МІНІСТЕРСТВО ОСВІТИ І НАУКИ УКРАЇНИ
ЧЕРКАСЬКИЙ ДЕРЖАВНИЙ ТЕХНОЛОГІЧНИЙ УНІВЕРСИТЕТ

МЕТОДИЧНІ РЕКОМЕНДАЦІЇ

ДО ВИКОНАННЯ ЛАБОРАТОРНИХ РОБІТ
ТА КОНТРОЛЮ ЗНАНЬ
З ДИСЦИПЛІНИ «БІОМЕТРІЯ»
для здобувачів освітнього рівня магістр
спеціальності 101 «Екологія» усіх форм навчання

Черкаси 2021

УДК 57.087.1(07)
М54

Затверджено вченою радою БФ,
протокол № 10 від 22 червня 2021 р.,
згідно з рішенням кафедри екології
протокол № 3 від 24.03.2021 р.

Упорядники: Єгорова О.В., *к.т.н., ст. викладач, Мислюк О.О., к.х.н., доцент*

Рецензент Лавданська О.В., *к.т.н., доцент кафедри інформаційних технологій проектування ЧДТУ*.

М54 Методичні рекомендації до виконання лабораторних робіт та контролю знань з дисципліни «Біометрія» для здобувачів освітнього рівня магістр спеціальності 101 «Екологія» усіх форм навчання [Текст] / Укл.: О.В. Єгорова, О.О. Мислюк; М-во освіти і науки України, Черкас. держ. технол. ун-т. – Черкаси: ЧДТУ, 2021. – 72 с.

Методичні рекомендації спрямовані для використання під час лабораторного практикуму та формування у здобувачів вищої світи практичних навичок проведення лабораторних досліджень, поглиблення та закріплення теоретичних знань.

УДК 57.087.1(07)

Навчальне електронне видання
мережного використовування

МЕТОДИЧНІ РЕКОМЕНДАЦІЇ
ДО ВИКОНАННЯ ЛАБОРАТОРНИХ РОБІТ ТА КОНТРОЛЮ ЗНАНЬ
З ДИСЦИПЛІНИ «БІОМЕТРІЯ»
для здобувачів освітнього рівня магістр
зі спеціальності 101 «Екологія»

Упорядники **Єгорова** Оксана В’ячеславівна
Мислюк Ольга Олександрівна

В авторській редакції

ЗМІСТ

ВСТУП.. 4

1 ЛАБОРАТОРНА РОБОТА №1. *Вступ. Основні поняття статистичної сукупності. Біологічні ознаки*.. 6

2 ЛАБОРАТОРНА РОБОТА №2. *Групування результатів спостереження*.. 10

3 ЛАБОРАТОРНА РОБОТА №3. *Визначення середньої арифметичної і біометричних показників ряду розподілу. Графіки розподілу*............. 22

4 ЛАБОРАТОРНА РОБОТА № 4. *Розрахунок помилок репрезентативності*.. 39

5 ЛАБОРАТОРНА РОБОТА № 5. *Визначення числових показників кореляції і їх достовірності*.. 43

6 ЛАБОРАТОРНА РОБОТА № 6. *Регресійний аналіз. Проста лінійна регресія. Визначення коефіцієнтів регресії*... 50

7 ЛАБОРАТОРНА РОБОТА №7. *Дисперсійний аналіз однофакторних рівномірних і нерівномірних комплексів малих груп*... 56

ПОРЯДОК ВИКОНАННЯ ТА ОЦІНЮВАННЯ ЛАБОРАТОРНИХ РОБІТ. 60

ТЕСТОВІ КОНТРОЛЬНІ ЗАВДАННЯ ДЛЯ ПЕРЕВІРКИ ЯКОСТІ ЗНАНЬ ЗДОБУВАЧІВ ВИЩОЇ ОСВІТИ.. 61

СПИСОК РЕКОМЕНДОВАНОЇ ЛІТЕРАТУРИ……...................................... 66

ДОДАТКИ.. 67

ВСТУП

Сьогодні статистичні методи і підходи при моделюванні складних явищ природи набули широкої популярності при прогнозуванні стану довкілля. У реалізації математично-статистичного підходу важливе місце займають експериментальні методи і методи багатовимірної статистики, які дозволяють дослідникам виявити закономірності за яких відбуваються ті чи інші природні процес, розкрити причинно-наслідкові зв’язки між елементами живої природи, зробити їх доступними для опису точними математичними моделями.

При наукових дослідженнях біологічних явищ найбільш ефективним є метод масових спостережень, використання якого передбачає проведення великого числа спостережень. Зібраний матеріал обробляють, аналізують, роблять відповідні висновки і встановлюють ті чи інші закономірності. Розглянутий шлях називається прямим індуктивним методом, коли від окремих фактів переходять до загальних положень. Однак точність і достовірність результатів біологічних експериментів, так само як і коректність сформульованих висновків, залежать не тільки від якостей експериментальних методик. Властивості самих біологічних об’єктів і явищ також чинять вплив на варіацію сукупності величин спостережень.

Застосування статистичного аналізу експериментально отриманих даних доповнює і поглиблює пізнання біологічних явищ природи, дозволяє об’єктивно оцінити отримані результати, знівелювавши суб’єктивізм дослідника і методичні помилки при постановці експерименту, страхує експериментатора від неточних і необґрунтованих висновків і висновків щодо досліджуваного явища.

Мета даних методичних рекомендацій, котрі містять інформацію про стандартні і найбільш вживані біометричні методи, – закріпити теоретичні знання і набуття досвіду у здобувачів вищої освіти математичної обробки даних спостереження. Лабораторний практикум містить практичні розрахункові роботи, в яких відпрацьовуються основні математичні методи обробки масових експериментальних матеріалів, повторюються ключові теоретичні моменти окремих розділів курсу «Біометрія», а також завдання для самостійної індивідуальної роботи. У посібнику дається оцінка методів математичної статистики з точки зору їх можливостей і меж застосування. При виконанні конкретних видів робіт студенти можуть використовувати в якості матриць експериментальних даних результати власних досліджень.

