

МІНІСТЕРСТВО ОСВІТИ І НАУКИ УКРАЇНИ
Київський національний університет будівництва і архітектури

**МОНІТОРИНГ ЯКОСТІ ДОВКІЛЛЯ
ТА СТАТИСТИЧНА ОБРОБКА
ЕКСПЕРИМЕНТАЛЬНИХ ДАНИХ
І РЕЗУЛЬТАТІВ НАУКОВИХ ДОСЛІДЖЕНЬ**

Методичні рекомендації
до виконання практичних робіт
для студентів спеціальності 183
«Технології захисту навколишнього середовища»

Київ 2023

УДК 574.4.556.18

М78

Укладачі: І. В. Сатін, канд. техн. наук, доцент;
О. Г. Жукова, канд. техн. наук, доцент

Рецензент О. С. Волошкіна, д-р техн. наук, професор

Відповідальний за випуск Т. М. Ткаченко, д-р техн. наук,
професор

*Затверджено на засіданні кафедри технологій захисту
навколишнього середовища та охорони праці, протокол № 11
від 31 травня 2023 року.*

В авторській редакції.

Моніторинг якості довкілля та статистична обробка
М78 експериментальних даних і результатів наукових досліджень :
методичні рекомендації до виконання практичних робіт / уклад.:
І. В. Сатін, О. Г. Жукова. – Київ : КНУБА, 2023. – 40 с.

Містять нормативні методики статистичної обробки
експериментальних даних щодо оцінки стану навколишнього
середовища внаслідок нераціонального використання природних
ресурсів.

Призначено для студентів освітнього рівня «магістр»
спеціальності 183 «Технології захисту навколишнього середовища».

© КНУБА, 2023

ЗМІСТ

ЗАГАЛЬНІ ПОЛОЖЕННЯ.....	4
Практична робота № 1.....	5
Практична робота № 2.....	12
Практична робота № 3.....	21
Практична робота № 4.....	23
Список літератури.....	38

ЗАГАЛЬНІ ПОЛОЖЕННЯ

Моніторинг якості довкілля невід'ємно пов'язаний з обробкою статистичних даних. Для реалізації наукового моніторингу довкілля, пошуку взаємозв'язків між чинниками та вивчення нових явищ, які виявлені під час здійснення спостережень за станом довкілля, доводиться проводити експерименти. У всіх випадках фахівці, які працюють у системі моніторингу якості довкілля, повинні мати знання та навички зі статистичної обробки експериментальних даних і результатів наукових досліджень.

Фахівці, які здійснюють спостереження за станом довкілля, повинні вміти проводити й контролювати експерименти, а також виключати вплив зовнішніх факторів. Для проведення наукових досліджень важливою характеристикою є точність вимірювань і точність отриманих даних. Кожен дослідник прагне зменшити число впливових факторів і змінних в експерименті, оскільки це прискорює його роботу та зменшує вартість наукових досліджень.

Мета завдання. За результатами виконання практичних робіт студенти повинні набути таких практичних навичок:

1. Навчитися здійснювати спостереження й експерименти в екології. Проводити статистичні спостереження. Оволодіти методами узагальнення статистичної інформації.
2. Проводити зведення та первинне оброблення статистичних даних і ранжування вибірових даних. Уявляти ряди розподілу та їх графічне зображення. Здійснювати статистичну оцінку екологічного стану довкілля.
3. Виконувати дисперсійний і кореляційний аналіз і надати статистичну оцінку істотності зв'язку. Використовувати моделі дисперсійного аналізу.
4. Планувати експеримент, вибирати параметри оптимізації. Обробляти результати експерименту.

Практична робота № 1

Вибірка та генеральна сукупність.

Основи дисперсійного аналізу

Дисперсійний аналіз – статистичний метод виявлення *впливу* окремих факторів на результат експерименту з метою вибору серед них найважливіших. Спочатку дисперсійний аналіз був запропонований англійським статистиком Рональдом Ейлмером Фішером у 1925 році для обробки результатів агрономічних дослідів з виявлення умов, за яких сорт випробовуваної сільськогосподарської культури дає максимальний урожай. Сучасне застосування дисперсійного аналізу охоплює широке коло завдань економіки, біології та техніки і трактується зазвичай у термінах статистичної теорії виявлення систематичних відмінностей між результатами безпосередніх вимірів, виконаних за тих чи інших мінливих умов.

Потрібно розрізняти однофакторний і багатофакторний дисперсійний аналіз. Під час однофакторного аналізу досліджується вплив одного параметра (x_i) на відгук, а в багатофакторному аналізі – кількох параметрів (x_{mn}):

$$y = f \begin{bmatrix} x_{11} & \dots & x_{1n} \\ \dots & \dots & \dots \\ x_{m1} & \dots & x_{mn} \end{bmatrix}.$$

Щоб отримати вичерпну інформацію про стан тієї чи іншої статистичної сукупності, потрібно врахувати її склад. Однак не завжди доводиться вдаватися до суцільного обстеження сукупностей, що вивчаються. Звідси походять поняття про генеральну (загальну) та вибіркочну сукупність ознак. Сукупність, з якої відбираються варіанти для спільного вивчення, називається генеральною (N), а відібрана з генеральної сукупності частина її членів називається вибіркою (n). За властивостями частини (вибірки) судять про чисельні характеристики цілого (генеральної сукупності). Правильно зроблена вибірка досить добре *репрезентує* (представляє) структуру генеральної сукупності. Основною умовою вибіркового методу є *принцип рандомізації*, тобто випадковий відбір варіант із генеральної сукупності з одночасним усуненням суб'єктивних впливів. У

багатьох випадках доводиться розглядати малі вибірки (до 30 спостережень). Уявлення про малі й великі вибірки пов'язані з дослідженнями Вільяма Госсета (1908 р.), який публікувався під псевдонімом *Student*.

Основними характеристиками вибіркової сукупності, що визначають рівень її репрезентативності, є середня арифметична – \bar{x} , вибіркова дисперсія – σ_x^2 та стандартне (середньоквадратичне) відхилення – σ_x .

Вибіркова дисперсія (від лат. *dispersio* – «розсіювання») – міра відхилення від середнього арифметичного значення. У математичній інтерпретації вибіркова дисперсія (σ_x^2) – це середнє арифметичне із квадратів відхилень величин x_i від їхнього середнього арифметичного. На вибірках великого обсягу різниця між (n) і $(n-1)$ практично не позначається, тому вибіркoву дисперсію можна обчислювати за такою формулою:

$$\sigma^2 = \frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^2, \quad (1.1)$$

$$\bar{x} = \frac{1}{n} (x_1 + x_2 + \dots + x_n). \quad (1.2)$$

На вибірках невеликого обсягу (< 30) вибіркова дисперсія має обчислюватися з урахуванням ступеня свободи $(n-1)$ за формулою:

$$\sigma^2 = \frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x})^2. \quad (1.3)$$

Для спрощення техніки проведення дисперсійного аналізу однофакторних комплексів використовують суму квадратів відхилень $D(x)$. При цьому потрібно дати статистичну оцінку за трьома рівнями варіації – усередині серії, між факторами та загальну. Сума квадратів відхилень між факторами визначається за виразом:

$$D_1(x) = n \sum (\bar{x}_i - \bar{x})^2, \quad (1.4)$$

сума квадратів відхилень усередині серії:

$$D_2(x) = \sum \sum (x_i - \bar{x}_i)^2, \quad (1.5)$$

загальну суму квадратів відхилень визначають за виразом:

$$D_3(x) = \sum (x_i - \bar{x})^2. \quad (1.6)$$

Майже у всіх випадках вибіркового спостереження параметри генеральної сукупності залишаються невідомими. Про них доводиться судити за вибірковими даними, які є випадковими величинами. Для оцінки величини генеральних параметрів за вибірковими показниками в науці використовується *нульова гіпотеза*, тобто припущення про те, що генеральні параметри, які визначаються за вибірковими даними, не відрізняються один від одного і що різниця, яка спостерігається між вибірковими показниками, має не систематичний, а виключно випадковий характер.

Для спростування чи підтвердження нульової гіпотези у статистичних сукупностях, які розподіляються за нормальним законом, можна скористатися параметричними критеріями – критеріями Стьюдента (t) і Фішера (F). Для оцінки розбіжностей між вибірками з іншими формами розподілу використовуються критерій Г Уайта і критерій Вілкоксона (Z).

Коефіцієнт варіації. Стандартне відхилення є основною мірою варіабельності (мінливості) ознак. Цей показник можна використовувати для порівняльної оцінки варіювання однорідних ознак. Однак його не можна застосовувати для порівняння варіабельності вибірок із різними одиницями вимірювання. Щоб стандартне відхилення могло бути використане як міра порівняння варіабельності ознак із різними одиницями вимірювання, його виражають у відсотках від середньої арифметичної величини. Отриманий показник є відносним числом, що виражає варіабельність у відсотках (або в частках); його називають коефіцієнтом варіабельності та розраховують за формулою:

$$CV = 100 \frac{\sigma}{\bar{x}}. \quad (1.7)$$

Якщо $CV > 40\%$, то розкид значень ознаки X у досліджуваній вибірці значний. Якщо $CV > 100\%$, то проведені спостереження неоднорідні.

Нормоване відхилення. Нормоване відхилення показує, на скільки сигм (тобто одиниць міри, якою слугує стандартне відхилення) варіанта сукупності відхиляється від середньої ознаки, що варіює $-\pm t\sigma$. Нормоване відхилення (критерій Стьюдента) визначається за виразом:

$$t = \frac{|x - \bar{x}|}{\sigma} \quad (1.8)$$

Будь-яка варіанта, що належить сукупності, яка розподіляється за нормальним законом (розподіл Лапласа – Гауса), може відхилитися від середньої арифметичної до $\pm 3\sigma$ з довірчою імовірністю $P = 99,7\%$.

