МІНІСТЕРСТВО ОСВІТИ І НАУКИ УКРАЇНИ
Одеський державний екологічний університет

МЕТОДИЧНІ ВКАЗІВКИ

до практичних занять з дисципліни
«АНАЛІЗ ЕКОЛОГІЧНОЇ ІНФОРМАЦІЇ»
для студентів спеціальності 101 «Екологія»
Освітньо-професійна програма:
«Екологічний контроль і аудит»

Одеса – 2020

МІНІСТЕРСТВО ОСВІТИ І НАУКИ УКРАЇНИ
Одеський державний екологічний університет

МЕТОДИЧНІ ВКАЗІВКИ
до практичних занять з дисципліни
«АНАЛІЗ ЕКОЛОГІЧНОЇ ІНФОРМАЦІЇ»
для студентів спеціальності 101 «Екологія»
Освітньо-професійна програма:
«Екологічний контроль і аудит»

Затверджено
на засіданні групи забезпечення спеціальності
Протокол № 2 від « 18 » 06 2020р.
Голова групи ______________Чугай А.В.

Затверджено
на засіданні кафедри екологічного права і контролю
Протокол № 11 від « 9 » 06. 2020р.
Зав. кафедри______________Лоєва І.Д.

Одеса – 2020

Методичні вказівки до практичних занять з дисципліни «Аналіз екологічної інформації» для студентів які навчаються за спеціальністю: 101 «Екологія», освітньо-професійною програмою «Екологічний контроль і аудит» /Лоєва І.Д., Бургаз О.А., - Одеса: ОДЕКУ, 2020. – 65 с., укр. мова.

ЗМІСТ

ПЕРЕДМОВА

Методичні вказівки призначені для виконання практичних робіт з дисципліни «Аналіз екологічної інформації».

Практичні роботи спрямовані на оволодіння студентами основних положень дисципліни, отримання системи теоретичних знань з методів обробки і аналізу екологічної інформації й навичок щодо використання відповідних алгоритмів для рішення прикладних задач екології.

У збірнику методичних вказівок наведена методика отримання вихідного ряду даних, розрахунків статистичних оцінок моментів, використання параметричних та непараметричних критеріїв перевірки статистичних гіпотез, розрахунку коефіцієнту кореляції.

Студент оформлює практичні роботи у письмовій формі, які містять алгоритм розрахунку, результати розрахунку і їх аналіз. Виконані роботи оцінюються у балах, згідно до вимог робочої програми дисципліни «Аналіз екологічної інформації».

ПРАКТИЧНА РОБОТА № 1
«ФОРМУВАННЯ РЯДУ ВИХІДНОЇ ІНФОРМАЦІЇ НА ОСНОВІ ДАНИХ АВТОМАТИЗОВАНОЇ СИСТЕМИ СПОСТЕРЕЖЕНЬ ЗА СТАНОМ ЗАБРУДНЕННЯ АТМОСФЕРИ»

1.1 Загальні положення

Великий обсяг даних спостережень за станом забруднення атмосфери, необхідність їх широкого використання в природоохоронній діяльності як загальнодержавного, так і регіонального рівнів потребує автоматизованої обробки інформації.

Первинною формою збирання результатів спостережень за концентрацією домішок та необхідними метеорологічними характеристиками є ТЗА – таблиці забруднення атмосфери, які містять характеристики забруднення атмосферного повітря за різні просторові та часові інтервали. Залежно від виду вимірювань розрізняють ТЗА – 1, ТЗА – 2, ТЗА – 3 та ТЗА – 4.

В ТЗА – 1 записують дані спостережень на мережі постійно діючих стаціонарних або маршрутних постів. В ТЗА – 2 записують дані спостережень під факелами промислових підприємств на різних відстанях від джерела забруднення. В ТЗА – 3 записують дані добових спостережень на стаціонарних постах. В ТЗА – 4 записують дані безперервних спостережень, які проводились за допомогою автоматичних газоаналізаторів [1].

Типи постів та їх позначення, що використовується при кодуванні результатів спостережень (код **К**), приведені в табл. 1.1

Таблиця 1.1 – Типи постів спостережень та характеристика коду **К**

Тип поста	**К**
Метеостанція без спостережень за забрудненням атмосфери	0
Опорний пункт на основі метеостанції	1
Опорний пост	2
Стаціонарний пост на основі метеостанції	3
Стаціонарний пост	4
Маршрутний пост	5
Стаціонарні (опорні) пости санітарно-епідеміологічної служби	6
Підфакельні спостереження	7
Автоматизований пост спостережень	8
Стаціонарний пост з добовою програмою спостережень	9

1.2 Розкодування результатів спостережень

Макет для розкодування спостережень має вигляд:

I	II	III	IV	V	
//ааГГ	**MMhhhP**	**φφφλλλ**	**PPRRKS**	**XXXXNN**	
1	**2**	**3**	**4**	**5**	**6**
((JJR$_1$	**GGYYY**	**bbrrr**	**+TTT0**	**ddvvψ**	**ffeee**

Макет призначений для розкодування спостережень за місяць на окремому стаціонарному (або маршрутному) посту або під факелом одного промислового підприємства. Він складається з п'яти шестизначних (групи I – V) та шести п'ятизначних груп (групи 1 – 6). Групи I – V містять загальну інформацію – ознаковий блок. Групи 1 – 6 використовують при кодуванні інформації в окремі строки спостережень і поділені на три блоки:

- просторово-часовий блок (групи 1 – 2);
- дані спостережень за концентраціями домішок (група 3);
- дані метеорологічних спостережень (групи 4 – 6).

Макет служить для розкодування результатів спостережень за місяць на окремому стаціонарному (маршрутному) посту, проведених за допомогою автоматичних засобів реєстрації. Групи відокремлюються пробілами, на які відводяться окремі позиції [1].

1.2.1 Розкодування ознакового блоку

Група I займає позиції 1 – 6.

// (або ##) – означає початок запису відомостей одного поста за один місяць, розташовується у позиціях 1 та 2 (символи // використовуються при записі даних на перфострічках, ## – використовуються при записі даних на магнітних стрічках);

аа – на позиціях 3 та 4 записується шифр виду інформації (для ТЗА-1 завжди 11);

ГГ – рік спостережень на позиціях 5 – 6 (дві останні цифри).

Позиція 7 – пробіл.

Група II займає позиції 8 – 13.

ММ – номер місяця, що знаходиться на позиціях 8 та 9, січень кодується як 01, грудень – 12;

hhh – висота станції над рівнем моря знаходиться на позиціях 10-12 і кодується у десятках метрів (наприклад, 57 м записується як 006);

P – позначення довготи (позиція 13):

P=0 – східна довгота менша за 100°;

P=1 – східна довгота більша або дорівнює 100°;

P=2 – західна довгота більша або дорівнює 100°;

P=3 – західна довгота менша за 100°.

Позиція 14 – пробіл.

Група III – координатний номер міста, знаходиться на 15 – 20 позиціях.

Позиція 21 – пробіл.

Група IV – позиції 22 – 27.

PP – загальна кількість постів у місті за даний місяць, записується на позиціях 22 – 23;

RR – загальна кількість видів домішок, концентрація яких була виміряна (позиції 24 – 25);

K – код типу поста (позиція 26). Відповідні позначення коду **K** приведені в табл. 1.1;

S – знак координат поста (позиція 27).

Позиція 28 – пробіл.

Група V – позиції 29 – 34.

XXXX – координати поста (позиції 29 – 32);

NN – номер поста (позиції 33 – 34).

Описана вище інформація кодується один раз на місяць. Після ознакового блоку розміщується блок результатів спостережень. Кожен новий рядок – інформація за окремий строк спостережень [1].

1.2.2 Розкодування результатів спостережень

Дані спостережень за всі строки об'єднані у п'ятизначні групи номер яких вказаний під позиціями, які вони займають. Результати спостережень об'єднані у два блоки – концентрації домішок (групи 1 – 3) й метеорологічних характеристик (групи 4 – 6). Блоки розділяються символом **::**.

Група 1 займає позиції 1 – 5:

((або **NN** – початок кодування інформації за окремий строк (позиції 1 – 2). Якщо дані записані на перфострічці запис кожного рядка починають з **((** , якщо на магнітній стрічці, то в позиціях 1 – 2 записується номер поста у місті;

JJ – дата спостережень (позиції 3 – 4);

$\mathbf{R_1}$ – кількість шифрів домішок, що записуються в групі 3 за даний строк, завершує групу.

Позиція 6 – пробіл.

Група 2 займає позиції 7 – 11:

GG – строк відбору проб (позиції 7 та 8);

YYY – позиції 9 – 11 не несуть інформації та використовуються для закінчення групи.

Позиція 12 – пробіл.

У групу 3 записуються дані про концентрації домішок (позиції 13 – 53):

bb – шифр домішки у АСОІЗА (табл. 1.2);

rrr – концентрація домішки, записана з заданою точністю.

Таблиця 1.2 – Перелік шифрів деяких домішок у АСОІЗА, точності запису їх концентрацій та середньодобові гранично допустимі концентрації домішок [1]

№ п/п	Назва домішки	Шифр у АСОІЗА	Точність запису концентрації	ГДК$_{сд}$
1	Пил неорганічний	01	0,1	0,15
2	Двоокис сірки SO_2	02	0,001	0,05
3	Оксид вуглецю CO	04	1	5
4	Двоокис азоту NO_2	05	0,01	0,04
5	Озон O_3	07	0,001	0,03
6	Сірководень H_2S	08	0,001	0,008
7	Фенол C_6H_6OH	10	0,01	0,01
8	Сажа	11	0,01	0,05
9	Фтористий водень HF	13	0,001	0,005
10	Формальдегід $HCHO$	22	0,001	0,003

1.2.3 Розкодування даних спостережень за метеопараметрами

Групи 4 і 5 – дані спостережень за метеопараметрами (позиції 55 – 65), які вимірювались у строк відбору проб на посту.

+ТТТ – температура повітря зі знаком з точністю до десятих (позиції 55 – 58). Наприклад значення 20,7 ºС записується як +207, а -5,0 – -050;

0 – на позиції 59 означає закінчення групи;

dd – напрямок вітру (позиції 61 – 62), закодовано у десятках градусів;

vv – швидкість вітру м/с (позиції 63 – 64), 8 м/с записують як 08, а 15 м/с – 15;

ψ – шифр атмосферних явищ (позиція 65)

Якщо в даний строк спостереження спостерігався штиль, то в позиціях 61 – 64 записують нулі. Шифр атмосферних явищ наведено у табл. 1.3.

Група 6 займає позиції 67 – 71 і закодована по даним для основної метеостанції:

ff – значення відносної вологості з точністю до 1%, вологість 100% закодована буквами ЭЭ, кодується в позиціях 67 та 68;

eee – парціальний тиск водяної пари (позиції 69 – 71) у гектопаскалях з точністю до десятих.

Якщо метеорологічні характеристики за даний строк відсутні то в позиціях 55 – 56 знаходиться символ **:: ::**.

Таблиця 1.3 – Характеристика стану погоди

ψ	Стан погоди, атмосферні явища	Ознаки
0		Атмосферних явищ шифру 2 – 9 немає
1	Ясно	На небі немає хмар
2	Імла	Помутніння повітря за рахунок завислих часток пилу, диму, гару. Повітря має синюватий відтінок
3	Серпанок	Слабке помутніння атмосфери за рахунок перенасичення повітря вологою. Повітря має сірий відтінок; видимість більша за 1 км
4	Дощ	Опади у вигляді рідких краплин
5	Мряка	Атмосферні опади у вигляді дрібних краплин, що майже непомітні для ока
6	Пилова буря	Погіршення видимості на великій території через пил, піднятий сильним вітром
7	Сніг	Опади у вигляді льодових кришталиків
8	Туман	Помутніння атмосфери при горизонтальній видимості меншій за 1 км
9	Туман (або серпанок) з опадами	Помутніння атмосфери за рахунок туману (або серпанку) за наявності опадів

1.3 Приклад розкодування

Фрагмент запису спостережень за забрудненням атмосферного повітря має вид:

##1102 010040 464308 070634 081108
08022 01шшш 02037 05004 :: -0800 30021 67022
08024 07шшш 01001 04002 07030 11000 :: -0900 25021 67021
08022 13шшш 02028 05005 :: -0310 22041 49024
08026 19шшш 01002 02018 04002 05003 07040 11001 :: -0360 23030 56026

Розкодування:

Ознаковий блок:

Вид інформації – ТЗА – 1

Рік спостережень – 2002 р.

Місяць – 01

Висота поста над рівнем моря – 40 м.

Східна довгота менша за 100°

Координатний номер – 464308

Загальна кількість постів у місті за даний місяць – 7

Кількість домішок – 6

Стаціонарний пост на основі метеостанції

Координати поста +08-11

Пост 08

Блок результатів спостережень:

Дата	Строк	Концентрації домішок, мг/м^3						T °C	Вітер		ψ	Вологість	
		01	02	04	05	07	11		DD	VV		FF	EE
2	1		0,37		0,004			-8,0	300	2	1	67	22
2	7	0,1		0,02		0,3	0	-9,0	250	2	1	67	21
2	13		0,28		0,005			-3,0	220	4	1	49	24
2	19	0,2	0,18	0,02	0,003	0,4	0,001	-3,6	230	3	0	56	26

1.4 Контрольні запитання

1. Які існують види таблиць забруднення атмосфери?
2. Яка інформація входить в ознаковий блок макету розкодування даних спостережень?
3. Яка інформація входить у просторово-часовий блок макету розкодування даних спостережень?
4. Яка інформація входить у блок метеорологічних параметрів макету розкодування даних спостережень?