Окремий розділ практикуму присвячений можливостям вирішення завдань прикладної статистики засобами Microsoft Excel. У лабораторних роботах наводяться приклади застосування стандартних функцій і пакета аналізу даних MS Excel для вирішення конкретних завдань з докладним описом алгоритму рішення. Додаток включає необхідні довідкові дані для виконання лабораторних робіт.

При підготовці методичних рекомендацій використаний досвід інших вузів, наукова і методична література, основний список якої наводиться.

Лабораторна робота №1
ВСТУП. ОСНОВНІ ПОНЯТТЯ СТАТИСТИЧНОЇ СУКУПНОСТІ. БІОЛОГІЧНІ ОЗНАКИ

Мета роботи: освоїти навички вибору і вимірювання мірних і рахункових ознак на різних біологічних об'єктах.

Матеріали і обладнання: листки деревних рослин або чагарників (3-4 види по 50 шт.), насіння квіткових культур (наважки – 1 г), лінійки, калькулятори.

ТЕОРЕТИЧНА ЧАСТИНА

Дослідження закономірності того чи іншого біологічного явища зазвичай проводиться не за окремими одиничними спостереженнями, які можуть виявитися випадковими, нетиповими, чи мають неповноту вираженості сутності даного явища, а на безлічі однорідних спостережень, що в свою чергу дає більш повну інформацію про досліджуваний об'єкт

Сукупність відносно однорідних одиниць спостереження, які об'єднані умовами експерименту, за тією або іншою ознакою, називається *статистичною сукупністю.*

Як правило, спостереження здійснюється на численних об'єктах, які складаються з певних одиниць. Кожна така одиниця спостереження називається *варіантою* і позначається $x_1, x_2, x_3 ... x_i$. Загальна кількість варіант статистичної сукупності має назву *обсяг вибірки* і позначається буквою N.

Значна частина статистичних досліджень пов'язана з описом великих сукупностей об'єктів. Якщо цікавлюча нас сукупність дуже чисельна або її елементи малодоступні, або існують інші причини, що не дозволяють вивчати відразу всі її елементи, то звертаються до вивчення якоїсь частини цієї сукупності. Така частина сукупності, вибрана для повного дослідження усіх елементів, називається *вибіркою або вибірковою сукупністю,* а уся множина досліджуваних елементів називається *генеральною сукупністю.* Бажано сформувати вибірку таким чином, щоб вона щонайкраще представляла б усю генеральну сукупність, тобто була б *репрезентативною.*

При виборі одного об'єкту з N вірогідність вибору кожного елементу рівна $\frac{1}{N}$. Вибір n об'єктів з N вважається чисто випадковим, якщо всі набори з n об'єктів мають однакову вірогідність бути вибраними. Чисто випадковий вибір n об'єктів випадкову вибірку об'єму n) можна отримати, витягуючи з генеральної сукупності по одному об'єкту послідовно і чисто випадково. Порушення принципів випадкового вибору може привести до непрезентативності вибірки.

Однією з найбільш поширених форм угруповань вибіркових даних служать статистичні таблиці. Статистичні таблиці мають ілюстративне значення, показуючи якісь загальні підсумки, положення окремих елементів в загальній серії спостережень тощо. Все це сприяє розумінню описуваних явищ, з'ясування того істотного, чим вони відрізняються один від одного.

ПРАКТИЧНА ЧАСТИНА

Завдання 1. Класифікація біологічних ознак.

1. На запропонованих об'єктах провести виділення мірних і рахункових біологічних ознак за категоріями (довжина стебла колоса, число зерен в колосі, ширина листа подорожника, забарвлення кошенят, забарвлення квітів, несучість курей тощо)

2. Провести систематизацію та групування отриманих даних

3. Оформити статистичні таблиці за зразком (таблиця 1.1).

Таблиця 1.1 – Результати вимірювань ознак біологічних об'єктів

№	Ознака			
	Висота рослини	Маса рослини	Кількість листів	Кількість суцвіть
1.				
2.				
3.				
….				
n				
Усереднене значення				

Завдання 2. Сортування даних за окремою ознакою

1. Створити базу даних в MS Excel, занісши отримані результати експериментальних даних у відповідні стовпці та рядки.

2. Відфільтрувати отримані дані по певній довжині, ширині жолудя, забарвленням раковини тощо.

3. У створеній базі вибрати модуль *Дані* та обрати вкладку *Фільтр* (рисунок 1.1).

4. Відсортувати дані по довжині, вибравши значення 3 см і більше (рисунок 1.2).

5. Аналогічно провести сортування даних за величиною діаметру, вибравши значення менше 1 см. Відсортувати дані за двома параметрами: довжина жолудя 2,5 см, діаметр 1,1 см.

6. Записати кількість варіант із заданими параметрами.

7. Зберегти складену базу в різних форматах (TXT, CSV), оскільки різні комп’ютерні програми аналізу даних вимагають для імпорту баз даних ці формати.