Приклад

Під час вивчення експлуатаційних якостей фільтрів доочищення питної води виділили один фактор – каламутність одержаної води, мг/дм³. Досліджувалося три завантаження сорбенту за однакових умов. У кожному випробуванні проводили чотири заміри каламутності (табл. 1.1). Чи можна вважати отримані дані за трьома вибірками такими, що належать одній генеральній сукупності? Дати статистичну оцінку значення каламутності 0,8 мг/дм³, отриманого в наступному дослідженні роботи завантаження № 2 фільтра.

Таблиця 1.1

Експериментальні дані каламутності води за трьома завантаженнями фільтра

Завантаження	Каламутність отриманої води, мг/дм ³			
	<i>1</i>	<i>2</i>	<i>3</i>	<i>4</i>
1	0,85	0,90	0,95	0,8
2	0,7	0,75	0,7	0,75
3	0,80	0,85	0,90	0,75

Розв'язання

1. Для зручності виконання роботи потрібно заповнити табл. 1.2,

Таблиця 1.2

Математична обробка отриманих даних

№ з/п	x_{i1}	$(x_{i1})^2$	x_{i2}	$(x_{i2})^2$	x_{i3}	$(x_{i3})^2$	x_{i4}	$(x_{i4})^2$	$\sum_{j=1}^{nk} x_{ij}$	$\left(\sum_{j=1}^{nk} x_{ij}\right)^2$	$\sum (x_{ij})^2$
...
Σ											

де i – номер серії спостережень; k – кількість серій спостережень, $k = 3$;
 j – номер спостереження у серії; n_i – кількість спостережень в i -ї серії,
 $n_i = 4$; x_{ij} – одержані під час випробування показники каламутності води,
 мг/дм³.

2. Визначаємо середнє арифметичне значення отриманих даних каламутності води в кожній серії випробувань, мг/дм³:

$$\bar{x}_i = \frac{1}{n} \sum_{j=1}^{n_i} x_{ij} . \quad (1.9)$$

3. Розраховуємо загальне середнє арифметичне значення всіх спостережень, мг/дм³:

$$\bar{x} = \frac{1}{nk} \sum_{i=1}^k (\bar{x}_i \cdot n_i) . \quad (1.10)$$

4. Визначаємо вибіркoву дисперсію одержаних результатів. Розрахунок виконуємо у формі табл. 1.3.

Таблиця 1.3

Дисперсійний аналіз результатів експерименту

Джерело варіації	Сума квадратів відхилень варіант, $D(x)$	Число ступенів свободи, f	Вибіркова дисперсія, σ^2
Між факторами	$D_1(x) = \frac{1}{n_i} \Sigma(\Sigma x_{i,j})^2 - \frac{1}{k \cdot n_i} (\Sigma(\Sigma x_{i,j}))^2$	$f_1 = k - 1$	$\frac{D(x)}{f}$
Усередині серії	$D_2(x) = \Sigma \Sigma x_{i,j}^2 - \frac{1}{n_i} \Sigma(\Sigma x_{i,j})^2$	$f_2 = k \cdot (n - 1)$	
Загальна	$D_3(x) = \Sigma \Sigma x_{i,j}^2 - \frac{1}{k \cdot n_i} (\Sigma(\Sigma x_{i,j}))^2$	$f_3 = n \cdot k - 1$	

5. Для перевірки достовірності нульової гіпотези H_0 використовують критерій Фішера:

$$F_{\text{набл}} = \frac{\sigma_3^2}{\sigma_2^2}. \quad (1.11)$$

6. За отриманими значеннями числа ступенів свободи f_1 і f_2 для рівня значущості гіпотези (задаємо $\alpha = 0,05$) визначаємо табличне значення критерію Фішера (додаток А):

$$F_{\alpha}(f_1; f_2).$$

7. Якщо $F_{\text{набл}} > F_{\alpha}(f_1; f_2)$, то нульова гіпотеза відкидається, різниця між порівнюваними величинами визнається статистично достовірною, а значення кожної з вибірок не можна вважати в середньому однаковими (тобто вибірки не належать до однієї генеральної сукупності). Якщо $F_{\text{набл}} \leq F_{\alpha}(f_1; f_2)$, то отримані значення належать до однієї генеральної сукупності, нульова гіпотеза зберігається, а розбіжності між вибірками є статистично недостовірними.

8. Розрахуємо коефіцієнт варіації та проаналізуємо варіабельність ознаки X (табл. 1.4).

Таблиця 1.4

Аналіз варіабельності ознаки X

Заміри	$\bar{x}, \text{мг/дм}^3$	$x_i - \bar{x}_i$	$(x_i - \bar{x}_i)^2$	σ	CV %
...					
$\Sigma(x_i - \bar{x}_i)^2$					

9. Розрахуємо нормоване відхилення для оцінки значення каламутності 0,8 мг/дм³ за формулою 1.8.

Після закінчення розрахунку дослідник робить висновок відповідно до поставленої мети, ґрунтуючись на отриманих статистичних показниках.

Індивідуальне завдання

Задача 1 (для варіантів 1–15). Досліджуються експлуатаційні показники повітророзподільних решіток. Під час аналізу виділяють один фактор – коефіцієнт місцевого опору. Було випробувано чотири повітророзподільні решітки. Зроблено по п'ять замірів для кожної розподільної решітки. Чи можна вважати отримані дані в середньому однаковими для всіх досліджень? Провести дисперсійний аналіз для рівнів значимості $\alpha = 0,01; 0,05; 0,10$. Дати оцінку варіабельності. Визначити відхилення коефіцієнта місцевого опору для ВР-2 $\xi = 6$. Зробити висновок.

Таблиця 1.5

Вихідні дані (для варіантів 1–15)

Тип повітророзподільних решіток	Коефіцієнти місцевих опорів, ξ				
	1	2	3	4	5
ВР-1	7,2	7,0	8,2	9,5	8,9
ВР-2	6,9	5,2	7,3	7,0	6,5
ВР-3	5,3	5,2	6,5	10,3	7,4
ВР-4	6,6	5,9	3,2	8,0	7,9

Задача 2 (для варіантів 16–30). Досліджуються експлуатаційні показники пристроїв захисного відключення (ПЗВ). Виділяється один фактор – час спрацьовування ПЗВ. Проведено випробування трьох апаратів із п'ятьма замірами в кожній серії випробувань. Чи можна вважати отримані дані в середньому однаковими для всіх досліджень? Провести дисперсійний аналіз для рівнів значимості $\alpha = 0,01; 0,05; 0,10$. Дати оцінку варіабельності. Визначити відхилення значення часу спрацьовування для ПЗВ-3 $k = 117$ мс. Зробити висновок.

Таблиця 1.6

Вихідні дані (для варіантів 16–30)

Номер пристрою	Час спрацьовування, мс				
	1	2	3	4	5
ПЗВ -1	210	220	190	150	100
ПЗВ -2	100	350	150	170	210
ПЗВ -3	250	310	100	440	200

Практична робота № 2
Основи однофакторного кореляційно-регресійного аналізу.
Лінійна модель регресії

Кореляція (від пізньолат. *correlatio* – «співвідношення») – термін, що застосовується в різних галузях науки і техніки для позначення взаємозалежності, взаємної відповідності. Кореляція – це ймовірнісна залежність. Вона виникає тоді, коли залежність однієї з ознак ускладнюється наявністю ряду випадкових факторів.

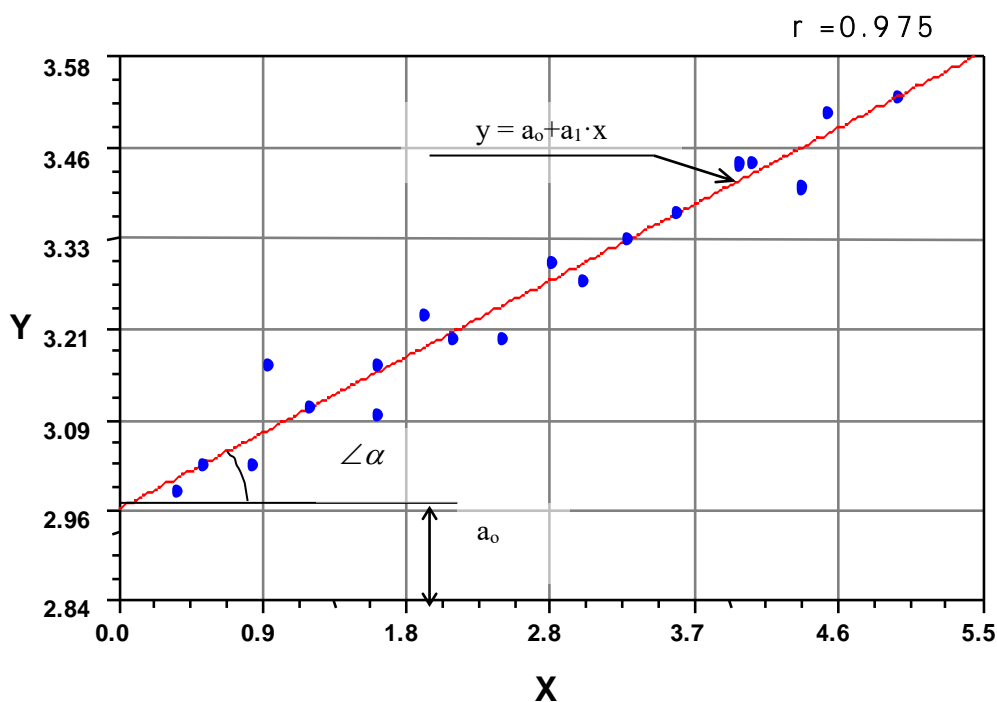


Рис. 2.1. Лінійна регресія

Кореляційний аналіз призначений для оцінювання значимості впливу чинників (x_1, \dots, x_i) на відгук (y_i). Кореляційний аналіз дає змогу встановити значущість зміни спостережень y_i та ступінь їхньої залежності від випадкової величини x_i . Найбільш уживаний в оцінці ступеня залежності лінійний коефіцієнт кореляції (коефіцієнт кореляції Пірсона) між ознаками X і Y за наявності лінійної залежності:

$$r = \frac{\overline{xy} - \bar{x} \cdot \bar{y}}{\sigma_x \sigma_y}, \quad (2.1)$$

де σ_x – стандартне відхилення варіаційного ряду факторної ознаки X ,
 σ_y – стандартне відхилення варіаційного ряду результативної ознаки Y .

$$\sigma_x = \sqrt{\frac{\sum_{i=1}^n x_i^2}{n} - \left(\frac{\sum_{i=1}^n x_i}{n}\right)^2} \quad \text{і} \quad \sigma_y = \sqrt{\frac{\sum_{i=1}^n y_i^2}{n} - \left(\frac{\sum_{i=1}^n y_i}{n}\right)^2}, \quad (2.2), (2.3)$$

де n – кількість отриманих предикторних змінних (кількість дослідів).

