**МІНІСТЕРСТВО ОСВІТИ І НАУКИ УКРАЇНИ**
**ЗАПОРІЗЬКИЙ НАЦІОНАЛЬНИЙ УНІВЕРСИТЕТ**

**МАТЕМАТИЧНИЙ ФАКУЛЬТЕТ**
**Кафедра прикладної математики і механіки**

# КВАЛІФІКАЦІЙНА РОБОТА МАГІСТРА

на тему: **«АЛГОРИТМІЗАЦІЯ ПРОЦЕСУ ОБРОБКИ РЕЗУЛЬТАТІВ ЕМПІРИЧНОГО ДОСЛІДЖЕННЯ БІОЛОГІЧНИХ ПРОЦЕСІВ МЕТОДАМИ МАТЕМАТИЧНОЇ СТАТИСТИКИ»**

Виконав(ла): студент(ка) 2 курсу, групи 8.1139
спеціальності 113 прикладна математика
(шифр і назва спеціальності)
освітньої програми прикладна математика
(назва освітньої програми)
С. Л. Лось
(ініціали та прізвище)

Керівник доцент кафедри прикладної математики і механіки, доцент, к.ф.-м.н. Леонтьєва В. В.
(посада, вчене звання, науковий ступінь, прізвище та ініціали)

Рецензент Завідувач кафедри фундаментльної математики, доцент, д.т.н. Гребенюк С. М.
(посада, вчене звання, науковий ступінь, прізвище та ініціали)

Запоріжжя
2020

**МІНІСТЕРСТВО ОСВІТИ І НАУКИ УКРАЇНИ**

**ЗАПОРІЗЬКИЙ НАЦІОНАЛЬНИЙ УНІВЕРСИТЕТ**

Факультет математичний

Кафедра прикладної математики і механіки

Рівень вищої освіти магістр

Спеціальність 113 прикладна математика
(шифр і назва)

Освітня програма Прикладна математика

**ЗАТВЕРДЖУЮ**

Завідувач кафедри прикладної математики і механіки, д.т.н., професор

_____________ Грищак В.З.
(підпис)

« 01 » 12 2020 р.

# З А В Д А Н Н Я

## НА КВАЛІФІКАЦІЙНУ РОБОТУ СТУДЕНТОВІ(СТУДЕНТЦІ)

Лось Сергію Леонідовичу
(прізвище, ім'я та по-батькові)

1. Тема роботи (проекту) Алгоритмізація процесу обробки результатів емпіричного дослідження біологічних процесів методами математичної статистики

керівник роботи (проекту) Доцент кафедри прикладної математики і механіки, доцент, к.ф.-м.н. Леонтьєва В. В.
(прізвище, ім'я та по-батькові, науковий ступінь, вчене звання)

затверджені наказом ЗНУ від « 20 » 05 2020 року № 576-с

2. Строк подання студентом роботи 14.12.2020

3. Вихідні дані до роботи 1. Постановка задачі.
2. Перелік літератури.

4. Зміст розрахунково-пояснювальної записки (перелік питань, які потрібно розробити)

5. Перелік графічного матеріалу (з точним зазначенням обов'язкових креслень)
Презентація

6. Консультанти розділів роботи

| Розділ | Прізвище, ініціали та посада консультанта | Підпис, дата | |
|---|---|---|---|
| | | завдання видав | завдання прийняв |
| | | | |
| | | | |
| | | | |

7. Дата видачі завдання 20.05.2020

## КАЛЕНДАРНИЙ ПЛАН

| № | Назва етапів кваліфікаційної роботи | Строк виконання етапів роботи | Примітка |
|---|---|---|---|
| 1. | Розробка плану роботи. | 29.05.2020 | |
| | | | |
| 2. | Збір вихідних даних. | 20.07.2020 | |
| | | | |
| 3. | Обробка методичних та теоретичних | 22.08.2020 | |
| | джерел. | | |
| | | | |
| 4. | Розробка першого розділу. | 18.09.2020 | |
| | | | |
| 5. | Розробка другого розділу. | 21.10.2020 | |
| | | | |
| 6. | Оформлення та нормоконтроль | | |
| | кваліфікаційної роботи. | 20.11.2020 | |
| | | | |
| 7. | Захист кваліфікаційної роботи. | 17.12.2020 | |

Студент ______________ С.Л. Лось
(підпис) (ініціали та прізвище)

Керівник роботи ______________ В.В. Леонтьєва
(підпис) (ініціали та прізвище)

**Нормоконтроль пройдено**

Нормоконтролер ______________ В.В. Леонтьєва
(підпис) (ініціали та прізвище)

## РЕФЕРАТ

Кваліфікаційна робота магістра «Алгоритмізація процесу обробки результатів емпіричного дослідження біологічних процесів методами математичної статистики»: 58с., 13 рис., 9 табл., 26 джерел.

БІОМЕТРІЯ, БІОЛОГІЧНА СТАТИСТИКА, ВАРІАЦІЙНИЙ РЯД, ЕМПІРИЧНИЙ РОЗПОДІЛ, КОРЕЛЯЦІЯ, МЕДІАНА РОЗПОДІЛУ, МЕТЕОРОЛОГІЯ, СТАТИСТИКА, СТАТИСТИЧНІ МЕТОДИ, СТОХАСТИЧНІСТЬ.

Об’єкт дослідження – біологічні процеси.

Предмет дослідження: математичне моделювання та оптимізація біологічних процесів методами математичної статистики.

Метод дослідження – аналітичний.

Мета роботи: Дослідити в біологічних процесах живі організми в взаємодії з комплексом різних факторів зовнішнього середовища. Розглянути питання про відповідність нормального розподілу ймовірності. Перевірити відповідність отриманих даних нормальному закону розподілу. За даними температур знайти основні статистичні характеристики. Розрахувати імовірності характеристики погоди.

У роботі розглядаються основні поняття та характеристики математичної статистики, медико-біологічні дослідження та їх статистична обробка, проводиться збір даних та підрахунок основних статистичних характеристик, обробка отриманих результатів та прогнозування подальшої метеорологічної ситуації.

# SUMMARY


Master's qualifying paper «Algoritmization of the process of processing results of an empirical study of biological processes using the methods of mathematical statistics»: 59 pages, 13 figures, 9 tables, 26 references.

BIOMETRY, BIOLOGICAL STATISTICS, VARIATION RANGE, EMPIRICAL DISTRIBUTION, CORRELATION, MEDIAN OF DISTRIBUTION, METEOROLOGY, STATISTICS, STATISTICAL METHODS, STOCHASTICITY.

Object of the study – biological processes.

Method of research – analytical.

Aim of the study: To study living organisms in biological processes in interaction with a complex of various environmental factors. Consider the correspondence of the normal probability distribution. Check the compliance of the obtained data with the normal distribution law. According to the temperature find the main statistical characteristics.

The paper considers the basic concepts and characteristics of mathematical statistics, medical and biological research and their statistical processing, data collection and calculation of basic statistical characteristics, processing of the results and forecasting the future meteorological situation.

# ЗМІСТ

# ВСТУП

Статистика – один із розділів математики, який найчастіше використовується для емпіричного дослідження в будь-якій галузі науки [7,20,24]. Наприклад, в географії, особливо в кліматології та метеорології, статистика використовується під час спостережень у вигляді екстремальних та середніх значень. Оцінки параметрів спостереження у метеорології – випадкові величини.

Для визначення взаємозв'язків суспільно-економічних явищ і процесів, які обумовлюють тенденції розвитку економіки, провідну роль відіграють вибір статистичних методів для вивчення залежностей між показниками економічного зростання та матеріального добробуту населення, їх комплексне оцінювання та прогнозування. Економічна статистика розкриває конкретну сторону і особливості цих явищ і процесів в умовах визначеного місця (території) та часу.

Спеціалістам в області психології трапляється працювати з практичною інформацією, якій притаманна випадковість та стохастичність [5,16]. Тому, математичні методи у психології реалізовуються у вигляді математичної статистики.

У біологію математична статистика стала впроваджуватися завдяки розвитку теорії інформації та кібернетики. Завдяки впровадженню математики у дослідження, почала розвиватися нова галузь науки – біометрія [3,4], яка є більш широкою, ніж біологічна статистика [8]. Багатьом біологічним процесам також властиві статистичні закономірності при вивченні сукупностей.

Зв'язок між біологічними явищами та процесами і показниками, що їх характеризують, за специфікою впливу поділяється на функціональний та стохастичний (статистичний). При функціональному, або жорстко детермінованому, зв'язку кожному можливому значенню незалежної ознаки х відповідає одне або декілька точно визначених значень залежної ознаки. Характерною особливістю функціональних зв’язків є те, що в кожному окремому випадку відомий повний перелік факторів, які визначають значення залежної результативної ознаки, а також точний механізм їх впливу, який виражається певним рівнянням. Так, функціональні зв’язки [7,8,10] характеризуються повною відповідністю між причиною і наслідком факторів і результативних ознак.

Найважливішими прикладними задачами статистики є:

а) побудова невідомих емпіричних та теоретичних функцій розподілу за емпіричними вибірковими спостереженнями;

б) оцінка невідомих параметрів за результатами спостереження або за відомою кривою розподілу;

в) вибір та апроксимація зв’язків;

г) статистична перевірка різних гіпотез, пов’язаних з точністю оцінок параметрів, зв’язків, однорідністю вхідної інформації і т.д.

Мета роботи: Дослідити в біологічних процесах живі організми в взаємодії з комплексом різних факторів зовнішнього середовища. Розглянути питання про відповідність нормального розподілу ймовірності. Перевірити відповідність отриманих даних нормальному закону розподілу. За даними температур знайти основні статистичні характеристики. Розрахувати імовірності характеристики погоди.

В першому розділі дається повна характеристика підрозділів математичної статистики, розглядається статистичний аналіз і методи. Також дається повна характеристика статистичного аналізу медико-біологічних даних, розглядається питання про нормальний розподіл ймовірності.

В другому розділі за даними денної та нічної температур знаходяться всі основні статистичні характеристики, будуються функції розподілу, полігони та гістограми частот, розраховуються додаткові параметри розподілу температур на основі отриманих результатів та за їх допомогою робиться короткостроковий прогноз погоди на грудень 2020 у місті Запоріжжя

Об’єктом даного дослідження є біологічні процеси. Предметом дослідження виступає математичне моделювання та оптимізація біологічних процесів методами математичної статистики.

# 1 ТЕОРЕТИКО-МЕТОДОЛОГІЧНА БАЗА ДОСЛІДЖЕННЯ БІОЛОГІЧНИХ ПРОЦЕСІВ

## 1.1 Основні поняття та характеристики математичної статистики

Математична статистика (від лат. «status» - положення, стан явищ) — це підрозділ математики, який збирає, вивчає, зберігає, обробляє та інтерпретує дані суспільних явищ [7,24]. Також математична статистика будує математичні моделі явищ, у яких наявна характеристика «ймовірності», а отже – теорія ймовірностей є теоретичною базою статистики [16]. Двома основними підрозділами статистики є статистичні висновки та описова статистика (див. рис. 1.1).

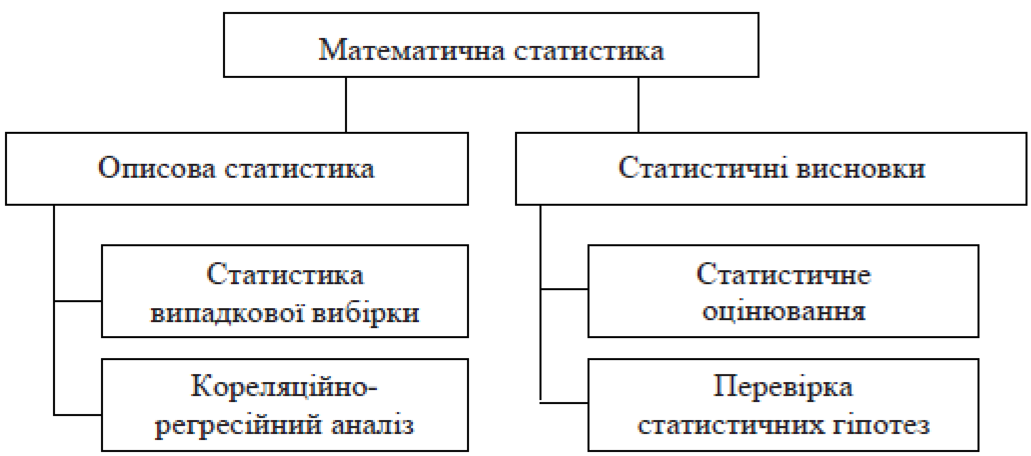



Рисунок 1.1 - Основні підрозділи математичної статистики

Завдання описової статистики [24]:

а) узагальнення показників випадкової вибірки;

б) кореляційно-регресійний аналіз, тобто встановлення взаємозв’язків між багатьма змінними.

Завдання статистичних висновків [20]:

а) оцінка ефективності, надійності та точності описової статистики;

б) статистичне оцінювання — виявлення похибок у процесі дослідження;

в) узагальнення параметрів генеральної сукупності або перевірка статистичних гіпотез.

Статистичне спостереження – це науково спланований та організований збір певних даних про різноманітні явища, процеси і величини. Якщо статистичне спостереження охоплює всі елементи певної множини, то воно називається суцільним, а якщо лише частину елементів – вибірковим.

Генеральна сукупність – це повні й кінцеві класи явищ, осіб, подій або ознак, які досліджуються. Об’ємом генеральної сукупності називають кількість її елементів.