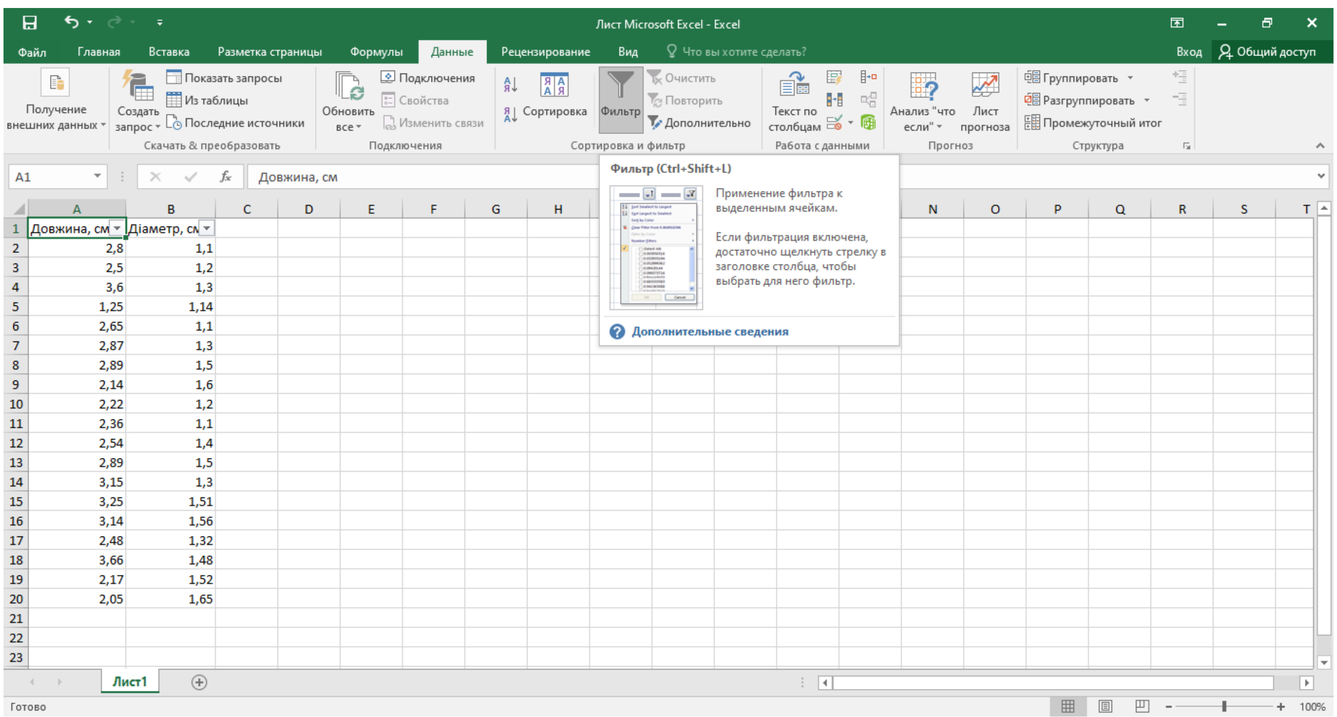


Рисунок 1.1 – Вибір вкладки «Фільтр» в модулі Дані

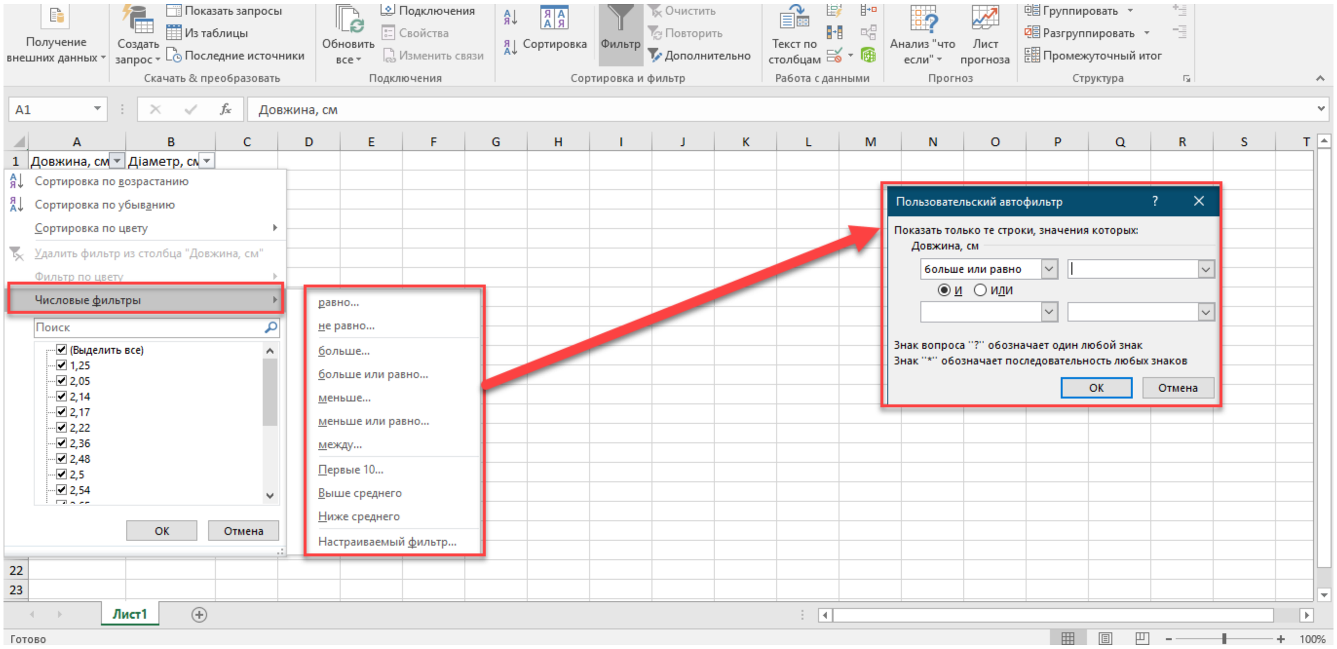


Рисунок 1.2 – Сортування даних по заданому показнику

Контрольні запитання

1. Біологічні ознаки, їх класифікація.
2. Помилки спостережень.
3. Поняття вибіркової сукупності і генеральної сукупності.
4. Статистична сукупність і її властивості.
5. Вимоги до вибірки.
6. Статистичні таблиці.

Лабораторна робота №2
ГРУПУВАННЯ РЕЗУЛЬТАТІВ СПОСТЕРЕЖЕННЯ

Мета роботи: навчитись систематизувати результати спостереження, виявляти властиві досліджуваним ознакам закономірності.