Лінійний коефіцієнт кореляції завжди набуває значення $r \in [-1; 1]$. Якщо $r > 0$, то зв'язок між X і Y прямий (у протилежному випадку – зворотний), якщо $0.3 \leq r \leq 0.5$, зв'язок визнається помірним, якщо ж $0.5 \leq r \leq 0.75$, кореляція вважається значною, якщо $r > 0.75$, то зв'язок можна вважати тісним. Якщо $r = 1$, зв'язок можна вважати функціональним.

Значимість лінійного коефіцієнта кореляції за малої кількості дослідів ($n < 100$) визначають за допомогою t -розподілу Стьюдента з $(n-2)$ ступенями свободи, для чого розраховуємо t -критерій Стьюдента, висуваючи нульову гіпотезу H_0 , що зв'язок не спостерігається між ознаками X і Y :

$$t_{\text{розр.}} = \sqrt{\frac{r^2}{1-r^2}} (n-2). \quad (2.4)$$

За таблицями критичних значень t -критерію Стьюдента (додаток Б) залежно від рівня значущості α та кількості ступенів свободи $f = n - 2$ визначаємо табличне значення коефіцієнта Стьюдента ($t_\alpha(n-2)$). Отримане значення ($t_{\text{розр.}}$) порівнюємо з табличним $t_\alpha(n-2)$. Якщо $t_{\text{розр.}} \geq t_\alpha(n-2)$, то відкидаємо нульову гіпотезу H_0 , тобто коефіцієнт кореляції значимий, і якщо $t_{\text{розр.}} \leq t_\alpha(n-2)$, то коефіцієнт кореляції неадекватно відображає ступінь залежності між ознаками Y і X , а нульова гіпотеза зберігається.

Адекватність коефіцієнта кореляції Пірсона за досить великої кількості спостережень ($n \geq 100$) визначається за формулою:

$$t_{\text{розр.}} = \frac{r}{\xi_r}, \quad (2.5)$$

де ξ_r – помилка коефіцієнта кореляції, яка визначається за формулою:

$$\xi_r = \frac{1-r^2}{\sqrt{n}}, \quad (2.6)$$

де n – кількість дослідів;

r – коефіцієнт кореляції Пірсона.

Регресія – залежність середнього значення будь-якої величини від деякої іншої або декількох величин. Спочатку термін «регресія» був ужитий англійським психологом і лікарем Френсісом Гальтоном (1886 р.) в теорії спадковості у спеціальному сенсі: *regression to mediocrity* («поверненням до середнього стану») було названо явище, яке полягає в тому, що діти тих батьків, ріст яких перевищує середнє значення на a одиниць, мають у середньому ріст, що перевищує середнє значення менше ніж на a одиниць.

Регресійний аналіз призначений для визначення аналітичної залежності відгуку у від кількісних чинників x_i з вхідного набору і визначає точні кількісні характеристики зміни y . Мета регресійного аналізу полягає у визначенні загального виду рівняння регресії, побудові оцінок невідомих параметрів, що входять до рівняння регресії, і перевірці статистичних гіпотез про регресію.

У статистиці виділяють два основні класи регресії – лінійні та нелінійні. Лінійна регресія описується за допомогою рівняння $\tilde{y} = a_0 + a_1 \cdot x_i$ (рис. 2.1), яка характеризується коефіцієнтами рівняння регресії a_0 і $a_1 = \text{tg}(\acute{\alpha})$. Пряму чи криву, побудовану з допомогою рівняння регресії називають прямою чи кривою регресії. У математичній статистиці коефіцієнти рівняння регресії знаходять за допомогою методу найменших квадратів:

$$S = \sum_{i=1}^n (y_i - \tilde{y})^2 \rightarrow \min, \quad (2.7)$$

$$S = \sum_{i=1}^n (y_i - a_0 - a_1 \cdot x_i)^2 \rightarrow \min, \quad (2.8)$$

тобто сума квадратів (S) має прямувати до нуля. Якщо продиференціювати рівняння (2.8) по a_0 і по a_1 , то отримаємо часткові похідні 1-го порядку:

$$\frac{\partial S}{\partial a_0} = 2 \sum (y_i - a_0 - a_1 \cdot x_i) \cdot (-1) = 0; \quad (2.9)$$

$$\frac{\partial S}{\partial a_1} = 2 \sum (y_i - a_0 - a_1 \cdot x_i) \cdot (-x_i) = 0. \quad (2.10)$$

Розв'язавши систему диференціальних рівнянь із частковими похідними 1-го порядку (2.9) та (2.10), отримаємо систему нормальних рівнянь для лінійної регресії:

$$\left\{ \begin{array}{l} n \cdot a_0 + a_1 \sum_{i=1}^n x_i = \sum_{i=1}^n y_i \\ a_0 \sum_{i=1}^n x_i + a_1 \sum_{i=1}^n x_i^2 = \sum_{i=1}^n y_i x_i \end{array} \right\}. \quad (2.11)$$

Отже, вирішуючи систему нормальних рівнянь, можна визначити коефіцієнти рівняння регресії.

Показник лінійності зв'язку. Оскільки коефіцієнт кореляції Пірсона характеризує лише лінійний зв'язок, а кореляційне відношення – будь-яку форму зв'язку, то за наявності лінійної залежності між ознаками X і Y має виконуватися рівність $\eta = r$. Тоді:

$$\gamma = \eta_{y/x}^2 - r^2, \quad (2.12)$$

де $\eta_{y/x}^2$ – кореляційне відношення Y по X , що визначається за формулою:

$$\eta_{y/x}^2 = \frac{\sum (\tilde{y} - \bar{y})^2}{\sum (y_i - \bar{y})^2}. \quad (2.13)$$

Критерієм достовірності показника лінійності зв'язку є його відношення до помилки:

$$t_\gamma = \frac{\gamma}{m_\gamma}, \quad (2.14)$$

$$m_\gamma = \frac{2\sqrt{\gamma - \gamma^2(2 - \eta^2 - r^2)}}{\sqrt{n}}. \quad (2.15)$$

На малих вибірках m_γ потрібно обчислювати з урахуванням ступенів свободи $n-1$ у підкореновому виразі. Отриманий критерій t_γ порівнюємо з табличним значенням $t_\alpha(n-2)$ (за таблицями розподілу Стюдента), для рівня значимості α і $f=n-2$. Якщо $t_\gamma \leq t_\alpha(n-2)$, то отриманий критерій є достовірним. Якщо $t_\gamma < 2.5$, то можна стверджувати про наявність прямолінійного зв'язку, а якщо $t_\gamma = 0$, є лише прямолінійна залежність.

Довірча зона лінійної регресії (рис. 2.2). Максимальна похибка репрезентативності вибіркової лінії регресії може бути виражена у вигляді такого рівняння:

$$\tilde{y} = (a_1 + a_2x) \pm t_\alpha(n-2) \cdot \sigma_{y/x}, \quad (2.16)$$

де $\sigma_{y/x}$ – похибка рівняння регресії:

$$\sigma_{y/x} = \sigma_y \sqrt{1-r^2}, \quad (2.17)$$

де $\psi = \sqrt{1-r^2}$ – коефіцієнт алієнації, тобто поправка на сполученість між ознаками X та Y; σ_y – стандартне відхилення ряду Y.

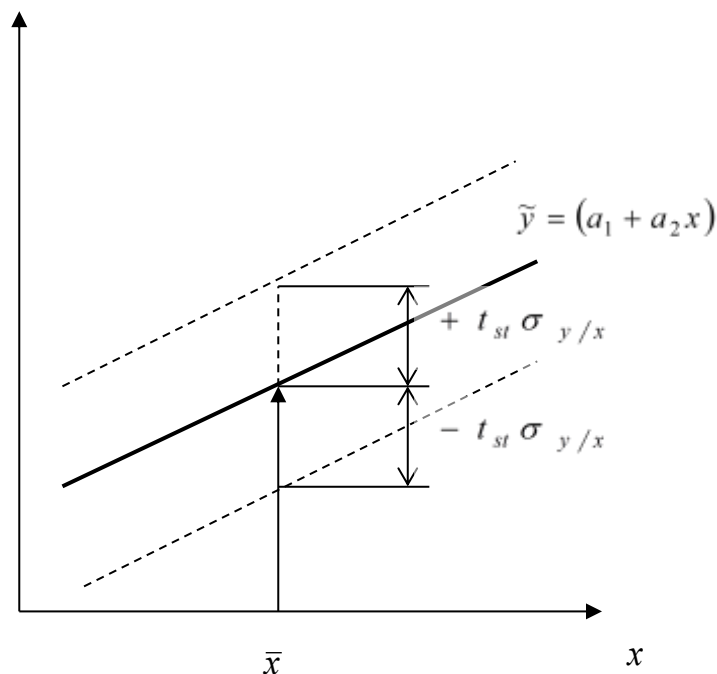


Рис. 2.2. Довірча зона лінійної регресії

Приклад

Нехай за результатами експерименту отримані дані залежності результату Y від фактора X (табл. 2.1). Потрібно розрахувати коефіцієнт кореляції Пірсона, визначити його значущість за t -критерієм Стюдента, скласти рівняння лінійної регресії, нанести на графік експериментальні точки та побудувати пряму рівняння регресії $\hat{y} = a_0 + a_1x_i$; перевірити адекватність отриманого рівняння за критерієм Фішера.

