-	-	-
5	90	92	-	-	-	-		-	-	-	-	-

8. На хімічному заводі розроблено два нові варіанти технологічного процесу. Щоб оцінити, як зміниться денна продуктивність під час переходу на роботу за новими варіантами технологічного процесу, завод протягом десяти днів працює за кожним варіантом, включаючи існуючий. Денна продуктивність заводу (в умовних одиницях) наводиться у таблиці:

Технологічний процес	Добова продуктивність									
	1	2	3	4	5	6	7	8	9	10
існуюча схема	46	48	73	52	72	44	66	46	60	48
Варіант I	74	82	64	72	84	68	76	88	70	60
Варіант II	52	63	72	64	48	70	78	68	79	54

9. Серед великої групи польових транзисторів із тижневим інтервалом було отримано три вибірки. Нижче наводяться результати вимірювання ємності затвора-стоку у цих транзисторів (пікофарадах):

Номер вибірки	Ємність (пФ)															
	1	2.8	3.2	2.9	3.5	3.3	3.7	3.9	3.1	3.2	3.1	3.4	3.0	3.6	3.1	3.2

2	3.1	3.2	3.3	3.4	3.7	3.4	3.0	3.1	2.9	3.5	3.2	3.2	-	-	-	-
3	3.6	2.8	3.0	3.2	3.0	3.7	3.2	3.2	3.6	3.4	3.1	3.2	-	-	-	-

10. З'ясувати чи залежить обсяг робіт, виконаних на будівництві за зміну, від працюючої бригади. Дані з чотирьох бригад наведені в наступній таблиці:

Номер бригади	Обсяг виконаної роботи					
1	140	144	142	145	146	140
2	150	149	152	150	-	-
3	150	149	146	147	148	150
4	150	155	154	152	157	-

11. Наведено два останніх десяткових знаки константи в експерименті визначення гравітаційної постійної $G = 6,67430(15) \cdot 10^{-11}$ ($\text{м}^3 \cdot \text{с}^{-2} \cdot \text{кг}^{-1}$). Наприклад, табличне значення 83 відповідає спостережене значення 6.683. Експеримент ставився з кулями, зробленими із золота, платини та скла.

Матеріал	Значення константи					
Золото	83	81	76	78	79	72
Платина	61	61	67	67	64	-
Стекло	78	71	75	72	74	-

12.

Номер вибірки	Спостереження									
1	92	78	60	67	53	66	-	-	-	-
2	83	96	98	60	99	78	77	103	93	107
3	66	97	100	96	96	-	-	-	-	-

13. Подано спроби довговічності електричних ламп, взятих із чотирьох партій.

Номер партії	Тривалість горіння в годинах							
1	1600	1610	1650	1680	1700	1700	1800	-
2	1580	1640	1640	1700	1750	-	-	-
3	1460	1550	1600	1620	1640	1660	1740	1820
4	1510	1520	1530	1570	1600	1680	-	-

14. Наведено зміни критерію чистоти поверхні металу для трьох

приладів.

Номер приладу	Відхилення від загальної медіани у сотих частках мікрона					
	1	-4	-2	-21	-4	-4
2	7	11	30	28	27	103
3	19	2	-13	-9	2	1

15. Результати 22 випробувань на чотирьох рівнях фактора:

Рівень фактора	Спостереження						
F ₁	1.38	1.45	1.38	1.42	1.42	1.44	1.39
F ₂	1.41	1.42	1.44	1.45	1.46	1.43	-
F ₃	1.32	1.33	1.34	1.31	1.35	-	-
F ₄	1.31	1.33	1.32	1.33	-	-	-

16. Проведено 22 випробування, результати яких представлені у таблиці.

Рівень фактора	Спостереження								
F ₁	30.56	32.66	34.78	35.50	36.63	40.20	42.28	41.76	35.17
F ₂	43.44	47.51	53.80	50.11	46.23	51.19	-	-	-
F ₃	31.36	36.20	36.38	42.20	35.13	39.93	34.72	-	-

17. Результати випробувань на трьох рівнях фактора наступні:

Рівень фактора	Спостереження											
F ₁	37	47	40	60	52	48	42	-	-	-	-	-
F ₂	60	86	67	92	90	95	98	103	89	91	95	97
F ₃	69	100	98	75	85	101	94	73	89	96	-	-

18. У наступній таблиці наведено рівні постачання сировини (в умовних одиницях) у серії з п'яти партій.

Партії	Рівень постачання сировини																			
1	62	66	64	64	63	62	64	64	66	64	66	63	65	63	63	63	61	56	64	65
2	66	65	65	66	67	66	69	70	68	69	63	65	64	65	64	-	-	-	-	-
3	62	64	62	62	65	64	65	62	62	63	64	-	-	-	-	-	-	-	-	-
4	65	64	63	62	65	63	64	63	-	-	-	-	-	-	-	-	-	-	-	-
5	65	64	67	62	65	62	64	64	64	65	-	-	-	-	-	-	-	-	-	-

19. Таблиця даних містить результати визначення октанового числа бензину, які отримані в чотирьох округах на північному сході США влітку 1953 року.

Ок-руг	Октанове число бензину												
	A	84.0	83.5	84.0	85.0	83.1	83.5	81.7	85.4	84.1	83.0	85.8	84.0
82.2		83.6	84.9	-	-	-	-	-	-	-	-	-	-
B	82.4	82.4	83.4	83.3	83.1	83.3	82.4	83.3	82.6	82.0	83.2	83.1	82.5
C	83.2	82.8	83.4	80.2	82.7	83.0	85.0	83.0	85.0	83.7	83.6	83.3	83.8
	85.1	83.1	84.2	80.6	82.3	-	-	-	-	-	-	-	-
D	80.2	82.9	84.6	84.2	82.8	83.0	82.9	83.4	83.1	83.5	83.6	86.7	82.6
	82.4	83.4	82.7	82.9	83.7	81.5	81.9	81.7	82.5	-	-	-	-

20. Наведено дві останні цифри чисел, що виражають швидкість світла, отримані Майкельсоном у його досвіді із шістьма круговими дзеркалами.

Номер дзеркала	Спостереження																			
	1	47	47	38	62	29	59	92	44	41	47	44	41	-	-	-	-	-	-	-
2	42	18	36	45	33	30	0	27	18	27	57	66	48	24	15	-	-	-	-	
3	3	39	27	67	48	15	3	7	27	27	42	37	69	24	63	15	30	27	42	60
4	6	21	27	33	9	24	6	39	42	18	12	63	-	-	-	-	-	-	-	
5	18	9	12	30	30	27	30	39	18	27	48	24	18	-	-	-	-	-	-	
6	30	21	33	18	12	33	24	23	57	39	44	33	30	24	24	30	-	-	-	-

21. Фруктовий сік зберігався протягом кількох місяців у цистернах чотирьох типів, після чого визначалося його якість виставленням чисельної оцінки. Нижче наведено результати випробувань.

Цистерна	Спостереження							
	A	6.14	5.72	6.90	5.80	6.23	6.06	5.42
B	6.55	6.29	7.40	6.40	6.28	6.26	6.22	6.76
C	5.54	5.61	6.60	5.70	5.31	5.58	5.57	5.84
D	4.81	5.09	6.61	5.03	5.15	5.05	5.77	6.17

22. Лікар рекомендував своїм пацієнтам, які скаржаться на зайву вагу, ліки А, В і С. При цьому він щоразу фіксував вагу пацієнта після

лікування в фунтах (1 фунт = 453.6 г), в результаті чого отримані наступні результати.

Ліки	Вага пацієнта											
А	147	183.5	150	167	180	216.5	127.5	222	132	167	221	203
В	180	161.5	157	155	146	131.5	163.3	160	162	225	159	-
С	216	172	140	154	161	-	-	-	-	-	-	-

23. Наступна таблиця містить спеціальні оцінки у балах, що відповідають одному з чотирьох експериментальних умов.

Умови	Оцінки									
1	0	1	3	3	5	10	13	17	26	
2	0	6	7	9	11	13	20	20	24	
3	0	5	8	9	11	13	16	17	20	
4	1	5	12	13	19	22	25	27	29	

24. Наведено вміст вологи (%) у зразках деякого продукту залежно від умов зберігання.

Умови зберігання	Вміст вологи (в %)														
1	7.8	7.7	7.4	7.9	8.3	8.2	8.0	7.6	7.4	7.7	8.4	8.3	8.5	8.3	8.2
2	5.4	5.3	5.2	5.5	5.6	7.4	7.3	7.5	7.1	7.0	6.9	-	-	-	-
3	8.1	8.0	7.9	8.2	8.3	6.4	6.3	6.5	-	-	-	-	-	-	-
4	7.9	8.0	7.8	8.1	7.9	9.5	9.6	9.4	10.1	10	9.9	-	-	-	-
5	7.1	6.9	7.0	7.3	7.2	-	-	-	-	-	-	-	-	-	-

25. У таблиці наведено середню кількість помилок при виконанні 12 різних завдань тваринами трьох видів.

Тварини	Середня кількість помилок											
Щури	1.5	1.1	1.8	1.9	4.3	2.0	8.4	6.6	2.4	6.5	2.6	6.5
Кролі	1.7	1.5	8.1	1.3	4.0	4.6	4.0	5.1	2.5	6.9	2.5	6.8
Кішки	0.3	1.0	3.6	0.0	0.6	5.5	1.0	3.1	0.1	1.6	4.3	1.0

26. Наведено результати дослідження тремтіння м'язів рук (тремор) у шести пацієнтів залежно від ваги браслета. Кожне табличне значення – середнє з п'яти експериментальних вимірювань частоти тремору (Гц).

Пацієнт	Частота тремору (в Гц)							
	1	2.58	2.63	2.62	2.59	2.85	3.01	2.96
2	2.70	2.83	3.15	3.43	3.47	-	-	-
3	2.78	2.71	3.02	2.89	3.14	3.01	3.35	-
4	2.36	2.49	2.58	2.86	2.93	3.10	-	-
5	2.67	2.96	3.02	3.08	3.32	3.41	-	-
6	2.43	2.50	2.85	3.06	3.07	-	-	-

27. У наступній таблиці наведено кількість вирішених завдань у шести однорідних групах із п'яти осіб. Завдання пропонувалися кожному випробуваному незалежно від інших. Групи відрізняються між собою величиною грошової винагороди за вирішуване завдання.

Група	Кількість вирішених завдань				
	1	10	11	9	13
2	8	10	16	13	12
3	12	17	14	9	16
4	12	15	16	16	19
5	24	16	22	18	20
6	19	18	27	25	24

28. Наведено кількість металевих заготовок певних форми та розміру, виготовлених робітниками трьох різних груп, що відрізняються різними уявленнями про мету роботи (I-відсутність інформації, II- загальні уявлення, III-точна інформація).

Інформація про мету роботи	Кількість оброблених заготовок					
	I	40	35	38	43	44
II	38	40	47	44	40	42
III	48	40	45	43	46	44

29. Дані таблиці представляють розривну міцність волокон бавовни

(в умовних одиницях) залежно від рівня калійних добрив, що вносяться до ґрунту.