Матеріали і обладнання: листки деревних рослин або чагарників (3-4 види по 50 шт), насіння квіткових культур (наважки – 1 г), лінійки, калькулятори.

ТЕОРЕТИЧНА ЧАСТИНА

В результаті проведення спостережень над біологічними об'єктами отримують велику кількість фактичного матеріалу. Перш за все його потрібно впорядкувати з метою отримання необхідної інформації. *Спостереження* – це збір первинних даних про об'єкт, який підлягає вивченню. Кожне окреме спостереження називають *варіантою* або датою. *Групування* – це процес систематизації результатів спостережень для отримання закладеної в них інформації, а також виявлення закономірностей, які властиві досліджуваним ознакам.

Крім статистичних таблиць формою первинного угрупування вибіркових даних є спосіб ранжування. Однією з форм групування є побудова *статистичних рядів* – рядів числових значень ознаки розміщених в певному порядку. Для обчислення статистичних показників числових сукупностей досить часто використовують ряд розподілу (*варіаційний ряд*) – подвійний ряд чисел, що складається з класових варіант і відповідних їм чисельностей, розмішених у порядку зростання або спадання. Для побудови ряду розподілу потрібно згрупувати всі варіанти у наперед встановлені інтервали (класи).

Статистичні розподіли можна класифікувати за ознакою типів вимірювань змінної на: варіаційні, ранжировані та атрибутивні.

Варіаційні розподіли базуються на даних, які виміряні за шкалою відношень або інтервалів. Ранжировані розподіли застосовуються у разі порядкових (рангових) типів вимірювання. Атрибутивні розподіли характеризують дані, які виміряні за номінальними шкалами або шкалами «найменувань». Основні види статистичних розподілів такі: диференціальні та інтегральні, які можуть складатися з абсолютних і відносних частот

За технологією побудови варіаційні розподіли поділяють на розподіли не згрупованих і згрупованих варіант. З метою лаконічності домовимося їх називати не згрупованими і згрупованими розподілами. Для незгрупованих розподілів частоти мають відношення до безпосередніх значень варіант з варіативного ряду; для згрупованих розподілів – до груп (або інтервалів) значень варіант.

Число класових інтервалів можна визначити двома способами:

а) за таблицею, де число інтервалів (класів), залежить від числа вибірки (таблиця 2.1):

Таблиця 2.1 – Число інтервалів варіаційного ряду за П.Ф. Рокицькому

Кількість спостережень	25–40	41–60	61–100	101–200	≥201
Кількість інтервалів	5-6	6-8	7-10	8-12	9-15

б) за формулою, де число інтервалів можна розрахувати логарифмічно – за формулою Г.А. Стерджеса (1926):

$$\lambda = \frac{x_{\max} - x_{\min}}{1 + 3.32 \ln n}, \qquad (2.1)$$

За початок першого часткового інтервалу, як правило, вибирається точка $x_0 = x_{\min} - \dfrac{h}{2}$.

Якщо $\lambda = 1$, то зібраний матеріал розподіляється в безінтервальний варіаційний ряд; якщо $\lambda \neq 1$, то данні необхідно розподіляти в інтервальний варіаційний ряд. Точність величини класового інтервалу (λ) повинна відповідати точності, прийнятої для вимірювання ознаки.

При побудові інтервального варіаційного ряду необхідно, щоб мінімальна варіанта сукупності ($x_{\min}$) попадала всередину першого класового ряду, тобто нижня межа варіаційного ряду:

$$x_н = \frac{x_{\min} \cdot \lambda}{2}, \qquad (2.2)$$

де $x_н$ – нижня межа першого класового інтервалу;

$x_{\min}$ – мінімальна варіанта сукупності;

λ – величина класового інтервалу.

Верхню межу першого класу одержують додаванням до нижньої межі величини:

$$x_в = x_н + \lambda. \qquad (2.3)$$

Потім поступовим додаванням λ визначають верхню межу другого класу, третього і т. д. до тих пір, поки не одержимо інтервал (межу), в який попадає максимальна варіанта сукупності.

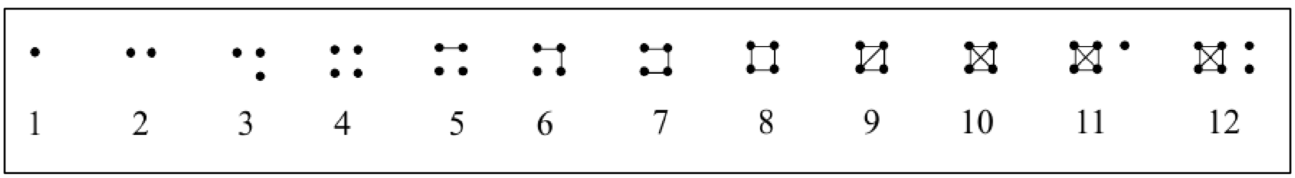
Визначивши класові інтервали, залишається розподілити по ним всі варіанти сукупності, тобто визначити частоти кожного класу. Розміщування класів найчастіше проводять у такій послідовності: варіанти, які збігаються з верхньою межею одного класу і нижньою межею другого класу, розташовують у наступному класі, а верхню межу класів зменшують на величину, що дорівнює точності виміру ознаки:

$$x_к = x_в - \varepsilon = (x_н + \lambda) - \varepsilon. \qquad (2.4)$$

де ε – похибка вимірювання.

Для точного визначення частот використовують допоміжну таблицю, в першу графу якої вносять класи, в другу – величину інтервалів, в третю –частоти f_i за допомогою допоміжних знаків (шифр частоти – таблиця 2.2), в четвертій графі частота представлена в цифрах (таблиця 2.3).