Таблиця 2.1

Дані експерименту

x	0	1	2	3	4	5
y	2,90	3,15	3,22	3,27	3,43	3,52

Розв'язання

1. Зобразимо експериментальні точки (рис. 2.3). Первинна оцінка на око розташування точок дає змогу припустити наявність лінійної залежності між X та Y .

2. Для зручності розрахунку коефіцієнта кореляції заповнюємо табл. 2.3.

Таблиця 2.3

Розрахунок допоміжних значень системи нормальних рівнянь

Номер досліду	x	y	x^2	xy	y^2
...					
Сума					
Середнє значення					

3. Визначається стандартне відхилення (формули 2.2 та 2.3) і коефіцієнт кореляції Пірсона (формула 2.1).

4. Перевіряємо значущість коефіцієнта кореляції за допомогою t -критерію Стюдента, визначаючи розрахункове значення коефіцієнта Стюдента (формула 2.4).

За таблицею критичних значень t -критерію Стюдента (додаток Б) залежно від рівня значущості (задаємося $\alpha = 0,05$) та кількості ступенів

свободи f визначаємо табличне значення коефіцієнта Стюдента $t_{\alpha}(f)$. Проводиться порівняння, робиться висновок.

5. Складаємо рівняння лінійної регресії $\tilde{y} = a_0 + a_1x$ і знаходимо числові значення його коефіцієнтів a_0 і a_1 , які розраховуються із системи нормальних рівнянь (2.11).

6. Нанесемо отримані експериментальним шляхом точки на графік (рис. 2.3) і в тій самій системі координат збудуємо пряму лінію $\tilde{y} = f(x)$. Для цього заповнимо табл. 2.4.

Таблиця 2.4

Розрахунок значень прямої регресії (за отриманим рівнянням регресії)

x	0	1	2	3	4	5
y	2,90	3,15	3,22	3,27	3,43	3,52
\tilde{y}

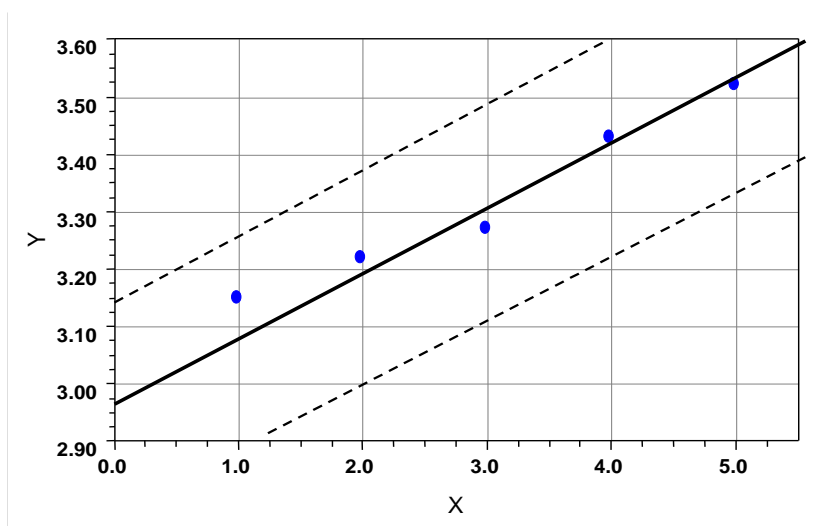


Рис. 2.3. Побудова прямої регресії з нанесенням експериментальних точок

7. Перевіряємо отримане рівняння регресії на адекватність реальним умовам за критерієм Фішера. Для цього заповнюємо табл. 2.5.

Таблиця 2.5

Допоміжні дані для розрахунку дисперсії

Номер досліду	x	y	\tilde{y}	$\Delta\acute{o} = \acute{o} - \grave{o}$	Δy^2	$y - \bar{y}$	$(y - \bar{y})^2$
...							
Сума				-		-	
Середнє	-		-	-	-	-	-

Обчислюємо залишкову дисперсію адекватності за формулою:

$$S_{ocm}^2 = \frac{m \sum (\Delta y^2)}{f}, \quad (2.11)$$

де f – число ступенів свободи, $f = n - (k+1)$; n – кількість дослідів; m – кількість дослідів у кожній серії спостережень; k – кількість коефіцієнтів рівняння регресії.

Визначаємо дисперсію стосовно середнього за формулою:

$$S_y^2 = \frac{\sum (y - \bar{y})^2}{n(k-1)}. \quad (2.12)$$

Розраховуємо критерій Фішера за виразом (якщо немає паралельних дослідів, інакше потрібно аналізувати за критерієм Кохрена):

$$F = \frac{S_{\tilde{m}\tilde{o}}^2}{S_{\tilde{o}}^2}. \quad (2.13)$$

За додатком А визначаємо табличне значення критерію Фішера $F_\alpha(f_1; f_2)$ (за рівня значущості $\alpha = 0,05$) залежно від ступенів свободи: $f_1 = n - 1$; $f_2 = n - k - 1$.

Аналізуємо отримані значення F і $F_\alpha(f_1; f_2)$ і робимо висновок щодо адекватності отриманого рівняння регресії реальним умовам.

8. Визначаємо показник лінійності зв'язку. Для цього зробимо розрахунок у формі табл. 2.6.

Таблиця 2.6

Розрахунок допоміжних даних для визначення показника лінійності

Номер дослідів	y	\tilde{y}	$(\tilde{y} - \bar{y})$	$(\tilde{y} - \bar{y})^2$	$y - \bar{y}$	$(y - \bar{y})^2$
...						
Сума			-		-	
Середнє		-	-	-	-	-

Далі для розрахунку використовуємо наведені формули 2.12–2.15. Робимо висновок про достовірність критерію та ступеня лінійності зв'язку.

10. Визначаємо довірчу зону лінійної регресії за формулами 2.16 і 2.17.

Нанесемо отриману зону графік (рис. 2.3).

Індивідуальне завдання

За результатами проведення експерименту отримано дані залежності результативної ознаки y від факторіального x (табл. 2.7). Потрібно розрахувати коефіцієнт кореляції Пірсона, вказати похибку, визначити його значущість за t -критерієм Стьюдента, скласти рівняння лінійної регресії, нанести на графік експериментальні точки та побудувати пряму рівняння регресії $\hat{y} = a_0 + a_1x_i$; перевірити адекватність отриманого рівняння за критерієм Фішера. Визначити ступінь лінійності зв'язку та довірчу зону лінійної регресії. Зробити загальний висновок.

Таблиця 2.7

Експериментальні дані

x	0 + В	1 + В	2 + В	3 + В	4 + В	5 + В
y	1,2 + В	1,9 + В	3,6 + В	4,8 + В	5,2 + В	7,2 + В

В – номер варіанта студента.

Практична робота № 3
Багатофакторний регресійний аналіз.
Множинна кореляція

Багатофакторний регресійний аналіз передбачає дослідження залежності результативної ознаки від двох і більше предикторних змінних. Рівняння залежності $\tilde{y} = f(x_1; x_2 \dots x_i)$ називається багатофакторним рівнянням регресії. Залежність між ознаками може бути лінійною та нелінійною.

Розглянемо лінійну залежність результативної ознаки у двох предикторних x_1 і x_2 . Рівняння регресії у цьому випадку:

$$\tilde{y} = a_0 + a_1x_1 + a_2x_2. \quad (3.1)$$

Для визначення коефіцієнтів рівняння регресії потрібно розв'язати систему трьох рівнянь:

$$\left\{ \begin{array}{l} na_0 + a_1 \sum_{i=1}^n x_1 + a_2 \sum_{i=1}^n x_2 = \sum_{i=1}^n y \\ a_0 \sum_{i=1}^n x_1 + a_1 \sum_{i=1}^n x_1^2 + a_2 \sum_{i=1}^n (x_1 \cdot x_2) = \sum_{i=1}^n y \cdot x_1 \\ a_0 \sum_{i=1}^n x_2 + a_1 \sum_{i=1}^n x_1 \cdot x_2 + a_2 \sum_{i=1}^n x_2^2 = \sum_{i=1}^n x_1 \cdot x_2 \end{array} \right. \quad (3.2)$$

Для оцінки ступеня впливу предикторних змінних на величину відгуку обчислюється сукупний коефіцієнт множинної кореляції за формулою:

$$R = \sqrt{\frac{r_{x_1x_2}^2 + r_{yx_2}^2 - 2 \cdot r_{yx_1} \cdot r_{yx_2} \cdot r_{x_1x_2}}{1 - r_{yx_1}^2}}. \quad (3.3)$$

Часткові лінійні коефіцієнти кореляції визначаються за співвідношенням:

$$r_{x_1x_2} = \frac{\sum((x_1 - \bar{x}_1)(x_2 - \bar{x}_2))}{n\sigma_{x_1}\sigma_{x_2}}. \quad (3.4)$$

$$r_{yx1} = \frac{\sum((x_1 - \bar{x}_1)(y - \bar{y}))}{n\sigma_{x1}\sigma_y}. \quad (3.5)$$

$$r_{yx2} = \frac{\sum((x_2 - \bar{x}_2)(y - \bar{y}))}{n\sigma_{x2}\sigma_y}. \quad (3.6)$$

Розрахунок виконується аналогічно до вищенаведених методик.

Індивідуальне завдання

За результатами проведення двофакторного експерименту отримані дані, наведені в табл. 3.1.

Таблиця 3.1

Вихідні дані

y	25,6	25,7	25,8	25,9	26,1	27,5	28,1	29,3	30,8	32,4	33,5	35,4	36,7	38,1	40,0
x₁	12,2	13,5	18,4	21,6	23,6	28,9	34,7	39,5	45,8	48,1	50,2	60,7	70,2	80,4	90,8
x₂	1,2	1,3	1,8	1,9	2,5	3,9	4,8	5,9	6,6	7,8	9,1	12,5	23,5	32,4	39,1

Потрібно визначити багатофакторне рівняння регресії та сукупний коефіцієнт кореляції. Зробити висновок.

Практична робота № 4

Математичне планування експерименту

Багатофакторний регресійний аналіз передбачає дослідження залежності результативної ознаки від двох і більше змінних.