1.5 Завдання до практичної роботи

1. Отримати у викладача варіант вихідних даних.
2. Провести формування ряду строкових спостережень за забрудненням атмосферного повітря діоксидом сірки (SO_2) на контрольно-вимірювальних постах (КВП) міста. Отримані результати занести до таблиці за формою (табл. 1.4).

Таблиця 1.4 – Результати розкодування записів спостережень за забрудненням атмосфери

Місяць	Дата	Строк	Концентрація домішки							
			КВП №8	КВП №10	КВП №15	КВП №16	КВП №17	КВП №18	КВП №19	КВП №20

ПРАКТИЧНА РОБОТА № 2
«АНАЛІЗ СТАТИСТИЧНОЇ ОДНОРІДНОСТІ ДАНИХ СТРОКОВИХ СПОСТЕРЕЖЕНЬ ЗА КРИТЕРІЄМ СТЬЮДЕНТА. ФОРМУВАННЯ ЧАСОВИХ РЯДІВ СЕРЕДНЬОДОБОВИХ ЗНАЧЕНЬ КОНЦЕНТРАЦІЇ ДОМІШКИ»

2.1 Перевірка однорідності членів статистичної сукупності

Поняття *статистична сукупність*, то має на увазі дві категорії:

- *генеральна сукупність;*
- *статистичний ряд (вибірка).*

Під **генеральною сукупністю** розуміють нескінченну множину незалежних значень випадкової величини, які підпорядковуються одному й тому ж закону розподілу.

Статистичний ряд (вибірка) – обмежена кількість значень випадкової величини, добута випадковим чином із генеральної сукупності. Тому статистичні ряди називають вибірками із генеральної сукупності. Вибірки випадкові та число їх безмежне.

Важливою ознакою ряду є його об'єм. "Об'єм" – кількість членів, що складають сукупність.

Однією з вимог, яким повинна відповідати статистична сукупність, є *вимога однорідності її членів. Це означає, що всі члени вибірки з заданою ймовірністю повинні належати до однієї генеральної сукупності.* Значення фізичних параметрів, що відрізняються від їх середнього рівня (норми) отримали назву «*викидів*». «*Викиди*» спричиняють похибки при статистичному оцінюванні моментів розподілу випадкових величин і їх треба вилучити з вибірки, але тільки після перевірки статистичної гіпотези. Висувається дві гіпотези – основна H_0 та альтернативна H_1 [2, 4].

Гіпотеза H_0 – максимальні і мінімальні значення членів вибірки ($x_{екстр}$) належать до тієї ж генеральної сукупності, що і всі інші її члени. Тобто, якщо екстремальні значення статистичного ряду є однорідними, то й інші члени цієї вибірки також будуть підпорядковуватися одному закону розподілу.

Гіпотеза H_0 перевіряється на заданому рівні значущості α.

Для перевірки гіпотези використовують *критерій Стьюдента* t у вигляді:

$$t = \frac{\left| x_{екстр} - \overline{x} \right|}{S_x}, \tag{2.1}$$

де: $x_{екстр}$ екстремальне (мінімальне, максимальне) значення ряду;

$\overline{x}$ середнє значення концентрації забруднюючої величини;

S_x середній квадратичний відхил концентрації забруднюючої величини.

Отримане за формулою (2.1) значення критерію (t) порівнюється з його критичним значенням – $t_{кр}(\alpha,\nu)$.

$t_{кр}(\alpha,\nu)$ отримують по таблицях (додаток А) для визначеного рівня значущості (наприклад, α=0.05) і числа степенів вільності $\nu = n-1$, де n – об'єм вибірки.

Якщо

$$t < t_{кр}(\alpha,\nu), \qquad (2.2)$$

то гіпотеза H_0 не відкидається. Тобто, екстремальний член сукупності, відносно котрого є підозра, залишається у вибірці. Навпаки, якщо

$$t > t_{кр}(\alpha,\nu), \qquad (2.3)$$

то гіпотеза H_0 відхиляється й приймається альтернативна гіпотеза H_1, про те, що $x_{екстр}$ на рівні значущості α не належить до тієї ж генеральної сукупності, що й інші члени вибірки, тобто є «викидом». У такому разі його слід вилучити з ряду значень перед розрахунками статистичних оцінок моментів розподілу випадкової величини.

Перевірку гіпотези H_0 треба повторювати до тих пір, поки на заданому рівні значущості члени статистичного ряду, що складають вибірку, будуть однорідними [3, 4].

2.2 Приклад розрахунку

Таблиця 2.1 – Середньодобова концентрація діоксиду сірки в атмосфері, мг/м^3 (Одеса, КВП №8, 1.03 – 31.05.2003 р.)

i	**0**	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**
0		0,40	0,43	0,45	0,50	0,42	0,42	0,50	0,48	0,47
10	0,45	0,54	0,44	0,43	0,39	0,46	0,43	0,41	0,43	0,51
20	0,44	0,46	0,40	0,43	0,45	0,39	0,40	0,38	0,42	0,39
30	0,43	0,46	0,41	0,46	0,39	0,45	0,46	0,47	0,44	0,39
40	0,39	0,31	0,37	0,35	0,35	0,35	0,42	0,43	0,39	0,44
50	0,38	0,41	0,43	0,41	0,40	0,48	0,48	0,51	0,57	0,46

60	0,44	0,49	0,51	0,53	0,56	0,48	0,60	0,57	0,49	0,51
70	0,48	0,45	0,47	0,48	0,53	0,41	0,46	0,48	0,45	0,47
80	0,50	0,48	0,49	0,44	0,50	0,44	0,45	0,46	0,41	0,46
90	0,44	0,44	0,41							

x_{min} = 0,31 мг/м^3.
x_{max} = 0,60 мг/м^3.

2.2.1 Перевірка однорідності членів статистичної сукупності

Формулюємо гіпотезу H_0: максимальні і мінімальні значення членів вибірки ($x_{екстр}$) належать до тієї ж генеральної сукупності, що і всі інші її члени.

Визначаємо рівень значущості α і число степенів вільності ν:
нехай $\alpha = 0{,}05$, $\nu = n - 1 = 92 - 1 = 91$
де n – об’єм вибірки.

З таблиці Додатку А знаходимо критичне значення критерію Стьюдента $t_{кр}\left(\alpha,\nu\right) = t_{кр}(0{,}05,91) = 1{,}99$

Розраховуємо значення критерію Стьюдента t на основі вихідного ряду за формулою (2.14), та порівнюємо їх зі значенням $t_{кр}$. Так як $x_{\min} = 0{,}31$ мг/м^3 то:

$$t = \frac{\left|x_{\min} - \bar{x}\right|}{S_x} = \frac{\left|0{,}31 - 0{,}44\right|}{0{,}051} = 2{,}55$$

Отже $t > t_{кр}$ і гіпотеза H_0 відхиляється, приймається альтернативна гіпотеза H_1: член ряду $x_{\min}$ на рівні значущості α=0.05 не належить до тієї ж генеральної сукупності, що й інші члени вибірки і його потрібно вилучити з вибірки.

Повторюємо таку процедуру для $x_{\max}$:
$x_{\max} = 0{,}60$ мг/м^3

$$t = \frac{\left|x_{\max} - \bar{x}\right|}{S_x} = \frac{\left|0{,}60 - 0{,}44\right|}{0{,}051} = 3{,}13$$

Так як $t > t_{кр}$ то в даному випадку також приймається альтернативна гіпотеза H_1, $x_{\max}$ потрібно вилучити з вибірки.

Таким чином ми встановили,що $x_{\min}$ та $x_{\max}$ являються «викидами».

Значень ряду концентрації діоксиду сірки в атмосфері (пост №8, весняний період 2003 р.), лишилось 90. На їх основі ми знову знаходимо екстремальні значення і розраховуємо середнє значення та середній квадратичних відхил:

$x_{\min} = 0,35$ мг/м^3, $x_{\max} = 0,57$ мг/м^3

$\overline{x} = 0,42 \qquad S_x = 0,045$

$t_{кр}(\alpha, \nu) = t_{кр}(0,05, 89) = 1,99$

Розраховуємо значення критерію Стьюдента t

$$t = \frac{|x_{\min} - \overline{x}|}{S_x} = \frac{|0,35 - 0,42|}{0,045} = 1,55$$

Отже $t < t_{кр}$ і гіпотеза H_0 приймається: член ряду $x_{\min}$ на рівні значущості α=0.05 належить до тієї ж генеральної сукупності, що й інші члени вибірки.

$$t = \frac{|x_{\max} - \overline{x}|}{S_x} = \frac{|0,57 - 0,42|}{0,045} = 3,3$$

Так як $t > t_{кр}$ то H_0 відкидається і приймається альтернативна гіпотеза H_1, $x_{\max}$ потрібно вилучити з вибірки.

Подібну процедуру повторюємо до тих пір доки не буде прийнята гіпотеза H_0.

Результати перевірки гіпотези про однорідність членів ряду концентрації діоксиду сірки в атмосфері занесені до таблиці 2.5.

Таким чином, на рівні значущості α=0.05 вибірка середньої концентрації діоксиду сірки в атмосфері утримує вісім «викидів», які треба перед розрахунками статистичних оцінок моментів розподілу з неї вилучити. Проте останні 84 є однорідними, тобто такими, що підпорядковуються одному й тому ж закону розподілу.

Як бачимо, після перевірки гіпотези $x_{\min} = 0,35$ мг/м^3, $x_{\max} = 0,51$ мг/м^3, $\overline{x} = 0,44$ мг/м^3, $S_x = 0,04$.

Після перевірки членів ряду на однорідність можна приступати до розрахунків середньодобових значень концентрації забруднюючої речовини.

Середньодобові концентрації розраховуються як середні арифметичні значення строкових концентрацій вмісту домішки.

Таблиця 2.5 – Перевірка гіпотези про однорідність членів ряду концентрації діоксиду сірки в атмосфері $t_{кр}(\alpha,\nu)=1{,}99$

n	$\overline{x}$	S_x	x_{max}	t	Гіпотеза	x_{min}	t	Гіпотеза
92	0,44	0,051	0,60	3,13	H_1	0,31	2,55	H_1
90	0,42	0,045	0,57	3,3	H_1	0,35	1,55	H_0
89	0,44	0,046	0,57	2,8	H_1			
88	0,44	0,043	0,56	2,8	H_1			
87	0,443	0,042	0,54	2,4	H_1			
86	0,442	0,041	0,53	2,1	H_1			
85	0,441	0,04	0,53	2,2	H_1			
84	0,44	0,04	0,51	1,75	H_0			

Якщо в результаті перевірки на однорідність, за певний день спостережень, вимозі однорідності відповідає тільки одне значення строкових спостережень – то воно приймається як середньодобове (рис. 2.1).

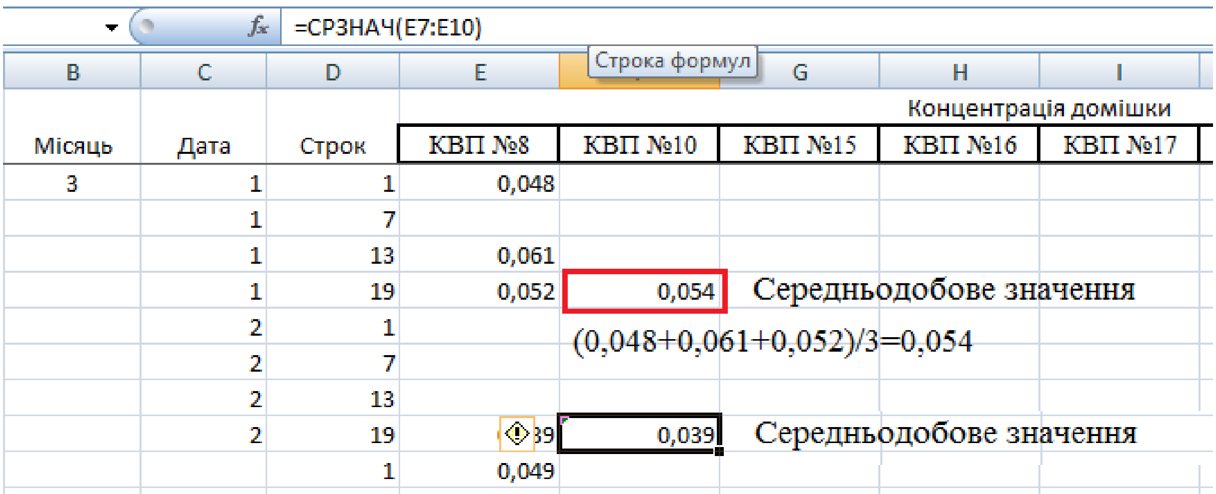
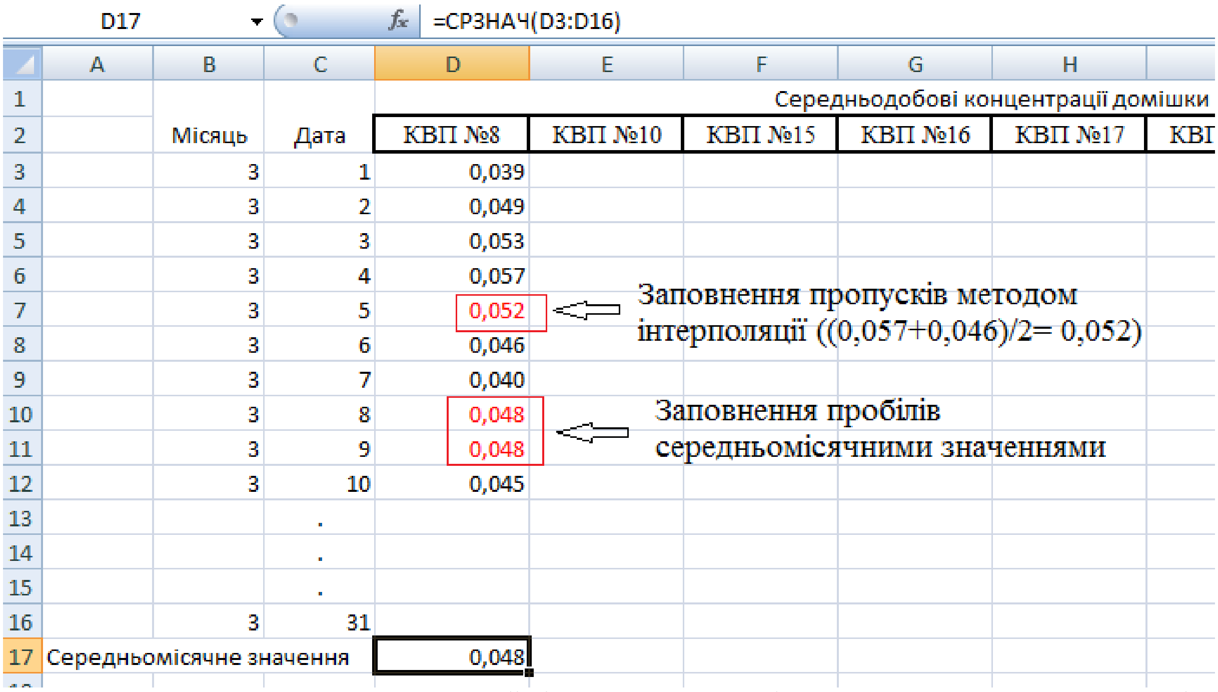