Таблиця 2.3 – Шифр частот



Таблиця 2.4 – Форма таблиці інтервального ряду

Клас k	Класові інтервали	Шифр частоти	Частота f_1
1k	3,12-3,30	•	1
2k	3,31-5,59	•	1
3k	5,60-8,68	•••	3

Перетворення виконується шляхом заміни класових інтервалів на їх центральні, тобто серединні варіанти. Серединні (центральні) значення класових інтервалів можна точно визначити по формулі:

$$x_i = x_н + x_k, \qquad (2.5)$$

де $x_н$ і x_k – нижня і кінцева межа класового інтервалу;

$x_k = x_н + 1 - \varepsilon$ – кінцева межа класового інтервалу – це початок наступного класу зменшеного на величину точності вимірювань.

Дані безінтервального варіаційного ряду формуються таблицею 1.2.

При вивченні варіаційних рядів використовують також поняття *накопиченої частоти* $m_i^{нак}$. Накопичена частота показує, скільки спостерігається варіантів із значенням ознаки, меншої за x. Відношення накопиченої частоти до загальної кількості спостережень називають накопиченою відносною частотою $p_i^{нак}$. Накопичені частоти для кожного інтервалу знаходять послідовним сумуванням частот всіх попередніх інтервалів, включаючи даний.

Кількість одиниць спостереження в одному класі чи інтервалі називають *чисельностями*, а значення чисельностей, виражене у відсотках – частковостями. Загальне число спостережень складає об'єм ряду (об'єм вибірки).

Для наочного виявлення закономірностей розподілу кількості варіант стосовно класів, числові ряди зображають графічно.

Найбільш вживаним способом графічного зображення числових рядів є гістограма, кумулята і огіва.

Числові показники, які описують мінливість ознак називають показниками варіацій. До них відносяться ліміти, розмах варіації, дисперсія, стандартне відхилення, коефіцієнт варіації (С).

Ліміти – найменша (x_{min}) і найбільша (x_{max}) варіанти.

Розмах варіації (R) – різниця між найбільшою та найменшою варіантами, яка характеризує мінливість ознаки.

Варіаційна крива – це графічне зображення варіаційного у вигляді лінійної кривої, ординати якої пропорційні частотам варіаційного ряду. Статистичний ряд може бути представлений у вигляді полігону, гістограми і кумуляти.

Полігоном частот або відносних частот називають ламану лінію, яка з'єднує точки дискретного ряду (x_i, m_i) або (x_i, p_i), відповідно (рисунок 2.1.а).

Для переривчастих варіаційних рядів графічне зображення зручніше давати у вигляді гістограми (рисунок 2.1, б).

Гістограма – це графічне зображення у вигляді стовпчастої діаграми результатів спостереження варіювання якоїсь кількісної ознаки.

Кумулятою називається крива накопичених частот (рисунок 2.1.в), яка має вигляд ламаної лінії, що з'єднує точки (x_i, $m_i^{нак}$) або (x_i, $p_i^{нак}$).

У залежності від характеру розподілу варіант варіаційні криві можуть бути: нормальні, біноміальні, асиметричні, пуасонівські, ексценсивні, плосковершинні, двовершинні, трансгресивні.

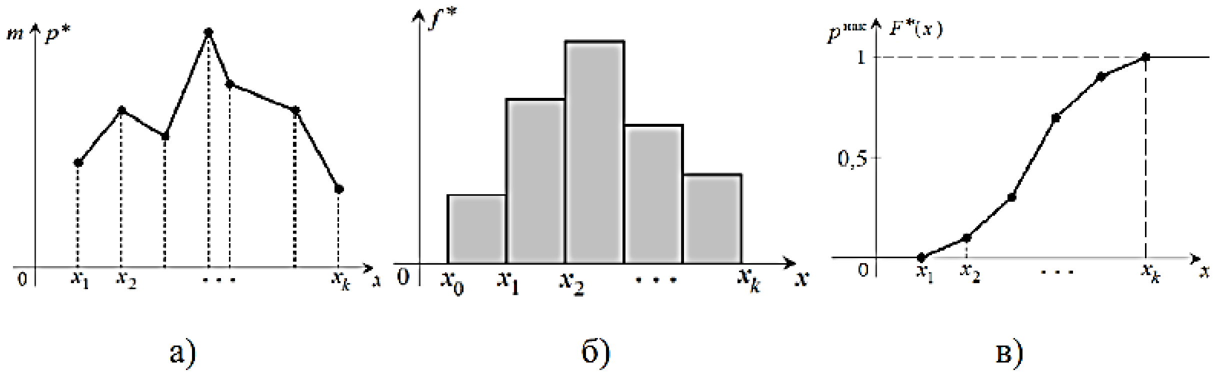


Рисунок 2.1 – Графічний вигляд статистичного ряду: полігон відносних частот (а); гістограма відносних частот (б); кумулята відносних частот (в)

Приклад розв'язання типового завдання

Завдання 1. *Побудувати варіаційний ряд за результатами вимірювань довжини жолудів: 19, 21, 19, 18, 27, 31, 29, 20, 21, 22, 23, 24, 31, 29, 25, 26, 27, 28, 29, 30, 19, 22, 25, 26, 19, 25, 22, 23, 18, 31, 19, 21, 19, 18, 27, 31, 29, 20, 21, 22, 23, 24, 31, 29, 25, 26, 27, 28, 29, 30, 19, 22, 25 мм. Охарактеризуйте як варіює досліджувана ознака. Визначити ліміти варіації, розмах варіації, величину класового інтервалу. На основі результатів розподілу побудувати гістограму.*

Розв'язання

1. Будуємо ранжирований варіаційний ряд:
18, 18, 18, 19, 19, 19, 19, 19, 19, 19, 20, 20, 21, 21, 21, 21, 22, 22, 22, 22, 22, 23, 23, 23, 24, 24, 25, 25, 25, 25, 25, 26, 26, 26, 27, 27, 27, 28, 28, 29, 29, 29, 29, 29, 30, 30, 31, 31, 31, 31, 31.