Вибірка – це частина генеральної сукупності, яка сформована за певною, науково обґрунтованою ознакою [5]. Вибірка має повторювати властивості генеральної сукупності, тобто бути репрезентативною. Саме факт того, що вибірка є «підмножиною» генеральної сукупності, є найголовнішою причиною для опису та оцінки характеристик генеральної сукупності.

За способом формування вибірки виділяють випадкову та невипадкову вибірку. В основі випадкової вибірки лежить показник ймовірності певного елементу генеральної сукупності. За типом відбору елемента у вибірку існує вибірка без повернення об’єкта у генеральну сукупність та вибірка з поверненням (див. рис. 1.2).

Повне вивчення елементів генеральної сукупності в практичних дослідженнях недоцільне або неможливе, тому для дослідження генеральної сукупності застосовують вибірковий підхід.

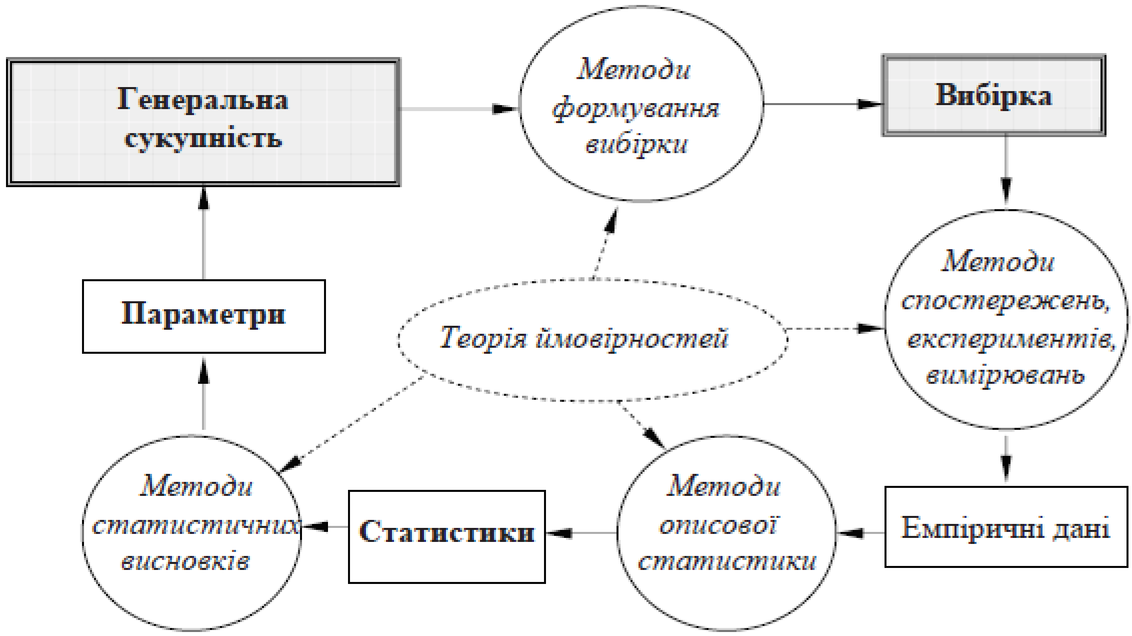



Рисунок 1.2 - Принцип вибіркового підходу

Алгоритм статистичних методів за вибірковим підходом [25]:

а) сформувати вибірку із генеральної сукупності на основі певної властивості(перехід до абстрактної математичної схеми);

б) отримати емпіричні дані для дослідження вибірки(проведення розрахункових дій);

в) обробити отримані дані за отриманими статистиками, тобто певними показниками вибірки(статистичні висновки щодо вибірки);

г) за допомогою статистичних висновків, отримати властиві до генеральної сукупності параметри (інтерпретація статистичних висновків для реального об’єкта).

Статистика як показник – це характеристика або система характеристик певної вибірки, наприклад, розподіл частот, вибіркові середні і т.д, які розкривають властивості вибірки (див. табл. 1.1).

Таблиця 1.1 – Класифікація статистичних показників

| Група розподілу | Приклади розподілів | Значення у дослідженні |
|---|---|---|
| Емпіричний розподіл | Варіаційний, атрибутивний, ранжирований | Наочно та швидко охарактеризувати структуру досліджуваної властивості |
| Вибірковий показник | Мода, медіана, розмах, ексцес тощо | Представляють чисельні значення типових властивостей вибірки |
| Кореляційно-регресійний показник | Коефіцієнт кореляції, коефіцієнт регресії | Дають можливість встановити приховані зв’язки та закономірності явищ, спрогнозувати розвиток досліджуваних процесів |

Варіаційний ряд – це емпірична послідовність кількісних показників (варіантів) досліджуваної ознаки (змінної) у порядку зростання чи спадання [10]. Впорядкування за величиною елементів вибірки називають ранжуванням. Побудова варіаційного ряду є підготовчим етапом для розрахунків та побудови статистичної моделі досліджуваного явища.

Випадкова величина – величина, яка приймає лише одне заздалегідь невідоме числове значення за випадкових причин в наслідок випробування. Позначають випадкові величини великими літерами: *A, B, X, Z* тощо, а можливі значення випадкової величини – $a_1$, $b_2$, $x_3$, $z_n$, тощо.

Дискретна випадкова величина – це величина, яка може приймати відокремлені один від одного числові значення з відповідними ймовірностями.

Неперервна випадкова величина – це величина яка може приймати необмежену кількість значень з будь-якого інтервалу *(a;b)*.

Статистичний розподіл – це математична модель реальних об’єктів у вигляді співвідношення значень змінної $X$ до частоти їх появи [25]. Отже, це емпіричний розподіл частот. За ознакою типів вимірювань, статистичні розподіли бувають варіаційні, атрибутивні та ранжировані.

Атрибутивні розподіли доцільно використовувати у разі категоріальних вимірювань властивостей певного об’єкта. Вони дають можливість оцінити властивості в абсолютних та відносних значеннях. Ранжировані розподіли використовують у разі рангових типів вимірювань

Закон розподілу випадкової величини – це співвідношення яке встановлює зв'язок між значеннями випадкової величини та ймовірностями, що їм відповідають. Закон розподілу можна задати таблично, графічно або аналітично.

За законом розподілу випадкової величини, статистичні розподіли бувають диференціальними та інтегральними, які можуть складатись з абсолютних та відносних частот.

Частотою варіанти $x_i$ статистичного розподілу називають кількість $n_i \geq 1$ разів, яку вона зустрічається у спостереженні. Відносною частотою варіанти $x_i$ називають відношення частоти варіанти до об’єму вибірки:

$$W_i = \frac{n_i}{n}, \tag{1.1}$$

при чому $\sum_{i=1}^{k} W_i = 1$.

Перелік варіант варіаційного ряду та відповідних їм частот чи відносних частот називають дискретним статистичним розподілом.

Аналітично закон дискретного розподілу випадкової величини можна задати за допомогою емпіричної функції розподілу вибірки [20,25]:

$$F^*(x) = W(X < x) = \frac{n_i}{n}, \qquad (1.2)$$

при чому $F^*(x) \in [0; 1]$ – неспадна функція.

Графічно закон дискретного розподілу випадкової величини можна задати за допомогою ламаної лінії. Вдовж осі абсцис відкладають варіанти варіаційного ряду, а вздовж осі ординат – частоти варіант або відносні частоти. Отримана фігура називається полігоном частот або полігоном відносних частот відповідно.

Закони розподілу дискретної випадкової величини надають повну характеристику величини, яка аналізується. Але на практиці не завжди вдається отримати закон розподілу, тому потрібно характеризувати дискретну випадкову величину за допомогою числових характеристик.

Міра центральної тенденції – це чисельні показники властивостей, які є типовими для емпіричних даних. Статистичні характеристики центру розподілу – це вибіркове середнє, мода та медіана.

Вибірковим середнім або математичним сподіванням дискретної випадкової величини називають величину, яка дорівнює сумі добутків усіх можливих варіант вибірки на відповідні їм частоти поділених на об’єм вибірки [20]:

$$x_{в} = \frac{1}{n}\sum_{i=1}^{k} x_i n_i. \qquad (1.3)$$

Модою $M_O$ розподілу називають варіанту, яка з’являється найчастіше серед емпіричних даних. Це не частота значення, а значення з найбільшою частотою. На графіку – це максимальна ордината полігону частот. Мод може бути декілька. Якщо розподіл має одну моду, то він називається одномодальним, якщо дві – бімодальним і т.д.

Медіаною $M_E$ розподілу називають варіанту, яка ділить варіаційний ряд навпіл за кількістю варіант. Якщо кількість варіант варіаційного ряду парна, то медіана знаходиться як півсума двох середніх варіант.

Мір центральної тенденції може бути недостатньо, адже вибірки можуть мати різні розподіли, але повністю ідентичні міри центральної тенденції. В такому випадку для дослідження використовують міри мінливості або показники розміру варіації. До них відносять: відхилення варіант, вибіркову дисперсію, середнє квадратичне відхилення, розмах варіації та вибірковий коефіцієнт варіації.

Різницю між варіантою вибірки і вибірковим середнім, помноженим на частоту цієї варіанти називають відхиленням варіанти: $(x_{\text{в}} - x_i)n_i$. Сума відхилень усіх варіант дорівнює 0.

Для того, щоб виміряти розсіювання варіант відносно вибіркового середнього використовується вибіркова дисперсія.

Вибірковою дисперсією варіаційного розподілу називають середнє арифметичне квадратів відхилень варіант щодо вибіркового середнього:

$$D_B = \frac{1}{n}\sum_{i=1}^{k} x_i^2 n_i - \overline{(x_B)}^2. \tag{1.4}$$

Варто зауважити, що $D(X) \geq 0$.

Середнім квадратичним відхиленням генеральної сукупності називають величину, квадратом якої є вибіркова дисперсія: $\sigma_B = \sqrt{D_B}$. Варто зауважити, що середнє квадратичне відхилення, на відміну від дисперсії, характеризує розсіювання варіант в тих самих одиницях, в яких вимірюється ознака $X$ [21].

Для більш грубої оцінки розсіювання варіант відносно математичного сподівання використовується розмах варіації. Це величина, яка є різницею між максимальною та мінімальною варіантами варіаційного ряду:

$$R = x_{max} - x_{min}. \quad (1.5)$$

Для порівняльного аналізу варіації середніх величин, які не дорівнюють нулю недостатньо визначити лише абсолютну величину варіації. Тому в таких випадках використовують відносний показник варіації як показник однорідності вибірки, який називається вибірковим коефіцієнтом варіації. Він обчислюється як відношення середнього квадратичного відхилення до математичного сподівання:

$$V_x = \frac{\sigma_B}{x_B} \cdot 100\%. \quad (1.6)$$

Генеральна сукупність є однорідною, якщо $V_x < 33\%$. Однорідність є передумовою застосування інших методів статистики.

Узагальненими числовими характеристиками для випадкових величин є центральні та початкові моменти *k*-го порядку. Початковим моментом *k*-го порядку випадкової величини *X* називають математичне сподівання від величини *k*-го порядку:

$$V_k = \sum_{i=1}^{n} x_i^k n_i. \quad (1.7)$$

Початковий момент першого порядку є вибірковим середнім. Центральним моментом k-го порядку випадкової величини *X* називають середнє арифметичне *k*-го степеня відхилення окремих значень ознаки від середнього:

$$\mu_k = \frac{1}{n} \sum_{i=1}^{n} (x_i - \overline{x}_B)^2 n_i. \quad (1.8)$$

Перший центральний момент дорівнює нулю. Центральним моментом другого порядку є вибіркова дисперсія. Початкові та центральні моменти більше ніж другого порядку дозволяють краще вплив розсіяних варіант на математичне сподівання.

В однорідних сукупностях розподіли є одномодальними. Полімодальність свідчить про те, що склад сукупності є неоднорідним, а також, що окремі складові сукупності є різнотиповими. В таких випадках дані перерозподіляються в більш однорідні групи. Серед одномодальних розподілів є симетричні, та скошені, гостро та плоско-модальні. При симетричному розподілі, значення ознаки, які рівновіддалені від центра мають однакову частоту, а при асиметричному розподілі – вершина зміщена. При зміщенні праворуч маємо лівосторонню асиметрію і навпаки. У симетричному розподілі міри центральної тенденції приблизно однакові. В разі лівосторонньої асиметрії: $\overline{x}_в < M_в < M_O$, а при правосторонній - $\overline{x}_в > M_в > M_O$.

Коефіцієнт асиметрії – це величина, яка характеризує напрям і міру скошеності в середині розподілу:

$$A = \frac{\overline{x}_B - M_O}{\sigma_B} = \frac{\mu_3}{\sigma_в^3}, \tag{1.9}$$

якщо $A < 0$, то маємо лівосторонню асиметрію, а при $A > 0$ – правосторонню.

Властивість зосередженості елементів генеральної сукупності навколо центра одномодальних розподілів є називають ексцесом:

$$E_k = \frac{\mu_4}{\sigma_в^4} - 3. \tag{1.10}$$

Він характеризує згладженість та опуклість розподілу вибірки. Якщо $E_k < 0$, то розподіл є відносно згладженим, а якщо $E_k > 0$, то – загостреним.

Своєрідним «стандартом» серед розподілів є нормальний розподіл, у якого коефіцієнти асиметрії та ексцесу дорівнюють нулю. Розподіл частот є нормальним, якщо коефіцієнти асиметрії та ексцесу менші в 3 чи більше раз за стандартні відхилення.

Якщо досліджувана ознака генеральної сукупності $X$ є неперервною, то варіант буде багато і їх доцільно об'єднувати в частинні інтервали. Інтервальним статистичним розподілом вибірки називають перелік, розміщений у порядку зростання, частинних інтервалів та відповідних їм частот чи відносних частот.