Рисунок 2.1 – Приклад розрахунку середньодобових значень концентрації домішки

Зверніть увагу на те, що вимірювання концентрацій домішок проводяться на постах в усі дні крім неділі та святкових днів. Таким чино є необхідність заповнити існуючі пробіли у рядах.

При наявності пропуску в один день він заповнюється середнім арифметичним значенням сусідніх значень (***методом інтерполяції***). Якщо у рядах відсутні спостереження за декілька днів – вони заповнюються середньомісячними значеннями концентрації домішки (рис. 2.2) [5].



D17 =СРЗНАЧ(D3:D16)

	A	B	C	D	E	F	G	H	
1						Середньодобові концентрації домішки			
2		Місяць	Дата	КВП №8	КВП №10	КВП №15	КВП №16	КВП №17	КВІ
3		3	1	0,039					
4		3	2	0,049					
5		3	3	0,053					
6		3	4	0,057					
7		3	5	0,052					
8		3	6	0,046					
9		3	7	0,040					
10		3	8	0,048					
11		3	9	0,048					
12		3	10	0,045					
13			.						
14			.						
15			.						
16		3	31						
17	Середньомісячне значення			0,048					

Рисунок 2.2 – Заповнення пробілів середньодобових значень концентрації домішки

2.3 Контрольні запитання

1. Дати визначення генеральної сукупності, вибіркової сукупності.
2. Яким вимогам повинні відповідати вибірки, по яких розраховуються статистичні оцінки параметрів розподілу?
3. Які члени статистичного ряду називаються однорідними?
4. Що називається статистичною гіпотезою та як формулюється основна гіпотеза по перевірці членів статистичного ряду на однорідність?
5. Які статистичні оцінки моментів розподілу випадкової величини необхідно розрахувати, щоб отримати фактичний критерій Стьюдента для знаходження «викидів» у статистичній сукупності?
6. Від яких величин залежить критичне значення критерію Стьюдента для перевірки членів статистичного ряду на однорідність?
7. Які випадкові величини в статистичній сукупності називаються «викидами»?
8. Якщо вибірка має «викиди», як необхідно діяти, щоб розрахувати статистичні оцінки параметрів розподілу випадкової величини?

2.4 Завдання до практичної роботи

1 В якості вихідних використовуються ряди значень вмісту домішок у атмосфері, що отримані у першій практичній роботі.

2 Розрахувати середнє значення випадкової величини та середній квадратичний відхил.

3 Провести перевірку членів ряду на однорідність.

4 Розрахувати середньодобові концентрації домішок в атмосфері на **усіх** постах спостережень.

5 Отримати безперервні ряди концентрації домішки.

ПРАКТИЧНА РОБОТА № 3
«РОЗРАХУНОК І АНАЛІЗ СТАТИСТИЧНИХ ХАРАКТЕРИСТИК СЕРЕДНЬОДОБОВИХ ЗНАЧЕНЬ КОНЦЕНТРАЦІЇ ДОМІШКИ. АНАЛІЗ ПРОСТОРОВОЇ ОДНОРІДНОСТІ ЗАБРУДНЕННЯ АТМОСФЕРНОГО ПОВІТРЯ МІСТА ЗА ДОПОМОГОЮ КРИТЕРІЮ ВІЛКОКСОНА»

3.1 Побудова та графічне представлення згрупованого ряду

3.1.1 Загальні положення

Поняття *статистична сукупність*, то має на увазі дві категорії:

- *генеральна сукупність;*
- *статистичний ряд (вибірка).*

Під **генеральною сукупністю** розуміють нескінченну множину незалежних значень випадкової величини, які підпорядковуються одному й тому ж закону розподілу.

Статистичний ряд (вибірка) – обмежена кількість значень випадкової величини, добута випадковим чином із генеральної сукупності. Тому статистичні ряди називають вибірками із генеральної сукупності. Вибірки випадкові та число їх безмежне.

Важливою ознакою ряду є його об’єм. «Об’єм» – кількість членів, що складають сукупність.

Вибірки характеризуються статистичними **оцінками параметрів**.

Значення параметра генеральної сукупності, яке розраховують на основі вибірки, називають статистичною оцінкою ($\hat{\theta}$) цього параметра θ і позначають символом "^".

Первинною формою зображення екологічної інформації є **простий** статистичний ряд, значення якого розташовуються в хронологічній послідовності.

Вихідні дані подаються у вигляді *простого* статистичного ряду головним чином у тих випадках, коли задача дослідження полягає у вивченні особливостей їх часової мінливості. Якщо така задача не ставиться, то ряди випадкових величин можуть зображатися у більш компактній формі – у вигляді ***згрупованого*** *ряду* [2, 3].

3.1.2 Побудова згрупованого ряду

Побудова згрупованого ряду на основі простого статистичного ряду проводиться наступним чином [4]:

1) визначають *область значень* величини X $\left[x_{min};x_{max}\right]$, де x_{min} – мінімальне, x_{max} – максимальне значення із вихідного ряду даних;
2) всі члени вихідного ряду ранжуються – розташовуються в напрямку їх збільшення (або зменшення);
3) знаходять **k** – *кількість часткових інтервалів (градацій)*, на які треба поділити область значень. Для цього використовється формула:

$$k = 5\lg n \ , \tag{3.1}$$

де n – об'єм ряду.

У випадку коли k не є цілим числом, його округлюють до цілого.

4) знаходиться *довжина часткового інтервалу* **c** за формулою:

$$c = \frac{x_{max} - x_{min}}{k} \ ; \tag{3.2}$$

5) визначається *значення* випадкової величини X *на межах часткових інтервалів*. Для i-того часткового інтервалу, значення величини X *на лівій межі* є $\left[x_{min} + (i-1)c\right]$, а *на правій* – $\left[x_{min} + ic\right]$. Можна позначати *ліву* межу часткового інтервалу як x_{i-1}, а *праву* – x_{i+1} $\left(i = \overline{1,k}\right)$.

Кінець попередньої і початок наступної градацій будуть повторюватися. Тому треба визначити закриті й відкриті межі градацій, тобто встановити, яку з величин враховувати в даній градації, щоб виключити повторення одних і тих же значень випадкової величини X, які дорівнюють значенню на межі градації [3, 4];

6) підраховують *кількість членів ряду*, що потрапляють до кожного i- того часткового інтервалу – m_i $\left(i = 1, 2, \ldots, k\right)$. Величини m_i називають *інтервальними емпіричними частотами*. Сума частот по всіх часткових інтервалах дорівнює об'єму вибірки *X*,
7) розраховують *інтервальні частості* p_i (*відносні інтервальні частоти*) за формулою:

$$p_i = \frac{m_i}{n}. \tag{3.3}$$

8) знаходять $\tilde{x}_i$ – значення випадкової величини X на *середині* кожного *часткового інтервалу* за формулою:

$$\tilde{x}_i = \frac{x_{i-1} + x_{i+1}}{2} \quad \left(i = \overline{1,k}\right). \tag{3.4}$$

Згрупованим називають такий *ранжований статистичний ряд,* який представляють сукупністю значень випадкової величини X на серединах часткових інтервалів $\tilde{x}_i$ і відповідних інтервальних частот m_i, або частостей p_i:

$$X : \begin{cases} \tilde{x}_1, \tilde{x}_2, \tilde{x}_3, ..., \tilde{x}_i, ..., \tilde{x}_k \\ m_1, m_2, m_3, ..., m_i, ..., m_k \end{cases} \left(i = \overline{1,k}\right). \tag{3.5}$$

Згрупований ряд, може зображатися за допомогою діаграм: *гістограми* чи *полігону*.

Гістограма – це система прямокутників, основи яких дорівнюють довжині часткового інтервалу, а висоти – відповідним інтервальним частотам (або частостям). Якщо точки з координатами $\left(\tilde{x}_i; m_i\right)$ або $\left(\tilde{x}_i; \hat{p}_i\right)$ з'єднати відрізками прямої, то отриману таким чином діаграму називають *полігоном* [3, 4].

3.2 Розрахунок статистичних оцінок параметрів розподілу

3.2.1 Основні положення

Основні властивості випадкових величин характеризуються початковими (ν), центральними (μ) та основними (r) моментами розподілу різних порядків (l). В дослідженнях, як правило, використовуються перелічені моменти перших чотирьох порядків ($l = \overline{1,4}$).

Оцінка **першого початкового моменту** розподілу $(\hat{\nu}_1)$ є оцінкою математичного сподівання $(\hat{m}_x)$ і дорівнює середньому значенню $(\overline{x})$ випадкової величини X:

$$\hat{\nu}_1 = \hat{m}_x = \overline{x} = \frac{1}{n} \sum_{i=1}^{k} \tilde{x}_i m_i \tag{3.6}$$

Центральні моменти розподілу оцінюються, починаючи з другого моменту (l=2), тому, що перший центральний момент завжди дорівнює нулю, як і його оцінка. Центральний момент другого порядку має сенс дисперсії випадкової величини: $\mu_2 = \sigma_x^2$.

Статистичної оцінка **центрального моменту розподілу другого** порядку на основі згрупованого ряду розраховується за формулою:

$$\hat{\mu}_2 = \hat{\sigma}_x^2 = \frac{1}{n}\sum_{i=1}^{k}\left(\tilde{x}_i - \overline{x}\right)^2 m_i\,, \tag{3.7}$$

Статистична оцінка другого центрального моменту розподілу, що розраховується за формулою (3.7), є зсуненою оцінкою дисперсії [3].

Незсунену, ефективну та умотивовану оцінку дисперсії випадкової величини X позначають S_x^2 і розраховують за формулою (3.8):

$$S_x^2 = \frac{1}{n-1}\sum_{i=1}^{k}(\tilde{x}_i - \overline{x})^2 m_i\,. \tag{3.8}$$

Статистична оцінка *середнього квадратичного відхилу* цієї величини розраховується за формулою:

$$S_x = \sqrt{S_x^2}\,. \tag{3.9}$$

Оцінка третього основного моменту характеризує асиметрію кривої розподілу інтервальних частостей (або частот) і називається ***коефіцієнтом асиметрії***: $\hat{r}_3 = As$.

$$As = \hat{r}_3 = \frac{1}{n}\sum_{i=1}^{k}\frac{(\tilde{x}_i - \overline{x})^3 m_i}{S_x^3} \tag{3.10}$$

Крива розподілу має *правосторонню асиметрію* за умови $As > 0$, і *лівосторонню* – за умови $As < 0$. Вона є *симетричною* відносно центру розподілу, якщо $As = 0$ (рис 3.1) [2].