Рівень добрив	Міцність волокон							
	I	7.46	7.68	7.21	7.17	7.57	7.80	7.87
II	7.76	7.73	7.74	8.14	8.15	7.87	-	-
III	7.62	8.00	7.93	7.54	8.11	-	-	-

4 ПЛАНУВАННЯ ЕКСПЕРИМЕНТУ

Планування експерименту дозволяє експериментатору, користуючись математичними методами, цілеспрямовано змінювати умови дослідів. Грунтуючись на наявних відомостях про досліджуваний процес, дослідник вибирає деяку оптимальну методику для управління експериментом. Процес дослідження зазвичай розбивається деякі етапи. Після кожного етапу дослідник отримує нову інформацію, що дозволяє змінювати методику дослідження.

Завдання всього дослідження формулюється наступним чином: потрібно отримати уявлення про функцію відгуку $y = \phi(x_1, x_2, \dots, x_m)$ де y - параметр процесу, що підлягає оптимізації; x_1, x_2, \dots, x_m - незалежні змінні - фактори, які можна варіювати при постановці експериментів.

Координатний простір з координатами x_1, x_2, \dots, x_m називається факторним простором, а геометричний образ, що відповідає функції відгуку, називається поверхнею відгуку, представленою рисунку 4.1 б.

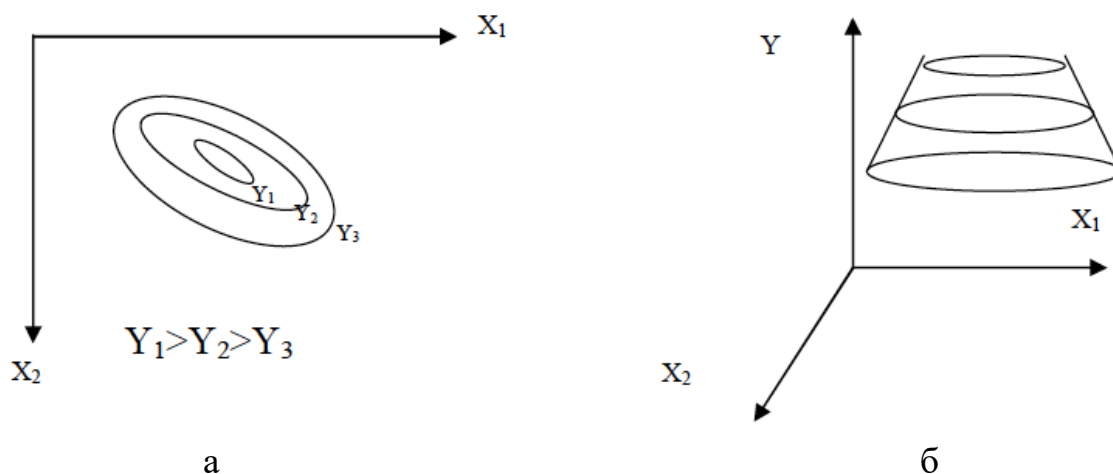


Рис 4.1 Поверхня відгуку

При двох факторах поверхню відгуку можна зобразити у вигляді проєкцій ліній рівного рівня на площину $X_1 X_2$ представленої на рисунку 4.1, а.

Завдання планування експерименту формулюється так: на кожному етапі дослідження потрібно вибрати оптимальне розташування точок у факторному просторі для того, щоб отримати деяке уявлення про поверхню відгуку. Оптимальність розташування точок визначається метою, якої прагне експериментатор цьому етапі.

Активний експеримент можна назвати двічі оптимальним: він визначає

найкраще співвідношення незалежних чинників щодо технологічного процесу, а методи проведення самого експерименту оптимальні тому, що дозволяють вирішити завдання з мінімальними витратами часу і коштів. Тим не менш, існує інша назва - екстремальні експерименти, хоча екстремум далеко не завжди є оптимумом. Ця назва прийнята у статистиці за традицією: до екстремальних завдань відносять завдання відшукування максимумів та мінімумів функції, якщо вони є оптимальними значеннями.

Перш ніж зайнятися побудовою функції відгуку, необхідно вирішити питання про те, скільки і які фактори слід прийняти в розгляд, а також провести так звані експерименти, що відсівають, мета яких — виділити домінуючі фактори серед великої кількості можливих. Після визначення домінуючих факторів необхідно розпочати безпосередньо експеримент.

4.1 Повний факторний експеримент

Математична модель технологічного процесу, яку необхідно оптимізувати за допомогою активного експерименту; є рівнянням регресії. Щоб отримати незалежні оцінки - коефіцієнти цього рівняння, потрібно пред'явити до фіксованих значень незалежних змінних деякі вимоги. Таблиця значень або рівнів факторів при проведенні кожного досвіду (надалі називатимемо її матрицею планування) повинна мати лінійно незалежні стовпці, а сума творів значень будь-яких двох факторів, тобто чисел двох стовпців, повинна дорівнювати нулю. Таку матрицю, яка називається ортогональною, має повний факторний експеримент.

Перед початком дослідів виходячи з апріорних даних, вибирається технологічний режим, який є базовим. Математичний опис об'єкта на околиці точки базового режиму x_{i0} може бути отримано варіюванням кожного з факторів X_i на двох рівнях, що відрізняються від базового рівня на величину кроку варіювання $\pm \Delta x_i$. Чинники повинні бути незалежні, тобто допускати зміну будь-якого з них на деяку величину без зміни при цьому всіх інших факторів.

Крок варіювання по кожній змінній вибирається таким, щоб збільшення величини вихідного параметра Y до базового значення при реалізації кроку можна було виділити на тлі «шуму» при невеликій кількості паралельних дослідів. Крок варіювання не повинен бути ні надто маленьким (не менше похибки вимірювання фактора), ні надто великим. Частка кроку або інтервалу варіювання для всієї області можливих значень цього фактора вибирається залежно від конкретних умов кожного етапу експерименту.

Повним факторним експериментом (ПФЕ) можна вважати

експеримент, що реалізує всі можливі комбінації рівнів незалежних змінних, що не повторюються, кожен з яких примусово (активно) варіюється на двох рівнях. Число цих комбінацій при n факторах дорівнює $N = 2^n$ та визначає тип планування.

Надалі розглянемо планування типу $N=2^3$, тобто на прикладі об'єкта з трьома факторами $x_1 x_2 x_3$.

При плануванні експерименту проводять перетворення (нормалізацію) незалежних змінних x_i ; за допомогою безрозмірних змінних z_i ($Z_i = (x_i - x_{i0}) / \Delta x_i$, що дає можливість легко побудувати ортогональну матрицю планування та полегшує подальші розрахунки, оскільки верхній та нижній рівні варіювання ($z_{i \text{ верх}}$ коли $x_{i0} < x_i$, $z_{i \text{ нижн}}$, коли $x_{i0} > x_i$) рівні відповідно $+1$ та -1 .

Для трифакторного завдання нормалізоване рівняння регресії має вигляд

$$y = b_0 + b_1 z_1 + b_2 z_2 + b_3 z_3 = b_0 + \sum (b_i z_i) \quad (4.1.1)$$

Після запровадження фіктивної змінної $z_0=1$, $y = \sum (b_i z_i)$.

Повний факторний експеримент дозволяє знайти окремі оцінки коефіцієнтів b_i .

Знаходження моделі методом повного факторного експерименту складається з планування експерименту, власне експерименту, перевірки відтворюваності (однорідності вибіркової дисперсії), отримання математичної моделі об'єкта з перевіркою статистичної значущості вибіркового коефіцієнта регресії та перевірки адекватності математичного опису

Побудуємо матрицю планування для трифакторного завдання. При цьому слід пам'ятати, що повний факторний експеримент - це весь можливий перебір комбінацій рівнів x_i .

Матрицю зручно почати з рядка (під рядком, як і у вихідному журналі первинних даних, розуміють умови та результат одного досвіду), де всі керовані змінні знаходяться на нижньому рівні, тобто $z_1 = -1$, $z_2 = -1$, $z_3 = -1$. Наступні рядки (багатомірні точки) вибираються за правилом: при построківому переборі всіх варіантів частота зміни знака керованих змінних для кожної наступної змінної вдвічі менша; ніж для попередньої. Оскільки всі змінні можуть набувати лише значень $+1$ і -1 , це дозволяє з метою спрощення записувати в матрицю тільки знаки «+» та «-». Фіктивна змінна z_0 у всіх рядках має знак "+". Кількість рядків $2^n = 2^3 = 8$.

Таблиця 4.1 - Матриця планування трифакторного експерименту

	Z ₀	Z ₁	Z ₂	Z ₃	Y
1	+	-	-	-	Y ₁
2	+	+	-	-	Y ₂
3	+	-	+	-	Y ₃
4	+	+	+	-	Y ₄
5	+	-	-	+	Y ₅
6	+	+	-	+	Y ₆
7	+	-	+	+	Y ₇
8	+	+	+	+	Y ₈

Число рядків матриці планування, а саме число дослідів, зростає за показовою функцією 2^n .

Побудований таким способом план експерименту має низку цінних властивостей:

1. Симетричність щодо центру експерименту: алгебраїчна сума елементів стовпців для кожного фактора дорівнює нулю $\sum z_{ji} = 0$ (де i – номер фактора (змінюється від 1 до n); j – номер досвіду (змінюється від 1 до N)).

2. Умова нормування: сума квадратів елементів кожного стовпця дорівнює числу дослідів $\sum z_{ij}^2 = N$.

3. Сума добутків будь-яких двох стовпців матриці планування дорівнює нулю: $\sum z_{ij}z_{ky} = 0$ $i \neq k, i, k = 0, 1, 2, \dots, n$.

Ця важлива властивість називається ортогональністю. Тому такі плани та його матриці називають ортогональними. З останньої якості випливає взаємна незалежність оцінок коефіцієнтів рівняння регресії.

При реалізації повного факторного експерименту, оскільки зміна вихідної величини y носить випадковий характер, доводиться в кожній багатовимірній точці проводити паралельних дослідів і результати спостережень $y_{j1}, y_{j2}, \dots, y_{jl}, \dots, y_{jm}$ усереднювати:

$$\bar{y}_j = \sum_{i=1}^m \frac{y_{ji}}{m}. \quad (4.1.2)$$

Весь експеримент у цьому випадку ділиться на m серій дослідів, у кожній з яких повністю реалізується матриця планування.

Перш ніж будувати математичну модель за результатами дослідів, необхідно перевірити, чи не було впливу на деякі з них невраховані фактори. Перевірка відтворюваності є перевіркою виконання передумови регресійного аналізу про однорідність вибірових дисперсій S^2_{yi}

Завдання полягає в перевірці гіпотези про рівність дисперсій $S^2_{y1} = S^2_{y2} = S^2_{y3} = \dots = S^2_{yn}$ при експериментах у всіх багатовимірних точках (8 точок для трьох факторів). Оцінки дисперсій знаходять за формулою

$$S_{yj}^2 = \frac{\sum_{i=1}^m (y_{ji} - \bar{y}_j)^2}{m - 1}. \quad (4.1.3)$$

Якщо кожен досвід проводився тричі, то дисперсії отримують за вибірками однакового об'єму $m=3$, а число ступенів свободи для всіх дисперсій однаково і дорівнює $m - 1=2$.