Інтервальний розподіл, як і дискретний, можна подати аналітично [23,25], за допомогою функції розподілу

$$F(x) = P(X < x) = P(-\infty < X < x), \qquad (1.11)$$

яка називається інтегральною функцією розподілу ймовірностей випадкової величини, чи графічно – за допомогою гістограми частот або гістограми відносних частот.

Диференціальний закон розподілу або щільність розподілу ймовірностей – це перша похідна від функції розподілу: $f(x) = F'(x)$. Крива розподілу або кумулята – це графік щільності розподілу.

Гістограмою частот називається ступінчата фігура з прямокутників, основою яких є часткові інтервали довжини $h_i$ та висотами, пропорційними частотам $n_i$ у разі рівних інтервалів чи щільністю частоти $\frac{n_i}{h_i}$ при нерівних інтервалах. Гістограмою відносних частот називається ступінчата фігура з прямокутників, основою яких є часткові інтервали довжини $h_i$ та висотами, пропорційними відносним частотам $\omega_i$ у разі рівних інтервалів чи щільністю відносної частоти $\frac{\omega_i}{h_i}$ при нерівних інтервалах.

Математичним сподіванням неперервної випадкової величини $X$ називають збіжний інтеграл:

$$M(X) = \int_{-\infty}^{+\infty} xf(x)dx. \tag{1.12}$$

У разі якщо значення $X$ належать скінченному проміжку [a;b], то:

$$M(X) = \int_{a}^{b} xf(x)dx. \tag{1.13}$$

Дисперсія неперервної випадкової величини $X$ знаходиться за формулою:

$$D(X) = \int_{-\infty}^{+\infty} x^2 f(x)dx - M^2(X). \tag{1.14}$$

Аналогічно, для скінченного проміжку:

$$D(X) = \int_{a}^{b} x^2 f dx - M^2(X). \tag{1.15}$$

Середньоквадратичне відхилення, як і для дискретної випадкової величини обчислюється за формулою:

$$\sigma_B = \sqrt{D_B}. \tag{1.16}$$

Медіана неперервної випадкової величини $X$ знаходиться за формулою:

$$M_{\text{в}} = x_{i-1} + \frac{0{,}5 - F(x_{i-1})}{F(x_i) - F(x_{i-1})} h, \tag{1.17}$$

де $h = x_i - x_{i-1}$ – крок до часткового інтервалу.

Модальним інтервалом неперервної випадкової величини $X$ називають інтервал з найбільшою частотою появи досліджуваної ознаки. Модою неперервної випадкової величини $X$ називають величину, яка обчислюється за формулою:

$$M_O = x_{i-1} + \frac{n_{M_O} - n_{M_{O-1}}}{2n_{M_O} - n_{M_{O-1}} - n_{M_{O+1}}} h, \qquad (1.18)$$

де: $x_{i-1}$ – початок модального інтервалу;

$n_{M_{O-1}}$ – частота домодального інтервалу;

$n_{M_O}$– частота інтервалу, який містить моду;

$n_{M_{O+1}}$– частота післямодального інтервалу;

$h$ – крок часткового інтервалу.

Незміщеною оцінкою вибіркової дисперсії або виправленою дисперсією називають величину:

$$S^2 = \frac{n}{n-1} D_B. \qquad (1.19)$$

Таку формулу використовують при $n < 30$.

Точковою оцінкою параметрів розподілу генеральної сукупності називають оцінку, що визначається одним числом [2]. Наприклад, точковою оцінкою математичного сподівання є вибіркове середнє $x_B$. Інтервальною оцінкою називають оцінку, що визначається кінцями інтервалу.

Довірчий інтервал для оцінки математичного сподівання $a$ при нормальному розподілі та відомого середньоквадратичного відхилення з ймовірністю $\gamma$ буде мати вигляд:

$$a \in \left(\overline{x}_B - \frac{t\sigma}{\sqrt{n}};\ \overline{x}_B + \frac{t\sigma}{\sqrt{n}}\right), \qquad (1.20)$$

де значення $t$ за заданого $\gamma$ береться із таблиці значень функції Лапласа. Якщо середньоквадратичне відхилення невідоме, то за даними можна побудувати випадкову величину з розподілом Стьюдента. Тоді довірчий інтервал матиме вигляд:

$$a \in \left(\overline{x}_B - \frac{t_\gamma \sigma}{\sqrt{n}};\ \overline{x}_B + \frac{t_\gamma \sigma}{\sqrt{n}}\right), \qquad (1.21)$$

де $t_\gamma(\gamma, n)$ знаходять за таблицею розподілу Стьюдента.

Довірчий інтервал для оцінки середньоквадратичного відхилення з надійністю $\gamma$ матиме вигляд:

$$\begin{cases} \sigma \in \big(s(1-q); s(1+q)\big), q < 1 \\ \sigma \in \big(0; s(1+q)\big), q > 1 \end{cases}, \qquad (1.22)$$

де $q$ знаходять за таблицею розподілу $\chi^2$.

Отже, в залежності від інформації про випадкову величину, для знаходження довірчого інтервалу використовуються різні розподіли. Але, описова статистика не лише систематизує емпіричні дані за допомогою описаних вище показників, а й виявляє зв'язок між змінними – функціональний, кореляційний чи недетермінований.

Кореляцією [1,17] називають статистичну залежність ймовірного характеру між випадковими величинами. Кореляція називається позитивною, якщо при підвищенні рівня однієї змінної підвищується і рівень іншої. У разі, коли підвищення рівня однієї змінної супроводжується зниженням рівня іншої, кореляція називається негативною. Якщо зв'язок між змінними відсутній, то маємо нульову кореляцію, яка свідчить про відсутність лінійних зв’язків.

Коефіцієнт кореляції може приймати значення з проміжку [-1;1]. Якщо $r > 0$ - маємо прямий зв'язок між змінними, а при $r < 0$ - обернений зв'язок. Більш того, якщо $r$ близький до нуля, то зв'язок відсутній, $r < 0{,}25$ - слабкий зв'язок, при $0{,}25 \leq r \leq 0{,}75$, маємо помірний тип зв’язку, а $r \geq 0{,}75$ - сильний [20]. Вибір методу кореляційного аналізу залежить від типів шкали, в яких виміряні змінні (див. рис. 1.3).

| Шкали ознаки $Y$ | Шкали ознаки $X$ | | |
|---|---|---|---|
| | Інтервальна (відношень) | Рангова | Номінальна |
| Інтервальна (відношень) | Коефіцієнт Пірсона $r_{xy}$; Дихотомічний коефіцієнт кореляції $\varphi$; Тетрахоричний коефіцієнт кореляції $r_{tet}$ | | |
| Рангова | Коефіцієнт Спірмена $r_s$ (при умові, якщо для $x$ шкалу інтервалів або відношень перетворити в рангову шкалу) | Коефіцієнти кореляції Спірмена $r_s$; $\tau$ Кендалла; Коефіцієнт конкордації $W$ | |
| Номінальна | Точково-бісеріальний коефіцієнт кореляції $r_{pb}$; Бісеріальний коефіцієнт кореляції $r_{bis}$ | Рангово-бісеріальний коефіцієнт кореляції $r_b$ | Коефіцієнт асоціації $\Phi$; Коефіцієнт контингенції Юла $Q$; Коефіцієнти спряженості Чупрова $K$ і Пірсона $C$ |

Рисунок 1.3 – Кореляційні методи в залежності від шкали

Також вибір методу кореляційного аналізу [17] залежить від нормальності розподілу даних (див. рис 1.4).

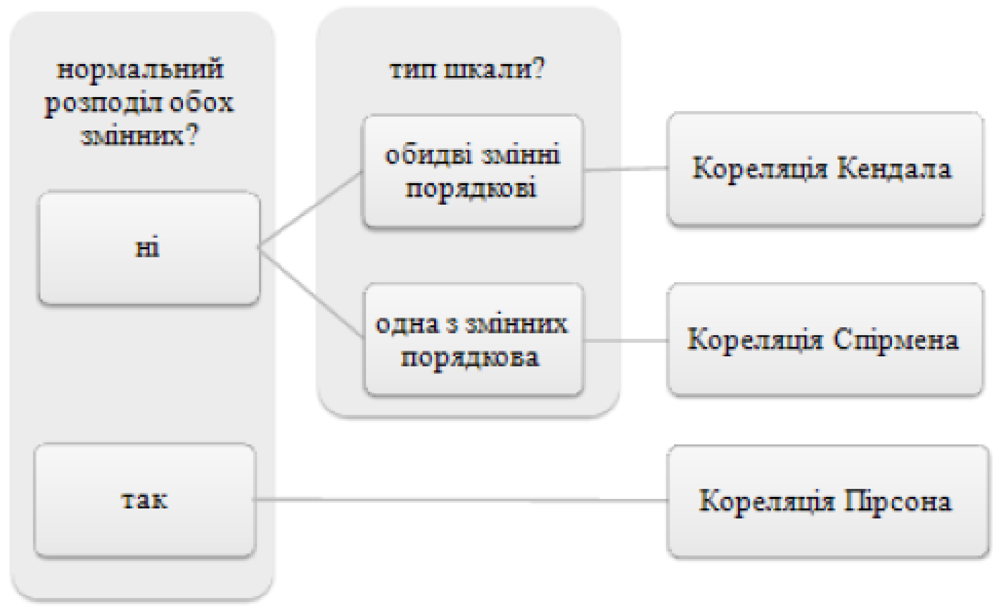



Рисунок 1.4 – Кореляція в залежності від розподілу

Випадкові величини $X$ та $Y$, для яких коефіцієнт кореляції є ненульовим називають корельованими, в іншому – некорельованими.

Коефіцієнт кореляції Пірсона оцінює лінійний зв'язок для емпіричних даних, виміряних за шкалою відношень чи інтервалів:

$$r_{xy} = \frac{\sum_{i=1}^{n}(x_i - \overline{x}_B)(y_i - \overline{y}_B)}{\sqrt{\sum_{i=1}^{n}(x_i - \overline{x}_B)^2 \cdot (y_i - \overline{y}_B)^2}} = \frac{cov(x,y)}{\sqrt{S_x^2 S_y^2}}, \tag{1.23}$$

де $cov(x, y) = M(XY) - M(X)M(Y)$.

Якщо $r_{xy} = \pm 1$, то випадкові величини $X$ та $Y$ пов'язані між собою залежністю $Y = aX + b$.

Отже, кореляційний аналіз встановлює зв'язок між величинами, а також оцінює інтенсивність та напрямок цього зв'язку.

Також, для встановлення статистичних зв'язків між величинами застосовують регресійний аналіз. На відміну від кореляційного аналізу, регресійний аналіз встановлює форму залежності змінних. Регресія [20] визначає структуру та параметри об'єктів дослідження, а отже будує математичну модель. Також вона дозволяє знаходити очікувані значення однієї ознаки за величиною іншої за наявності кореляційних зв'язків.

Пряма $Y = aX + b$ називається прямою регресії *Y* на *X*, якщо її коефіцієнти підібрані так, що $\sigma(ax + b) \to min$. Коефіцієнти прямої визначаються за формулами: $a = \frac{\sigma_y}{\sigma_x} r_{xy}$ та $b = M(Y) - aM(X)$. Коефіцієнт $a$ називається коефіцієнтом регресії *Y* на *X*. Слід зауважити, що одне з найважливіших завдань регресійного аналізу полягає у знаходженні прямої регресії.

Отже, найважливішими підрозділами математичної статистики є описова статистика та статистичні висновки [7,24]. Статистика ґрунтується на базі теорії ймовірності та застосовується як в природничому, так і у гуманітарному циклі наук. Основною базою для використання статистики у емпіричних дослідженнях є генеральна сукупність. Шляхом виділення репрезентативної вибірки як частини генеральної сукупності, за певною, науково обґрунтованою, ознакою науковці проводять подальші дослідження та перехід до абстрактної моделі. Першим математичним кроком у дослідженні є побудова варіаційного ряду, а також підрахунок мір центральної тенденції – вибіркового середнього, моди та медіани й мір мінливості – вибіркової дисперсії, середньоквадратичного відхилення, коефіцієнту варіації, асиметрії та ексцесу. На основі мір мінливості та мір центральної тенденції, можна продовжити статистичний аналіз застосувавши інші статистичні методи – побудова довірчого інтервалу, перевірка гіпотез, кореляційний та регресійний аналіз тощо. Останнім кроком є математичні висновки та їх інтерпретація щодо реальних об'єктів чи процесів дійсності.

## 1.2 Медико-біологічні дослідження та їх статистична обробка

Об'єктом досліджень в медико-біологічних науках є живі організми [8,22], як правило - в зрізі їх взаємодії з комплексом різних факторів зовнішнього середовища, частина з яких відноситься до хвороботворним. Об'єкт дослідження з точки зору математики - це досліджуваний феномен з усіма відносяться до справи зовнішніми компонентами, включаючи простір (поширення) і час (динаміка). Інші фактори спрямовані на профілактику або лікування хвороби. Статистичному аналізу може бути піддана не тільки ефективність нового методу лікування, а й ефективність роботи самого лікаря. Основним підходом у цьому напрямку як і раніше залишається експеримент, а основною його метою - встановлення закономірностей виникнення різних хвороб, механізмів їх розвитку, розробка і перевірка ефективності нових методів профілактики і лікування.