Крім асиметрії, крива розподілу, порівняно з кривою нормального розподілу, може бути *витягнутою* або *сплюснутою* (рис 3.2). Мірою цього є ***коефіцієнт ексцесу E***:

$$E = r_4 - 3 = \left(\frac{1}{n}\sum_{i=1}^{k}\frac{(\tilde{x}_i - \overline{x})^4 m_i}{S_x^4}\right) - 3\,. \tag{3.11}$$

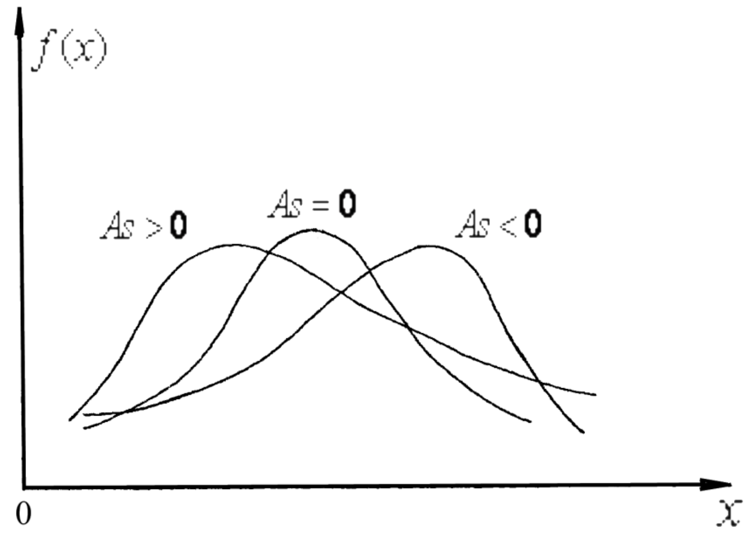


Рисунок 3.1 – Форми кривих розподілу при різних значеннях коефіцієнту асиметрії

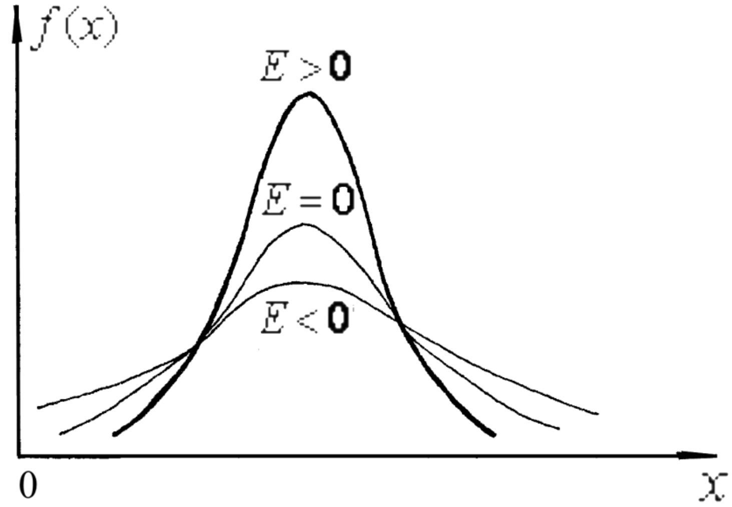


Рисунок 3.2 – Форми кривих розподілу при різних значеннях коефіцієнту ексцесу

Інколи при статистичних дослідженнях рядів екологічної інформації необхідно визначити модальне значення Mo та медіану Me.

Мода – значення випадкової величини, що зустрічається частіше усього, тобто має максимальну ймовірність для дискретної величини або максимум функції щільності ймовірності в даній точці для безперервної випадкової величини.

Якщо ряд є *згрупованим* та *ранжування* проводилося *в бік зростання* значень випадкової величини, то моду визначають за формулою:

$$M_o = x_0 + c\frac{(m_i - m_{i-1})}{(2m_i - m_{i-1} - m_{i+1})}, \tag{3.12}$$

де x_0, c, m_i – відповідно початок, довжина та емпірична частота модального інтервалу;

m_{i-1}, m_{i+1} – частоти попереднього і наступного за модальним часткових інтервалів.

3.3 Перевірка гіпотези про однорідність двох рядів за допомогою непараметричного критерію Вілкоксона

Непараметричні критерії перевірки основної гіпотези H_0 про однорідність рядів випадкових величин X та Y використовуються в тому випадку, коли ряди не підпорядковуються нормальному закону розподілу, або інформація про закон розподілу відсутня.

Одним із непараметричних критеріїв є критерій Вілкоксона. Цей критерій використовується в двох варіантах: перший базується на підрахунках числа інверсій, другий – ранговий [2, 3].

Основна гіпотеза H_0 формулюється таким чином: ряди випадкових величин X і Y належать до однієї генеральної сукупності на рівні значущості α.

3.3.1 Інверсійний критерій Вілкоксона

Для перевірки гіпотези ряди випадкових величин об'єднують у загальну послідовність і проводять операцію ранжування.

Якщо якому-небудь значенню x у загальній послідовності передує деяке значення y, то кажуть, що ця пара утворює інверсію. Сумарна кількість інверсій через позначається через U.

В однорідних рядах, кожний з котрих має не менше 10 членів, число інверсій розподіляється приблизно за нормальним законом. Його параметри розраховуються за формулами:

$$m_U = \frac{m \cdot n}{2} \tag{3.13}$$

і

$$\sigma_U^2 = \frac{m \cdot n}{12}(m + n + 1), \tag{3.14}$$

де n і m - кількість членів у першій та другій вибірках [3].

Для перевірки гіпотези H_0 необхідно знайти межі допустимих значень U, що відділяють область прийняття гіпотези H_0 від критичної області, яка є двосторонньою. Якщо значення критерію, що отримане за даними спостережень, попаде до області прийняття гіпотези, то гіпотеза H_0 на цьому рівні значущості не відхиляється. У протилежному випадку приймається альтернативна гіпотеза H_1.

Область прийняття гіпотези H_0 визначається виразом:

$$m_U - t_{кр}(\alpha,\nu)\sigma_U \le U \le m_U + t_{кр}(\alpha,\nu)\sigma_U, \qquad (3.15)$$

а критична – нерівностями:

$$U_{кр_1} < m_U - t_{кр}(\alpha,\nu)\sigma_U, \qquad (3.16)$$

$$U_{кр_2} > m_U + t_{кр}(\alpha,\nu)\sigma_U. \qquad (3.17)$$

де $\sigma_U = \sqrt{\sigma_U^2}$ – середній квадратичний відхил числа інверсій; $t_{кр}(\alpha,\nu)$ – критерій Стьюдента для рівня значущості α і числа степенів вільності $\nu = m + n - 2$ (Додаток А) [4].

3.3.2 Ранговий варіант критерію Вілкоксона

Перевірка гіпотези H_0 за допомогою рангового критерію Вілкоксона відбувається на основі об’єднаної ранжованої (у бік зростання) послідовності рядів випадкових величин.

Кожному члену об’єднаної послідовності присвоюється порядковий номер (ранг). Якщо однакові значення належать різним вибіркам, то їм присвоюють один і той же ранг (середній) [4].

Наступним кроком є визначення в об’єднаному ряді суми рангів членів вибірки, об’єм якої менший. Дана сума позначається як $W_{виб.}$.

Нижню критичну точку визначають за формулою:

$$W_{кр.н}\left(m,n,\alpha/2\right) = \frac{(m+n+1)\cdot m - 1}{2} - Z_{кр}\sqrt{\frac{m\cdot n\cdot(m+n+1)}{12}}, \qquad (4.6)$$

де m і n – об’єми вибірок випадкових величин X і Y;
α – рівень значущості;

$Z_{кр}$ – визначається по таблицях функції Лапласа за правилом:

$$\Phi\left(Z_{кр}\right)=\frac{1-\alpha}{2}. \tag{4.7}$$

За умови α=0.05, функція Лапласа $\Phi\left(Z_{кр}\right)=\frac{1-\alpha}{2}=0.475$ і $Z_{кр}=1.96$. Верхня критична точка розраховується за формулою:

$$W_{кр.в}=\left(m+n+1\right)\cdot m-W_{кр.н}. \tag{4.8}$$

Гіпотеза H_0 на заданому рівні значущості α не відкидається, якщо

$$W_{кр.н}<W_{виб.}<W_{кр.в}. \tag{4.9}$$

За умов $W_{виб.}<W_{кр.н}$ або $W_{виб.}>W_{кр.в}$ приймається альтернативна гіпотеза H_1 про те, що ряди X та Y є неоднорідними [4].

3.4 Приклад розрахунку статистичних оцінок моментів розподілу

Статистичні оцінки моментів розподілу розраховуються на основі ряду концентрації діоксиду сірки в атмосфері (мг/м^3) члени якого пройшли перевірку на однорідність (табл. 3.1).

Таблиця 3.1 – Ранжований ряд середньодобової концентрації діоксиду сірки в атмосфері, мг/м^3

i	**0**	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**
0		0,35	0,35	0,35	0,37	0,38	0,38	0,39	0,39	0,39
10	0,39	0,39	0,39	0,39	0,40	0,40	0,40	0,40	0,41	0,41
20	0,41	0,41	0,41	0,41	0,42	0,42	0,42	0,42	0,43	0,43
30	0,43	0,43	0,43	0,43	0,43	0,43	0,43	0,44	0,44	0,44
40	0,44	0,44	0,44	0,44	0,44	0,44	0,45	0,45	0,45	0,45
50	0,45	0,45	0,45	0,46	0,46	0,46	0,46	0,46	0,46	0,46
60	0,46	0,46	0,47	0,47	0,47	0,47	0,48	0,48	0,48	0,48
70	0,48	0,48	0,48	0,48	0,48	0,49	0,49	0,49	0,50	0,50
80	0,50	0,51	0,51	0,51	0,51					

На основі методики, що викладена в п. 3.1 отримаємо:

$x_{\min} = 0{,}35$ мг/м^3, $x_{\max} = 0{,}51$ мг/м^3 $n = 84$, $k = 8$, $c \approx 0{,}02$.

Результати розрахунків приводяться у табл. 2.7

Таблиця 3.2 – Згрупований ряд середньодобової концентрації діоксиду сірки в атмосфері, мг/м^3

i	Градації	m_i	m_i^*	p_i	$\tilde{x}_i$,
1	**2**	**3**	**4**	**5**	**7**
1	[0,35…0,37 [	4	4	0,05	0,36
2	[0,37…0,39 [	9	13	0,11	0,38
3	[0,39…0,41 [	10	23	0,12	0,40
4	[0,41…0,43 [	13	36	0,15	0,42
5	[0,43…0,45 [	16	52	0,19	0,44
6	[0,45…0,47 [	13	65	0,15	0,46
7	[0,47…0,49 [	12	77	0,14	0,48
8	[0,49…0,51 [	7	84	0,08	0,50

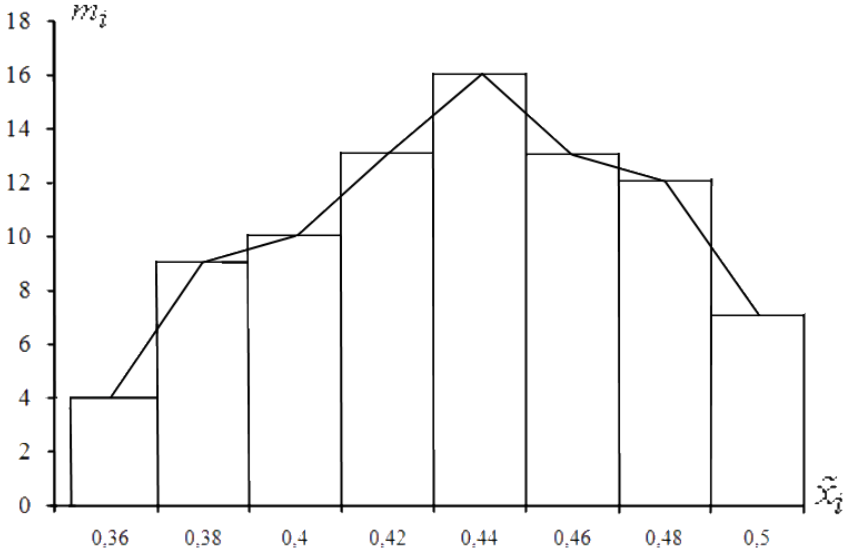


Рисунок 3.3 – Гістограма та полігон розподілу ряду середньодобової концентрації діоксиду сірки в атмосфері (Одеса, КВП №8, 1.03 – 31.05.2003 р.)

Використавши формули (3,6 – 3,11) отримаємо:

$$\bar{x} = 0{,}44$$

$$S_x^2 = \frac{n}{n-1}\hat{\mu}_2 = \frac{1}{n-1}\sum_{i=1}^{k}(\tilde{x}_i - \overline{x})^2 m_i = 0,00155 .$$

$$S_x = \sqrt{S_x^2} = \sqrt{0,00155} = 0,04 .$$

$$As = \hat{r}_3 = \frac{\hat{\mu}_3}{(S_x)^3} = \frac{-2,8 \cdot 10^{-5}}{0,04^3} = -0,14 .$$

$$E = \hat{r}_4 - 3 = 2,05 - 3 = -0,95 .$$

Обчислення модального значення випадкової величини проводиться за формулою (3,12):

$$M_0 = x_0 + c\frac{(m_i - m_{i-1})}{2m_i - m_{i-1} - m_{i+1}} = 0,43 + 0,02\frac{(16-13)}{2\times 16 - 13 - 13} = 0,44 ,$$

де $x_0 = 0,43$, $c = 0,02$, $m_i = 16$, $m_{i-1} = 13$, $m_{i+1} = 13$.

3.5 Приклад визначення однорідності двох рядів

Перевірку статистичної гіпотези будемо проводити на основі рядів середньодобової концентрації двоокису сірки в атмосфері (мг/м^3) в Одесі на КВП №15 (ряд X) та КВП №17 (ряд Y) влітку 2001 року (табл. 3.3 – 3.4). Об’єми вибірок складають по 92 члени.