Для перевірки гіпотези про однорідність оцінок дисперсій слід користуватися критерієм Кохрена G , який ґрунтується на законі розподілу відношення максимальної отриманої дисперсії до суми всіх дисперсій. Розрахункове значення критерію набувають за формулою:

$$G_p = \frac{S_{uj \text{ макс}}^2}{\sum S_{yj}^2}. \quad (4.1.4)$$

Для знаходження табличного значення критерію Кохрен розраховують $f_1 = m-1$; $f_2 = N$. Якщо обчислене значення критерію G_p виявиться менше його табличного значення G_p для обраного рівня значущості (зазвичай $\alpha=0,05$), то гіпотеза про однорідність дисперсій приймається. Це означає, що різниці між дисперсіями не повинно бути.

У цьому випадку можна розрахувати генеральну дисперсію відтворюваності

$$S_y^2 = \frac{\sum_{j=1}^n S_{yj}^2}{N} = S_{\text{воспр}}^2. \quad (4.1.5)$$

що має $f_3 = N(m - 1)$ ступенів свободи.

Якщо перевірка на відтворюваність дала негативний результат, залишається визнати експеримент що не відтворюється щодо керованих змінних внаслідок флуктуації некерованих і неконтрольованих змінних, створюють на виході великий рівень «шуму». При цьому слід повторити експеримент із збільшеним числом паралельних дослідів.

Побудова математичної моделі об'єкта починається з визначення незалежних оцінок – коефіцієнтів b_i за формулою:

$$b_i = \frac{\sum_{i=A}^n z_i \bar{y}_i}{N}. \quad (4.1.6)$$

Після цього необхідно перевірити гіпотезу про значущість коефіцієнтів b_i (перевірка нуль-гіпотези $b_i = 0$).

Перевірка гіпотези проводиться за допомогою t-критерію, який у даному випадку формується як

$$t_p = \frac{|b_i|}{S_{bi}} \quad (4.1.7)$$

де S_{bi} - середньоквадратична помилка, визначення коефіцієнта b_i при повному факторному експерименті, рівна

$$S_{bi} = \sqrt{\frac{S_y^2}{Nm}}. \quad (4.1.8)$$

Якщо знайдена величина параметра t перевищує значення $t_{кр}$, визначене за таблицями Стьюдента для числа ступенів свободи $f = N(m - 1)$ при заданому рівні значущості, тобто $sign(t_p - t_{кр}) = "+"$, то гіпотеза відкидається і коефіцієнт b_i визнається значущим.

Інакше, якщо $sign(t_p - t_{кр}) = «-»$, нуль-гіпотеза приймається і коефіцієнт b_i - вважають статистично не значущим (тобто рівним нулю).

Значимість коефіцієнтів можна оцінити за допомогою довірчого інтервалу:

$$\Delta b_j = \pm(N(m - 1), \alpha/2) \cdot S_{bj}. \quad (4.1.9)$$

Коефіцієнт вважається значним, якщо виконується нерівність:

$$Abs(b_j) \geq \Delta b_j$$

де $abs(b_j)$ – значення коефіцієнтів по модулю.

Статистична незначимість коефіцієнта b_i може бути зумовлена такими причинами:

- 1) рівень базового режиму x_{i0} близький до точки приватного екстремуму змінної X_i (або за твором змінних);
- 2) крок варіювання Δx_i – обраний малим;
- 3) дана змінна (твір змінних) не має функціонального зв'язку з вихідним параметром y , тобто $b_i = 0$ за сутністю;
- 4) велика помилка експерименту внаслідок флуктуації некерованих та неконтрольованих змінних.

Оскільки ортогональне планування дозволяє визначати незалежні оцінки коефіцієнтів, якщо будь-який з коефіцієнтів виявиться незначним, може бути відкинуто без перерахунку всіх інших.

Після цього математична модель об'єкта складається у вигляді рівняння зв'язку вихідного параметра i змінних x_i , Що включає тільки значущі коефіцієнти.

Щоб перевірити гіпотезу про адекватність результатів експерименту знайденим за рівнянням зв'язку, достатньо оцінити відхилення передбаченої рівнянням регресії вихідної величини y_{ip} від результатів експерименту y_i у

точках факторного простору.

Розкид результатів експерименту щодо лінії регресій у разі порівнюється з розкидом точок між собою. Перший розкид характеризується дисперсією адекватності $S^2_{ад}$, оцінка якої знаходиться за формулою

$$S^2_{ад} = \frac{\sum_{i=1}^n (\bar{y}_i - y_{ip})^2}{N - d}. \quad (4.1.10)$$

де d - число значущих коефіцієнтів лінійного рівняння при числі ступенів свободи $f_{ад} = N - n - 1$.

Другий (розподіл точок) — дисперсією відтворюваності $S^2_{відтв}$ з числом ступенів свободи $f_{відтв} = N(m - 1)$. Дисперсія відтворюваності визначається за формулою (2.2.6).

Якщо $S^2_{ад} < S^2_{ад}$, то отримана математична модель адекватно представляє результати експерименту. Якщо $S^2_{ад} > S^2_{ад}$ то потрібно з'ясувати, чи випадково або не випадково розрізняються ці дисперсії, і якщо не випадково, то робиться висновок про неадекватність рівняння.

Перевірка гіпотези щодо адекватності проводиться з використанням критерію Фішера.

$$F_{екс} = \frac{S^2_{ад}}{S^2_y}. \quad (4.1.11)$$

Якщо $F_{екс} < F_{кр}$ для відповідних ступенів свободи $f_{ад} = N - d = N - n - 1$ і $f_{відтв} = N(m - 1)$ при заданому рівні значущості, то нуль-гіпотеза приймається, тобто рівняння адекватне при $sgn(F_{експ} - F_{кр}) = "-"$. В іншому випадку гіпотеза відкидається і опис визнається неадекватним об'єкту.

Перевірка адекватності можлива при $f_{ад} > 0$.

Якщо гіпотеза про адекватність відкидається, необхідно переходити до складнішої форми рівняння або (якщо це можливо) проводити експеримент із меншим кроком варіювання Δx_i .

Якщо математична модель адекватна, перевіряють її працездатність. Перевірка працездатності здійснюється за допомогою коефіцієнта детермінації. Розрахунок коефіцієнта детермінації може бути здійснено за формулою. Якщо значення $R^2 > 0,7$, то математична модель є працездатною. Інакше математичну модель не можна використовуватиме прогнозування результатів, т.к. вона не є працездатною.

4.2 Побудова експерименту в пакеті Statgraphics

Для побудови експерименту виконуються такі дії:

1. Відкривається пакет STATGRAPHICS.

2. Вибирається меню Special, підменю Experimental Design , команду Create Design .

3. У діалоговому вікні в області Design Class , що відкрилося, встановлюється маркер навпроти команди Screening. У рядку No. of Response Variables встановлюється 1. У рядку No. of Experimental Factors встановлюється 3. Натискається кнопка Ok .

4. Для кожного з трьох факторів виставляються значення нижнього (-1,0) та верхнього (1,0) рівнів. Даються назви кожному фактору.

5. У наступному діалоговому вікні у рядку Name дається ім'я вихідного показника, на який впливають фактори, зазначені у попередньому вікні.

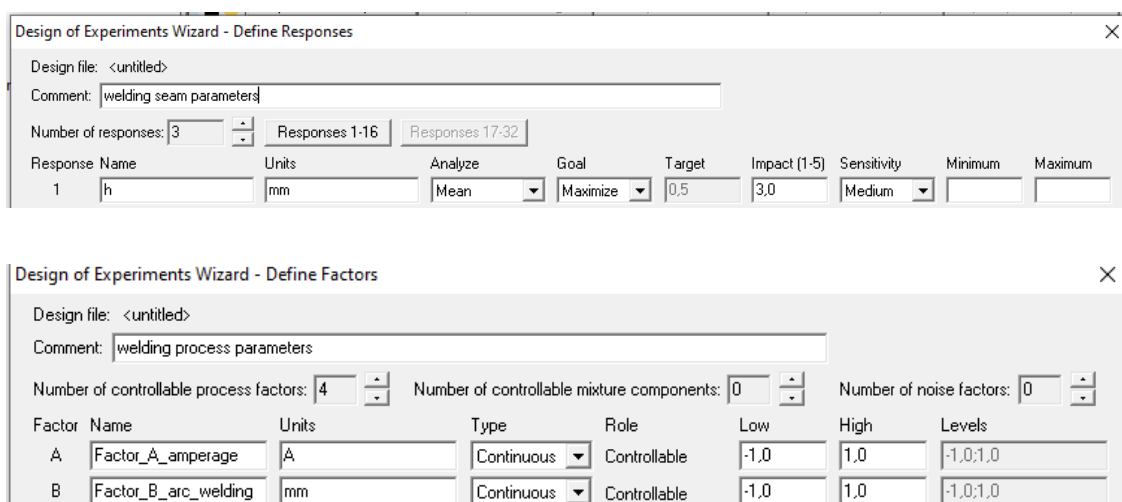


Рис. 4.1 Діалогове вікно визначення параметрів та факторів ПФЕ в програмі Statgraphics

6. У наступному діалоговому вікні вибирається, який експеримент необхідно провести (ПФЕ).

7. У наступному діалоговому вікні забирається галочка навпроти команди Randomize.

8. Після чого відкриваємо розрахункову таблицю, у яку вносяться середні значення вихідного показника Y.

9. Зберігаємо документ.

10. У вікні Screening Design Attributes вибирається на панелі інструментів кнопка Tabular Options .

11. У діалоговому вікні, що відкрилося, натискається кнопка All .

12. Вибирається меню Special , підменю Experimental Design команда Analyze Design .

13. У діалоговому вікні Analyze Design вибирається параметр Y.

14. У вікні Analyze Experimental на панелі інструментів вибирається кнопка Graphical Options . У діалоговому вікні, що відкрилося, натискається кнопка All. В результаті у вікні програми відобразяться всі розрахунки та графічні залежності.

4.3 Аналіз математичної моделі та результатів експерименту

Коефіцієнти при незалежних змінних вказують на силу впливу факторів: чим більша чисельна величина коефіцієнта, тим більше впливає фактор. Якщо коефіцієнт має знак «+», то зі збільшенням значення фактора параметр оптимізації у збільшується, а якщо «-» - зменшується. Величина коефіцієнта відповідає вкладу даного фактора величину параметра оптимізації при переході фактора з нульового рівня на верхній або на нижній.

Іноді зручно оцінювати внесок фактора під час переходу від нижнього рівня до верхнього (або навпаки), а не до нульового. Вклад, визначений в такий спосіб, називається ефектом чинника (іноді його називають основним чи головним ефектом). Він чисельно дорівнює подвоєному коефіцієнту. Для якісних факторів, що варіюються на двох рівнях, основний (нульовий) рівень часто не має фізичного сенсу. Тому поняття «ефект фактора» тут є природним.

Відомо, що більшість технічних систем та технологічних процесів в машинобудування підпорядковуються ступеневої ($y=x_m^{an}$) або експоненціальній ($y = a \exp x$) закономірностям.