Планування експерименту – це процедура вибору числа й умов проведення дослідів, потрібних і достатніх для вирішення поставленого завдання з відповідною точністю. При цьому проводиться таке:

- 1) мінімізація загальної кількості дослідів;
- 2) одночасне варіювання всіма змінними, що визначають процес, за спеціальними алгоритмами;
- 3) використання математичного апарата, який формалізує багато дій експериментатора (отримання інтерполяційної формули);
- 4) вибір чіткої стратегії, що дає змогу приймати обґрунтовані рішення після кожної серії експериментів.

Пошук оптимальних умов, побудова інтерполяційних формул, вибір істотних факторів, оцінка й уточнення констант теоретичних моделей – ось приклади завдань, для вирішення яких застосовується планування експерименту.

Розглянемо застосування методів планування експерименту на прикладі вивчення процесу корозії сталевих трубопроводів. Рекомендується така послідовність для дослідження поверхні відгуку з метою отримання інтерполяційної формули.

Попереднє планування експерименту

1. Загальна інформація про процес. З виробничого досвіду чи досвіду інших учених виділяється проблема, у цьому випадку проблема корозії водопровідних трубопроводів. Аналіз інформації з друкованих, рукописних, актових та недокументальних джерел дає змогу визначити суть процесу, можливі методи дослідження, точність вимірювань, переваги та недоліки попередніх робіт. Дослідження показують, що на пришвидшення процесу корозії впливають такі чинники: збільшення концентрації кисню у воді, підвищення температури середовища, підвищення концентрації кислот і ангідридів кислот (CO_2 ; SO_2 ; H_2S), поперемина дія рідких і газових середовищ; збільшення концентрації сульфідів, солевмісту та хлоридів, зниження рН води, різка зміна режиму руху середовища, а також життєдіяльність мікроорганізмів.

2. *Вибір факторів.* Спочатку дослідник аналізує безліч чинників, відомих за літературними даними чи з досвіду експериментів. Можна виділити основні умови, фізичний зміст яких зрозумілий. Потрібно включати до розгляду всі відомі фактори. Значущість того чи іншого фактору визначається експериментальним методом, методами апріорного ранжування, випадкового балансу чи планами Плакетта – Бермана. Якщо якийсь суттєвий фактор виявиться неврахованим і надалі флюктує (приймає випадкові значення), то це призведе до збільшення похибки досліду. Фіксація ж неврахованого суттєвого фактора може призвести до помилкового уявлення про екстремальні значення отриманої залежності. Для планування експерименту фактори повинні мати такі параметри: *керовані* – тобто параметри, які експериментатор може змінювати; *однозначні* – тобто не бути результатом функції з інших вимірювань; *сумісні* – комбінація параметрів можлива та безпечна; *незалежні* – тобто кореляція між факторами повинна бути відсутня (достатньо, щоб зв'язок був нелінійним).

Отже, ми відібрали такі фактори, що варіюються, які впливають на швидкість корозії бетону: x_1 – температура середовища, $^{\circ}\text{C}$; x_2 – концентрація сульфатів, $\text{мг}/\text{дм}^3$; x_3 – концентрація хлоридів, $\text{мг}/\text{дм}^3$. Відгуком є швидкість корозії – y , $\text{мг}/(\text{см}^2 \cdot \text{доб})$.

3. *Вибір параметра оптимізації та методики експерименту.* У цьому випадку оптимізація не проводиться. Метод проведення експерименту через його трудомісткість і складність не описуємо.

Вибір умов проведення дослідів

Визначаємо область визначення факторів, у якій доцільно проводити дослідження. У нашому випадку область визначення факторів повинна відображати реальні умови роботи водопровідної мережі. За даними досліджень, температура середовища (x_1) у водопровідній мережі варіюється протягом року від 5 до 25°C ; концентрація сульфат-іонів (x_2) – від 150 до 230 $\text{мг}/\text{дм}^3$; концентрація хлоридів (x_3) – від 60 до 120 $\text{мг}/\text{дм}^3$.

Вибираємо рівні й інтервали варіювання. Основний рівень доцільно встановлювати в центрі діапазону варіювання значень фактора. Вузкий інтервал варіювання (не більше ніж 10 % зони визначення) призначають за точних обчислень, середній інтервал варіювання (від 10 до 30%) – за

вимірювань, виконаних із середньою точністю. Результати розрахунку занесено до табл. 4.1.

Таблиця 4.1

Рівні та інтервали варіювання факторів

Рівні	Фактори		
	x_1	x_2	x_3
Основний	15	190	90
Інтервал варіювання	10	40	30
Верхній (+1)	25	230	120
Нижній (-1)	5	150	60

Складання плану експерименту

Допустимо, що у вибраній зоні серії дослідів лінійна модель виявиться адекватною. Тоді в нашому випадку шуканий план має бути дворівневим. Цій вимозі задовольняє факторний експеримент 2^3 (має бути проведено $2^3 = 8$ дослідів). Матрицю планування наведено в табл. 4.2.

Таблиця 4.2

Матриця планування дослідів і результати дослідів

Номер дослідів	Кодовані фактори			Натуральні фактори			Результати експерименту		
	x_1	x_2	x_3	\tilde{x}_1	\tilde{x}_2	\tilde{x}_3	y_{i1}	y_{i2}	y_{i3}
1	–	–	–	5	150	60	0,05	0,06	0,04
2	+	–	–	25	150	60	0,25	0,25	0,27
3	–	+	–	5	230	60	0,08	0,09	0,09
4	+	+	–	25	230	60	0,30	0,31	0,33
5	–	–	+	5	150	120	0,07	0,06	0,07
6	+	–	+	25	150	120	0,29	0,27	0,27
7	–	+	+	5	230	120	0,11	0,12	0,14
8	+	+	+	25	230	120	0,36	0,38	0,37

Умови проведення дослідів у натуральних змінних зазначені в табл. 4.2. Перехід від натуральних значень до кодованих задається формулою:

$$x_i = \frac{\tilde{x}_i - x_0}{\Delta x}, \quad (4.1)$$

де \tilde{x}_i – значення факторів на межі інтервалу варіювання;

x_0 – значення основного рівня;

Δx – інтервал варіювання фактора.

$$x_1 = \frac{\tilde{x}_1 - 15}{10}; \quad x_2 = \frac{\tilde{x}_2 - 190}{40}; \quad x_3 = \frac{\tilde{x}_3 - 90}{30}.$$

Обробка результатів експерименту

Обробку результатів проводимо в такій послідовності:

а) оцінка дисперсії в кожній серії вимірювань (рядку матриці):

$$s_i^2 = \frac{\sum_{i=1}^n (y_i - \bar{y}_i)^2}{(m-1)}, \quad (4.2)$$

де m – кількість паралельних дослідів;

Таблиця 4.3

Розрахунок дисперсії середнього арифметичного значення

Номер дослідів	y_i			\bar{y}_i	$y_{i,m} - \bar{y}_i$			$\Sigma (y_{i,m} - \bar{y}_i)^2$	s_i^2
	$y_{i,1}$	$y_{i,2}$	$y_{i,3}$						
1	0,05	0,06	0,04	0,05	0	0,01	-0,01	0,0002	0,0001
2	0,25	0,25	0,27	0,257	-0,007	-0,007	0,013	0,000267	0,000134
3	0,08	0,09	0,09	0,0825	-0,003	0,008	0,008	0,000137	0,000059
4	0,30	0,31	0,33	0,313	-0,013	-0,003	0,017	0,000467	0,000234
5	0,07	0,06	0,07	0,066	0,004	-0,006	0,004	0,000068	0,000034
6	0,29	0,27	0,27	0,279	0,011	-0,009	-0,009	0,000283	0,000142
7	0,11	0,12	0,14	0,123	-0,013	-0,003	0,017	0,000467	0,000234
8	0,36	0,38	0,37	0,37	-0,01	0,01	0	0,0002	0,0001
									0,00104

б) перевірка гіпотези про однорідність дисперсій за допомогою критерію Кохрена:

$$G = \frac{s_{\max}}{\sum_1^N s_i^2}, \quad (4.3)$$

$$G = \frac{0,000234}{0,00104} = 0,225.$$

За додатком Ж визначаємо табличне значення критерію Кохрена $G(8;3-1) = 0.6152$. Оскільки розрахункове значення критерію Кохрена менше за табличне, гіпотеза про однорідність дисперсій підтверджується;

в) якщо дисперсії однорідні, то наводиться розрахунок оцінки дисперсії відтворюваності

$$S_{\{\bar{y}\}}^2 = \frac{\sum_1^N s_i^2}{n}, \quad (4.4)$$

$$S_{\{\bar{y}\}}^2 = \frac{0,00104}{8} = 0.00013.$$

Кількість ступенів свободи дисперсії відтворюваності:

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

г) визначення коефіцієнтів регресії

$$b_j = \frac{\sum \bar{y}_i x_{ji}}{n}, \quad (4.5)$$

$$b_0 = \frac{0,05 + 0,257 + 0,0825 + 0,313 + 0,066 + 0,279 + 0,123 + 0,37}{8} = 0,193$$

$$b_1 = \frac{-0,05 + 0,257 - 0,0825 + 0,313 - 0,066 + 0,279 - 0,123 + 0,37}{8} = 0,112$$

$$b_2 = \frac{-0,05 - 0,257 + 0,0825 + 0,313 - 0,066 - 0,279 + 0,123 + 0,37}{8} = 0,03$$

$$b_3 = \frac{-0,05 - 0,257 - 0,0825 - 0,313 + 0,066 + 0,279 + 0,123 + 0,37}{8} = 0,017$$

$$b_{12} = \frac{+0,05 - 0,257 - 0,0825 + 0,313 + 0,066 - 0,279 - 0,123 + 0,37}{8} = +0,007$$

$$b_{13} = \frac{+0,05 - 0,257 + 0,0825 - 0,313 - 0,066 + 0,279 - 0,123 + 0,37}{8} = 0,003$$

$$b_{23} = \frac{+0,05 + 0,257 - 0,0825 - 0,313 - 0,066 - 0,279 + 0,123 + 0,37}{8} = 0,007$$

$$b_{123} = \frac{-0,05 + 0,257 + 0,0825 - 0,313 + 0,066 - 0,279 - 0,123 + 0,37}{8} = 0,001$$