Медико-біологічне дослідження включає п'ять самостійних, але взаємопов'язаних етапів [18]:

а) планування дослідження (вибір об'єкта дослідження, формулювання мети, розробка завдань, програми і плану дослідження);

б) статистичне спостереження (збір матеріалу для його подальшої статистичної обробки);

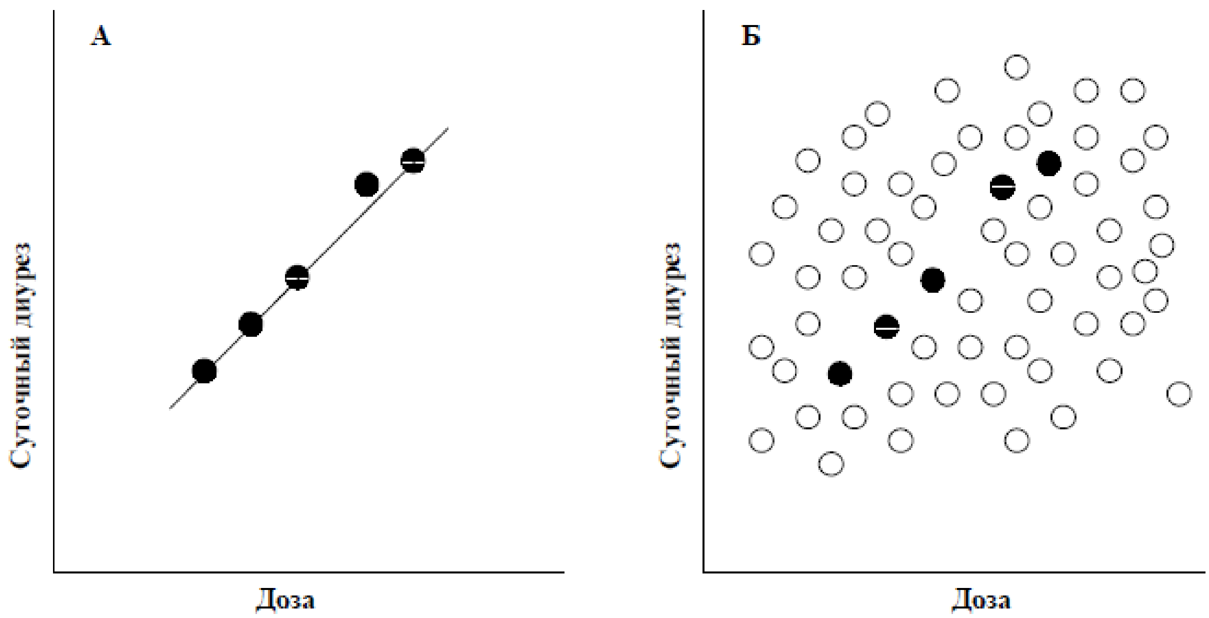
в) статистична угруповання і зведення матеріалів спостереження;

г) первинна статистична обробка даних;

д) науково-статистичний аналіз, графічне і літературне оформлення результатів дослідження.

Наведемо один із прикладів важливості застосування математичної статистики у медицині [6,15].

**Приклад 1.1.** Винахідники препарату припускають, що він збільшує діурез пропорційно прийнятій дозі. Для перевірки цього припущення вони призначають п'яти добровольцям різні дози препарату. За результатами спостережень будують графік залежності діурезу від дози (див. рис. 1.5, а)). Залежність видно неозброєним оком. Дослідники вітають один одного з відкриттям, а світ - з новим діуретиком.




а) б)

Рисунок 1.5 – Статистика та клінічна практика:
а) залежність діурезу від дози п’яти добровольців; б) залежність діурезу від дози п’яти і більше добровольців

Насправді ж, дані дозволяють достовірно стверджувати лише те, що існує залежність (лінійна регресія) діурезу від дози спостерігалася у цих п'яти добровольців. Але, те, що ця залежність проявиться у всіх людей, які будуть приймати препарат, - не більше ніж припущення (гіпотеза). На рис. 1.5, б) могли б побачити ми, якби могли досліджувати зв'язок дози і діурезу у всіх людей: залежності немає. П'ятеро людей, що увійшли в первинне дослідження, позначені чорним. В даному випадку уявна залежність породжена випадковістю.

За допомогою статистичних методів можна оцінити ймовірність подібної помилки. В результаті застосування статистичного методу ми отримуємо не істину в останній інстанції, а всього лише оцінку ймовірності того чи іншого припущення. Розглянутий приклад підтверджує наступний висновок: будь-яке загальне лише приблизно охоплює всі окремі предмети. Будь-яке окреме неповно входить у загальне.

Характерною властивістю ознак біологічних і екологічних об'єктів є варіювання їх значень в певних межах при переході від однієї одиниці спостереження до іншої, що відповідає термінам варіація, дисперсія, варіабельність, розсіювання варіант, розкид, мінливість [3,22].

В даний час важко назвати сферу людської діяльності, яка не залежала б від погодних умов. Особливе значення метеорологічні умови мають для авіації. Висота хмар і дальність видимості в пункті зльоту і посадки, конвективні явища, турбулентність, швидкість і напрям вітру на маршруті польоту - ось далеко не повний перелік метеорологічних величин, що впливають на безпеку і економічну ефективність польотів. До прийняття рішення інформація, одержувана в результаті спостереження, зазнає, як правило, істотні перетворення. Найбільш характерними етапами таких перетворень в більшості сфер людської діяльності є вимірювання і обробка результатів вимірювань. Вимірювання полягає в порівнянні спостерігається величини (властивості об'єкта) з еталоном і отриманні в результаті цього її чисельного значення. При візуальному вимірі дальності видимості спостерігач визначає найвіддаленіший видимий об'єкт і, користуючись схемою видимості орієнтирів, визначає видимість. При приладових вимірах реєструється значення вимірюваного параметра.

Математичною моделлю обробки інформації називається сукупність математичних залежностей, що залучаються для опису досліджуваного процесу або явища. За допомогою моделі можна встановити зв'язки між вимірюваними величинами, між що спостерігаються і прогнозованими характеристиками стану атмосфери і т.д.

Математичні моделі ґрунтуються на використанні двох видів зв'язків між величинами *X* і *Y*: функціональної і статистичної [8]. Функціональний зв'язок встановлює однозначна відповідність: кожному значенню *x* відповідає одне і тільки одне значення *y*. Статистична зв'язок встановлює неоднозначну відповідність між величинами: кожному значенню *x* ставиться у відповідність один і тільки один розподіл повторюваності значення *y*.

У метеорології, особливо в синоптиків, особлива роль відводиться побудові різних прогностичних зв'язків на основі регресійного аналізу та теорії розпізнання образів. За умовами використання статистична обробка підрозділяється на оперативну і неоперативну [26].

Оперативна обробка проводиться безпосередньо в міру надходження інформації та здійснюється з метою прийняття найбільш важливих на даний період рішень. До таких рішень відносяться: видача інформації про погоду в пункті посадки, розробка прогнозів погоди на польоти по існуючим алгоритмам і т.д.

Неоперативна обробка призначена для використання інформації протягом тривалих проміжків часу. До такого виду обробки відносяться: розробка авіаційно-кліматичних описів і довідок, побудова різних прогностичних алгоритмів і оцінка їх ефективності і т.д.

Природно, що оперативна і неоперативна обробка вимагають різних обсягів інформації. Для розробки прогностичного алгоритму необхідно всебічно проаналізувати досліджуваний процес, вибрати фактори, які найбільш на нього впливають, побудувати прогностичні зв'язки та оцінити їх і т.д. Тому тут використовується вся інформація, що отримується в процесі спостереження, а сама обробка називається повною. При розробці прогнозів по наявним алгоритмам потрібно вибрати найбільш підходящий для даної ситуації спосіб і вже для цього способу взяти необхідну інформацію. У цьому випадку використовується тільки частина інформації про стан об'єктів, і така обробка називається неповною.

**Приклад 1.2** Лікарям відомо, що найбільш сильний вплив на бронхіально-легеневі захворювання людей надає чадний газ - оксид вуглецю. Поставивши за мету визначити цю залежність, фахівці з медичної статистики проводять збір даних. Вони збирають відомості з різних міст про середню концентрації чадного газу в атмосфері і про захворюваність астмою (число хронічних хворих на 1000 жителів). Отримані дані можна звести в таблицю, а також представити у вигляді точкової діаграми (див. табл. 1.2 та рис. 1.6) [6,15].

Таблиця 1.2 – Статистичні дані дослідження

| С (мг/м$^3$) | Р (бол./тис) |
|---|---|
| 2 | 19 |
| 2,5 | 20 |
| 2,9 | 32 |
| 3,2 | 34 |
| 3,6 | 51 |
| 3,9 | 55 |
| 4,2 | 90 |
| 4,6 | 108 |
| 5 | 171 |

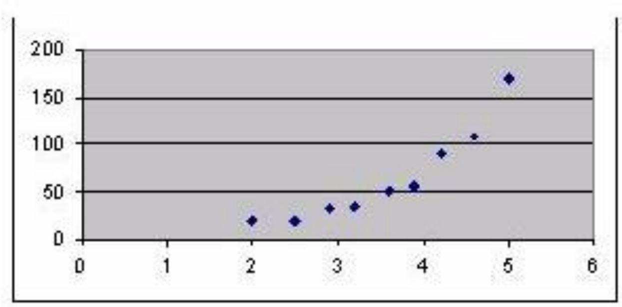



Рисунок 1.6 – Статистичні дані дослідження

Проаналізувавши результати, можна зробити висновок, що при концентрації чадного газу до 3 мг / м3 його вплив на захворюваність астмою несильно.  З подальшим зростанням концентрації настає різке зростання захворюваності.

У клінічних дослідженнях, коли вивчається ряд тих чи інших показників у хворих, особливо при невеликому обсязі вибірці (менше 30), найчастіше немає підстав вважати розподіл нормальним, тому статистичний аналіз проводиться за допомогою непараметричних методів, які не потребують нормальності розподілу (див. табл. 1.3).

Таблиця 1.3 - Статистичні критерії у медицині

| Порівнювальні показники | Параметричні тести | Непараметричні тести | Методи порівняння часткою |
|---|---|---|---|
| Дві незалежні групи | Критерій Стьюдента (двохвибірковий) | Критерій Манна-Уітні, критерій Уілкоксона, медіанний критерій | Критерій $\chi^2$, точний критерій Фішера |
| Більше двох незалежних груп | Дисперсійний аналіз, критерій Стьюдента для множинних порівнянь, критерій Даннета | Критерій Краксела-Уоллиса, медіанний критерій | Критерій $\chi^2$ |
| Одна група, зв’язані виміри | Критерій Стьюдента для зв’язаних пар (парний тест), дисперсійний аналіз | Т-критерій Уілкоксона, критерій знаків | Критерій Мак-Німара |
| Одна група, декілька зв’язаних вимірів | Дисперсійний аналіз повторних вимірів, критерій Шеффе для залежних вибірок | Критерій Фрідмана | Критерій Кокрена |

*Продовження табл. 1.3*

| Порівнювальні показники | Параметричні тести | Непараметричні тести | Методи порівняння часткою |
|---|---|---|---|
| Два і більше спряжених рядів чисел | Коефіцієнт кореляції Пірсона | Коефіцієнт рангової кореляції Спірмена | |

Отже, статистичний аналіз медико-біологічних даних зазвичай проходить кілька етапів [7,22,25]. Першим з них є проведення заходів щодо описової статистикою, тобто розрахунок кількісних статистичних показників. Тут же розглядається питання про відповідність розподілу нормальному розподілу ймовірності. Перевірити відповідність отриманих даних нормальному закону розподілу можна двома способами. Перший з них - за допомогою побудови графіків на нормальній ймовірнісної папері, другий - за допомогою критеріїв згоди. Другим етапом є статистичне оцінювання, тобто проводиться оцінка параметрів генеральної сукупності за параметрами вибіркової сукупності. На цьому етапі важливо знання нормальності аналізованої статистичної сукупності. Наступним етапом є перевірка статистичних гіпотез. У цих випадках результати вибірки використовуються для перевірки припущень (гіпотез) щодо тих чи інших властивостей розподілу генеральної сукупності або порівнянні параметрів різних вибірок.

# 2 ВИКОРИСТАННЯ МАТЕМАТИЧНОЇ СТАТИСТИКИ ДЛЯ МЕТЕРЕОЛОГІЧНИХ ПРОГНОЗІВ

## 2.1 Збір даних та підрахунок основних статистичних характеристик

Для прогнозування погоди в Запоріжжі методами математичної статистики, наведемо метеорологічні дані за листопад 2020. Спочатку наведемо показники денної температури (див. табл. 2.1):

Таблиця 2.1 - Запорізька погода вдень за листопад

| Дата листопада | Погода ℃ | Швидкість вітру, м/с |
|---|---|---|
| 1 | 9 | 5 |
| 2 | 9 | 4 |
| 3 | 10 | 4 |
| 4 | 11 | 5 |
| 5 | 10 | 4 |
| 6 | 10 | 5 |
| 7 | 10 | 5 |
| 8 | 10 | 5 |
| 9 | 10 | 5 |
| 10 | 10 | 5 |
| 11 | 10 | 5 |
| 12 | 9 | 4 |
| 13 | 9 | 5 |
| 14 | 7 | 5 |
| 15 | 5 | 4 |
| 16 | 5 | 4 |
| 17 | 5 | 4 |
| 18 | 6 | 4 |
| 19 | 6 | 5 |
| 20 | 6 | 6 |
| 21 | 7 | 6 |
| 22 | 6 | 6 |
| 23 | 4 | 5 |

*Продовження табл. 2.1*

| Дата листопада | Погода ℃ | Швидкість вітру, м/с |
|---|---|---|
| 24 | 3 | 4 |
| 25 | 3 | 4 |
| 26 | 4 | 5 |
| 27 | 4 | 5 |
| 28 | 6 | 5 |
| 29 | 5 | 5 |
| 30 | 5 | 6 |

Тепер наведемо таблицю даних за вечірню погоду (див. табл. 2.2):

Таблиця 2.2 - Запорізька погода вночі за листопад

| Дата | Погода, ℃ |
|---|---|
| 01.11 | 4 |
| 02.11 | 4 |
| 03.11 | 6 |
| 04.11 | 7 |
| 05.11 | 6 |
| 06.11 | 6 |
| 07.11 | 6 |
| 08.11 | 6 |
| 09.11 | 6 |
| 10.11 | 7 |
| 11.11 | 6 |
| 12.11 | 6 |
| 13.11 | 6 |
| 14.11 | 6 |
| 15.11 | 4 |
| 16.11 | 3 |

*Продовження табл. 2.2*

| Дата | Погода, ℃ |
|---|---|
| 17.11 | 2 |
| 18.11 | 3 |
| 19.11 | 3 |
| 20.11 | 3 |
| 21.11 | 4 |
| 22.11 | 4 |
| 23.11 | 2 |
| 24.11 | 0 |
| 25.11 | 1 |
| 26.11 | 2 |
| 27.11 | 1 |
| 28.11 | 3 |
| 29.11 | 3 |
| 30.11 | 3 |

На основі отриманих даних побудуємо варіаційні ряди. В першому рядку позначимо всі варіанти температур за місяць, у другому рядку – кількість повторень такої температури, у третьому – відносну частоту температур, а у четвертому – накопичену частоту (див. табл. 2.3; 2.4).