Водночас жодних труднощів у описів цих закономірностей рівняннями як поліномів і фізичної інтерпретації результатів немає. Відомо, що «вирівнювання», або лінеаризація статичного рівняння, проводиться логарифмуванням його обох частин, а експоненційного рівняння - тільки правої його частини. Після цього фактори замінюють логарифмами та проводять усі розрахунки як для звичайного рівняння регресії. Відгук у першому випадку також замінюється на його логарифм. Вирівнювати можна більшість інших закономірностей.

Таким чином, можливості фізичної інтерпретації рівнянь у поліномах ширші за інші, оскільки вони дозволяють врахувати взаємний вплив, або ефекти взаємодії факторів.

Найповнішою формою аналізу результатів є геометрична інтерпретація результатів ПФЕ.

Відомо, що графічно можна показати лише безперервний зв'язок між двома змінними. Залежності функції з інших аргументів можна зобразити дискретно, тобто параметрично, чи безперервно, але у спотвореному вигляді

за допомогою аксонометрических побудов. При інтерпретації результатів ПФЕ використовуються обидва ці методи, але перевага надається першому внаслідок його простоти.

Наприклад, вивчався вплив на геометричні параметри зварного шва внаслідок процесу зварювання його параметрів таких як сила струму та довжина дуги зварювання. Фактори були прийняті сила струму x_1 і довжина дуги x_2 . За функцію відгуку прийнята ширина зварного шва.

Рівняння регресії, отримане в результаті ПФЕ, є неповним квадратним рівнянням: $179.75 - 18.25z_1 - 8.75z_2 - 25.75z_1z_2$. Це рівняння також задовольнило всі вимоги статистичного аналізу (перевірка значущості коефіцієнтів за допомогою (t-критерію, перевірка адекватності рівняння за F-критерієм)).

Геометрична інтерпретація представлена рис. 4.2. По координатним осям відкладені у розмірному вираженні сила струму та довжина дуги зварювання. Значення функції відгуку відкладені у вигляді сімейства кривих (мм): 1 – 16; 2 – 18; 3-18,5; 4 – 18,7; 5 – 19; 6 – 20.

Ці криві мають вигляд рівнобічних гіперболів з асимптотами, які збігаються у напрямку з координатними осями. Геометричний аналіз кривих показує, що між значеннями функції відгуку 18,5 та 18,7 вони повинні виродитись у крапку. Ця особлива точка називається «мінімаксом». У ній перетинаються асимптоти, що ділять координатне поле на квадранти з різними властивостями (у I та III при віддаленні від мінімаксу функція відгуку зменшується, у II та IV – зростає).

Вибір II або IV квадранту для оптимізації процесу зварювання в даному випадку повинен бути визначений з умов оптимальності іншого (додаткового) критерію. Оскільки зі збільшенням струму та зменшенням значення довжини дуги зварювання (IV квадрант) не відбувається зменшення ширини шва, що є основним показником досліджуваного процесу, оптимум слід шукати саме у цьому квадранті.

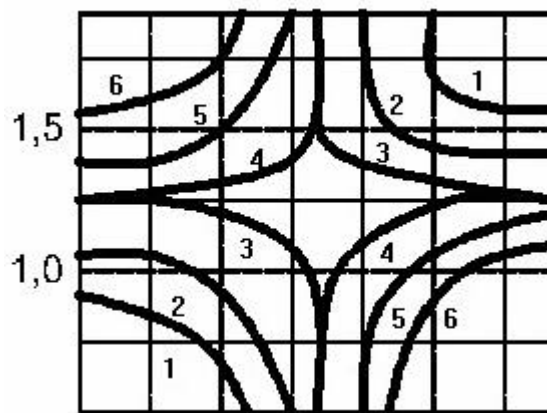


Рис.4.2 Параметрична діаграма

Другий спосіб аналізу результатів за допомогою аксонометричних побудов дозволяє визначити оптимальну область проведення експерименту за допомогою максимуму або мінімуму на поверхні відгуку або проекції ліній рівного рівня, рис. 4.2.

Значимість коефіцієнтів моделі визначають також за допомогою картки Парето. Карта дозволяє оцінити значущість коефіцієнтів, а відповідно і ступінь впливу факторів на вихідний параметр Y . Карта Парето є гістограмою, приклад представлений на рис 4.3.

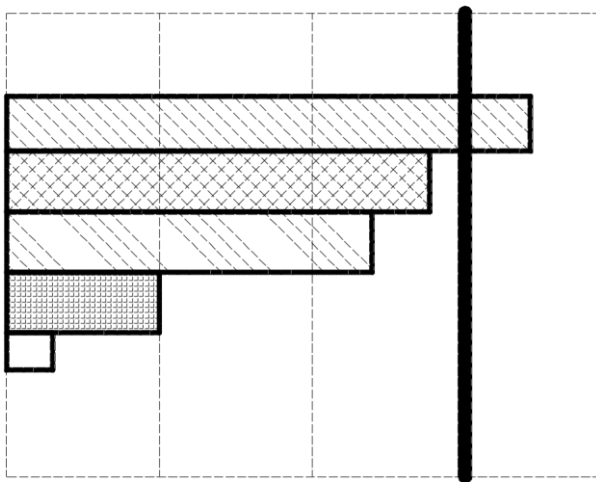


Рис. 4.3 Карта Парето

4.4 Алгоритм реалізації ПФЕ

1. Відповідно до завдання з додатка 4 будується матриця планування.
2. Виробляється усереднення значень вихідного параметра Y за формулою (4.1.2).
3. Перебувають оцінки дисперсій за формулою (4.1.3).
4. Визначається розрахункове значення критерію Кохрена за формулою (40). Робиться висновок про однорідність дисперсій.
5. Визначається дисперсія відтворюваності за формулою (4.1.4).
6. Визначаються коефіцієнти математичної моделі (4.1.5).
7. Визначається значення коефіцієнтів математичної моделі за допомогою формули (45).
8. Складається математична модель з урахуванням лише значних коефіцієнтів.
9. Визначається середньоквадратична помилка S_{bi} (4.1.8).

10. Розраховується дисперсія адекватності $S_{ад}$ (4.1.10).
11. Визначається критерій Фішера (4.1.11). Робиться висновок про адекватність математичної моделі.
12. Визначається працездатність адекватної моделі.
13. Здійснюється побудова ПФЕ пакеті Statgraphics.
14. Проводиться аналіз математичної моделі та результатів експерименту.

5 ДОСЛІДЖЕННЯ ПРОЦЕСУ ЗВАРЮВАННЯ ТА РЕАЛІЗАЦІЯ РЕГРЕСІЙНОЇ МОДЕЛІ

5.1 Аналіз існуючих регресійних моделей параметрів зварного шва

При дослідженні процесу формування шва враховувався вплив наступних параметрів режиму зварювання: I – сила зварювального струму; U – напруга на дузі; L_d – довжина дуги ($L_d = L_{д} + f_d$, де f_d – збуджуючі впливи по довжині дуги, викликані особливостями конструкції виробів, що зварюються); V – швидкість зварювання; $V_{пд}$ – швидкість подачі присадочного дроту; d_e – діаметр вольфрамового електроду; $d_{пд}$ – діаметр присадочного дроту.

Для отримання аналітичних залежностей форми шва і математичної моделі позначимо розміри поперечного перерізу зварного шва (див. малюнок):

e – ширина зварного шва, h – висота зварного шва, g – глибина проплавлення.

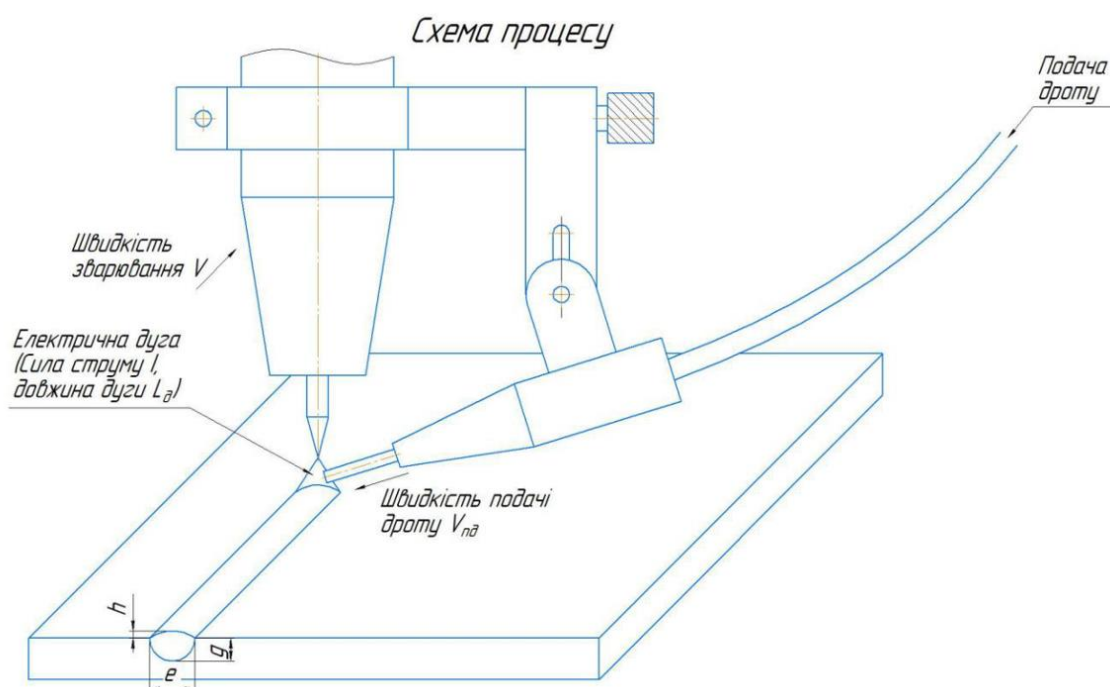


Рис. 5.1 Схема процесу зварювання

Розглянемо, розроблені останнім часом, детерміновано-статистичні моделі форми шва для механізованого зварювання в захисних газах

плавким присадочним дротом [1].