Перевірка значущості коефіцієнтів регресії. Для цього потрібно розрахувати дисперсію визначення коефіцієнтів регресії:

$$s_{\{b_j\}}^2 = \frac{S_{\{\bar{y}\}}^2}{n}, \quad (4.6)$$

$$s_{\{b_j\}}^2 = \frac{0.00013}{8} = 1.63 \cdot 10^{-5}$$

Визначаємо похибку коефіцієнтів регресії за формулою:

$$\Delta b_j = \pm t s_{\{b_j\}}, \quad (4.7)$$

$$\Delta b_j = \pm 1,75 \cdot \sqrt{1,63 \cdot 10^{-5}} = \pm 0,008.$$

Якщо поверхня відгуку дійсно є площиною, то коефіцієнти за взаємодій факторів будуть незначні. У цьому разі похибки у визначенні коефіцієнтів регресії $\Delta b_j = \pm 0,008$. Усі коефіцієнти менші за розрахункову похибку вважаються незначними. Тоді рівняння набуває вигляду:

$$\tilde{y} = 0.193 + 0.112x_1 + 0.03x_2 + 0.017x_3.$$

Величини коефіцієнтів регресії показують ступінь впливу кожного чинника на параметр оптимізації. У нашому випадку найбільший вплив має температура – 0,112. Знак коефіцієнта показує напрямок впливу (плюс – прямо пропорційна залежність, мінус – обернено пропорційна залежність);

д) перевірка адекватності моделі за F-критерієм Фішера. На початку визначаємо дисперсію адекватності за формулою:

$$s_{\hat{a}\hat{a}}^2 = \frac{\sum_{i=1}^N (\bar{y}_i - \tilde{y}_i)^2}{n - (k + 1)}, \quad (4.8)$$

де k – кількість невідомих коефіцієнтів рівняння регресії, $k = 3$ (x_1, x_2, x_3).

$$s_{\hat{a}\hat{a}}^2 = \frac{0,000936}{8 - (3 + 1)} = 0,000234.$$

Розрахунок дисперсії адекватності

Номер досліджу	\bar{y}_i	\tilde{y}_i	$\bar{y}_i - \tilde{y}_i$	$(\bar{y}_i - \tilde{y}_i)^2$
1	0,05	0,034	0,016	0,000256
2	0,257	0,258	-0,001	0,000001
3	0,0825	0,094	-0,0115	0,000132
4	0,313	0,318	-0,005	0,000025
5	0,066	0,068	-0,002	0,000004
6	0,279	0,292	-0,013	0,000169
7	0,123	0,128	-0,005	0,000025
8	0,37	0,352	0,018	0,000324
			Сума	0,00093625

Розрахункове значення F-критерію Фішера визначається за формулою:

$$F = \frac{s_{ad}^2}{s_{восн}^2}, \quad (4.9)$$

$$F = \frac{0,000234}{0,00013} = 1,8$$

Обчислене значення F-критерію Фішера порівнюється з табличним для числа ступенів свободи $f = n(m-1) = 8(3-1) = 16$ та рівні значущості 0,01. $F_{0,01}(4;16) = 4,77$.

Оскільки розрахункове значення F-критерію не перевищує табличне, то можна вважати, що рівняння регресії адекватно описує експериментальні дані.

Індивідуальне завдання

Задача 1. Потрібно застосувати методи планування експерименту на прикладі вивчення процесу корозії бетону. За результатами досліджень відібрали такі фактори, що варіюються, які впливають на швидкість корозії бетону: x_1 – температура середовища, $^{\circ}C$; x_2 – час перебування в анаеробних умовах, год; x_3 – концентрація сульфатів, мг/дм³. Відгуком є швидкість корозії – y , мм/рік. За даними досліджень Дрозда Г. Я.,

температура середовища (x_1) в каналізаційних колекторах (господарсько-побутового призначення) варіюється протягом року від 10 до 40 $^{\circ}\text{C}$; час перебування в анаеробних умовах (x_2) – від 0 до 5 годин; концентрація сульфатів (x_3) – від 50 до 400 мг/дм³.

Задача 2. Оптимізується процес коагуляції води. Варіювання факторів: x_1 – відношення коагулянту до кількості води, від 0,5 до 0,9 г/л; x_2 – температура від 20 до 30 $^{\circ}\text{C}$; x_3 – час реакції, від 15 до 45 хв. Параметр оптимізації – відсоток очищення води за показником каламутності.

Таблиця 4.5

Вихідні дані

Варіант 1			Варіант 2		
y_{i1}	y_{i2}	y_{i3}	y_{i1}	y_{i2}	y_{i3}
5,3	5,6	5,4	46,8	46,3	46,7
5,2	5,2	5,1	20,7	20,4	20,3
3,0	3,2	3,0	16,8	16,9	16,5
9,6	9,5	9,7	5,1	5,3	5,2
17,9	17,6	17,8	24,5	24,3	24,5
10,1	10,2	10,3	8,9	8,8	8,7
5,4	5,4	5,7	16,6	17,0	16,8
9,2	8,8	9,0	46,5	46,4	46,4

Додаток А

Таблиця А.1

**Таблиця критичних значень критерію ??? Фішера $F(\alpha; f_1; f_2)$;
(для $\alpha = 0,01$)**

f_2	f_1										
	1	2	3	4	5	6	8	10	12	24	INF
1	4052.1	4999.5	5403.3	5624.5	5763.6	5858.9	5981.0	6055.8	6106.3	6234.6	6365.8
2	98.503	99.000	99.166	99.249	99.299	99.333	99.374	99.399	99.416	99.458	99.499
3	34.116	30.817	29.457	28.710	28.237	27.911	27.489	27.229	27.052	26.598	26.125
4	21.198	18.000	16.694	15.977	15.522	15.207	14.799	14.546	14.374	13.929	13.463
5	16.258	13.274	12.060	11.392	10.967	10.672	10.289	10.051	9.888	9.466	9.020
6	13.745	10.925	9.780	9.148	8.746	8.466	8.102	7.874	7.718	7.313	6.880
7	12.246	9.547	8.451	7.847	7.460	7.191	6.840	6.620	6.469	6.074	5.650
8	11.259	8.649	7.591	7.006	6.632	6.371	6.029	5.814	5.667	5.279	4.859
9	10.561	8.022	6.992	6.422	6.057	5.802	5.467	5.257	5.111	4.729	4.311
10	10.044	7.559	6.552	5.994	5.636	5.386	5.057	4.849	4.706	4.327	3.909
11	9.646	7.206	6.217	5.668	5.316	5.069	4.744	4.539	4.397	4.021	3.602
12	9.330	6.927	5.953	5.412	5.064	4.821	4.499	4.296	4.155	3.780	3.361
13	9.074	6.701	5.739	5.205	4.862	4.620	4.302	4.100	3.960	3.587	3.165
14	8.862	6.515	5.564	5.035	4.695	4.456	4.140	3.939	3.800	3.427	3.004
15	8.683	6.359	5.417	4.893	4.556	4.318	4.004	3.805	3.666	3.294	2.868
16	8.531	6.226	5.292	4.773	4.437	4.202	3.890	3.691	3.553	3.181	2.753
17	8.400	6.112	5.185	4.669	4.336	4.102	3.791	3.593	3.455	3.084	2.653
18	8.285	6.013	5.092	4.579	4.248	4.015	3.705	3.508	3.371	2.999	2.566
19	8.185	5.926	5.010	4.500	4.171	3.939	3.631	3.434	3.297	2.925	2.489
20	8.096	5.849	4.938	4.431	4.103	3.871	3.564	3.368	3.231	2.859	2.421
21	8.017	5.780	4.874	4.369	4.042	3.812	3.506	3.310	3.173	2.801	2.360
22	7.945	5.719	4.817	4.313	3.988	3.758	3.453	3.258	3.121	2.749	2.305
23	7.881	5.664	4.765	4.264	3.939	3.710	3.406	3.211	3.074	2.702	2.256
24	7.823	5.614	4.718	4.218	3.895	3.667	3.363	3.168	3.032	2.659	2.211
25	7.770	5.568	4.675	4.177	3.855	3.627	3.324	3.129	2.993	2.620	2.169
26	7.721	5.526	4.637	4.140	3.818	3.591	3.288	3.094	2.958	2.585	2.131
27	7.677	5.488	4.601	4.106	3.785	3.558	3.256	3.062	2.926	2.552	2.097
28	7.636	5.453	4.568	4.074	3.754	3.528	3.226	3.032	2.896	2.522	2.064
29	7.598	5.420	4.538	4.045	3.725	3.499	3.198	3.005	2.868	2.495	2.034
30	7.562	5.390	4.510	4.018	3.699	3.473	3.173	2.979	2.843	2.469	2.006
40	7.314	5.179	4.313	3.828	3.514	3.291	2.993	2.801	2.665	2.288	1.805
60	7.077	4.977	4.126	3.649	3.339	3.119	2.823	2.632	2.496	2.115	1.601
120	6.851	4.787	3.949	3.480	3.174	2.956	2.663	2.472	2.336	1.950	1.381
INF	6.635	4.605	3.782	3.319	3.017	2.802	2.511	2.321	2.185	1.791	1.000