Таблиця 3.3 – Ряд X

i	1	2	3	4	5	6	7	8	9	10
0	0,31	0,58	0,67	0,57	0,77	0,77	0,78	0,77	0,84	0,76
10	0,72	0,66	0,6	0,53	0,63	0,68	0,58	0,52	0,67	0,59
20	0,51	0,6	0,48	0,52	0,61	0,38	0,39	0,39	0,4	0,58
30	0,33	0,3	0,24	0,29	0,34	0,41	0,38	0,32	0,37	0,37
40	0,36	0,36	0,41	0,43	0,43	0,42	0,26	0,3	0,35	0,39
50	0,42	0,48	0,72	0,48	0,47	0,45	0,39	0,4	0,49	0,41
60	0,53	0,57	0,6	0,49	0,53	0,38	0,53	0,41	0,45	0,48
70	0,53	0,55	0,46	0,7	0,74	0,75	0,76	0,69	0,54	0,61
80	0,48	0,6	0,59	0,58	0,59	0,6	0,52	0,43	0,46	0,4
90	0,33	0,51								

Таблиця 3.4 – Ряд *Y*

i	1	2	3	4	5	6	7	8	9	10
0	0,54	0,53	0,52	0,5	0,58	0,56	0,58	0,57	0,57	0,57
10	0,58	0,61	0,6	0,57	0,59	0,6	0,62	0,58	0,57	0,6
20	0,57	0,57	0,57	0,53	0,54	0,47	0,54	0,51	0,59	0,47
30	0,59	0,56	0,52	0,56	0,57	0,58	0,6	0,56	0,56	0,54
40	0,57	0,57	0,56	0,56	0,6	0,64	0,57	0,55	0,58	0,6
50	0,62	0,63	0,6	0,64	0,64	0,6	0,59	0,58	0,62	0,61
60	0,57	0,62	0,58	0,58	0,58	0,69	0,51	0,56	0,7	0,62
70	0,58	0,54	0,6	0,56	0,64	0,58	0,54	0,54	0,55	0,6
80	0,57	0,57	0,58	0,58	0,59	0,61	0,62	0,58	0,62	0,58
90	0,61	0,64								

Після об'єднання рядів та ранжування отримаємо (члени ряду *Y* підкреслені):

0,24	0,26	0,29	0,30	0,30	0,31	0,32	0,33	0,33	0,34
0,35	0,36	0,36	0,37	0,37	0,38	0,38	0,38	0,39	0,39
0,39	0,39	0,40	0,40	0,40	0,41	0,41	0,41	0,41	0,42
0,42	0,43	0,43	0,43	0,45	0,45	0,46	0,46	0,47	$\underline{0,47}$
$\underline{0,47}$	0,48	0,48	0,48	0,48	0,48	0,49	0,49	$\underline{0,50}$	0,51
$\underline{0,51}$	0,51	$\underline{0,51}$	0,52	0,52	$\underline{0,52}$	$\underline{0,52}$	0,52	0,53	$\underline{0,53}$
0,53	0,53	0,53	$\underline{0,53}$	0,53	$\underline{0,54}$	$\underline{0,54}$	0,54	$\underline{0,54}$	$\underline{0,54}$
$\underline{0,54}$	$\underline{0,54}$	$\underline{0,54}$	0,55	$\underline{0,55}$	$\underline{0,55}$	$\underline{0,56}$	$\underline{0,56}$	$\underline{0,56}$	$\underline{0,56}$
$\underline{0,56}$	$\underline{0,56}$	$\underline{0,56}$	$\underline{0,56}$	$\underline{0,56}$	0,57	$\underline{0,57}$	$\underline{0,57}$	$\underline{0,57}$	$\underline{0,57}$
$\underline{0,57}$	$\underline{0,57}$	$\underline{0,57}$	$\underline{0,57}$	$\underline{0,57}$	$\underline{0,57}$	$\underline{0,57}$	$\underline{0,57}$	$\underline{0,57}$	0,57
$\underline{0,57}$	$\underline{0,57}$	$\underline{0,58}$	$\underline{0,58}$	$\underline{0,58}$	$\underline{0,58}$	$\underline{0,58}$	$\underline{0,58}$	0,58	$\underline{0,58}$
$\underline{0,58}$	0,58	$\underline{0,58}$	0,58	0,58	$\underline{0,58}$	$\underline{0,58}$	$\underline{0,58}$	$\underline{0,58}$	$\underline{0,58}$
$\underline{0,58}$	$\underline{0,58}$	$\underline{0,59}$	$\underline{0,59}$	0,59	$\underline{0,59}$	$\underline{0,59}$	$\underline{0,59}$	0,59	0,59
0,60	$\underline{0,60}$	$\underline{0,60}$	$\underline{0,60}$	$\underline{0,60}$	$\underline{0,60}$	0,60	$\underline{0,60}$	0,60	$\underline{0,60}$
0,60	0,60	$\underline{0,60}$	$\underline{0,60}$	$\underline{0,60}$	0,61	$\underline{0,61}$	$\underline{0,61}$	0,61	$\underline{0,61}$
$\underline{0,61}$	$\underline{0,62}$	$\underline{0,62}$	$\underline{0,62}$	$\underline{0,62}$	$\underline{0,62}$	$\underline{0,62}$	$\underline{0,62}$	0,63	$\underline{0,63}$
$\underline{0,64}$	$\underline{0,64}$	$\underline{0,64}$	$\underline{0,64}$	$\underline{0,64}$	0,66	0,67	0,67	0,68	$\underline{0,69}$
0,69	$\underline{0,70}$	0,70	0,72	0,72	0,74	0,75	0,76	0,76	0,77
0,77	0,77	0,78	0,84						

3.5.1 Використання інверсійного критерію Вілкоксона

1. Розраховуємо параметри нормального розподілу числа інверсій за формулами (3.13 – 3.14):

$m_U = \frac{92 \cdot 92}{2} = 4232$

$\sigma_U^2 = \frac{92 \cdot 92}{12}\left(92 + 92 + 1\right) = 130486,7$

$\sigma_U = \sqrt{\sigma_U^2} = 361,3$

2. Підраховуємо число інверсій:

По X : U_x = 0 + 0 + 0 + 0 + 0 + 0 + 0 + 0 + 0 + 0 + 0 + 0 + 0 + 0 + 0 + 0 + 0 + 0 +0 + 0 + 0 + 0 + 0 + 0 + 0 + 0 + 0 + 0 + 0 + 0 + 0 + 0 + 0 + 0 + 0 + 0 + 0 + 0 + 0 + 0 + 2 + 2 + 2 + 2 + 2 + 2 + 2 + 3 + 4 + 5 + 5 + 7 +7 + 8 + 8 + 8 + 9 + 11 + 16 + 27 + 40 + 48 + 50 + 51 + 51 + 60 + 63 + 63 + 63 + 68 + 69 + 70 + 70 + +73+ 75 + 84 + 90 + 90 + 90 + 90 + 91 + 92 + 92 + 92 + 92 + 92 + 92 + 92 + 92 + 92 + 92 + 92 + 92 = 2685

По Y :U_y= 39 + 39 + 46 + 47 + 48 + 50 + 50 + 52 + 55 + 56 + 56 + 57 + 57 + 57 + 57 + 57 + 58 + 58 + 58 + 58 + 58 + 58 + 58 + 58 + 58 + 58 + 58 + 59 + 59 + 59 + 59 + 59 + 59 + 59 + 59 + 59 + 59 + 59 + 59 + 59 + 60 + 60 + 60 + 60 + 60 + 60 + 60 + 60 + 61 + 61 + 62 + 64 + 64 + 64 + 64 + 64 + 64 + 64 + 64 + 64 + 65 + 65 + 65 + 68 + 68 + 68 + 68 + 68 + 69 + 70 + 72 + 72 + 72 + 73 + 73 + 74 + 74 + 74 + 74 + 74 + 74 + 74 + 74 + 74 + 75 + 75 + 75 + 75 + 75 + 75 + 79 + 80 + =5779

3. За формулами (4.4 – 4.5) знаходимо межі критичної області прийняття гіпотези H_0:

Із додатку А – $t_{кр}\left(\alpha,\nu\right) = t_{кр}\left(0{,}05,\ 182\right) = 0{,}77$

$U_{кр_1} = 4232 - 0{,}77 \cdot 361{,}3 = 3953{,}85$,

$U_{кр_2} = 4232 + 0{,}77 \cdot 361{,}3 = 4510{,}15$

4. Перевіряємо чи потрапляє число інверсій по X та по Y в область прийняття гіпотези H_0 (4.3):

3953.85 > 2685 <4510.15

3953.85 < 5779 > 4510.15

Як бачимо жодне з чисел інверсій не потрапляє до критичної області, а значить на рівні значущості $\alpha = 0{,}05$ ряди не належать до однієї генеральної сукупності.

3.5.2 Використання рангового критерію Вілкоксона

1. Виконуємо розрахунок кількості рангів у загальному ряду:

$W_{виб.} = 6963$

2. За формулами (4.6, 4.8) знаходимо значення нижньої ($W_{кр.н}$) та верхньої ($W_{кр.в}$) критичних точок:

$W_{кр.н} = 7801{,}49$

$W_{кр.в} = 9218{,}51$

3. Перевіряємо чи виконується нерівність (4.9):

$W_{виб.} < W_{кр.н} < W_{кр.в}$

$6963 < 7801{,}49 < 9218{,}51$

Отже на заданому рівні значущості $\alpha = 0{,}05$ ряди не належать до однієї генеральної сукупності.

3.6 Контрольні запитання

9. Дати визначення генеральної сукупності, вибіркової сукупності.
10. Перелічити форми зображення статистичних рядів (вибірок).
11. Дати визначення простої статистичної сукупності? згрупованої?
12. Від яких величин залежить кількість градацій у згрупованій сукупності?
13. Від яких величин залежить розмір градації (довжина часткового інтервалу)?
14. Дати визначення інтервальної емпіричної частоти. Як за допомогою цієї величини знайти об’єм вибірки?
15. Дати визначення об’єму вибірки.
16. Дати визначення інтервальної частості. Чому дорівнює сума частостей по всіх градаціях?
17. Як графічно можна представити згруповані ряди?
18. Як розрахувати середнє значення випадкової величини на основі простої згрупованого ряду?
19. Статистичною оцінкою якого моменту розподілу є середнє значення випадкової величини (вибірки)?
20. Який сенс дисперсії випадкової величини та з яким моментом розподілу вона має зв'язок?
21. Як розрахувати незсунену, ефективну та умотивовану оцінку дисперсії на основі згрупованого ряду?
22. Як розрахувати коефіцієнт асиметрії та в чому полягає його сенс?
23. Як називають криву розподілу за умови $A_S = 0$? $A_S > 0$? $A_S < 0$?
24. Як розрахувати коефіцієнт ексцесу та в чому полягає його сенс?
25. Який вид кривої розподілу відносно нормального закону будемо мати за умови $E = 0$? $E > 0$? $E < 0$?
26. Яким вимогам повинні відповідати вибірки, по яких розраховуються статистичні оцінки параметрів розподілу?

3.7 Завдання до практичної роботи

1 Побудувати згрупований ряд. В якості вихідних використовуються ряди середньодобових значень вмісту домішок у атмосфері, що отримані у попередній практичній роботі.

2 Розрахувати статистичні оцінки параметрів розподілу випадкової величини на основі згрупованого ряду.

3 Сформулювати основну гіпотезу H_0.

4 Провести перевірку гіпотези H_0 за допомогою інверсійного критерію Вілкоксона (для усіх постів). Зробити висновки щодо прийняття гіпотези H_0.

5 Провести перевірку гіпотези H_0 за допомогою рангового критерію Вілкоксона (для усіх постів). Зробити висновки щодо прийняття гіпотези H_0.

ПРАКТИЧНА РОБОТА № 4
«ВИЯВЛЕННЯ ПРИХОВАНИХ ПЕРІОДИЧНОСТЕЙ, ЗГЛАДЖУВАННЯ ЧАСОВИХ РЯДІВ КОНЦЕНТРАЦІЙ ДОМІШОК І АНАЛІЗ ДИНАМІКИ ДЕТЕРМІНІРОВАНОЇ КОМПОНЕНТИ ЧАСОВОГО РЯДУ»

4.1 Основні теоретичні положення

Часові ряди метеорологічних величин, як показали численні дослідження, утримують періодичні компоненти, обумовлені хвильовою природою атмосферних процесів.

Існує ряд методів дослідження періодичностей, що містяться в часових рядах. Їх називають «прихованими» періодичностями. Зручним для реалізації на ПЕОМ є метод, оснований на інтегральному перетворенню Фур'є. Він дає можливість без будь-яких додаткових досліджень отримати частоти, амплітуди та початкові фази періодичних компонент, «прихованих» у часовій послідовності [4].

Часовий ряд $X(t)$, заданий на інтервалі $t \in [-\tau, \tau]$, можна розглядати як кусково-гладку функцію часу. Таку функцію у відповідності до теореми Діріхлє можна виразити суперпозицією простих гармонік:

$$X(t) = \sum_{k=0}^{\infty} A_k \sin(\omega_k t + \varphi_k) \ , \tag{4.1}$$

де A_k – амплітуда k – тої гармоніки,

ω_k – її частота,

φ_k – початкова фаза.