Таблиця 2.3 – Дискретний варіаційний ряд денної температури в Запоріжжі за листопад

| $x_i$, ℃ | 3 | 4 | 5 | 6 | 7 | 9 | 10 | 11 |
|---|---|---|---|---|---|---|---|---|
| $n_i$ | 2 | 3 | 5 | 5 | 2 | 4 | 8 | 1 |
| $w_i$ | 1/15 | 1/10 | 1/6 | 1/6 | 1/15 | 2/15 | 4/15 | 1/30 |

*Продовження табл. 2.3*

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| Частота | 2 | 5 | 10 | 15 | 17 | 21 | 29 | 30 |

Таблиця 2.4 – Дискретний варіаційний ряд нічної температури в Запоріжжі за листопад

| $x_i$, ℃ | 0 | 1 | 2 | 3 | 4 | 6 | 7 |
|---|---|---|---|---|---|---|---|
| $n_i$ | 1 | 2 | 3 | 7 | 5 | 10 | 2 |
| $w_i$ | 1/30 | 1/15 | 1/10 | 7/30 | 1/6 | 1/3 | 1/15 |
| Накопичена частота | 1 | 3 | 6 | 13 | 18 | 28 | 30 |

На основі дискретних варіаційних рядів побудуємо полігони частот (див. рис. 2.1; 2.2):

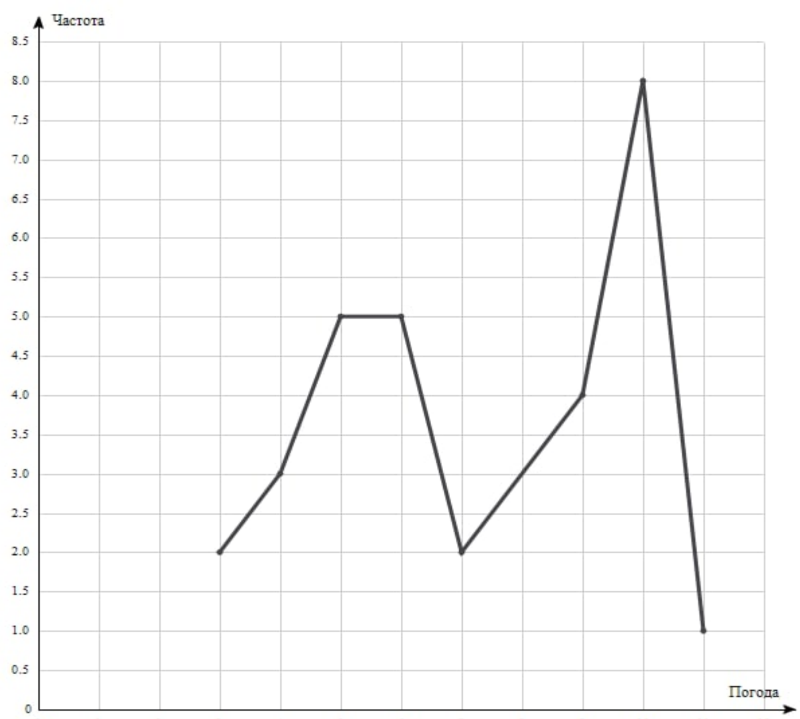



Рисунок 2.1 – Графік денної температури

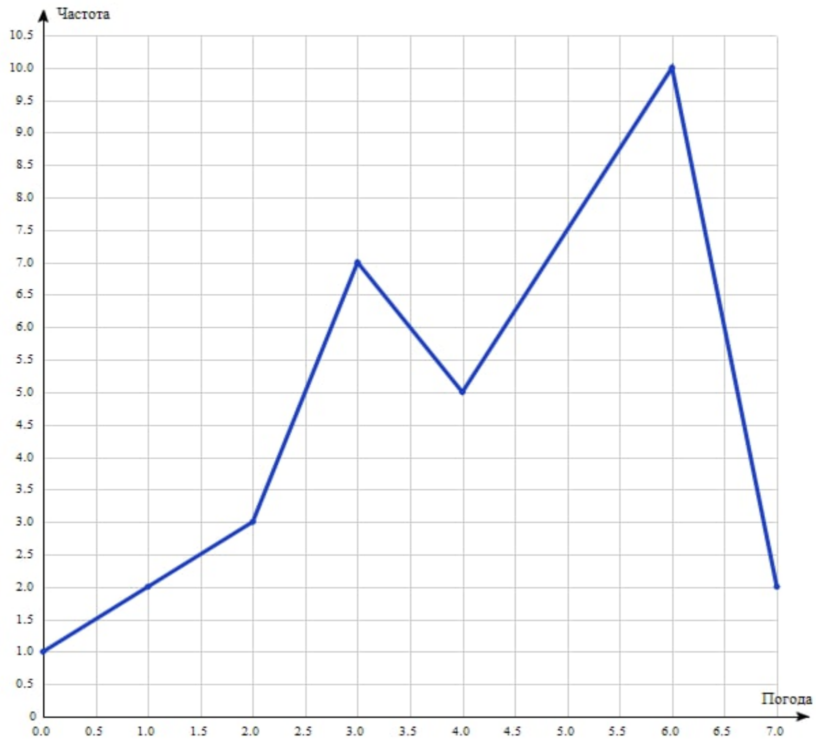



Рисунок 2.2 – Графік нічної температури

На основі накопиченої частоти, функція розподілу денної температури має вигляд:

$$F_{\text{денна}}(x) = \begin{cases} 0, x \leq 3 \\ \frac{2}{30}, 3 < x \leq 4 \\ \frac{5}{30}, 4 < x \leq 5 \\ \frac{10}{30}, 5 < x \leq 6 \\ \frac{15}{30}, 6 < x \leq 7 \\ \frac{17}{30}, 7 < x \leq 9 \\ \frac{21}{30}, 9 < x \leq 10 \\ \frac{29}{30}, 10 < x \leq 11 \\ 1, x > 11 \end{cases} \tag{2.1}$$

А функція розподілу нічної температури:

$$F_{\text{нічна}}(x) = \begin{cases} 0, x \leq 0 \\ \frac{1}{30}, 0 < x \leq 1 \\ \frac{3}{30}, 1 < x \leq 2 \\ \frac{6}{30}, 2 < x \leq 3 \\ \frac{13}{30}, 3 < x \leq 4 \\ \frac{18}{30}, 4 < x \leq 6 \\ \frac{28}{30}, 6 < x \leq 7 \\ 1, x \geq 7 \end{cases} \tag{2.2}$$

Графік першої функції має вигляд (див. рис. 2.3):

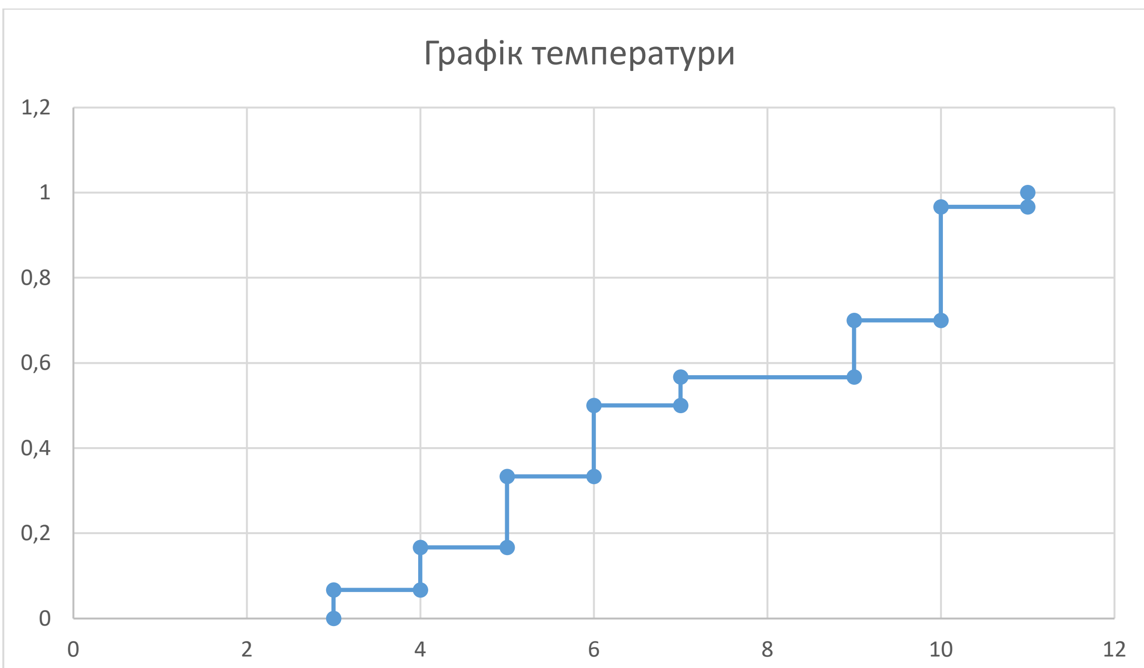



Рисунок 2.3 – Графік функції розподілу денної температури

Графік другої функції (див. рис. 2.4):

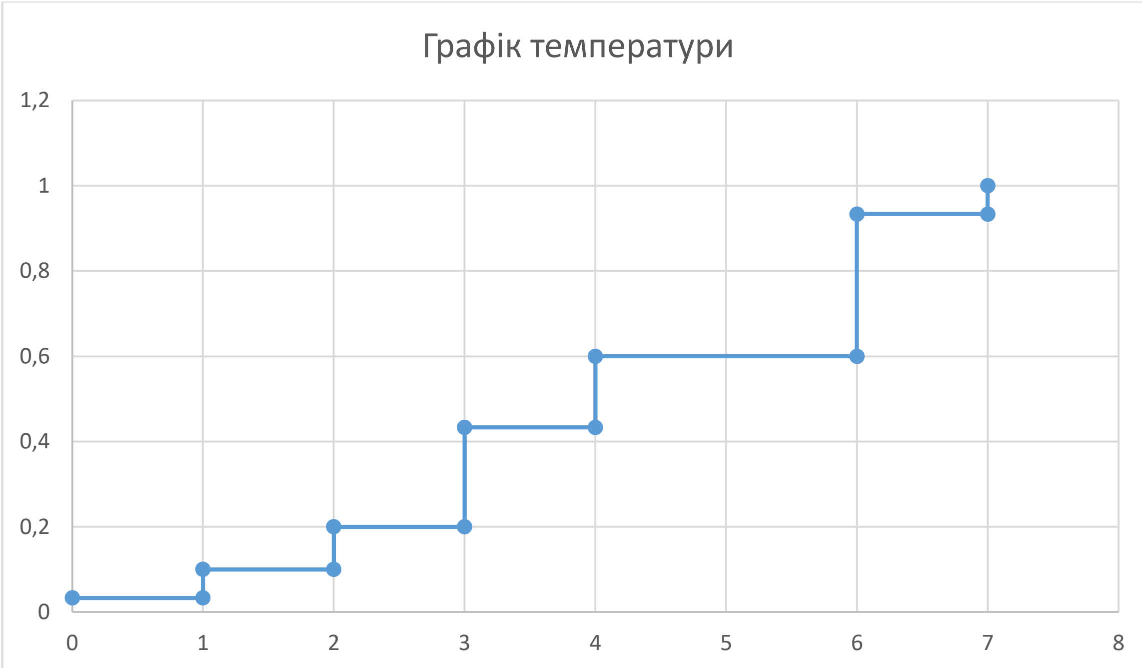



Рисунок 2.4 – Графік функції розподілу нічної температури

Побудуємо інтервальні таблиці частот, поділивши проміжок температур на 4 частини та на їх основі побудуємо гістограму частот.

Для денної температури розмах вибірки $R=t_{max}-t_{min}=11-3=8$. Довжина інтервалу $l_i=8:4=2$ (див. табл. 2.5).