**Таблиця критичних значень критерію ? Фішера $F(\alpha; f_1; f_2)$;
(для $\alpha = 0,05$)**

f_2	f_1									
	1	2	3	4	5	6	8	12	24	INF
1	161,45	199,50	215,72	224,57	230,17	233,97	238,89	243,91	249,04	254,32
2	18,512	18,999	19,163	19,248	19,298	19,329	19,371	19,414	19,453	19,496
3	10,129	9,552	9,276	9,118	9,014	8,941	8,844	8,744	8,638	8,527
4	7,710	6,945	6,591	6,388	6,257	6,164	6,041	5,912	5,774	5,628
5	6,607	5,786	5,410	5,192	5,050	4,950	4,818	4,678	4,527	4,365
6	5,987	5,143	4,756	4,534	4,388	4,284	4,147	4,000	3,841	3,669
7	5,591	4,737	4,347	4,121	3,972	3,866	3,725	3,574	3,410	3,230
8	5,317	4,459	4,067	3,838	3,688	3,580	3,438	3,284	3,116	2,928
9	5,117	4,256	3,863	3,633	3,482	3,374	3,230	3,073	2,900	2,707
10	4,965	4,103	3,708	3,478	3,326	3,217	3,072	2,913	2,737	2,538
11	4,844	3,982	3,587	3,357	3,204	3,094	2,948	2,788	2,609	2,405
12	4,747	3,885	3,490	3,259	3,106	2,999	2,848	2,686	2,505	2,296
13	4,667	3,805	3,410	3,179	3,025	2,915	2,767	2,604	2,420	2,207
14	4,600	3,739	3,344	3,112	2,958	2,848	2,699	2,534	2,349	2,131
15	4,543	3,683	3,287	3,056	2,901	2,790	2,641	2,475	2,288	2,066
16	4,494	3,634	3,239	3,007	2,853	2,741	2,591	2,424	2,235	2,010
17	4,451	3,592	3,197	2,965	2,810	2,699	2,548	2,381	2,190	1,961
18	4,414	3,555	3,160	2,928	2,773	2,661	2,510	2,342	2,150	1,917
19	4,381	3,522	3,127	2,895	2,740	2,629	2,477	2,308	2,114	1,878
20	4,351	3,493	3,098	2,866	2,711	2,599	2,447	2,278	2,083	1,843
21	4,325	3,467	3,072	2,840	2,685	2,573	2,421	2,250	2,054	1,812
22	4,301	3,443	3,049	2,817	2,661	2,549	2,397	2,266	2,028	1,783
23	4,279	3,422	3,028	2,795	2,640	2,528	2,375	2,203	2,005	1,757
24	4,260	3,403	3,009	2,777	2,621	2,508	2,355	2,183	1,984	1,733
25	4,242	3,385	2,991	2,759	2,603	2,490	2,337	2,165	1,965	1,711
26	4,225	3,369	2,975	2,743	2,587	2,474	2,321	2,148	1,947	1,691
27	4,210	3,354	2,961	2,728	2,572	2,459	2,305	2,132	1,930	1,672
28	4,196	3,340	2,947	2,714	2,558	2,445	2,292	2,118	1,915	1,654
29	4,183	3,328	2,934	2,702	2,545	2,432	2,278	2,104	1,901	1,638
30	4,171	3,316	2,922	2,690	2,534	2,421	2,266	2,092	1,887	1,622
40	4,085	3,232	2,899	2,606	2,449	2,336	2,180	2,004	1,793	1,509
60	4,001	3,151	2,758	2,525	2,368	2,254	2,097	1,918	1,700	1,389
120	3,920	3,072	2,680	2,447	2,290	2,175	2,106	1,843	1,608	1,254
INF	3,841	2,996	2,605	2,372	2,214	2,098	1,938	1,752	1,517	1,000

Таблиця критичних значень критерію ? Фішера $F(\alpha; f_1; f_2)$;
(для $\alpha = 0,10$)

f_2	f_1									
	1	2	3	4	5	6	8	12	24	INF
1	39.8634	49.5000	53.5932	55.8329	57.2400	58.2044	59.4389	60.7052	62.0020	63.3281
2	8.52632	9.00000	9.16179	9.24342	9.29263	9.32553	9.36677	9.40813	9.44962	9.49122
3	5.53832	5.46238	5.39077	5.34264	5.30916	5.28473	5.25167	5.21562	5.17636	5.13370
4	4.54477	4.32456	4.19086	4.10725	4.05058	4.00975	3.95494	3.89553	3.83099	3.76073
5	4.06042	3.77972	3.61948	3.52020	3.45298	3.40451	3.33928	3.26824	3.19052	3.10500
6	3.77595	3.46330	3.28876	3.18076	3.10751	3.05455	2.98304	2.90472	2.81834	2.72216
7	3.58943	3.25744	3.07407	2.96053	2.88334	2.82739	2.75158	2.66811	2.57533	2.47079
8	3.45792	3.11312	2.92380	2.80643	2.72645	2.66833	2.58935	2.50196	2.40410	2.29257
9	3.36030	3.00645	2.81286	2.69268	2.61061	2.55086	2.46941	2.37888	2.27683	2.15923
10	3.28502	2.92447	2.72767	2.60534	2.52164	2.46058	2.37715	2.28405	2.17843	2.05542
11	3.22520	2.85951	2.66023	2.53619	2.45118	2.38907	2.30400	2.20873	2.10001	1.97211
12	3.17655	2.80680	2.60552	2.48010	2.39402	2.33102	2.24457	2.14744	2.03599	1.90361
13	3.13621	2.76317	2.56027	2.43371	2.34672	2.28298	2.19535	2.09659	1.98272	1.84620
14	3.10221	2.72647	2.52222	2.39469	2.30694	2.24256	2.15390	2.05371	1.93766	1.79728
15	3.07319	2.69517	2.48979	2.36143	2.27302	2.20808	2.11853	2.01707	1.89904	1.75505
16	3.04811	2.66817	2.46181	2.33274	2.24376	2.17833	2.08798	1.98539	1.86556	1.71817
17	3.02623	2.64464	2.43743	2.30775	2.21825	2.15239	2.06134	1.95772	1.83624	1.68564
18	3.00698	2.62395	2.41601	2.28577	2.19583	2.12958	2.03789	1.93334	1.81035	1.65671
19	2.98990	2.60561	2.39702	2.26630	2.17596	2.10936	2.01710	1.91170	1.78731	1.63077
20	2.97465	2.58925	2.38009	2.24893	2.15823	2.09132	1.99853	1.89236	1.76667	1.60738
21	2.96096	2.57457	2.36489	2.23334	2.14231	2.07512	1.98186	1.87497	1.74807	1.58615
22	2.94858	2.56131	2.35117	2.21927	2.12794	2.06050	1.96680	1.85925	1.73122	1.56678
23	2.93736	2.54929	2.33873	2.20651	2.11491	2.04723	1.95312	1.84497	1.71588	1.54903
24	2.92712	2.53833	2.32739	2.19488	2.10303	2.03513	1.94066	1.83194	1.70185	1.53270
25	2.91774	2.52831	2.31702	2.18424	2.09216	2.02406	1.92925	1.82000	1.68898	1.51760
26	2.90913	2.51910	2.30749	2.17447	2.08218	2.01389	1.91876	1.80902	1.67712	1.50360
27	2.90119	2.51061	2.29871	2.16546	2.07298	2.00452	1.90909	1.79889	1.66616	1.49057
28	2.89385	2.50276	2.29060	2.15714	2.06447	1.99585	1.90014	1.78951	1.65600	1.47841
29	2.88703	2.49548	2.28307	2.14941	2.05658	1.98781	1.89184	1.78081	1.64655	1.46704
30	2.88069	2.48872	2.27607	2.14223	2.04925	1.98033	1.88412	1.77270	1.63774	1.45636
40	2.83535	2.44037	2.22609	2.09095	1.99682	1.92688	1.82886	1.71456	1.57411	1.37691
60	2.79107	2.39325	2.17741	2.04099	1.94571	1.87472	1.77483	1.65743	1.51072	1.29146
120	2.74781	2.34734	2.12999	1.99230	1.89587	1.82381	1.72196	1.60120	1.44723	1.19256
INF	2.70554	2.30259	2.08380	1.94486	1.84727	1.77411	1.67020	1.54578	1.38318	1.00000

Додаток Б

Таблиця Б.1

Критичні області для t-розподілу Стьюдента

<i>f</i>	Рівень значимості, α							
	0.40	0.25	0.10	0.05	0.025	0.01	0.005	0.0005
1	0.324920	1.000000	3.077684	6.313752	12.70620	31.82052	63.65674	636.6192
2	0.288675	0.816497	1.885618	2.919986	4.30265	6.96456	9.92484	31.5991
3	0.276671	0.764892	1.637744	2.353363	3.18245	4.54070	5.84091	12.9240
4	0.270722	0.740697	1.533206	2.131847	2.77645	3.74695	4.60409	8.6103
5	0.267181	0.726687	1.475884	2.015048	2.57058	3.36493	4.03214	6.8688
6	0.264835	0.717558	1.439756	1.943180	2.44691	3.14267	3.70743	5.9588
7	0.263167	0.711142	1.414924	1.894579	2.36462	2.99795	3.49948	5.4079
8	0.261921	0.706387	1.396815	1.859548	2.30600	2.89646	3.35539	5.0413
9	0.260955	0.702722	1.383029	1.833113	2.26216	2.82144	3.24984	4.7809
10	0.260185	0.699812	1.372184	1.812461	2.22814	2.76377	3.16927	4.5869
11	0.259556	0.697445	1.363430	1.795885	2.20099	2.71808	3.10581	4.4370
12	0.259033	0.695483	1.356217	1.782288	2.17881	2.68100	3.05454	4.3178
13	0.258591	0.693829	1.350171	1.770933	2.16037	2.65031	3.01228	4.2208
14	0.258213	0.692417	1.345030	1.761310	2.14479	2.62449	2.97684	4.1405
15	0.257885	0.691197	1.340606	1.753050	2.13145	2.60248	2.94671	4.0728
16	0.257599	0.690132	1.336757	1.745884	2.11991	2.58349	2.92078	4.0150
17	0.257347	0.689195	1.333379	1.739607	2.10982	2.56693	2.89823	3.9651
18	0.257123	0.688364	1.330391	1.734064	2.10092	2.55238	2.87844	3.9216
19	0.256923	0.687621	1.327728	1.729133	2.09302	2.53948	2.86093	3.8834
20	0.256743	0.686954	1.325341	1.724718	2.08596	2.52798	2.84534	3.8495
21	0.256580	0.686352	1.323188	1.720743	2.07961	2.51765	2.83136	3.8193
22	0.256432	0.685805	1.321237	1.717144	2.07387	2.50832	2.81876	3.7921
23	0.256297	0.685306	1.319460	1.713872	2.06866	2.49987	2.80734	3.7676
24	0.256173	0.684850	1.317836	1.710882	2.06390	2.49216	2.79694	3.7454
25	0.256060	0.684430	1.316345	1.708141	2.05954	2.48511	2.78744	3.7251
26	0.255955	0.684043	1.314972	1.705618	2.05553	2.47863	2.77871	3.7066
27	0.255858	0.683685	1.313703	1.703288	2.05183	2.47266	2.77068	3.6896
28	0.255768	0.683353	1.312527	1.701131	2.04841	2.46714	2.76326	3.6739
29	0.255684	0.683044	1.311434	1.699127	2.04523	2.46202	2.75639	3.6594
30	0.255605	0.682756	1.310415	1.697261	2.04227	2.45726	2.75000	3.6460
inf	0.253347	0.674490	1.281552	1.644854	1.95996	2.32635	2.57583	3.2905