В роботі наведено теоретичне обґрунтування необхідності застосування

- подібних математичних інструментів і розроблені три моделі для зварювання плавким електродним дротом в CO₂. Для дроту діаметром 2 мм моделі мають вигляд:

- Модель М1:

$$h = 0,07 \frac{I^{1,415}}{d_{min}^{0,33} U^{0,262} V^{0,289}},$$
$$e = 0,063 \frac{d^{0,020} I^{0,297} U^{1,513}}{V^{0,812}},$$
$$g = 0,954 \frac{I^{1,141}}{d_{\pi n}^{0,593} U^{1,251} V^{0,399}}$$

- Модель М2:

$$h = 0,007 \frac{I^{1,444}}{U^{0,339} V^{0,342}}$$
$$e = 0,313 \frac{I^{0,082} U^{1,453}}{V^{0,886}}$$
$$g = 0,609 \frac{I^{0,792}}{U^{0,649} V^{0,401}}$$

- Модель М3:

$$h = 0,005 \frac{I^{1,308}}{V^{0,398}}$$
$$e = 0,354 \frac{V^{1,554}}{V^{0,889}}$$
$$g = 0,410 \frac{I^{1,499}}{V^{0,469}}$$

5.2 Проведення досліджень та розробка математичної моделі

При адаптації існуючих моделей для аргонодугового зварювання електродом, що не плавиться, необхідно включити в них специфічні параметри режиму зварювання і врахувати вплив режиму на форму зварного шва. [1]

Для моделі висоти зварного шва в чисельник моделі необхідно

помістити швидкість подачі присадного дроту (V_{III}) і діаметр присадочного

$$\begin{aligned}
 h &= 0,042 \frac{I^{0,950}}{V^{0,486}} (1+b)^{0,658} \\
 g &= 1,047 \frac{I^{0,413}}{V^{0,638}} \frac{1}{(1+b)^{0,453}} \\
 h &= 0,020 \frac{I^{0,094}}{V^{0,409}} 1,284^b \\
 g &= 1,062 \frac{I^{0,409}}{V^{0,778}} 0,692^b \\
 h &= \frac{I^{3,752} U^{3,245} f^{-0,53} 1,373^b}{L^{4,439} V^{2,383}} \\
 e &= \frac{I^{3,597} U^{3,828} f^{-0,278} 0,69^b}{L^{4,439} V^{2,383}}, \\
 g &= \frac{I^{4,353} U^{4,65} f^{-0,579} 0,616^b}{L^{5,659} V^{2,845}} \\
 h &= 0,0034 \frac{I^{0,942} U^{0,045} f^{0,29} 1,373^b}{V^{0,09}}, \\
 e &= 0,015 \frac{I^{0,679} U^{0,71} f^{0,298} 0,69^b}{V^{0,134}} \\
 g &= 0,025 \frac{I^{0,95} f^{0,125} 0,616^b}{U^{0,23} V^{0,194}}
 \end{aligned} \tag{5.2.1}$$

дроту (d_m), так як при збільшенні цих параметрів висота зварного шва зростає. У знаменник слід помістити швидкість зварювання (V), діаметр вольфрамового електрода (d_e), силу зварювального струму (I) і довжину дуги (L_d), так як зі збільшенням цих параметрів висота шва падає:

$$h = \frac{V_{\text{III}}^x d_{\text{III}}^y}{V^0 d_e^j L_{\mu}^n}$$

В моделі ширини зварного шва параметри сили зварювального струму (I) і довжини дуги (L_d) слід помістити в чисельник - з їх збільшенням зростає ширина шва:

$$e = \frac{I^j L_{\mu}^n V_{\text{III}}^x d_{\text{III}}^y}{V^0 d_e^e}$$

Для моделі глибини проплавлення величину довжини дуги (L_d) необхідно помістити в знаменник - при її збільшенні відбувається різке падіння глибини проплавлення:

$$g = \frac{I^j V_{\text{ПП}}^x d_{\text{ПП}}^y}{V^o d_e^e L_{\mu}^n}$$

Числові значення коефіцієнтів x , y , 0 , e , j та n були визначені обробкою накопичених експериментальних даних.

Нижче наведені моделі форми зварного шва, отримані після статистичної обробки експериментальних даних методом регресивного аналізу. [1]

$$h = \frac{V_{\text{ПП}}^{1,465} d_{\text{ПП}}^{0,92}}{V^{1,9} d_e^{0,078} I^{0,12} L_{\mu}^{0,05}}$$

$$e = \frac{I^{0,584} L_{\mu}^{0,433} V_{\text{ПП}}^{0,095} d_{\text{ПП}}^{0,010}}{V^{0,832} d_e^{0,782}}$$

$$g = \frac{I^{0,659} V_{\text{ПП}}^{0,247} d_{\text{ПП}}^{0,02}}{V^{0,851} d_e^{2,587} L_{\mu}^{0,415}}$$

Аналіз отриманих моделей в формі статичних виразів показує, що вони відображають детерміновані залежності розмірів шва від основних режимів аргонодугового зварювання і не суперечать існуючим уявленням про формування шва. Комбінуючи параметри режиму в моделях, можна отримати залежності, що відображають сутність фізичних процесів з хорошою точністю.

Попередні експериментальні дослідження дозволили виявити основні фактори, що впливають на геометричні параметри зварного шва:

$$h=f(V_{\text{нд}}, d_{\text{д}}, V, I, d_e, L_d)$$

$$e=f(V_{\text{нд}}, d_{\text{д}}, V, I, d_e, L_d)$$

$$g=f(V_{\text{нд}}, d_{\text{д}}, V, I, d_e, L_d)$$

При плануванні необхідної кількості експериментів для функції відгуку – за допомогою планування багатофакторного експерименту виду 2^4 методом Бокса-Уілсона необхідно скласти рівняння регресії.

При виборі діапазонів варіювання факторів функцій, передбачених планом експерименту необхідно прослідкувати, щоб будь-яка їх сукупність могла бути реалізована і не приводила до протиріч. Для досягнення цих умов було проведено пошукові експерименти, щоб визначити область, в якій необхідні нас сполучення рівнів факторів були б реалізовані.

В табл. 5.1. Приведені істинні значення факторів, які встановлені на основі проведення пошукових експериментів.

Таблиця 5.1. Рівні факторів та інтервали варіювання

Фактори	Рівні факторів	
	-1	+1
X_1 – сила струму, А	155	215
X_2 – довжина дуги, мм	3	7
X_3 – швидкість зварювання, мм/с	2,78	4,2
X_4 – швидкість подачі дроту, мм/с	6	9

Для проведення планування на основі повнофакторного експерименту виду 2^4 було складено матрицю планування експерименту, яку наведено в табл.5.2.

Таблиця 5.2. Матриця планування експерименту

№ дослідів	X1	X2	X3	X4
1	-	+	-	+
2	-	-	+	+
3	+	-	-	-
4	+	-	+	-

5	-	+	-	-
6	-	-	-	+
7	-	+	+	-
8	-	-	-	-
9	-	+	+	+
10	+	+	+	+
11	+	-	-	+
12	+	+	-	+
13	+	-	+	+
14	+	+	-	-
15	-	-	+	-
16	+	+	+	-

Результати експерименту розраховувалися за формула (5.2.1) і наведені в табл 5.3.

Таблиця 5.3. Результати експериментів

№ дослідю	X1	X2	X3	X4	h	e	g
1	-	+	-	+	4,54	11,48	0,85
2	-	-	+	+	2,16	5,64	0,85
3	+	-	-	-	2,52	9,26	1,36

4	+	-	+	-	1,15	6,57	0,96
5	-	+	-	-	2,51	11,04	0,77
6	-	-	-	+	4,34	7,95	1,21
7	-	+	+	-	1,15	7,84	0,54
8	-	-	-	-	2,62	7,65	1,09
9	-	+	+	+	2,07	8,14	0,60
10	+	+	+	+	1,99	9,86	0,74
11	+	-	-	+	4,56	9,63	1,50
12	+	+	-	+	4,37	13,89	1,06
13	+	-	+	+	2,08	6,83	1,06
14	+	+	-	-	2,41	13,37	0,96
15	-	-	+	-	1,19	5,43	0,77
16	+	+	+	-	1,10	9,49	0,67

5.3. Аналіз результатів дослідження

В цьому розділі проведено аналіз отриманих даних в програмі STATGRAPHICS. Результати експериментів для висоти зварювального шва h

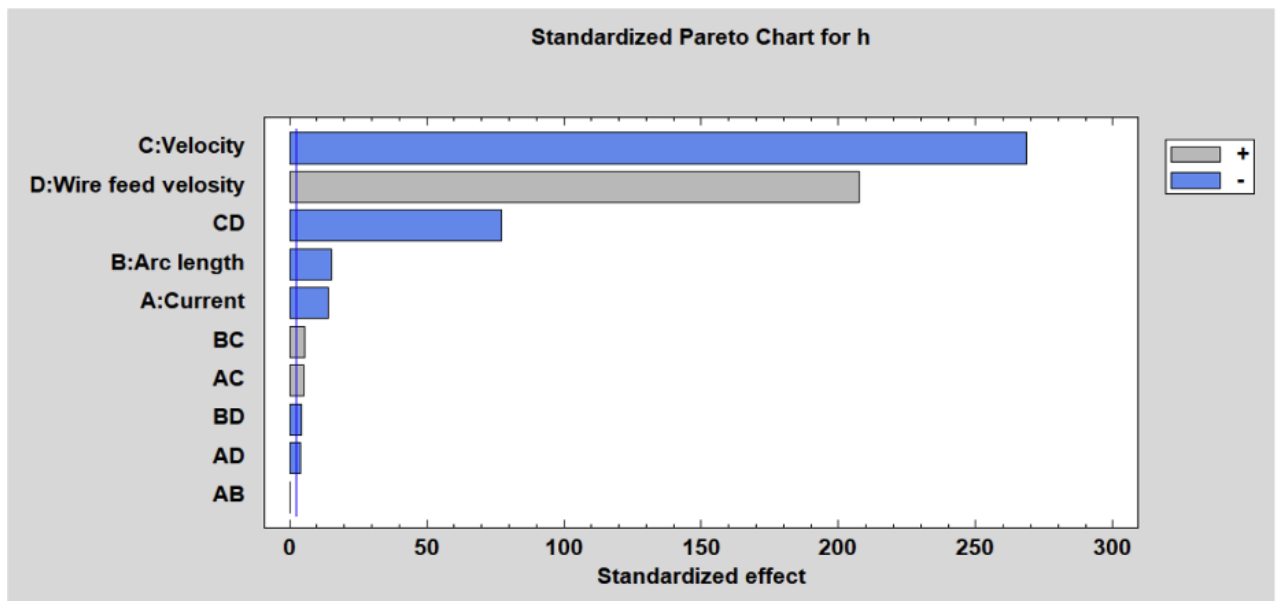


Рис. 5.2 Діаграма Парето для цільової функції – «висота зварного шва»

З діаграми Парето видно що найбільший вплив на висоту зварювального шва здійснює швидкість, швидкість подачі дроту та комбінація цих двох факторів. Зі збільшенням швидкості висота зварювального шва буде зменшуватися, а зі збільшенням швидкості подачі дроту – збільшуватися.