Рівність (4.1) може бути переписаною таким чином:

$$X(t) = \sum_{k=0}^{\infty} [a_k \cos \omega_k t + b_k \sin \omega_k t] \ , \tag{4.2}$$

якщо позначити

$$a_k = A_k \sin \varphi_k \ , \tag{4.3}$$

$$b_k = A_k \cos \varphi_k \ . \tag{4.4}$$

Як свідчать формули (4.3) та (4.4), початкова фаза розраховується за допомогою рівняння (4.5) [4]:

$$\varphi_k = arctg\frac{a_k}{b_k} \ . \tag{4.5}$$

Для кусково-гладкої функції $X(t)$, заданої на нескінченному інтервалі, справедливим є перетворення Фур'є:

$$F(i\omega) = \int_{-\infty}^{\infty} X(t)e^{-i\omega t}dt \ , \tag{4.6}$$

де $i = \sqrt{-1}$ – уявна одиниця.

Але часові ряди фізичних величин визначені на обмеженому інтервалі і можуть бути апроксимовані таким чином [4]:

$$X(t) = \begin{cases} x(t), & за \ умови \quad t \in [-\tau, \tau] \ ; \\ 0, & за \ умови \quad |t| > \tau \ . \end{cases} \tag{4.7}$$

Для такої функції перетворення Фур'є має вигляд:

$$F_\tau(i\omega) = \frac{1}{\tau}\int_{-\tau}^{\tau} X(t)\, e^{-i\omega t}dt \tag{4.8}$$

Якщо впровадити відому формулу Ейлера, то інтеграл (4.8) приймає таку форму:

$$F_\tau(i\omega) = U(\omega) - iV(\omega) \ , \tag{4.9}$$

де

$$U(\omega) = \frac{1}{\tau}\int_{-\tau}^{\tau} X(t)\cos\omega t dt \ , \tag{4.10}$$

$$V(\omega) = \frac{1}{\tau}\int_{-\tau}^{\tau} X(t)\sin\omega t dt \ . \tag{4.11}$$

Рівності (4.10) та (4.11) є відповідно косинус- і синус-перетворення Фур'є функції $X(t)$, що апроксимована виразом (4.7). Якщо частоти гармонічних компонент, які утримуються в $X(t)$, не є близькими, то $U(\omega)$ і $V(\omega)$ мають вигляд кривих з різко вираженими піками в точках

$\omega = \omega_k$. Висота піків приблизно дорівнює амплітудам парної (a_k) і непарної (b_k) складових періодичного коливання з частотою ω_k, «прихованого» в процесі $X(t)$. На тих самих частотах ω_k будуть спостерігатися піки амплітуд $A_k = A(\omega_k)$, оскільки $a_k \approx U(\omega_k)$ і $b_k \approx V(\omega_k)$ [4].

Амплітуду k-тої гармоніки в такому разі розраховують за формулою:

$$A(\omega_k) = [U^2(\omega_k) + V^2(\omega_k)]^{\frac{1}{2}}. \qquad (4.12)$$

З метою поліпшення селективних якостей перетворень Фур'є (4.10) та (4.11) в них вводять множники («вікна»), які зменшують вплив значень $X(t)$, заданих поблизу меж інтервалу визначення функції. Одним з таких «вікон» є множник Гіббса [4]:

$$g(t) = \frac{\sin\frac{\pi t}{\tau}}{\frac{2t}{\tau}} \quad . \qquad (4.13)$$

Очевидно, $g(\tau) = g(-\tau) = 0$. Максимум цієї функції спостерігається за умови $\tau = 0$ і дорівнює:

$$g(0) = \frac{\pi}{2} \lim_{t \to 0} \frac{\sin\frac{\pi t}{\tau}}{\frac{\pi t}{\tau}} = \frac{\pi}{2} \quad .$$

Таким чином, графік функції $g(t)$ на інтервалі $[-\tau, \tau]$ має вигляд, зображений на рис. 4.1.

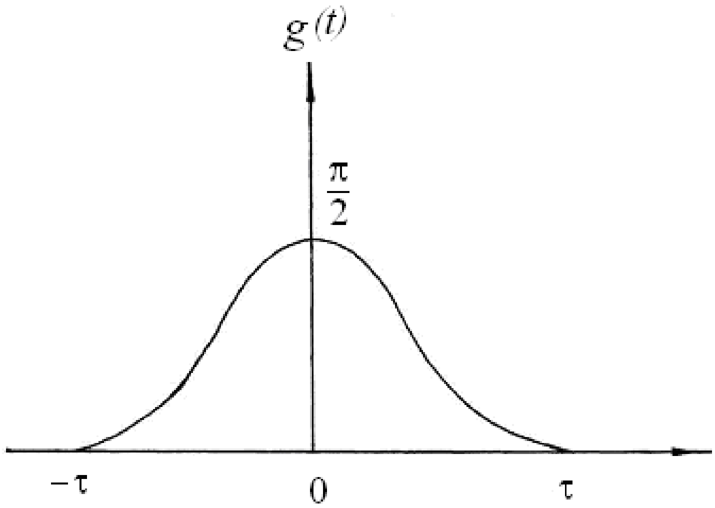


Рис. 4.1 – «Вікно» Гіббса

Отже, з врахуванням «вікна» Гіббса косинус- та синус-перетворення Фур’є описуються рівняннями (4.14) і (4.15) відповідно [4]:

$$U(\omega)=\frac{1}{\tau}\int_{-\tau}^{\tau}\frac{\sin\frac{\pi t}{\tau}}{\frac{2t}{\tau}}X(t)\cos\omega t dt\ , \qquad (4.14)$$

$$V(\omega)=\frac{1}{\tau}\int_{-\tau}^{\tau}\frac{\sin\frac{\pi t}{\tau}}{\frac{2t}{\tau}}X(t)\sin\omega t dt\ . \qquad (4.15)$$

Інтеграли (4.14) і (4.15) обчислюються одним із наближених методів.

Із-за обмеженості інтервалу й скінченної кількості точок задання функції, інформації про функцію недостатньо для визначення параметрів гармонік з періодом $T > 2\tau$ і $T < \frac{\tau}{m}$.

Отже, мінімально і максимально можливі гармоніки, які можуть бути виявлені, мають частоти, що розташовані в інтервалі $(\omega_{\min}, \omega_{\max})$, де $\omega_{\min} = \frac{\pi}{\tau}$, $\omega_{\max} = \frac{m\pi}{\tau}$.

Інтервал дискретності частоти $\Delta\omega$ при чисельному інтегруванні вибирається з урахуванням властивостей множника Гіббса. При його впровадженні у перетворення Фур’є можна гарантувати, що вплив амплітуд сусідніх за частотою гармонік не перебільшує 0,05 від амплітуди, за умови $\Delta\omega\tau \ge 4.5$. Звідси випливає, що крок при обчисленнях $U(\omega)$ та $V(\omega)$ визначається рівністю [4]:

$$\Delta\omega = \frac{4.5}{\tau}\ . \qquad (4.16)$$

Періодичності, що утримуються у випадковому процесі, визначаються по піках амплітуд $A(\omega_i)$ на періодограмі (амплітудно-частотній характеристиці), приклад якої наводиться у типовій задачі (рис. 4.2). На періодограмі існує ряд малозабезпечених піків, ускладнюючих аналіз. Для їх ліквідації застосовують фільтр Тьюккі :

$$\tilde{A}(\omega_i) = 0.25A(\omega_{i-1}) + 0.5A(\omega_i) + 0.25A(\omega_{i+1})\ . \qquad (4.17)$$

Визначення статистично значущих періодичностей, характерних для процесу $X(t)$, основується на побудові верхньої довірчої межі для амплітуд з заданою ймовірністю за умови, що амплітуди підпорядковуються нормальному розподілу. Періоди T_k гармонік ω_k, які відповідають пікам амплітуд, що виходять за довірчу межу, ототожнюються з періодами гармонічних коливань, які утримуються у випадковому процесі $X(t)$. Для кожного з них знаходять початкову фазу [4]:

$$\varphi_k = arctg \frac{U(\omega_k)}{V(\omega_k)} \ . \tag{4.18}$$

Початкова фаза дає можливість знайти точку h_k на осі часу, яка є початком коливання, тобто [4]:

$$h_k = \begin{cases} \dfrac{(\pi + \varphi_k)}{2\pi} \cdot T_k & \text{за умови} \quad \varphi_k \le 0; \\ \dfrac{\varphi_k \cdot T_k}{2\pi} & \text{за умови} \quad \varphi_k > 0, \end{cases} \tag{4.19}$$

де φ_k – початкова фаза гармонічного коливання в радіанах;

T_k – період гармонічного коливання в одиницях часу, який розраховується за формулою:

$$T_k = \frac{2\pi}{\omega_k}. \tag{4.20}$$

Треба мати на увазі, що у деяких випадках на амплітудно-частотних характеристиках спостерігаються сплески амплітуд (найчастіше на низьких частотах), які значно відрізняються від загального рівня коливань амплітуд гармонік. Це означає, що ці гармоніки характеризуються найбільшою, порівняно до інших, енергією. У таких випадках насамперед треба перевірити гіпотезу про те, що такі сплески амплітуд належать до тієї ж генеральної сукупності, що й амплітуди інших гармонік. У разі неприйняття цієї гіпотези такі сплески при розрахунках середнього значення амплітуди $\overline{A}$ й середнього квадратичного відхилу амплітуд σ_A треба виключити із сукупності амплітуд і після цього розрахувати для неї довірчу межу з тією чи іншою ймовірністю для визначення статистично значущих періодичностей [4].

4.2 Приклад визначення прихованих періодичностей часового ряду

Для часових рядів середньомісячних концентрацій формальдегіду на КВП міста, за методикою наведеною вище, визначені приховані періодичності.

Отримані такі характеристики: амплітуда i-тої гармоніки (A_i), її частота (ω_i), початкова фаза (φ_i), косинус- та синус-перетворення Фур'є – $U(t)$ та $V(t)$ відповідно (табл.4.1).

Таблиця 4.1 – Амплітудно-частотні характеристики періодичностей в часових рядах середньомісячних концентрацій формальдегіду, м.Одеса (КВП №10, 2008–2017 рр)

i	$\omega_i, міс^{-1}$	$T, міс$	A_i, мкг/м^3	$\varphi_i, рад.$	$U(t)$	$V(t)$
4	0,358	17,535	0,385	-0,275	-0,124	0,439
5	0,435	14,457	0,76	0,423	0,086	0,19
6	0,511	12,299	1,389	0,738	-1,456	-1,602
7	0,587	10,701	1,189	1,401	-1,005	-0,173
8	0,663	9,471	0,671	-1,118	-0,496	0,241
9	0,74	8,494	0,481	0,143	0,08	0,557
10	0,816	7,7	0,406	-1,411	-0,242	0,039
11	0,892	7,042	0,5	-0,26	0,147	-0,552
12	0,969	6,488	0,611	0,173	-0,105	-0,601
13	1,045	6,014	0,71	1,057	-0,567	-0,32
14	1,121	5,605	0,793	0,65	-0,56	-0,737
15	1,197	5,248	0,683	0,912	-0,531	-0,411
16	1,274	4,933	0,512	1,26	-0,443	-0,142
17	1,35	4,655	0,509	0,838	-0,332	-0,298
18	1,426	4,406	0,55	1,242	-0,642	-0,219
19	1,502	4,182	0,459	-0,13	-0,051	0,394
20	1,579	3,98	0,404	1,224	-0,341	-0,123
21	1,655	3,797	0,476	1,291	-0,473	-0,136
22	1,731	3,629	0,494	1,471	-0,553	-0,056
23	1,807	3,476	0,509	1,097	-0,332	-0,17
24	1,884	3,335	0,616	1,138	-0,666	-0,308
25	1,96	3,206	0,717	1,208	-0,581	-0,221
26	2,036	3,086	0,697	1,063	-0,778	-0,433
27	2,113	2,974	0,581	1,333	-0,375	-0,091
28	2,189	2,871	0,473	-0,703	-0,427	0,504
29	2,265	2,774	0,365	-0,239	0,043	-0,179

Продовження табл. 4.1

30	2,341	2,684	0,424	0,823	-0,315	-0,292
31	2,418	2,599	0,616	0,603	-0,37	-0,538
32	2,494	2,519	0,701	0,717	-0,479	-0,55
33	2,57	2,445	0,684	1,268	-0,661	-0,206
34	2,646	2,374	0,556	-1,548	-0,623	0,014
35	2,723	2,308	0,414	0,896	-0,224	-0,179
36	2,799	2,245	0,341	1,429	-0,453	-0,065
37	2,875	2,185	0,339	-0,313	0,05	-0,155
38	2,952	2,129	0,478	0,249	0,142	0,557
39	3,028	2,075	0,622	-0,642	-0,36	0,481
40	3,104	2,024	0,684	-1,061	-0,622	0,348
41	3,18	1,976	0,71	1,042	-0,613	-0,358

Розраховані статистики амплітуди: $\overline{A}$ = 0,6 мкг/м^3 та σ_A = 0,21 мкг/м^3 Використовуючи наведені в табл. 4.1 дані (2-й та 4-й стовпчики), була побудована амплітудно-частотна характеристика випадкового процесу, що розглядається (рис. 4.2).

Дискретність вихідних даних складає 1 місяць. Визначення статистично значущих періодичностей, характерних для цього процесу, основувалися на побудові верхньої довірчої межі для амплітуд з ймовірністю 68% та за умови, що амплітуди підпорядковуються нормальному розподілу.

Як свідчать характеристики, в часовому ряді середньомісячної концентрації формальдегіду на КВП №10 м Одеса максимальній амплітуді відповідає періодичність, що близька до одного року (табл. 4.2).