Відповідно, $x_{\max} = 31, x_{\min} = 18$, $N = 53$.

2. Встановлюємо значення класового інтервалу користуючись формулою 2.1:

$$\lambda = \frac{x_{\max} - x_{\min}}{1 + 3.32 \ln n} = \frac{31 - 18}{1 + 3{,}32 \ln 53} = 0{,}9167$$

У даному випадку $\lambda \neq 1$, тому данні необхідно розподіляти в інтервальний варіаційний ряд.

3. Визначаємо межі варіаційних рядів:

Для нашого прикладу $x_{\min} = 18$ і $\lambda = 0,92$, отже нижня межа першого класу буде:

$$x_{н} = 18 - \frac{0,92}{2} = 17,54$$

Шляхом додавання до $x_{н}$ $\lambda = 0,92$ значення визначимо верхню межу першого класу:

$$x_{в_1} = 17,54 + 0,92 = 18,46$$

Тепер визначимо верхню межу другого класу:

$$x_{в_2} = 18,46 + 0,92 = 19,38$$

І так далі до тих пір, поки не одержимо інтервал, в який попадає максимальна варіанта сукупності $x_{\max} = 31,8$. Таким чином:

Верхня межа І-го класу	17,54+0,92=18,46
ІІ-го	18,46+0,92=19,38
ІІІ-го	19,38+0,92=20,30
IV-го	20,30+0,92=21,22
V-го	21,22+0,92=22,14
VI-го	22,14+0,92=23,06
VII-го	23,06+0,92=23,98
VIII-го	23,98+0,92=24,90
IX-го	24,90+0,92=25,82
X-го	25,82+0,92=26,74
XI-го	26,74+0,92=27,66
XII-го	27,66+0,92=28,58
XIII-го	28,58+0,92=29,50
XIV-го	29,50+0,92=30,42

3. Використовуючи формулу (2.4) знайдемо межі класів:

$$1k = 18{,}46 - 0{,}01 = 18{,}45$$

Тобто 1k межа 17,54-18,45. І так повторюється для кожного класу.

4. Перетворимо інтервальний варіаційний ряд в безінтервальний. Визначаємо серединні (центральні) значення варіант для кожного класу, використовуючи формулу (2.5) або $x_i = x_н + \frac{\lambda}{2}$. Тоді отримуємо:

$$x_1 = 17{,}54 + \frac{0{,}92}{2} = 18$$

$$x_2 = 18{,}46 + \frac{0{,}92}{2} = 18{,}91$$

$$x_3 = 19{,}38 + \frac{0{,}92}{2} = 19{,}83$$

$$x_4 = 20{,}30 + \frac{0{,}92}{2} = 20{,}75$$

……

$$x_n = 29{,}50 + \frac{0{,}92}{2} = 29{,}96$$

Завдання 2. *Провести розрахунок числових характеристик вибірки та побудувавати графіки розподілу із використання функцій та пакету «Аналіз даних» MS Excel.*

1. Підготовка до роботи із MS Excel. Перед початком виконання роботи встановити «Пакет анализа данных». Для цього виконати наступні дії: «Файл» → «Параметры» → «Надстройки» → «Пакет анализа данных» і натиснути кнопку «Перейти» (рисунок 2.2). Обрати «Пакет анализа» і натиснути кнопку «ОК». Створити файл MS Excel.

2. Заповнити дані відповідно до номеру варіанта.

Сформувати таблицю «Вхідні дані». У даному прикладі 10 стовпців на 10 рядків. Сформувати з отриманих даних таблицю даних та скласти варіаційний ряд. Приклад виконання зображено на рисунку 2.2.

3. Визначення розмаху варіаційного ряду.

Для того аби знайти мінімальне x_{min} та максимальне x_{max} значення вибірки необхідно скористуватись вбудованою функцією «МИН» та «МАКС», відповідно (рисунку 2.3). Для визначення x_{max} комірці D14 присвоїти значення «МАКС(D4:M13)», а для визначення x_{min} комірці D15 присвоїти значення «МИН(D4:M13)». (D4:M13) – масив варіаційного ряду.

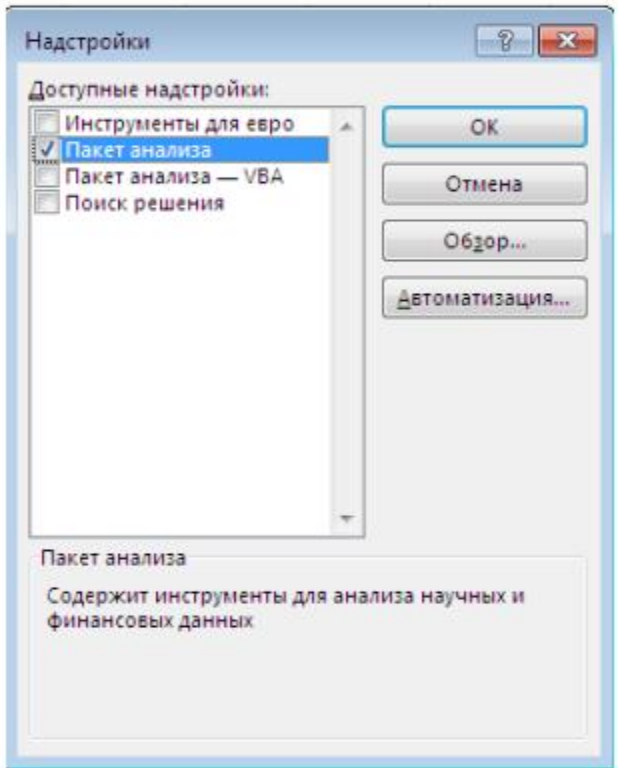


Рисунок 2.2 – Вікно налаштування «Пакета анализа»

4. Визначення часткових інтервалів.