Додаток В

Таблиця В.1

Максимальні значення коефіцієнтів варіабельності

n	$\alpha = 0,10$	$\alpha = 0,05$	$\alpha = 0,01$	n	$\alpha = 0,10$	$\alpha = 0,05$	$\alpha = 0,01$
3	1.41	1.41	1.41	26	2.55	2.73	3.09
4	1.64	1.69	1.72	27	2.57	2.75	3.11
5	1.79	1.87	1.96	28	2.58	2.76	3.12
6	1.89	2.00	2.13	29	2.60	2.78	3.14
7	1.97	2.09	2.26	30	2.61	2.79	3.16
8	2.04	2.17	2.37	31	2.62	2.80	3.17
9	2.10	2.24	2.46	32	2.63	2.82	3.18
10	2.15	2.29	2.54	33	2.65	2.83	3.20
11	2.19	2.34	2.61	34	2.66	2.84	3.21
12	2.23	2.39	2.66	35	2.67	2.85	3.22
13	2.26	2.43	2.71	36	2.68	2.86	3.24
14	2.30	2.46	2.76	37	2.69	2.87	3.25
15	2.33	2.49	2.80	38	2.70	2.88	3.26
16	2.35	2.52	2.84	39	2.71	2.89	3.27
17	2.38	2.55	2.87	40	2.72	2.90	3.28
18	2.40	2.58	2.90	41	2.73	2.91	3.29
19	2.43	2.60	2.93	42	2.74	2.92	3.30
20	2.45	2.62	2.96	43	2.74	2.93	3.31
21	2.47	2.64	2.98	44	2.75	2.94	3.32
22	2.49	2.66	3.01	45	2.76	2.95	3.33
23	2.50	2.68	3.03	46	2.77	2.96	3.34
24	2.52	2.70	3.05	47	2.78	2.96	3.35
25	2.54	2.72	3.07	48	2.78	2.97	3.35

Додаток Г

Таблиця Г.1

Таблиця значень інтегральної функції Лапласа, $2\Phi(t) = P$

<i>t</i>	$2\Phi(t)$	<i>t</i>	$2\Phi(t)$	<i>t</i>	$2\Phi(t)$
0.0	0.000	1.2	0.770	2.4	0.984
0.1	0.080	1.3	0.806	2.5	0.988
0.2	0.159	1.4	0.839	2.6	0.991
0.3	0.236	1.5	0.866	2.7	0.993
0.4	0.311	1.6	0.890	2.8	0.995
0.5	0.383	1.7	0.911	2.9	0.996
0.6	0.452	1.8	0.928	3.0	0.997
0.7	0.516	1.9	0.943	3.1	0.998
0.8	0.576	2.0	0.955	3.2	0.9986
0.9	0.632	2.1	0.964	3.3	0.9990
1.0	0.683	2.2	0.972	3.4	0.9993
1.1	0.729	2.3	0.979	3.5	0.9995

Додаток Д

Таблиця Д.1

Критичні значення критерію Кохрена (рівень значущості – 1 %)

<i>n</i>	Критичне значення критерію Кохрена у числі ступенів свободи, <i>v</i>									
	1	2	3	4	5	10	15	20	30	50
3	0,9933	0,9423	0,8831	0,8335	0,7933	0,6743	0,6145	0,5775	0,5327	0,4872
4	0,9676	0,8643	0,7814	0,7212	0,6761	0,5536	0,4964	0,4620	0,4213	0,3808
5	0,9279	0,7885	0,6957	0,6329	0,5875	0,4697	0,4168	0,3855	0,3489	0,3131
6	0,8828	0,7218	0,6258	0,5635	0,5195	0,4084	0,3597	0,3312	0,2982	0,2661
7	0,8376	0,6644	0,5685	0,5080	0,4659	0,3616	0,3167	0,2907	0,2606	0,2316
8	0,7945	0,6152	0,5209	0,4627	0,4227	0,3248	0,2832	0,2592	0,2316	0,2052
9	0,7544	0,5727	0,4810	0,4251	0,3870	0,2950	0,2563	0,2340	0,2086	0,1842
10	0,7175	0,5358	0,4469	0,3934	0,3572	0,2704	0,2342	0,2135	0,1898	0,1673
11	0,6837	0,5036	0,4175	0,3663	0,3318	0,2497	0,2157	0,1963	0,1742	0,1532
12	0,6528	0,4751	0,3919	0,3428	0,3099	0,2321	0,2000	0,1818	0,1611	0,1414
13	0,6245	0,4498	0,3695	0,3223	0,2909	0,2169	0,1865	0,1693	0,1498	0,1313
14	0,5985	0,4272	0,3495	0,3043	0,2741	0,2036	0,1748	0,1585	0,1400	0,1226
15	0,5747	0,4069	0,3318	0,2882	0,2593	0,1919	0,1645	0,1490	0,1315	0,1150
20	0,4799	0,3297	0,2654	0,2288	0,2048	0,1496	0,1274	0,1150	0,1010	0,0879
25	0,4130	0,2782	0,2220	0,1904	0,1699	0,1230	0,1043	0,0939	0,0822	0,0713
30	0,3632	0,2412	0,1914	0,1635	0,1455	0,1046	0,0885	0,0794	0,0694	0,0600
35	0,3247	0,2134	0,1685	0,1435	0,1274	0,0912	0,0769	0,0690	0,0601	0,0519
40	0,2940	0,1916	0,1507	0,1281	0,1136	0,0809	0,0681	0,0610	0,0531	0,0457
45	0,2690	0,1740	0,1364	0,1158	0,1025	0,0727	0,0611	0,0547	0,0475	0,0409
50	0,2481	0,1596	0,1248	0,1057	0,0935	0,0661	0,0555	0,0496	0,0431	0,0370
60	0,2151	0,1371	0,1068	0,0902	0,0796	0,0561	0,0469	0,0419	0,0363	0,0311
70	0,1903	0,1204	0,0935	0,0788	0,0695	0,0487	0,0407	0,0363	0,0314	0,0269
80	0,1709	0,1075	0,0832	0,0701	0,0617	0,0431	0,0360	0,0320	0,0277	0,0236
90	0,1553	0,0972	0,0751	0,0631	0,0555	0,0387	0,0322	0,0287	0,0248	0,0211
100	0,1424	0,0888	0,0685	0,0575	0,0505	0,0351	0,0292	0,0260	0,0224	0,0191

Ці значення являють собою апроксимації, розраховані за допомогою нерівності Бонферроні як верхня $0,01/n$ фракталя бета-розподілу. Проміжні значення колонки для n оцінок дисперсії можуть бути отримані за допомогою лінійної інтерполяції зворотних величин табульованих значень. Проміжні значення для ступенів свободи можуть бути отримані за допомогою інтерполяції другого порядку (квадратичної) для зворотних величин табульованих значень.

Список літератури

1. *Бідюк П. І.* Прикладна статистика : конспект лекцій / П. І. Бідюк. – Київ : НТУУ «КПІ», 2009. – 220 с.
2. *Гаркавий В. К.* Математична статистика : навч. посіб. / В. К. Гаркавий, В. В. Ярова. – Київ : Професіонал, 2004. – 384 с.
3. *Бобик О. І.* Теорія ймовірностей і математична статистика / О. І. Бобик, Г. І. Берегова, Б. І. Копитко. – Київ : Професіонал, 2007. – 560 с.
4. *Тарасова В. В.* Екологічна статистика (з блочно-модульною формою контролю знань) : підручник. – Київ: Центр учбової літератури, 2008. – 392 с.
5. *Статистика* : підручник / С. С. Герасименко, А. В. Головач, А. М. Єріна та ін. – 3-тє вид. – Київ : КНЕУ, 2015. – 478 с.
6. *Моделювання і прогнозування стану довкілля* : підручник / В. І. Лаврик, В. М. Боголюбов, Л. М. Полетаєва, С. М. Юрасов, В. Г. Ільїна; за ред. В. І. Лаврика. – Київ : ВЦ Академія, 2010. – 400 с.

Для нотаток

Навчально-методичне видання

**МОНІТОРИНГ ЯКОСТІ ДОВКІЛЛЯ
ТА СТАТИСТИЧНА ОБРОБКА
ЕКСПЕРИМЕНТАЛЬНИХ ДАНИХ
І РЕЗУЛЬТАТІВ НАУКОВИХ ДОСЛІДЖЕНЬ**

Методичні рекомендації
до виконання практичних робіт
для студентів спеціальності 183 «Технології захисту навколишнього
середовища»

Укладачі: **САТІН** Ігор Валентинович,
ЖУКОВА Олена Григорівна

Випусковий редактор *Т. В. Івченко*
Комп'ютерне верстання *Д. М. Ніколаєвич*

Підписано до друку 19.12.2023. Формат 60 x 84_{1/16}
Ум. друк. арк. 2,32. Обл.-вид. арк. 2,5.
Електронний документ. Вид. № 117/III-23

Видавець і виготовлювач:
Київський національний університет будівництва і архітектури

Повітрофлотський проспект, 31, Київ, Україна, 03037

Свідоцтво про внесення до Державного реєстру суб'єктів
видавничої справи ДК № 808 від 13.02.2002