Часовий ряд гармонічних коливань складається з детермінованої та випадкової компонент. Детермінована основа процесу відокремлюється шляхом фільтрації або згладжування вихідного часового ряду з урахуванням максимальної періодичності.

Таблиця 4.2 – Амплітудно-частотні характеристики максимальної періодичності часового ряду середньомісячної концентрації формальдегіду на КВП №10 м. Одеса (2008–2017 рр.)

№ КВП	А, мкг/м^3	ω, ч$^{-1}$	Т, міс	$\varphi_к$	$h_к$
10	1,389	0,511	12,299	0,74	1,44

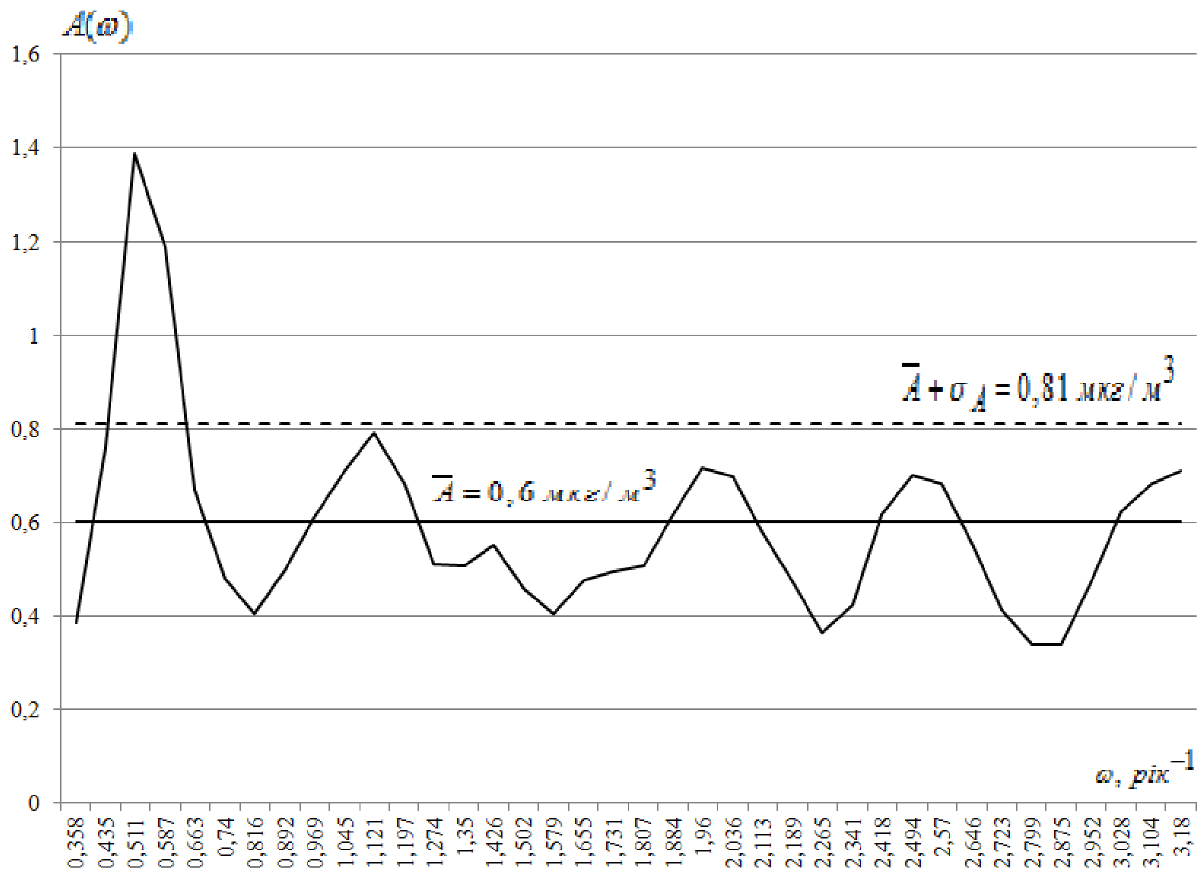


Рисунок 4.2 – Амплітудно-частотні характеристики часового ряду середньомісячної концентрації формальдегіду на КВП №10 м. Одеса (2008–2017 рр.)

Таким чином, часовий ряд має тренд, на який накладаються періодичні коливання, що обумовлені добовим, сезонним або річним ходом метеорологічних величин у пограничному шарі атмосфери. Характер трендів часових рядів свідчить про зміни антропогенного навантаження. Порівнювання детермінованих основ часових рядів концентрацій інгредієнту на окремих КВП дає змогу визначити райони міста, які є найбільш забрудненими, а також – коли це відбувається [6].

На рисунку 4.3 представлені вихідний та згладжений ряди середньомісячної концентрації формальдегіду на КВП №10 м. Одеса (2008–2017 рр.)

Як бачимо з рисунку, період дослідження можна розділити на два відрізки: перший – з січня 2008 по грудень 2011, другий – з січня 2012 по грудень 2017 року.

В перший період, на КВП №10 немає чіткого тренду зміни концентрацій формальдегіду.

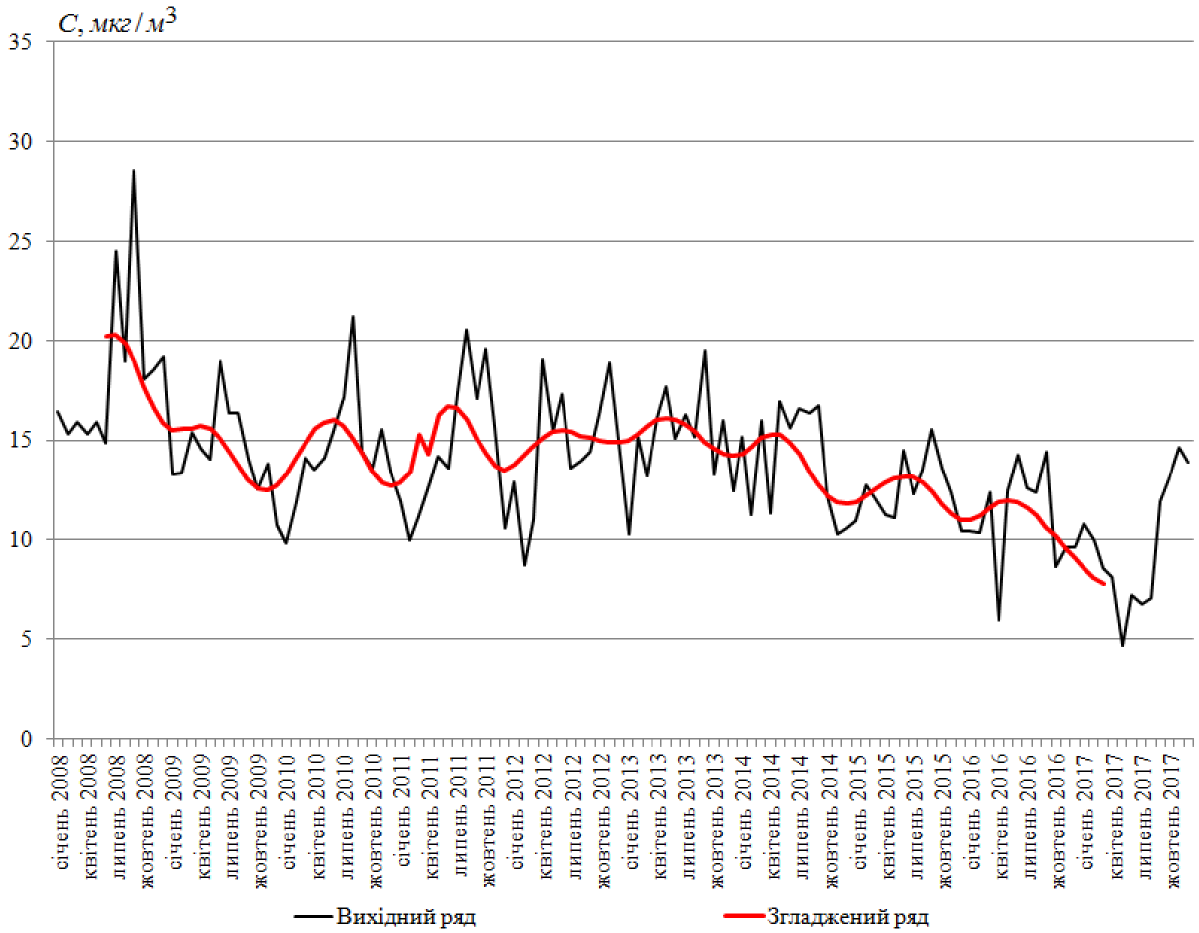


Рисунок 4.3 – Вихідний та згладжений ряди середньомісячної концентрації формальдегіду на КВП №10 м. Одеса (2008–2017 рр.)

Починаючи з січня 2012 року спостерігається тренд зменшення середньомісячних концентрацій формальдегіду, на тлі річних коливань вмісту домішки. При цьому, концентрації забруднюючої речовини стабільно вищі за ГДК$_{с.д.}$ (3 мкг/м^3).

4.3 Контрольні запитання

1. Чому процеси в атмосфері вважаються нестаціонарними часовими процесами?
2. Які методи використовуються при дослідженні періодичностей, що містяться в часових рядах?
3. Якими перевагами володіє інтегральне перетворення Фур’є для дослідження «прихованих» періодичностей?
4. З якою метою в перетворення Фур’є вводять множники («вікна»)?

5. В якому частотному інтервалі лежать мінімально і максимально можливі гармоніки, які можуть бути виявлені в нестаціонарному випадковому процесі за допомогою перетворення Фур'є?
6. Як по амплітудно-частотній характеристиці визначити з заданою ймовірністю статистично значущі періодичності у випадковому процесі?
7. Які характеристики «прихованих» гармонік у випадковому процесі можна отримати за допомогою перетворення Фур'є?
8. Дати визначення випадкової функції, випадкового процесу, випадкової послідовності.
9. Якими ймовірносними характеристиками описують випадкові функції?
10. Яка функція є стаціонарною випадковою функцією та якими ймовірносними характеристиками вона володіє?
11. Як розрахувати статистичні оцінки ймовірносних характеристик випадкової функції?
12. Що називають статистичною гіпотезою та в чому полягає основний принцип її перевірки?
13. Що називають рівнем значущості та довірчою ймовірністю?
14. Як розрахувати початкову фазу (в радіанах та в одиницях часу) періодичного коливання, що міститься у часовій послідовності?
15. Як отримати період гармонічного коливання, що міститься у випадковому процесі?
16. Як розрахувати амплітуду k-тої гармоніки, яка притаманна випадковому процесу?
17. За допомогою яких відомих методів можна отримати періодичні складові у випадковому процесі?

4.4 Завдання до практичної роботи

1 В якості вихідних використовуються ряди середньодобових значень вмісту домішок у атмосфері, що отримані у другій практичній роботі.

2 Виявити приховані періодичності, що містяться у вихідних рядах.

3 Визначити статистично значущі періодичності.

4 Визначити детерміновані компоненти часових рядів.

5 Побудувати згладжені графіки часових рядів концентрації домішки.

ПРАКТИЧНА РОБОТА № 5
«ФОРМУВАННЯ ЧАСОВИХ РЯДІВ СЕРЕДНЬОМІСЯЧНИХ КОНЦЕНТРАЦІЙ ДОМІШКИ. РОЗРАХУНОК І АНАЛІЗ СТАТИСТИЧНИХ ХАРАКТЕРИСТИК І МАТРИЦЬ КОРЕЛЯЦІЇ»

5.1 Загальні положення

У природі досить часто виникають ситуації, коли між деякими двома випадковими величинами проявляються статистичні зв'язки. Вони можуть бути функціональними або стохастичними.

Функціональною залежністю між двома випадковими величинами називається така залежність, коли можливому значенню однієї випадкової величини відповідає тільки одне значення іншої.

Стохастичним називають такий зв'язок між випадковими величинами, коли зміна однієї з них приводить до зміни закону розподілу іншої.

Якщо при змінюванні однієї з випадкових величин змінюється умовне математичне сподівання іншої, то такий зв'язок між цими випадковими величинами називається кореляційним [2, 4].

Кореляційну залежність можна трактувати як функціональну залежність умовного математичного сподівання однієї випадкової величини від значення іншої:

$$m_{y/x} = f(x). \tag{5.1}$$

Кореляційний зв'язок може бути прямим і оберненим. Якщо при збільшенні (зменшенні) однієї випадкової величини відбувається збільшення (зменшення) іншої, то такий зв'язок називають прямим. Коли збільшенню (зменшенню) однієї випадкової величини відповідає зменшення (збільшення) іншої, то це свідчить про обернений зв'язок між цими випадковими величинами.

Коефіцієнт кореляції розраховується на основі статистичних сукупностей за формулою:

$$r_{xy} = \frac{\sum_{i=1}^{n}(x_i - \overline{x})(y_i - \overline{y})}{n \cdot S_x \cdot S_y}. \tag{5.2}$$

Він може приймати значення із множини від -1 до +1.

Коефіцієнт кореляції характеризує степінь наближення кореляційного зв'язку між випадковими величинами X та Y до лінійної функціональної залежності.