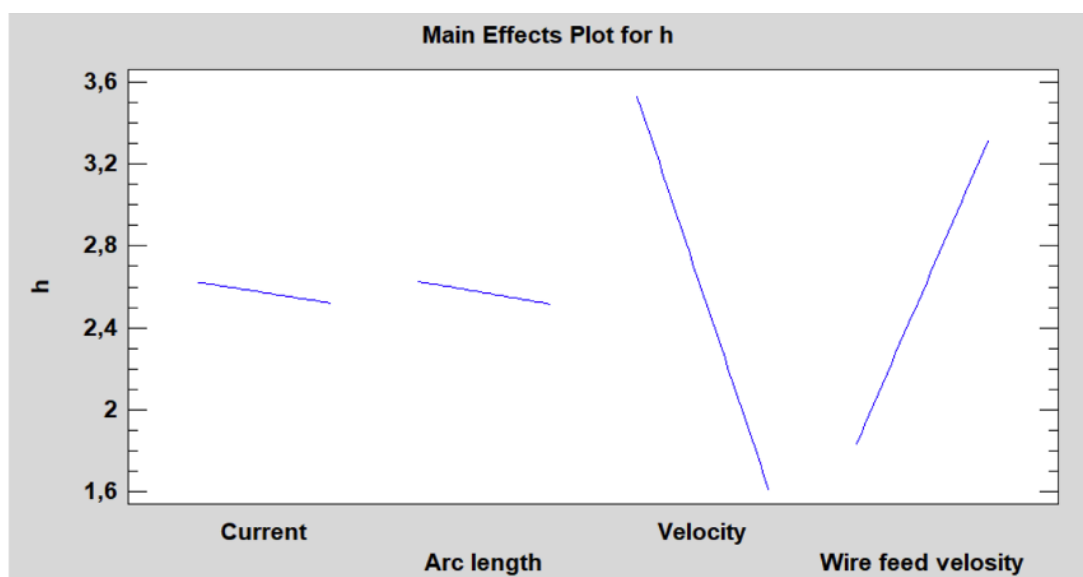


Рис. 5.3 Графік впливу факторів на цільову функцію - «висота зварного шва»

За рис. 5.3 видно що зі збільшенням сили струму, довжини дуги та швидкості висота зварного шва зменшується, а зі збільшенням швидкості подачі дроту.

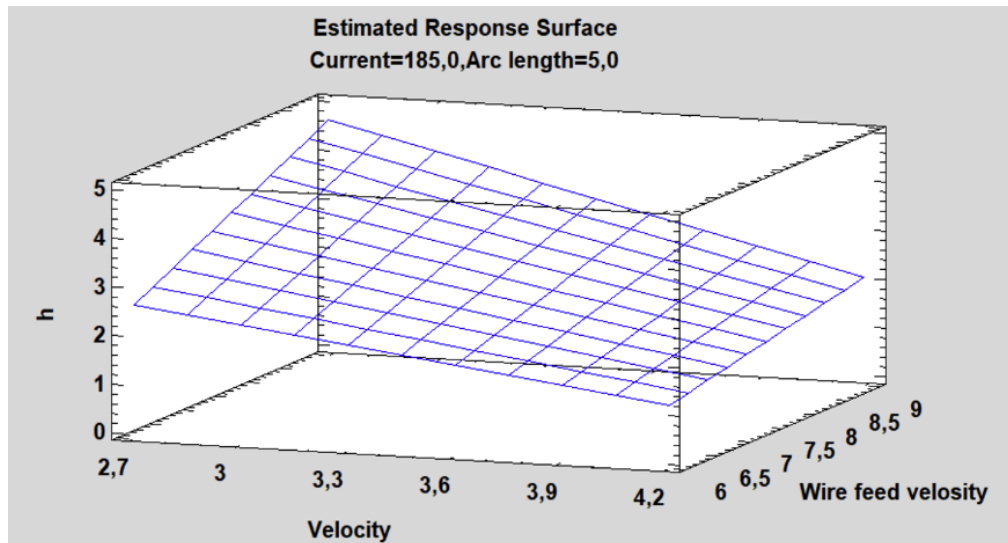


Рис. 5.4 Поверхня відгуків цільової функції – «висота зварного шва»

На рис. 5.4 показано поверхню відгуків цільової функції – «висота зварного шва» та їх двомірні перерізи в площинах параметрів впливу, які дозволяють наглядно проілюструвати залежність даної цільової функції від окремих параметрів впливу.

Рівняння регресії для функції відгуку – «висоти зварного шва», згідно проведеного багатofакторного експерименту має вигляд:

$$\begin{aligned}
 h = & -2,59376 - 0,0024017 * \text{Current} - 0,0410349 * \text{Arc length} + \\
 & 0,365082 * \text{Velocity} + 1,48785 * \text{Wire feed velocity} + \\
 & 0,0000166667 * \text{Current} * \text{Arc length} + \\
 & 0,000874413 * \text{Current} * \text{Velocity} - 0,000322222 * \text{Current} * \text{Wire feed} \\
 & \text{velocity} + 0,0143486 * \text{Arc length} * \text{Velocity} - 0,00525 * \text{Arc length} * \text{Wire feed} \\
 & \text{velocity} - 0,259977 * \text{Velocity} * \text{Wire}
 \end{aligned}$$

feed velocity

Результати експериментів для ширини зварювального шва e

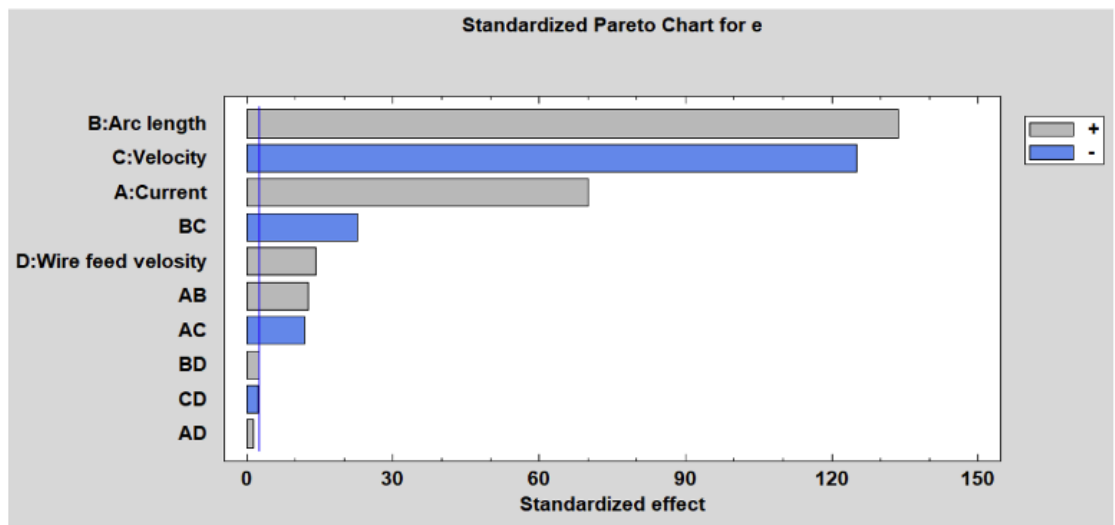


Рис. 5.5 Діаграма Парето для цільової функції – «ширина зварного шва»

На ширину зварювального шва помітний вплив мають три фактори: довжина дуги, швидкість та сила струму. Найбільший вплив має довжина дуги та швидкість.

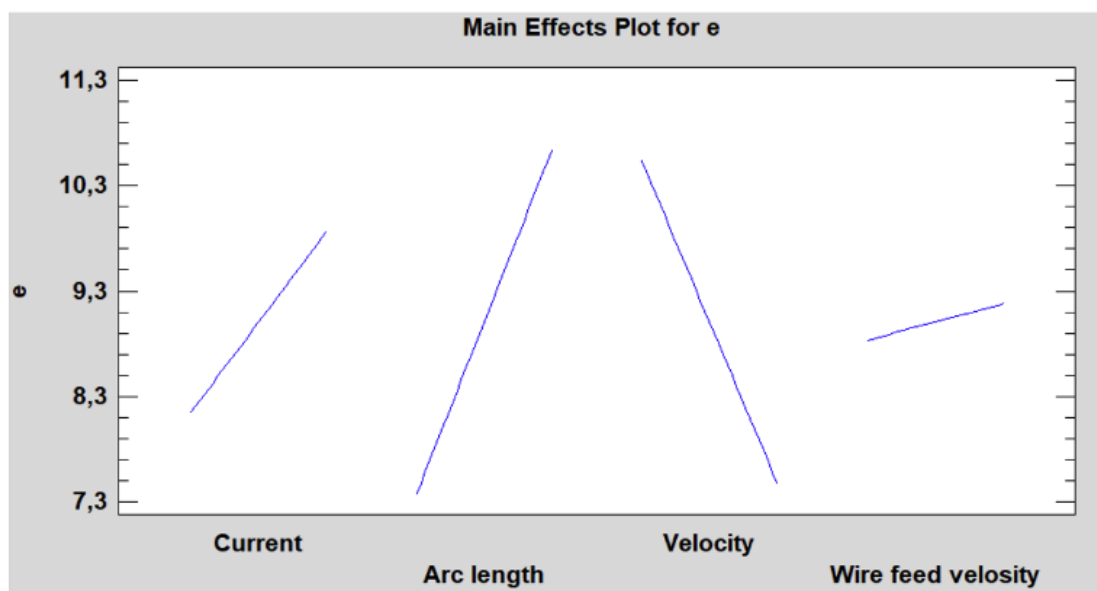


Рис. 5.6 Графік впливу факторів на цільову функцію - «висота зварного шва»

За графіком на рис. 5.6 видно що збільшення сили струму, довжини дуги та швидкості подачі дроту призводить до збільшення ширини зварювального шва, а збільшення швидкості – до зменшення.

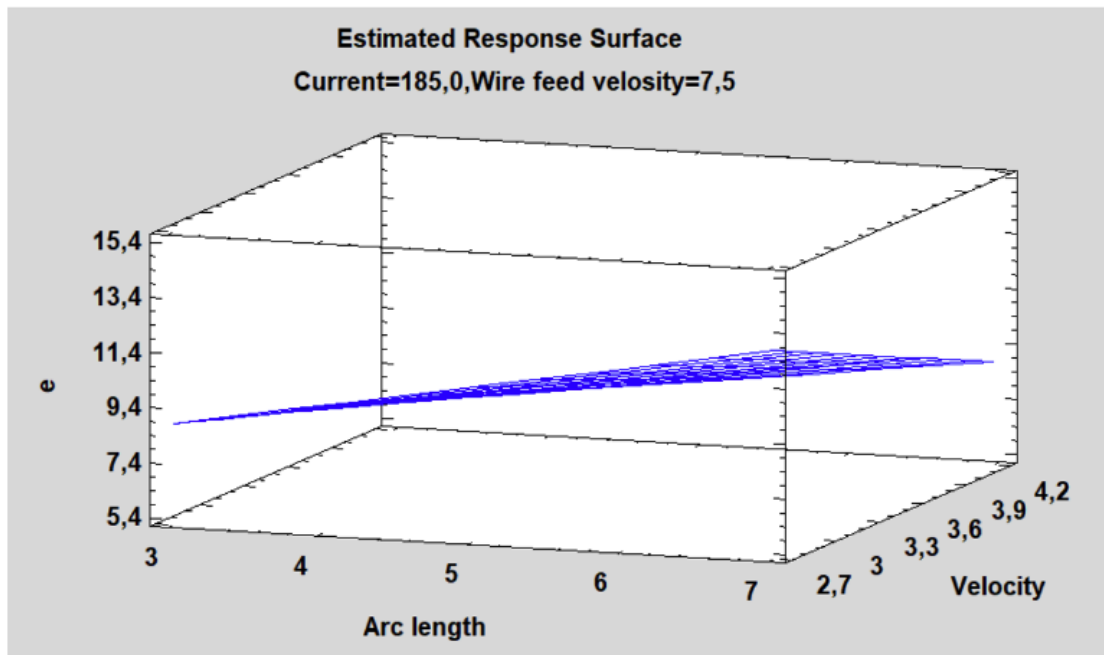


Рис. 5.7 Поверхня відгуку цільової функції – «ширина зварного шва»

На рис. 5.7 показано поверхню відгуків цільової функції – «ширина зварного шва» та їх двомірні перерізи в площинах параметрів впливу, які дозволяють наглядно проілюструвати залежність даної цільової функції від окремих параметрів впливу.