Таблиця 2.5 – Інтервальний варіаційний ряд денної температури в Запоріжжі за листопад

| $[z_{i-1}; z_i]$ | [3; 5) | [5; 7) | [7; 9) | [9; 11] |
|---|---|---|---|---|
| $n_i$ | 5 | 10 | 2 | 13 |
| $w_i$ | 5/30 | 10/30 | 2/30 | 13/30 |

Порахуємо висоти гістограми частот для кожного з інтервалів (див. рис. 2.5):

$$h_1 = \frac{5}{2} = 2{,}5;\ h_2 = \frac{10}{2} = 5;\ h_3 = \frac{2}{2} = 1;\ h_4 = \frac{13}{2} = 6{,}5. \qquad (2.3)$$

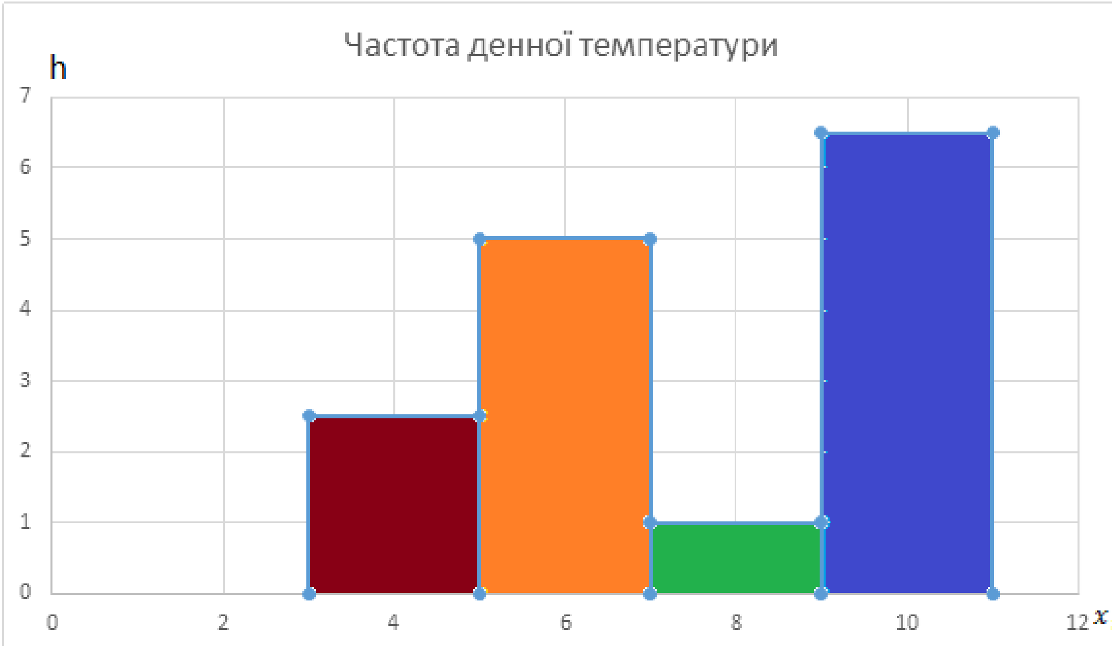



Рисунок 2.5 – Гістограма частот денної температури

Для нічної температури розмах вибірки $R=t_{max}-t_{min}=7-0=7$. Довжина інтервалу $l_i=7:4=1,75$ (див. табл. 2.6).

Таблиця 2.6 – Інтервальний варіаційний ряд денної температури в Запоріжжі за листопад

| $[z_{i-1}; z_i]$ | [0; 1,75) | [1,75; 3,5) | [3,5; 5,25) | [5,25; 7] |
|---|---|---|---|---|
| $n_i$ | 3 | 10 | 5 | 12 |
| $w_i$ | 3/30 | 10/30 | 5/30 | 12/30 |

Порахуємо висоти гістограми частот для кожного з інтервалів (див. рис. 2.6):

$$h_1 = \frac{3}{1,75} = 1,71;\ h_2 = \frac{10}{1,75} = 5,71;\ h_3 = \frac{5}{1,75} = 2,86;\ h_4 = \frac{12}{1,75} = 6,86. \qquad (2.4)$$

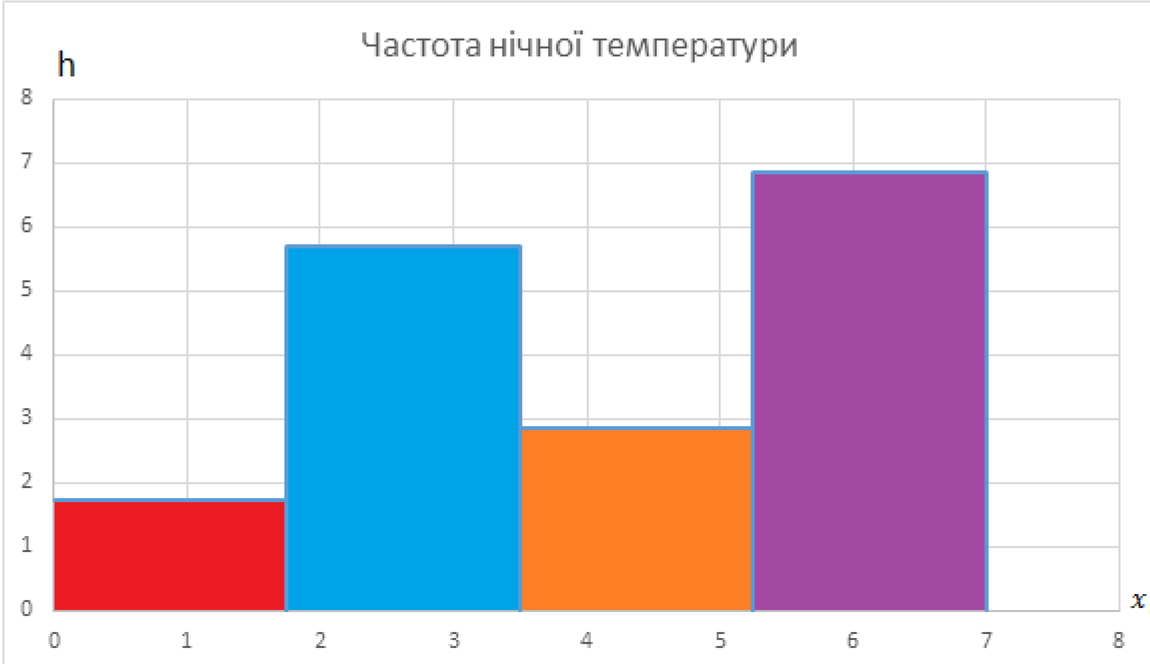



Рисунок 2.6 – Гістограма частот нічної температури

Розрахуємо основні статистичні характеристики для денної та нічної температур:

$$x_в = \frac{1}{30}(3 \cdot 2 + 4 \cdot 3 + 5 \cdot 5 + 6 \cdot 5 + 7 \cdot 2 + 9 \cdot 4 + 10 \cdot 8 + 11 \cdot 1) = \frac{1}{30} \cdot 214 = 7{,}13. \qquad (2.5)$$

$$y_в = \frac{1}{30}(0 \cdot 1 + 1 \cdot 2 + 2 \cdot 3 + 3 \cdot 7 + 4 \cdot 5 + 6 \cdot 10 + 7 \cdot 2) = \frac{1}{30} \cdot 123 = 4{,}1 \qquad (2.6)$$

Таким чином, середня температура вдень у листопаді була 7,13 ℃, а вночі – 4,1 ℃.

$$D_x = \frac{1}{30}(3^2 \cdot 2 + 4^2 \cdot 3 + 5^2 \cdot 5 + 6^2 \cdot 5 + 7^2 \cdot 2 + 9^2 \cdot 4 + 10^2 \cdot 8 + 11^2 \cdot 1) - (7{,}13)^2 =$$
$$= \frac{1}{30} \cdot 1714 - 50{,}84 = 6{,}3 \qquad (2.7)$$

$$D_y = \frac{1}{30}(0^2 \cdot 1 + 1^2 \cdot 2 + 2^2 \cdot 3 + 3^2 \cdot 7 + 4^2 \cdot 5 + 6^2 \cdot 10 + 7^2 \cdot 2) - (4{,}1)^2 = \frac{1}{30} \cdot 615 -$$
$$-16{,}81 = 3{,}7 \qquad (2.8)$$

Отже, міра розсіювання температур відносно середньої величини вдень 6,3 °C, а вночі – 3,7 °C.

$$\sigma_x = \sqrt{6{,}3} = 2{,}51; \;\; \sigma_y = \sqrt{3{,}7} = 1{,}9 \qquad (2.9)$$

Оскільки середнє квадратичне відхилення невелике, то можна стверджувати, що денна та нічна температури зосереджені навколо своїх середніх значень.

Оскільки, температура у градусах по Цельсію є відносною величиною, то для прогнозування коефіцієнт варіації $V_x = \frac{2{,}51}{7{,}13} \cdot 100\% = 35{,}2\%$ та $V_y = \frac{1{,}9}{4{,}1} \cdot 100\% = 46{,}3\%$ буде ненадійним показником.

Мода серед денних температур дорівнює 10 ℃, а серед нічних – 6 ℃.

Знайдемо медіани температур:

Денні температури: 3; 4; 5; 6; 7; 9; 10; 11. Тому

$$M_x = \frac{6+7}{2} = 6{,}5 \text{ °C}. \qquad (2.10)$$

Нічні температури: 0; 1; 2; 3; 4; 6; 7. Тобто

$$M_y = 3\text{°C}. \qquad (2.11)$$

Для розрахунку вибіркової асиметрії та ексцесу розрахуємо вибіркові емпіричні центральні моменти 3 та 4 порядку для денних та нічних температур.

Центральний момент третього порядку для денних температур:

$$\mu_3 = \frac{1}{30}((3-7{,}13)^2 \cdot 2 + (4-7{,}13)^3 \cdot 3 + (5-7{,}13)^3 \cdot 5 + (6-7{,}13)^3 \cdot 5 \ + (7 - \\ -7{,}13)^3 \cdot 2 + (9-7{,}13)^3 \cdot 4 + (10-7{,}13)^3 \cdot 8 + (11-7{,}13)^3) = \frac{1}{30}(-140{,}9 - 92 - \\ -48{,}3 - 7{,}2 - 0{,}004 + 26{,}2 + 189{,}1 + 58) = \frac{1}{30} \cdot (-15{,}104) = -0{,}5. \quad (2.12)$$

Центральний момент четвертого порядку для денних температур:

$$\mu_4 = \frac{1}{30}((3-7{,}13)^4 \cdot 2 + (4-7{,}13)^4 \cdot 3 + (5-7{,}13)^4 \cdot 5 + (6-7{,}13)^4 \cdot 5 + (7 - \\ -7{,}13)^4 \cdot 2 + (9-7{,}13)^4 \cdot 4 + (10-7{,}13)^4 \cdot 8 + (11-7{,}13)^4) = \frac{1}{30}(581{,}9 + 287{,}9 + \\ +102{,}9 + 8{,}2 + 0{,}0006 + 48{,}9 + 542{,}8 + 224{,}3) = \frac{1}{30} \cdot 1796{,}9 = 59{,}9. \quad (2.13)$$

Центральний момент третього порядку для нічних температур:

$$\mu_3 = \frac{1}{30}((0-4{,}1)^3 \cdot 1 + (1-4{,}1)^3 \cdot 2 + (2-4{,}1)^3 \cdot 3 + (3-4{,}1)^3 \cdot 7 + (4-4{,}1)^3 \cdot 5 + \\ +(6-4{,}1)^3 \cdot 10 + (7-4{,}1)^3 \cdot 2) = \frac{1}{30}(-68{,}9 - 59{,}6 - 27{,}8 - 9{,}3 - 0{,}005 + 68{,}6 + \\ +48{,}8) = \frac{1}{30} \cdot (-48{,}205) = -1{,}6. \quad (2.14)$$

Центральний момент четвертого порядку для нічних температур:

$$\mu_4 = \frac{1}{30}((0-4{,}1)^4 \cdot 1 + (1-4{,}1)^4 \cdot 2 + (2-4{,}1)^4 \cdot 3 + (3-4{,}1)^4 \cdot 7 + (4-4{,}1)^4 \cdot 5 + \\ +(6-4{,}1)^4 \cdot 10 + (7-4{,}1)^4 \cdot 2) = \frac{1}{30}(282{,}6 + 184{,}7 + 58{,}3 + 10{,}2 + 0{,}0005 + 130{,}3 + \\ +141{,}5) = \frac{1}{30} \cdot 807{,}6 = 26{,}9. \quad (2.15)$$

Розрахуємо коефіцієнт асиметрії для денних та нічних температур:

$$A_x = \frac{-0{,}5}{2{,}51^3} = -0{,}03; \ \ A_y = \frac{-1{,}6}{1{,}9^3} = -0{,}44. \quad (2.16)$$

Таким чином, для обох випадків $A < 0$, отже, маємо лівосторонню асиметрію.

Тепер розрахуємо коефіцієнт ексцесу для денних та нічних температур:

$$E_x = \frac{59{,}9}{2{,}51^4} = 1{,}51; \;\; E_y = \frac{26{,}9}{1{,}9^4} = 2{,}06. \tag{2.17}$$

Для обох випадків $E > 0$, тобто емпіричний розподіл денних та нічних температур є високим (або гостровершинним) відносно нормального розподілу.

Знайдемо незміщені дисперсії температур:

$$S_x^2 = \frac{30}{29} \cdot 6{,}3 = 6{,}5; \;\; S_y^2 = \frac{30}{29} \cdot 3{,}7 = 3{,}8. \tag{2.18}$$

Отже, в даному розділі за даними денної та нічної температур знайдено всі основні статистичні характеристики, побудовано функції розподілу, полігони та гістограми частот.