Розрахунок кількість часткових інтервалів за проводимо відповідно до формули Стерджеса (2.1), яка в MS Excel задається наступним чином: «=ОКРУГЛВВЕРХ((1+3,32*LOG10(СЧЁТ(D4:M13)));0)», де функція «СЧЁТ(D4:M13)» визначає об’єм вибірки, а «0» – вказує на число знаків після коми. Довжину інтервалу знаходимо за формулою (2.1), яка в MS Excel задається наступним чином: «=ОКРУГЛВВЕРХ(((D14-D15)/G14);0)».

Початок інтервалу знаходимо вказавши наступну команду в MS Excel: «=D15-(G15/2)». Якщо кількість інтервалів була цілою, то $x_0 = x_{min}$.

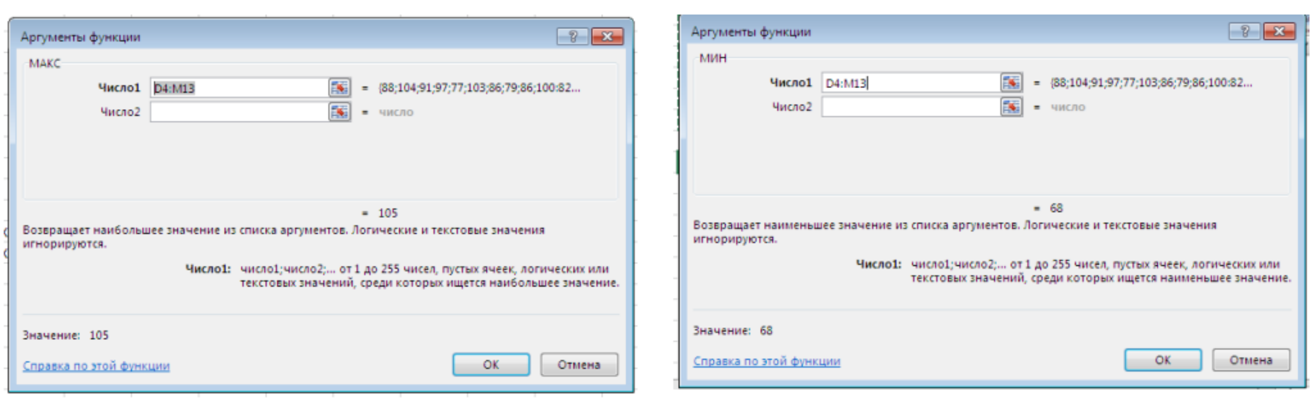
5. Відповідно до отриманих даних складаємо таблицю варіаційного ряду та відносних частот (рисунок 2.4)

6. Заповнюємо значення k від 1 до 10.

7. Заповнюємо значення інтервалів (x_{i-1};x_i). Для цього виконуємо наступні дії: комірці D18 присвоїти значення J14 (x_0); комірці E18 присвоїти значення D18+G15 (h), скопіювати його та поширити на наступні комірки (E19:E27); D19 присвоїти значення E18, скопіювати та поширити на наступні комірки (D20:B27).

Вхідні дані:														
	C	D	E											
4		88	104	91	97	77	103	86	79	86	100			
5		82	68	71	87	89	89	81	81	70	79			
6		84	91	87	83	90	69	83	96	79	94			
7		93	86	81	83	84	92	93	85	84	88			
8		77	85	93	85	87	100	76	79	90	91			
9		84	74	76	75	93	103	80	96	72	95			
10		81	102	75	80	90	85	82	77	94	102			
11		87	95	99	83	80	93	90	79	93	105			
12		95	85	84	90	93	95	98	88	79	91			
13		86	88	93	80	88	88	90	68	89	90			
14	Xmax	105		k=	10		x0=	66						
15	Xmin	68		h=	4		n=	100						
16														
17	*k*	*Інтервал (x(i-1);xi]*		*Середина інтервалу*	*Частота mi*	*Відносна частота p**	*f**	*F**		**Завдання 4.4.**Числові характеристики —				
18	1	66	70	68	4	0,04	0,01	0,04		Середнє значення, $X_в$			86,74	
19	2	70	74	72	3	0,03	0,0075	0,07		Дисперсія, D			70,37616	
20	3	74	78	76	7	0,07	0,0175	0,14		Середнє відхилення, б			8,38905	
21	4	78	82	80	16	0,16	0,04	0,3		Коефіцієнт асиметрії, A			-0,05492	
22	5	82	86	84	18	0,18	0,045	0,48		Ексцес, E			-0,27417	
23	6	86	90	88	20	0,2	0,05	0,68		Мода, Mo			93	
24	7	90	94	92	15	0,15	0,0375	0,83		Медіана, Me			87	
25	8	94	98	96	8	0,08	0,02	0,91						
26	9	98	102	100	5	0,05	0,0125	0,96						
27	10	102	106	104	4	0,04	0,01	1						
	сума				100	1								

Рисунок 2.2 – Приклад виконання типового завдання з побудови варіаційного ряду



а) б)

Рисунок 2.3 – Вікно функцій: «МАКС» (а) та «МИН» (б)

k	Інтервал, $(x_{(i-1)};x_i]$	Середина інтервалу, $\tilde{x}_i$	Частота, m_i	Відносна частота, p_i*	f*	F*

Рисунок 2.4 – Інтервальний варіаційний ряд

8. Для знаходження значення середини інтервалу комірці F18 присвоїти значення «=(E18+D18)/2», скопіювати та поширити на наступні комірки (F19: F27).

9. Для знаходження частоти необхідно виділити відповідний масив розміщення частот (G18:G27) та за допомогою «Майстра функцій» задати функцію «ЧАСТОТА()» (рис.2.5). У полі «Массив_данных» задати дані вибірки (D4:M13), в у полі «Массив_интервалов» – масив визначених інтервалів (D18:E27). Натиснути разом клавіші CTRL+SHIFT+ENTER.