Рівняння регресії для функції відгуку – «ширини зварного шва», згідно проведеного багатofакторного експерименту має вигляд:

$$\begin{aligned}
 e = & 1,03241 + 0,0367745 * \text{Current} + 0,940488 * \text{Arc length} + \\
 & 0,296303 * \text{Velocity} + 0,0919214 * \text{Wire feed velocity} + \\
 & 0,00259375 * \text{Current} * \text{Arc length} - \\
 & 0,00684859 * \text{Current} * \text{Velocity} + 0,000366667 * \text{Current} * \text{Wire feed velocity} - \\
 & 0,195511 * \text{Arc length} * \text{Velocity} + 0,0105 * \text{Arc length} * \text{Wire feed velocity} - \\
 & 0,0276995 * \text{Velocity} * \text{Wire} \\
 & \text{feed velocity}
 \end{aligned}$$

Результати експериментів для глибини проплаву зварювального шва g

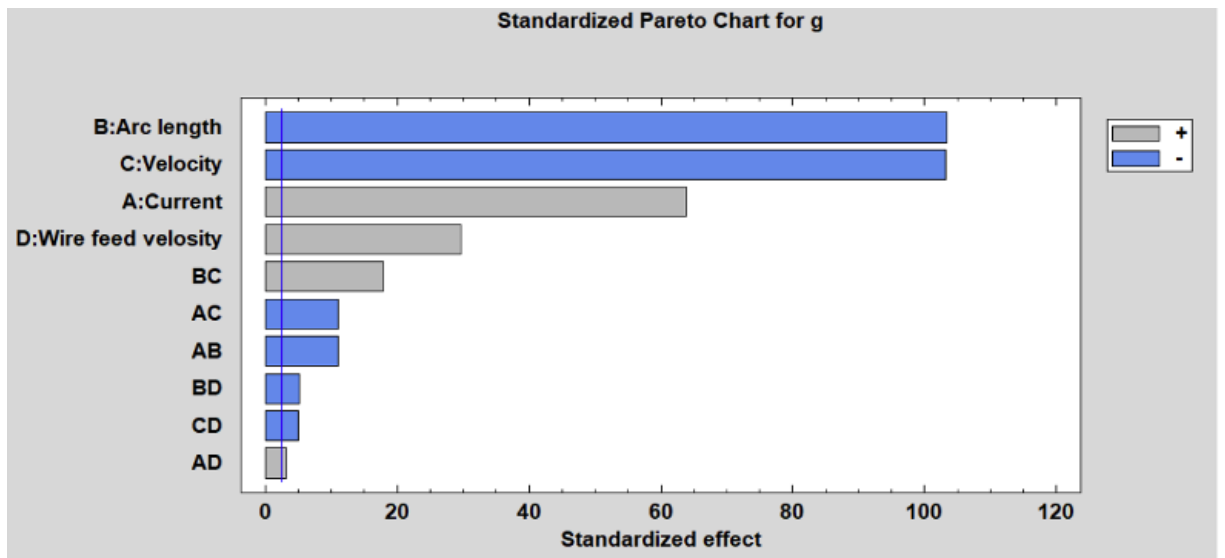


Рис. 5.8 Діаграма Парето для цільової функції – «глибина проплавлення зварного шва»

На діаграмі Парето, зображеній на рис.5.8, видно, що найбільший вплив глибину проплаву зварного шва мають довжина дуги, швидкість та сила струму. При збільшенні довжини дуги та швидкості глибина проплаву зварного шва зменшується, а при збільшенні сили струму – збільшується.

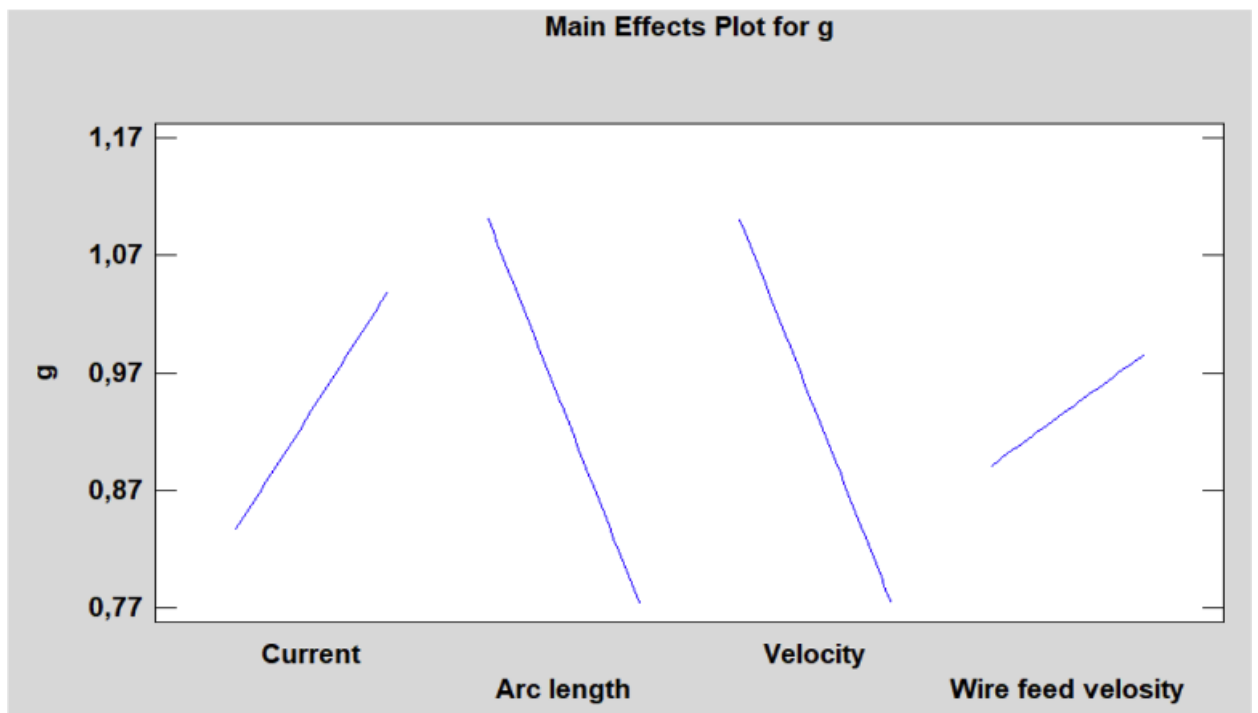


Рис. 5.9 Графік впливу факторів на цільову функцію - «висота зварного шва»

На рис. 5.9 відображений вплив чотирьох факторів на глибину зварного шва.

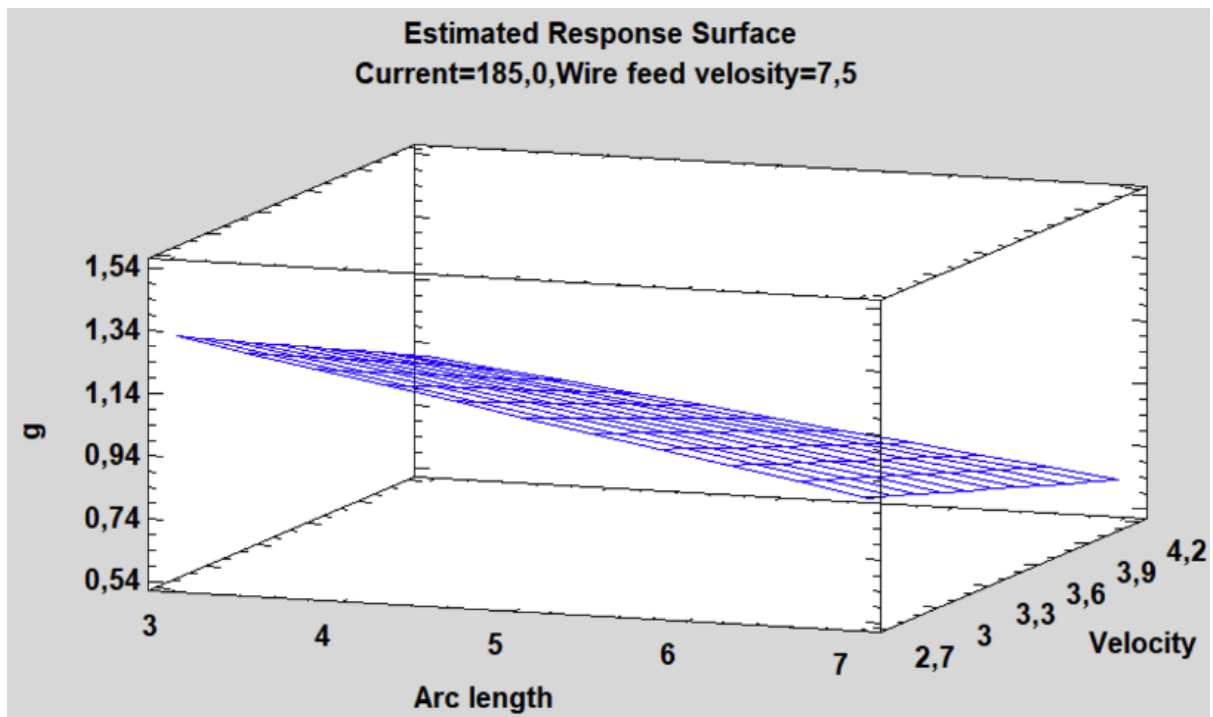


Рис. 5.10 Поверхня відгуку цільової функції – «глибина проплавлення зварного шва»

На рис. 5.10 показано поверхню відгуків цільової функції – «глибина проплавлення зварного шва» та їх двомірні перерізи в площинах параметрів впливу, які дозволяють наглядно проілюструвати залежність даної цільової функції від окремих параметрів впливу.

Рівняння регресії для функції відгуку – «глибина проплавлення зварного шва», згідно проведеного багатofакторного експерименту має вигляд:

$$g = 0,688702 + 0,0068816 * \text{Current} - 0,0763948 * \text{Arc length} - 0,12007 * \text{Velocity} + 0,0506604 * \text{Wire feed velocity} - 0,00029375 * \text{Current} * \text{Arc length} -$$

$$0,000827465 * \text{Current} * \text{Velocity} + 0,000111111 * \text{Current} * \text{Wire feed velocity} + 0,0199824 * \text{Arc length} * \text{Velocity} - 0,00275 * \text{Arc length} * \text{Wire feed velocity} - 0,00751174 * \text{Velocity} * \text{Wire feed velocity}$$

Завдання з лабораторної роботи. За допомогою розглянутих процедур пакету STATGRAPHICS вирішити завдання дослідження фізичних процесів або технологічних систем з реалізацією регресійної моделі, за певними напрямками:

- Дослідження гідроструминних та струминно-абразивних систем;
- Дослідження засобів для 3-д друку;
- Дослідження металообробних систем із паралельними кінематичними ланцюгами;
- Дослідження несучих конструкцій та елементів із композиційних матеріалів обробного обладнання;
- Дослідження засобів термдеформаційного формування поверхневих шарів та кластерів;
- Дослідження поведінки виробів, виготовлених за функціонально-орієнтованим підходом.