Додатні значення коефіцієнта кореляції $\left(r_{xy} > 0\right)$ позначають наявність прямого лінійного зв'язку між випадковими величинами X і Y. Обернений лінійний кореляційний зв'язок характеризується від'ємними значеннями коефіцієнта кореляції $\left(r_{xy} < 0\right)$. Чим більшим за модулем є коефіцієнт кореляції $\left|r_{xy}\right|$, тим тіснішим буде лінійний кореляційний зв'язок між випадковими величинами, і навпаки. Якщо $r_{xy} = 0$, то це означає, що лінійний кореляційний зв'язок між величинами X і Y відсутній, але це не означає, що відсутнім є кореляційний зв'язок іншої форми.

Якісне уявлення про тісноту та форму кореляційного зв'язку між величинами X і Y можна отримати, побудувавши кореляційний графік на площині в координатах $(x; y)$.

5.2 Приклад розрахунку

Розрахунок коефіцієнту кореляції будемо проводити на основі рядів середньодобової концентрації двоокису сірки в атмосфері (мг/м^3) в Одесі, отриманих на двох КВП (табл. 5.1, 5.2).

Таблиця 5.1 – Середньодобова концентрація двоокису сірки в атмосфері, мг/м^3 (Одеса, КВП №15, 1.06 – 30.08.2001 р.)

i	1	2	3	4	5	6	7	8	9	10
0	0,031	0,058	0,067	0,057	0,077	0,077	0,078	0,077	0,084	0,076
10	0,072	0,066	0,06	0,053	0,063	0,068	0,058	0,052	0,067	0,059
20	0,051	0,06	0,048	0,052	0,061	0,038	0,039	0,039	0,04	0,058
30	0,033	0,03	0,024	0,029	0,034	0,041	0,038	0,032	0,037	0,037
40	0,036	0,036	0,041	0,043	0,043	0,042	0,026	0,03	0,035	0,039
50	0,042	0,048	0,072	0,048	0,047	0,045	0,039	0,04	0,049	0,041
60	0,053	0,057	0,06	0,049	0,053	0,038	0,053	0,041	0,045	0,048
70	0,053	0,055	0,046	0,07	0,074	0,075	0,076	0,069	0,054	0,061
80	0,048	0,06	0,059	0,058	0,059	0,06	0,052	0,043	0,046	0,04
90	0,033	0,051								

Щоб мати уявлення про тісноту та форму кореляційного зв'язку між величинами X і Y необхідно побудувати кореляційний графік на площині в координатах $(x; y)$ (рис 5.1).

Таблиця 5.2 – Середньодобова концентрація двоокису сірки в атмосфері, мг/м^3 (Одеса, КВП №17, 1.06 – 30.08.2001 р.)

i	1	2	3	4	5	6	7	8	9	10
0	0,054	0,053	0,052	0,05	0,058	0,056	0,058	0,057	0,057	0,057
10	0,058	0,061	0,06	0,057	0,059	0,06	0,062	0,058	0,057	0,06
20	0,057	0,057	0,057	0,053	0,054	0,047	0,054	0,051	0,059	0,047
30	0,059	0,056	0,052	0,056	0,057	0,058	0,06	0,056	0,056	0,054
40	0,057	0,057	0,056	0,056	0,06	0,064	0,057	0,055	0,058	0,06
50	0,062	0,063	0,06	0,064	0,064	0,06	0,059	0,058	0,062	0,061
60	0,057	0,062	0,058	0,058	0,058	0,069	0,051	0,056	0,07	0,062
70	0,058	0,054	0,06	0,056	0,064	0,058	0,054	0,054	0,055	0,06
80	0,057	0,057	0,058	0,058	0,059	0,061	0,062	0,058	0,062	0,058
90	0,061	0,064								

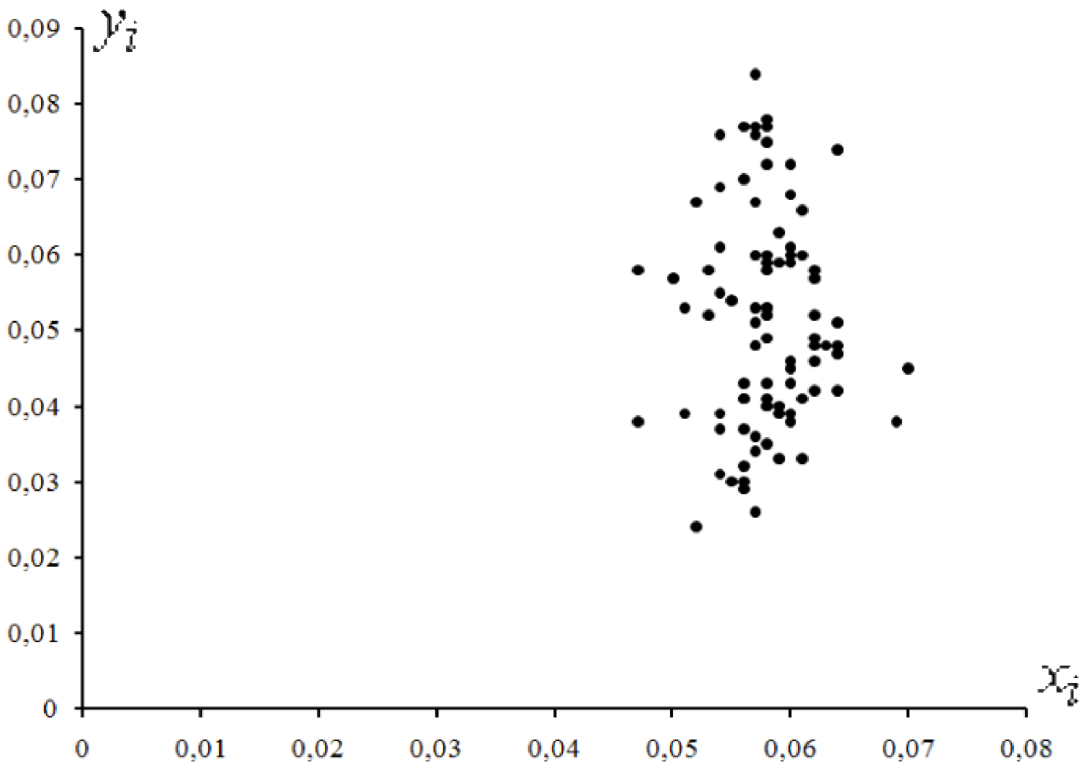


Рисунок 5.1 – Кореляційний графік середньодобової концентрації двоокису сірки в атмосфері (мг/м^3)

Розраховуємо статистичні оцінки моментів розподілу вихідних рядів:
$\overline{x} = 0{,}040$ мг/м^3, $S_x = 0{,}0083$ мг/м^3.
$\overline{y} = 0{,}055$ мг/м^3, $S_y = 0{,}0024$ мг/м^3.

Об'єм кожної вибірки складає 92 члени.
Використовуючи формулу (5.2) отримаємо:

$$r_{xy} = \frac{-0,0001}{0,183} = -0,0005$$

Отримане значення $r_{xy} = -0,0005$ свідчить про те, що між двома вибірками відсутня лінійна кореляційна залежність

5.3 Контрольні запитання

1. Які види зв’язків можуть спостерігатися між двома випадковими величинами?
2. Яка залежність між випадковими величинами називається функціональною? стохастичною?
3. Дайте визначення кореляційної залежності між двома випадковими величинами.
4. Що є якісною характеристикою тісноти та форми кореляційного зв’язку між двома випадковими величинами ?
5. Що є кількісною мірою лінійного кореляційного зв’язку між двома випадковими величинами ?
6. Статистичні оцінки яких моментів розподілу випадкових величин використовуються при розрахунках коефіцієнта кореляції?
7. У яких межах змінюється коефіцієнт кореляції?
8. У якому випадку лінійний кореляційний зв'язок буде тіснішим: за умови $r_{xy} = -0.82$ чи $r_{xy} = 0.82$?
9. Який буде зв'язок між двома випадковими величинами за умови $r_{xy} = 0$?
10. Який буде зв'язок між двома випадковими величинами за умови $|r_{xy}| = 1$?

5.4 Завдання до практичної роботи

1. Отримати у викладача варіант вихідних даних.
2. Побудувати кореляційний графік.
3. Розрахувати коефіцієнт кореляції. Порядок розрахунку статистичних оцінок, що необхідні для знаходження коефіцієнту кореляції занести до таблиці за формою (табл. 5.3)
4. Зробити висновок щодо тісноти кореляційної залежності між випадковими величинами.

Таблиця 5.3 – Розрахунки точкових статистичних оцінок моментів розподілу на основі рядів середньодобової концентрації діоксиду сірки в атмосфері, мг/м^3 (Одеса, КВП №15 – №17, 1.06 – 30.08.2001 р.)

№ п/п	x_i	y_i	$(x_i - \overline{x})$	$(y_i - \overline{y})$	$(x_i - \overline{x})^2$	$(y_i - \overline{y})^2$	$(x_i - \overline{x})(y_i - \overline{y})$

Для отримання заліку за практичну роботу №5 необхідно надати в письмовому вигляді правильні результати розрахунків коефіцієнта кореляції, таблиці за вказаною формою та аналіз отриманих результатів. Графічно представити кореляційний графік.

Максимальна кількість балів за виконання практичного завдання визначається робочою програмою.

ЛІТЕРАТУРА

1. Полєтаєва Л.М., Юрасов С.М. Моніторинг навколишнього середовища. – Одеса, ОДЕКУ, 2003. – 130 с.
2. Школьний Є.П., Лоєва І.Д., Гончарова Л.Д. Обробка та аналіз гідрометеорологічної інформації. – Одеса, 1999.
3. Школьний Є. П., Гончарова Л. Д., Миротворська Н. К. Методи обробки та аналізу гідрометеорологічної інформації (збірник задач і вправ): Навчальний посібник. – Одеса, 2000. – 420 с.
4. Гончарова Л. Д., Школьний Є. П. Методи обробки та аналізу гідрометеорологічної інформації (збірник задач і вправ): Навчальний посібник. – Одеса: Екологія, 2007. – 464 с.
5. Методическое указание «Регулирование выбросов при неблагоприятных метеорологических условиях». РД 52.04.52-85. - Л.: Гидрометеоиздат,1987.-550с.
6. Лоєва І.Д., Снісаренко В.В. Часові зміни концентрації діоксиду азоту в атмосферному повітрі м. Одеса. Науковий вісник Херсонського державного університету. Серія географічні науки. Херсон. 2017 р. Вип. 7. С 137 – 178.

ДОДАТКИ

Додаток А – Значення критерію Стьюдента для рівня значущості α і числа степенів вільності ν

ν	α							
	0.2	**0.1**	**0.05**	**0.02**	**0.01**	**0.005**	**0.002**	**0.001**
6	1.44	1.94	2.45	3.14	3.71	4.32	5.21	5.96
7	1.41	1.89	2.36	3.00	3.50	4.03	4.79	5.41
8	1.40	1.86	2.31	2.90	3.36	3.83	4.50	5.04
9	1.38	1.83	2.26	2.82	3.25	3.69	4.30	4.78
10	1.37	1.81	2.23	2.76	3.17	3.58	4.14	4.59
11	1.36	1.80	2.20	2.72	3.11	3.50	4.02	4.44
12	1.36	1.78	2.20	2.68	3.05	3.43	3.93	4.32
13	1.35	1.77	2.16	2.65	3.01	3.37	3.85	4.22
14	1.34	1.76	2.14	2.62	2.98	3.33	3.79	4.14
15	1.34	1.75	2.13	2.60	2.95	3.29	3.73	4.07
16	1.34	1.75	2.12	2.58	2.92	3.25	3.69	4.02
17	1.33	1.74	2.11	2.57	2.90	3.22	3.65	3.97
18	1.33	1.73	2.10	2.55	2.88	3.20	3.61	3.92
19	1.33	1.73	2.09	2.54	2.86	3.17	3.58	3.88
20	1.33	1.72	2.09	2.53	2.85	3.15	3.55	3.85
21	1.32	1.72	2.08	2.52	2.83	3.14	3.53	3.82
22	1.32	1.72	2.07	2.51	2.82	3.12	3.51	3.79
23	1.32	1.71	2.07	2.50	2.81	3.10	3.48	3.77
24	1.32	1.71	2.06	2.49	2.90	3.09	3.47	3.75
25	1.32	1.71	2.06	2.49	2.79	3.08	3.45	3.73
26	1.32	1.71	2.06	2.48	2.78	3.07	3.44	3.71
27	1.31	1.70	2.05	2.47	2.77	3.06	3.42	3.69
28	1.31	1.70	2.05	2.47	2.76	3.05	3.41	3.67
29	1.31	1.70	2.05	2.46	2.76	3.04	3.40	3.66
30	1.31	1.70	2.04	2.46	2.75	3.03	3.39	3.65
40	1.30	1.68	2.02	2.42	2.70	2.97	3.31	3.55
60	1.30	1.67	2.00	2.39	2.66	2.91	3.23	3.46
120	1.29	1.66	1.98	2.36	2.62	2.85	3.16	3.37
∞	1.28	1.64	1.96	2.33	2.58	2.81	3.09	3.29

МЕТОДИЧНІ ВКАЗІВКИ

до практичних занять

з дисципліни «АНАЛІЗ ЕКОЛОГІЧНОЇ ІНФОРМАЦІЇ»

Укладачі: Лоєва І.Д., Бургаз О.А.

Підписано до друку Формат 60 х 84/16 Папір офс.
Умовн. друк. арк.. Тираж Зам. №
Надруковано з готового оригінал – макета

Одеський державний екологічний університет,
65016, м. Одеса, вул. Львівська, 15