## 2.2 Обробка отриманих результатів та прогнозування подальшої метеорологічної ситуації

Оскільки погодні умови у листопаді та взимку останні роки майже не відрізняється, то знайдемо довірчі інтервали для майбутньої погоди з ймовірністю $\gamma = 0{,}95$.

Значення функції інтегральної Лапласа $\phi(t) = \frac{\gamma}{2} = \frac{0{,}95}{2} = 0{,}475$, а отже $t = 1{,}96$. У грудні у нас 31 день, отже $n = 31$. У якості середньоквадратичного відхилення візьмемо знайдені результати $\sigma_x = 2{,}51$ і $\sigma_y = 1{,}9$, а вибіркового середнього $x_B = 7{,}13$ та $y_B = 4{,}1$ для денної і нічної температури відповідно.

Довірчий інтервал для денної температури грудня, причому візьмемо лівий інтервал зі знаком мінус, адже згідно з даними погоди у попередні роки, температура у грудні була як мінусовою, так і плюсовою:

Найменша температура:

$$-\left(x_{\text{в}} - \frac{t\sigma}{\sqrt{n}}\right) = -\left(7{,}13 - \frac{1{,}96 \cdot 2{,}51}{\sqrt{31}}\right) = -6{,}25. \qquad (2.19)$$

Найбільша температура:

$$x_{\text{в}} + \frac{t\sigma}{\sqrt{n}} = 7{,}13 + \frac{1{,}96 \cdot 2{,}51}{\sqrt{31}} = 8. \qquad (2.20)$$

Таким чином, згідно з початковим прогнозом, денна температура грудня буде варіюватися у межах: $-6{,}25 < a < 8$.

Аналогічно знайдемо довірчий інтервал для нічної температури грудня:

Найменша температура:

$$-\left(x_{\text{в}} - \frac{t\sigma}{\sqrt{n}}\right) = -\left(4{,}1 - \frac{1{,}96 \cdot 1{,}9}{\sqrt{31}}\right) = -3{,}5. \qquad (2.21)$$

Найбільша температура:

$$x_{\text{в}} + \frac{t\sigma}{\sqrt{n}} = 4{,}1 + \frac{1{,}96 \cdot 1{,}9}{\sqrt{31}} = 4{,}8. \qquad (2.22)$$

Таким чином, згідно з початковим прогнозом, нічна температура грудня буде варіюватися у межах: $-3{,}5 < b < 4{,}8$.

Але даний прогноз не є максимально точним, адже для грудня ми не маємо середнього квадратичного відхилення. Для уточнення результатів спробуємо використати незміщену дисперсію $S_x^2 = 6{,}5$; $S_y^2 = 3{,}8$ за листопад та розподіл Стьюдента з $\gamma = 0{,}95$ та $n = 31$.

$$S_x = \sqrt{S_x^2} = \sqrt{6{,}5} = 2{,}55. \quad (2.23)$$

$$S_y = \sqrt{S_y^2} = \sqrt{3{,}8} = 1{,}95. \quad (2.24)$$

За даними розподілу Стьюдента маємо:

$$t_\gamma = t(n,\gamma) = t(31; 0{,}95) = 2{,}04. \quad (2.25)$$

Найменша температура:

$$-\left(x_в - \frac{t_\gamma S}{\sqrt{n}}\right) = -\left(7{,}13 - \frac{2{,}04 \cdot 2{,}55}{\sqrt{31}}\right) = -6{,}2. \quad (2.26)$$

Найбільша температура:

$$x_в + \frac{t_\gamma S}{\sqrt{n}} = 7{,}13 + \frac{2{,}04 \cdot 2{,}55}{\sqrt{31}} = 8{,}1. \quad (2.27)$$

Таким чином, згідно з уточненим прогнозом, денна температура грудня буде варіюватися у межах: $-6{,}2 < a < 8{,}1$.

Для нічної температури маємо:

Найменша температура:

$$-\left(x_в - \frac{t_\gamma \sigma}{\sqrt{n}}\right) = -\left(4{,}1 - \frac{2{,}04 \cdot 1{,}95}{\sqrt{31}}\right) = -3{,}4. \quad (2.28)$$

Найбільша температура:

$$x_{\text{в}} - \frac{t_{\gamma}\sigma}{\sqrt{n}} = 4{,}1 + \frac{2{,}04 \cdot 1{,}95}{\sqrt{31}} = 4{,}8. \tag{2.29}$$

Таким чином, згідно з уточненим прогнозом, нічна температура грудня буде варіюватися у межах: $-3{,}4 < b < 4{,}8$ (див. рис. 2.7).

Перший і другий прогнози майже не відрізняються один від одного.

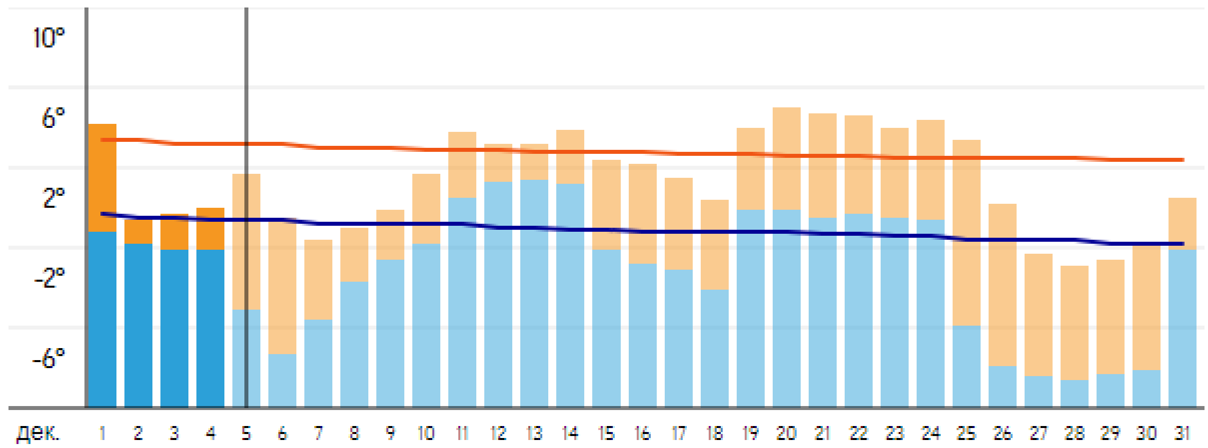



Рисунок 2.7 – Прогноз погоди у Запоріжжі на грудень

У якості основного математичного апарату для уточнення результатів прогнозу, візьмемо систему розподілів Пірсона. Це дозволить, зокрема, здійснити перевірку гіпотез однорідності і згоди вихідної інформації при підборі апроксимованих розподілів. Розподіл Пірсона повністю визначається першими чотирма центральними моментами випадкової величини. Центральним моментом другого порядку є дисперсія.

Для подальшого дослідження, у нас не вистачає лише центрального моменту першого порядку.

Знайдемо його для денної температури:

$$\mu_1 = \frac{1}{30}\big((3-7{,}13)\cdot 2 + (4-7{,}13)\cdot 3 + (5-7{,}13)\cdot 5 + (6-7{,}13)\cdot 5 + (7-7{,}13)\cdot \\ \cdot 2 + (9-7{,}13)\cdot 4 + (10-7{,}13)\cdot 8 + (11-7{,}13)\big)\frac{1}{30}\cdot(-8{,}26-9{,}39- \\ -10{,}65-5{,}65-0{,}26+7{,}48+22{,}96+3{,}87) = \frac{1}{30}\cdot 0{,}1 = 0{,}003. \quad (2.30)$$

Для нічної температури:

$$\mu_1 = \frac{1}{30}\big((0-4{,}1)\cdot 1 + (1-4{,}1)\cdot 2 + (2-4{,}1)\cdot 3 + (3-4{,}1)\cdot 7 + (4-4{,}1)\cdot 5 + (6- \\ -4{,}1)\cdot 10 + (7-4{,}1)\cdot 2\big) = \frac{1}{30}\cdot(-4{,}1-6{,}2-6{,}3-7{,}7-0{,}5+ \\ +19+5{,}8) = \frac{1}{30}\cdot 0 = 0. \quad (2.31)$$

Таким чином маємо такі дані:

Для денної температури:

$$\mu_1 = 0{,}003;\ \mu_2 = 6{,}3;\ \mu_3 = -0{,}5;\ \mu_4 = 59{,}9. \quad (2.32)$$

Для нічної температури:

$$\mu_1 = 0;\ \mu_2 = 3{,}7;\ \mu_3 = -1{,}6;\ \mu_4 = 26{,}9 \quad (2.33)$$

Розрахуємо параметри розподілу Пірсона:

$$A_x = 10\mu_4\mu_2 - 18\mu_2^3 - 12\mu_3^2 = 10\cdot 59{,}9\cdot 6{,}3 - 18\cdot 6{,}3^3 - 12\cdot \\ \cdot(-0{,}5)^2 = -730{,}15. \quad (2.34)$$

$$A_y = 10\mu_4\mu_2 - 18\mu_2^3 - 12\mu_3^2 = 10\cdot 26{,}9\cdot 3{,}7 - 18\cdot 3{,}7^3 - 12\cdot \\ \cdot(-1{,}6)^2 = 52{,}83. \quad (2.35)$$

Коефіцієнти розраховуються за формулою:

$$a_0 = \frac{\mu_3(\mu_4+3\mu_2^2)}{A}; \tag{2.36}$$

$$b_0 = -\frac{\mu_2(4\mu_2\mu_4-3\mu_3^2)}{A}; \tag{2.37}$$

$$b_1 = -\frac{\mu_3(4\mu_2\mu_4+3\mu_2^2)}{A}; \tag{2.38}$$

$$b_2 = -\frac{2\mu_2\mu_4-3\mu_3^2-6\mu_2^3}{A}. \tag{2.39}$$

Для денної температури:

$$a_0 = \frac{-0{,}5(59{,}9+3\cdot(6{,}3)^2)}{-730{,}15} = 0{,}25; \tag{2.40}$$

$$b_0 = -\frac{6{,}3(4\cdot 6{,}3\cdot 59{,}9-3\cdot(-0{,}5)^2)}{-730{,}15} = 13{,}02; \tag{2.41}$$

$$b_1 = -\frac{-0{,}5(4\cdot 6{,}3\cdot 59{,}9+3\cdot 6{,}3^2)}{-730{,}15} = -1{,}12; \tag{2.42}$$

$$b_2 = -\frac{2\cdot 6{,}3\cdot 59{,}9-3\cdot(-0{,}5)^2-6\cdot 6{,}3^2}{-730{,}15} = 0{,}71; \tag{2.43}$$

Для нічної температури:

$$a_0 = \frac{-1{,}6(26{,}9+3\cdot 3{,}7^2)}{52{,}83} = -2{,}06; \tag{2.44}$$

$$b_0 = -\frac{3{,}7(4\cdot 3{,}7\cdot 26{,}9-4\cdot(-1{,}6)^2)}{52{,}83} = -27{,}17; \tag{2.45}$$

$$b_1 = -\frac{-1{,}6(4\cdot 3{,}7\cdot 26{,}9+3\cdot 3{,}7^2)}{52{,}83} = -13{,}3; \tag{2.46}$$

$$b_2 = -\frac{2\cdot 3{,}7\cdot 26{,}9-3\cdot(-1{,}6)^2-6\cdot 3{,}7^3}{52{,}83} = 2{,}13. \tag{2.47}$$

Отже, для денної температури маємо диференціальне рівняння щільності розподілу:

$$\frac{\partial p(x)}{\partial x} = \frac{x - 0{,}25}{13{,}02 - 1{,}12x + 0{,}71x^2}. \qquad (2.48)$$

Для нічної температури маємо диференціальне рівняння щільності розподілу:

$$\frac{\partial q(x)}{\partial x} = \frac{x + 2{,}06}{-27{,}17 - 13{,}3x + 2{,}13x^2}. \qquad (2.49)$$

Тепер знайдемо корені тричлена $b_2x^2 + 2b_1x + b_0$.

Для денної температури:

$$0{,}71x^2 - 2{,}24x + 13{,}02 = 0;$$
$$D = b^2 - 4ac = (-2{,}24)^2 - 4 \cdot 0{,}71 \cdot 13{,}02 = 5{,}0176 - 36{,}9768 = -31{,}9592; \qquad (2.50)$$
$$\sqrt{D} = \pm 5{,}65i;$$
$$x_{1,2} = \frac{2{,}24 \pm 5{,}65i}{2 \cdot 0{,}71} = 1{,}58 \pm 3{,}98i.$$

Знайдемо $D$:

$$D = b_0b_2 - b_1^2 = 13{,}02 \cdot 0{,}71 - 1{,}12^2 = 7{,}99 > 0. \qquad (2.51)$$

Тепер знайдемо коефіцієнт $\lambda$:

$$\lambda = \frac{b_1^2}{b_0b_2} = \frac{1{,}12^2}{13{,}02 \cdot 0{,}71} = 0{,}14 > 0. \qquad (2.52)$$

Але 0,14 < 1. Таким чином, для денних температур маємо четвертий тип розподілу Пірсона.