Лабораторна робота складається з наступних етапів її виконання:

1. Грунтуючись на наявних відомостях про досліджуваний процес або систему, необхідно визначитись з методикою для управління експериментом із отриманням уявлення про функцію відгуку, навести схему експерименту та визначити експериментальні дані.
2. За отриманими експериментальними даними відгуку процесу або технологічної системи здійснити аналіз повно факторного експерименту в математичному пакеті STATGRAPHICS, виконати математичну модель у вигляді регресійного рівняння цільової функції відгуку, навести діаграми Парето, графіків впливу факторів та поверхні відгуків цільової функції.

СПИСОК РЕКОМЕНДОВАНОЇ ЛІТЕРАТУРИ

Основна рекомендована література

1. В. В. Мюллер, К. Н. Омеляненко «Синтез демерминированно-статистических моделей формы шва при аргонодуговой сварке неплавящимся электродом», вестн. самар. гос. техн. ун-та. сер. Технические науки. 2017. №3 (55)
Романов В.Н. Техника анализа сложных систем – СПб, - СЗТУ, -2007. -228 с.
2. Соколов В.Ю. Інформаційні системи та технології: навч. посіб./ В.Ю. Соколов; К.: ДУІКТ, 2010. – 138 с.
3. Дубровська Г.М. Системи сучасних технологій./ Г.М. Дубровська, А.П. Ткаченко; - К.: Центр навч. л-ри, 2004.
4. Душинский В.В. Основы научных исследований. – К.: КПИ, 1998 – 240 с.
5. Горопашная А.В. Методы анализа сложных технических систем – Дисс.к.т.н.. – СПб, 2009. – 160 с.
6. Методы параметрической оптимизации структурно-сложных технических системы .- А.А. МУСАЕВ, М.С. СКВОРЦОВ - Санкт-Петербургский институт информатики и автоматизации РАН, 20АО «СПИК СЗМА» 1СПИИРАН, 14-я линия ВО, д. 39, Санкт-Петербург, 199178; 20АО «СПИК СЗМА», пер. Каховского, д. 10, Санкт-Петербург, 199155
7. Пальчевський Б.О. Дослідження технологічних систем (моделювання, проектування, оптимізація): Навч. посібник/ Б.О. Пальчевський. – Львів: Світ, 2001. – 232 с
8. Дибкова Л.М. Інформатика і комп'ютерна техніка: навч. посібник / Л.М. Дибкова. -2-ге вид. перероб., доп.. – К.: Академвидав, 2005. – 416 с.
9. Лук'янова В.В. Комп'ютерний аналіз даних: навч. посібник / В.В. Лук'янова. – Київ: Видавничий центр " Академія", 2003. – 344 с.
10. Гужва В. М. Інформаційні системи і технології на підприємствах: навч. посібник/ В.М. Гужва. — К.: КНЕУ, 2001. — 400 с.
11. Павленко П.М. Інформаційні системи і технології : навч. посіб. / П. М. Павленко, С. Ф. Філоненко, К. С. Бабіч та ін. – К. : НАУ, 2013. – 324 с.

12. Павленко Л.В. Сучасні інформаційні технології: навч. посібник/ Л.В. Павленко, М.П. Павленко, С.В. Хоменко. – Бердянськ: БДПУ, 2017. – 401 с.
13. Дюк В. Обработка данных на ПК в примерах. – С-Пб., 2001. – 240 с. – ил.

Допоміжна

14. Тимейчук О. Ю. Дослідження робочих процесів машин і методи оптимізації: навч. посібник/С. В. Кравець, О. П. Лук'янчук, О. Ю. Тимейчук. – Рівне : НУВГП, 2011. – 240 с.
15. Бочков А.П. Модели и методы управления развитием технических систем / А.П Бочков. – СПб., 2003. – 288 с.
16. Синаторов С.В. Информационные технологии : учеб. пособие / С. В. Синаторов. - М. : Альфа-М, 2011. - 334 с.

ДОДАТОК А
Приклад оформлення титульного аркуша

НАЦІОНАЛЬНИЙ ТЕХНІЧНИЙ УНІВЕРСИТЕТ УКРАЇНИ
«КИЇВСЬКИЙ ПОЛІТЕХНІЧНИЙ ІНСТИТУТ
імені ІГОРЯ СІКОРСЬКОГО»
Механіко-машинобудівний інститут
Кафедра «Конструювання машин»

Звіт з лабораторних робіт

з дисципліни: Наукові дослідження складних технічних систем
(назва модуля (курсу))

Виконав (-ла): студент (-ка) _____ курсу, групи _____
(шифр групи)

(прізвище, ім'я, по батькові) (підпис)

Керівник _____
(посада, науковий ступінь, вчене звання, прізвище та ініціали) (підпис)

Оцінка _____

Кількість балів: _____

Київ – 20__ рік

ДОДАТОК Б
Приклад проведення регресійного аналізу
процесу зварювання

МВ-01м.01КП.001

Схема процесу

Функції геометричних параметрів зварювального шва

$$h = f(V_{пд}, d_{д}, V, I, d_e, L_{д})$$

$$e = f(V_{пд}, d_{д}, V, I, d_e, L_{д})$$

$$g = f(V_{пд}, d_{д}, V, I, d_e, L_{д})$$

Фактори та рівні факторів

Фактори	Рівні факторів	
	-1	+1
X_1 – сила струму, А	155	215
X_2 – довжина дуги, мм	3	7
X_3 – швидкість зварювання, мм/с	2,78	4,2
X_4 – швидкість подачі дроту, мм/с	6	9

Кількість дослідів визначалася за формулою

$$N = 2^k$$

Формули для визначення форми зварного шва

$$h = \frac{V_{пд}^{1,465} d_{д}^{0,92}}{V^{1,9} d_e^{0,078} I^{0,12} L_{д}^{0,05}} \text{ (мм)}$$

$$e = \frac{I^{0,584} L_{д}^{0,433} V_{пд}^{0,095} d_{д}^{0,01}}{V^{0,832} d_e^{0,782}} \text{ (мм)}$$

$$g = \frac{I^{0,659} V_{пд}^{0,247} d_{д}^{0,02}}{V^{0,851} d_e^{2,587} L_{д}^{0,415}} \text{ (мм)}$$

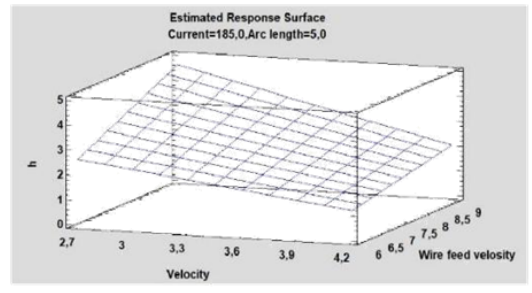
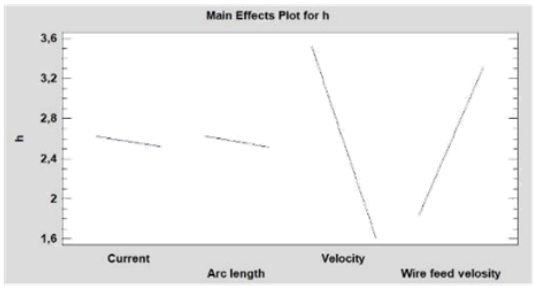
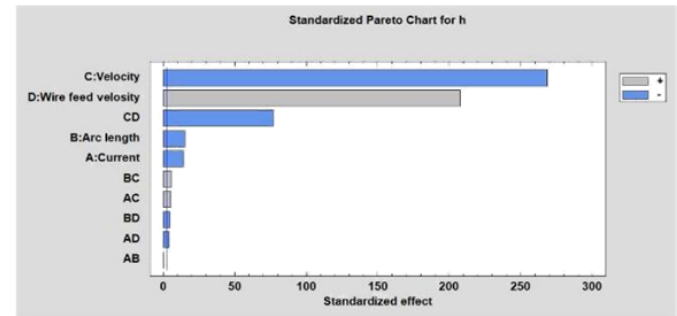
МВ-01м.01КП.001									
№ п/п	№	№	№	№	№	№	№	№	№
Відп.	Відп.	Відп.	Відп.	Відп.	Відп.	Відп.	Відп.	Відп.	Відп.
Відп.	Відп.	Відп.	Відп.	Відп.	Відп.	Відп.	Відп.	Відп.	Відп.
Відп.	Відп.	Відп.	Відп.	Відп.	Відп.	Відп.	Відп.	Відп.	Відп.
Відп.	Відп.	Відп.	Відп.	Відп.	Відп.	Відп.	Відп.	Відп.	Відп.
Відп.	Відп.	Відп.	Відп.	Відп.	Відп.	Відп.	Відп.	Відп.	Відп.
Відп.	Відп.	Відп.	Відп.	Відп.	Відп.	Відп.	Відп.	Відп.	Відп.
Відп.	Відп.	Відп.	Відп.	Відп.	Відп.	Відп.	Відп.	Відп.	Відп.
Відп.	Відп.	Відп.	Відп.	Відп.	Відп.	Відп.	Відп.	Відп.	Відп.
Відп.	Відп.	Відп.	Відп.	Відп.	Відп.	Відп.	Відп.	Відп.	Відп.

Схема процесу та математичний опис

Матриця планування експерименту

№ досліду	X ₁	X ₂	X ₃	X ₄	h	e	g
1	-	+	-	+	4,54	11,48	0,85
2	-	-	+	+	2,16	5,64	0,85
3	+	-	-	-	2,52	9,26	1,36
4	+	-	+	-	1,15	6,57	0,96
5	-	+	-	-	2,51	11,04	0,77
6	-	-	-	+	4,34	7,95	1,21
7	-	+	+	-	1,15	7,84	0,54
8	-	-	-	-	2,62	7,65	1,09
9	-	+	+	+	2,07	8,14	0,60
10	+	+	+	+	1,99	9,86	0,74
11	+	-	-	+	4,56	9,63	1,50
12	+	+	-	+	4,37	13,89	1,06
13	+	-	+	+	2,08	6,83	1,06
14	+	+	-	-	2,41	13,37	0,96
15	-	-	+	-	1,19	5,43	0,77
16	+	+	+	-	1,10	9,49	0,67

Результати експериментів для висоти зварного шва



MB-01m.01K1.002			
№	Дата	Відомості	Висновок
1		Матриця планування експерименту та результати	
2			
3			
4			
5			
6			
7			
8			
9			
10			
11			
12			
13			
14			
15			
16			

MB-01m.01K1.002
 Дата: 15.05.2014
 Автор: [Name]
 Інженер: [Name]
 Підпис: [Signature]

Електронне навчальне видання

Саленко Олександр Федорович
Вакуленко Сергій Валентинович

НАУКОВІ ДОСЛІДЖЕННЯ СКЛАДНИХ ТЕХНІЧНИХ СИСТЕМ

Навчальний посібник

Ум. рук. листів 3,07

Київський політехнічний інститут

ім. Ігоря Сікорського

Київ – 2023