Для нічної температури:

$$2{,}13x^2 - 26{,}6x - 27{,}17 = 0;$$

$$D = b^2 - 4ac = (-26{,}6)^2 - 4 \cdot 2{,}13 \cdot (-27{,}17) = 707{,}56 + 231{,}4884 = 939{,}0484;$$

$$x_1 = \frac{26{,}6 - \sqrt{939{,}0484}}{2 \cdot 2{,}13} = \frac{1330}{213} - \frac{1}{213}\sqrt{2347621} \approx -0{,}95; \tag{2.53}$$

$$x_2 = \frac{26{,}6 + \sqrt{939{,}0484}}{2 \cdot 2{,}13} = \frac{1330}{213} + \frac{1}{213}\sqrt{2347621} \approx 13{,}47.$$

Аналогічно розрахуємо $D$:

$$D = b_0 b_2 - b_1^2 = -27{,}17 \cdot 2{,}13 - 13{,}3^2 = -234{,}8. \tag{2.54}$$

Розрахуємо коефіцієнт $\lambda$:

$$\lambda = \frac{b_1^2}{b_0 b_2} = \frac{13{,}3^2}{-27{,}17 \cdot 2{,}13} = -3{,}06. \tag{2.55}$$

Маємо дійсні корені різного знаку та лямбда менше нуля, отже для нічних температур маємо перший тип розподілу або бета розподіл першого роду.

Розрахуємо формулу щільності розподілу денних температур для уточнення прогнозу.

Згідно з формулами розподілу Пірсона 4 типу, маємо:

$$y = x + \frac{b_1}{2b_2} = x + \frac{-1{,}12}{2 \cdot 0{,}71} = x + \frac{-1{,}12}{1{,}42} = x - 0{,}79; \tag{2.56}$$

$$a = \frac{\sqrt{4b_2 b_0 - b_1^2}}{2b_2} = \frac{\sqrt{4 \cdot 0{,}71 \cdot 13{,}02 - 1{,}12^2}}{1{,}42} = 4{,}21; \tag{2.57}$$

$$f(x) = b_2(y^2 + a^2) = 0{,}71((x - 0{,}79)^2 + 17{,}7) = 0{,}71(x^2 - 1{,}58x + 18{,}35); \tag{2.58}$$

$$m = \frac{1}{2b_2} = \frac{1}{1{,}42} = 0{,}7; \tag{2.59}$$

$$V = -\frac{2b_2 a - b_1}{2b_2^2 a} = -\frac{2 \cdot 0{,}71 \cdot 0{,}25 + 1{,}12}{2 \cdot 0{,}71^2 \cdot 4{,}21} = -\frac{1{,}475}{4{,}24} = -0{,}35. \tag{2.60}$$

Щільність розподілу знаходять за формулою:

$$p(x) = \frac{\left|\frac{\Gamma\left(m+\frac{V}{2}i\right)}{\Gamma(m)}\right|^2}{\alpha B(m-\frac{1}{2},\frac{1}{2})}\left[1+\left(\frac{x-\lambda}{\alpha}\right)^2\right]^{-m} exp\left[-Varctg\left(\frac{x-\lambda}{\alpha}\right)\right]. \quad (2.61)$$

$$p(x) = \frac{\left|\frac{\Gamma(0,7+0,175i)}{\Gamma(0,7)}\right|^2}{4,21\cdot\frac{\Gamma(0,2)\Gamma(0,5)}{\Gamma(0,7)}}\cdot\left[1+\left(\frac{x-2,6}{4,21}\right)^2\right]^{-0,7} exp\left[0,35arctg\left(\frac{x-2,6}{4,21}\right)\right]. \quad (2.62)$$

З урахуванням гамма розподілу, більш точний довірчий інтервал математичного сподівання для нічних температур відрізняється на:

$$M(X) = \frac{V}{m} = \frac{-0,35}{0,7} = -0,5. \quad (2.63)$$

Отже, уточнений прогноз для температур грудня має вигляд:

$$-6,25 - 0,5 < a < 8,1 - 0,5 \Rightarrow -6,75 < a < 7,6. \quad (2.64)$$

Можна стверджувати, що мінімальна денна температура буде майже -7 °C, а максимальна температура вдень може сягнути майже 8°C.

Для нічних температур маємо щільність ймовірності:

$$g = \frac{b_1 - x_1}{b_2(x_2 - x_1)}; \quad (2.65)$$

$$g = \frac{-13,3+0,95}{2,13(13,47+0,95)} = \frac{-12,35}{30,715} = -0,4 < 0; \quad (2.66)$$

$$h = \frac{x_2 - b_1}{b_2(x_2 - x_1)}; \quad (2.67)$$

$$h = \frac{13,47+13,3}{30,715} = \frac{26,77}{30,715} = 0,87 > 0. \quad (2.68)$$

Маємо коефіцієнти різних знаків, отже має $J$ - образну форму.

$$p(x) = \frac{(x+0{,}95)^{-0{,}4}(13{,}47-x)^{0{,}37}}{B(0;6;1{,}87)\cdot 14{,}47^{-0{,}4+0{,}87+1}} = \frac{(x+0{,}95)^{-0{,}4}(13{,}47-x)^{0{,}37}}{\frac{\Gamma(0{,}6)\Gamma(1{,}87)}{\Gamma(2{,}47)}\cdot 14{,}47^{1{,}47}} =$$

$$= \frac{(13{,}47-x)^{0{,}87}}{50{,}8\cdot\frac{1{,}49\cdot 0{,}95}{1{,}3}(x+0{,}95)^{0{,}4}} = \frac{(13{,}47-x)^{0{,}87}}{54{,}44\cdot(x+0{,}95)^{0{,}4}}. \quad (2.69)$$

Отже, більш точний довірчий інтервал математичного сподівання для нічних температур відрізняється на:

$$M(X) = \frac{g}{g+h} = \frac{-0{,}4}{-0{,}4+0{,}87} = -0{,}85. \quad (2.70)$$

Дисперсія відрізняється на:

$$D(X) = \frac{gh}{(g+h)^2(g+h+1)} = \frac{-0{,}4\cdot 0{,}87}{0{,}47^2\cdot 1{,}47} = -1{,}07. \quad (2.71)$$

З урахуванням отриманих результатів, уточнений прогноз для нічних температур грудня має наступний вигляд:

$$-3{,}4 - 0{,}85 < b < 4{,}8 - 0{,}85 \Rightarrow -4{,}25 < b < 4. \quad (2.72)$$

З урахуванням похибки приблизно в 1 градус, можна стверджувати, що мінімальна нічна температура буде -5 °С, а максимальна температура вночі може сягнути 5°С.

Отже, в даному підрозділі обчислено додаткові параметри розподілу температур на основі отриманих результатів у пункті 2.1 та за їх допомогою отримано короткостроковий прогноз погоди на грудень 2020 у Запоріжжі.

Таким чином, для апроксимації розподілів погоди і розрахунку імовірнісних характеристик погоди найзручніше використовувати змішане сімейство розподілів Пірсона. Змішане сімейство розподілів еквівалентно по апроксимується можливостям родин Пірсона, проте дозволяє уникнути обчислювальних труднощів визначення параметрів розподілів.

# ВИСНОВКИ

Найважливішими підрозділами математичної статистики є описова статистика та статистичні висновки. Статистика ґрунтується на базі теорії ймовірності та застосовується як в природничому, так і у гуманітарному циклі наук. Основною базою для використання статистики у емпіричних дослідженнях є генеральна сукупність. Шляхом виділення репрезентативної вибірки як частини генеральної сукупності, за певною, науково обґрунтованою, ознакою науковці проводять подальші дослідження та перехід до абстрактної моделі. Першим математичним кроком у дослідженні є побудова варіаційного ряду, а також підрахунок мір центральної тенденції – вибіркового середнього, моди та медіани й мір мінливості – вибіркової дисперсії, середньоквадратичного відхилення, коефіцієнту варіації, асиметрії та ексцесу. Но основі мір мінливості та мір центральної тенденції, можна продовжити статистичний аналіз застосувавши інші статистичні методи – побудова довірчого інтервалу, перевірка гіпотез, кореляційний та регресійний аналіз тощо. Останнім кроком є математичні висновки та їх інтерпретація щодо реальних об’єктів чи процесів дійсності.

Статистичний аналіз медико-біологічних даних зазвичай проходить кілька етапів. Першим з них є проведення заходів щодо описової статистикою, тобто розрахунок кількісних статистичних показників. Тут же розглядається питання про відповідність розподілу нормальному розподілу ймовірності. Перевірити відповідність отриманих даних нормальному закону розподілу можна двома способами. Перший з них - за допомогою побудови графіків на нормальній ймовірнісної папері, другий - за допомогою критеріїв згоди. Другим етапом є статистичне оцінювання, тобто проводиться оцінка параметрів генеральної сукупності за параметрами вибіркової сукупності. На цьому етапі важливо знання нормальності аналізованої статистичної сукупності. Наступним етапом є перевірка статистичних гіпотез. У цих випадках результати вибірки використовуються для перевірки припущень (гіпотез) щодо тих чи інших властивостей розподілу генеральної сукупності або порівнянні параметрів різних вибірок.

В підрозділі 2.1 за даними денної та нічної температур знайдено всі основні статистичні характеристики, побудовано функції розподілу, полігони та гістограми частот.

В підрозділі 2.2 обчислено додаткові параметри розподілу температур на основі отриманих результатів у пункті 2.1 та за їх допомогою отримано короткостроковий прогноз погоди на грудень 2020 у Запоріжжі.

Таким чином, для апроксимації розподілів погоди і розрахунку імовірнісних характеристик погоди найзручніше використовувати змішане сімейство розподілів Пірсона. Змішане сімейство розподілів еквівалентно по апроксимується можливостям родин Пірсона, проте дозволяє уникнути обчислювальних труднощів визначення параметрів розподілів.

# ПЕРЕЛІК ПОСИЛАНЬ


1. Рокицкий П. Ф. Биологическая статистика. Минск : Высшая школа, 1973. 320 с.

2. Исаев А. А. Статистика в метеорологии и климатологии. Москва : Издательство Московского университета, 1988. 246 с.

3. Лакин Г. Ф. Биометрия. Москва : Высшая школа, 1990. 352 с.

4. Артюхов В. Г., Пантявин А. А. Математические методы в биологии: учебно-методическое пособие для вузов. Воронеж : Издательско-полиграфический центр Воронежского государственного университета, 2007. 28 с.

5. Теорія ймовірностей і математична статистика: навчально-методичний посібник. Київ : Київський національний економічний університет. 349 с.

6. Новиков Д. А., Новочадов В. В. Статистические методы в медико-биологическом эксперименте. Волгоград : Волгоградский государственный медицинский университет, 2005. 84 с.

7. Мятлев В. Д., Панченко Л. А. Высшая математика и ее приложения к биологии. Москва : Академия, 2009. 315 с.

8. Муравьева В. Н., Максименко Л. Л. Статистические методы анализа в биологии. Ставрополь : Юность, 2014. 298 с.

9. Приданникова Ю. Є. Можливості статистичних методів для дослідження взаємозв’язків суспільно-економічних явищ і процесів. *Науковий вісник національної академії статистики, обліку та аудиту*. 2017. №3. С. 16-24.

10. Ліхоузова Т. А. Теорія ймовірностей та математична статистика. Курс лекцій. К.: КПІ ім. Ігоря Сікорського, 2018. 300 с.

11. Кобченко Ю. Ф., Резуненко В. А. *Проблеми безперервної географічної освіти і картографії*. 2009. Вип. 10. С. 102-106.

12. Карпов И. Г., Грибков А. Н. Модернизация распределений Пирсона для аппроксимации двухсторонних законов распределения экспериментальных данных. *Известия Томского политехнического университета*. №2. 2014. С. 5-10.

13. Кобченко Ю. Ф., Кобченко О. Ю. Просторово-часові тенденції зміни температури повітря у період потепління клімату на території України. *Вісник Харківського національного університету*. №1157. С. 88-94.

14. Сиделев С. И. Математические методы в биологии и экологии: введение в элементарную биометрию: учебное пособие. Ярославль : Яросл. гос. ун-т им. П. Г. Демидова, 2012. 140 с.

15. Скоринкин А. И. Математическое моделирование биологических процессов. Казань : Казан. ун-т, 2015. 86 с.

16. Поліщук Н. В., Орел Б. П., Штабалюк П. І. Теорія ймовірностей та математична статистика: навч. посіб. Київ. НТУУ «КПІ», 2014. 212 с.

17. Стентон Гланц. Медико-биологическая статистика. Пер. с англ. Москва : Практика, 1998. 459 с.

18. Глушанко В. С. Основы медицинской статистики: учеб.-метод. пособие. Витебск : ВГМУ, 2012. 155 с.

19. Лапина С. Н., Иванова Г. Ф., Семенова Н. В. Лабораторний практикум по курсу «Обработка и анализ метеорологической информации». Саратов : Саратовский государственный университет им. Н. Г. Чернышевского, 2011. 37 с.

20. Руденко В. М. Математична статистика. Навч. посіб. К.: Центр учбової літератури, 2012. 304 с.

21. Тимошенко С. И. Использование смешанного семейства распределений Пирсона и Джонсона для расчета вероятностных характеристик.

*Интернет журнал «Науковедение»*, том 9, №3. URL: http://naukovedenie.ru/PDF/03TVN317.pdf (2017).

22. Зверев А. А., Зефиров Т. Л. Статистические методы в биологии: учебно-методическое пособие. Казань, КФУ, 2013. 42 с.

23. Дегтярев А. С., Драбенко В. А. Статистические методы обработки метеорологической информации. Учебник. СПб: ООО «Андреевский издательский дом», 2015. 225 с.

24. Опря А. Т. Статистика (модульний варіант з програмованою формою контролю знань). Навч. посіб. К.: Центр учбової літератури, 2012. 448 с.

25. Данілов В. Я. Статистична обробка даних: навч. посіб. Київський національний університет ім. Тараса Шевченка. Київ, 2019. 256 с.

26. Кобченко Ю. Ф., Кобченко О. Ю. Кількісні методи дослідження фітопогодного комплексу. *Вісник Харківського національного університету*, №864. С. 144-149.