10. Для знаходження відносної частоти необхіно комірці H18 присвоїти значення «=G18/J15», де G18 – частота, J15 – об'єм вибірки.

11. Визначення частоти накопичення для побудови емпіричного графік проводять наступним чином: комірці J18 присвоїти значення відносної частоти H18. Комірці J19 присвоїти значення «=J18+ H19». Скопіювати J19 та вставити до кінця ряду (J18: J27). Зробити перевірку. Знайти суму частот та відносних частот.

12. Побудувати гістограму. Для цього виконати наступну послідовність: «Вставка» → «Диаграммы» → «Гистограма» → «Выбрать данные» (рис.2.6.а). Натиснути кнопку «Добавить» і у вікні «Изменение ряда» в полі «Имя ряда» напистати «Гістограма», в полі «Значения – вставити ряд із значеннями f * (I18:I27) (2.7.а). Натиснути кнопку «ОК» і у вікні «Выбору источника данных» натиснути кнопку «Изменить» та ввести значення ряду середини інтервалу (F18:F27) (2.7.б). Перейти в «Экспресс-макет» і обрати «Макет 8».

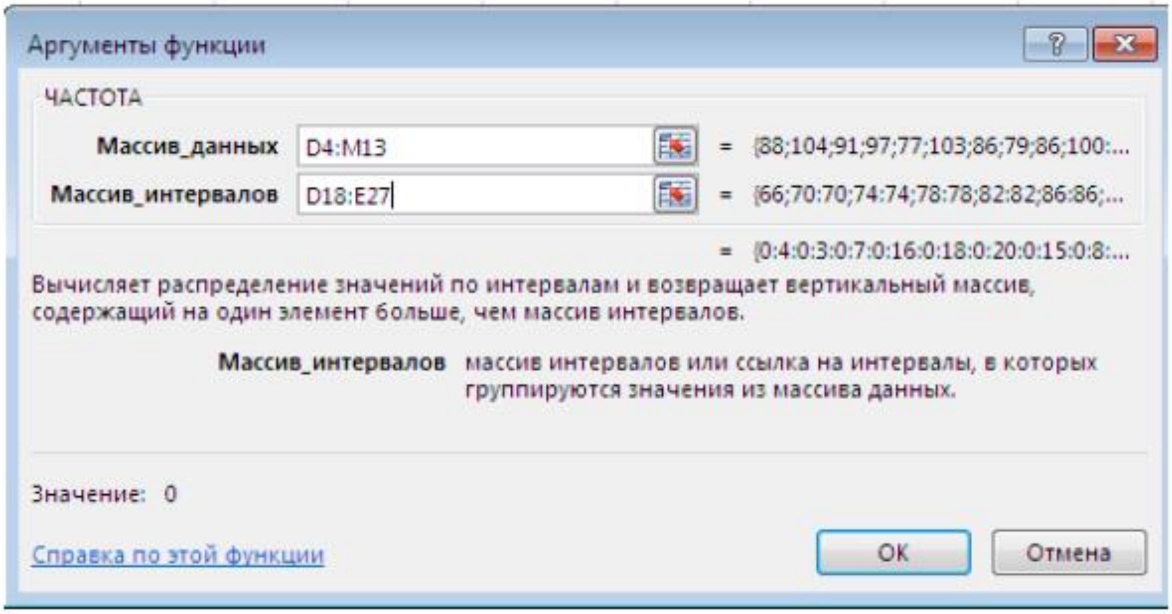
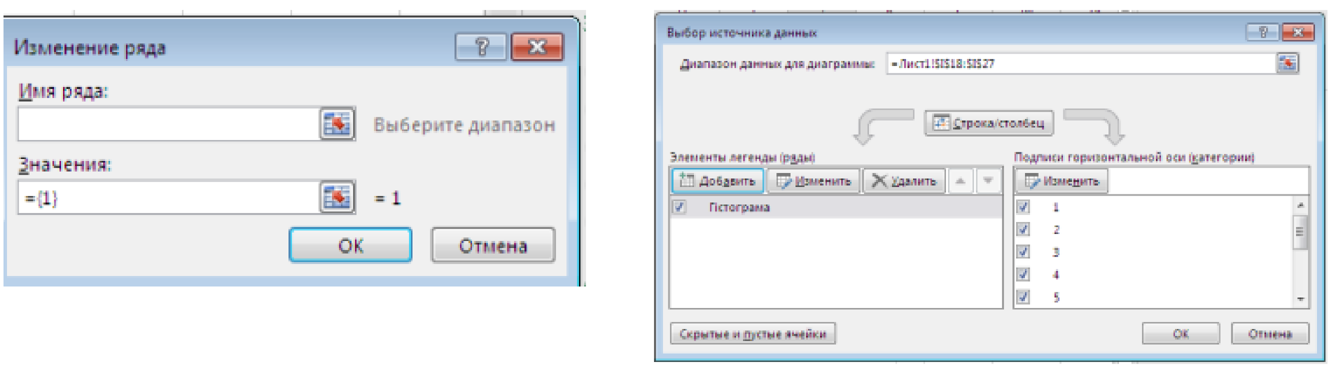


Рисунок 2.5 – Вікно функції «ЧАСТОТА»



а) б)

Рисунок 2.7 – Вікна вибору даних для побудови гістограми

ПРАКТИЧНА ЧАСТИНА

Ознайомитись з теоретичною частиною роботи і виконати наступні завдання:

Завдання 1. Побудова варіаційного ряду.

Визначали висоту рослин травостою на лучному газоні через тиждень після скошування. Отримали наступні результати: 22; 23; 22; 22; 17; 23; 20; 20; 21; 25; 27; 24; 22; 21; 16; 23; 18; 21; 24; 18; 21; 22; 25; 23; 21; 20; 25; 18; 21; 21; 24; 25; 19; 18; 22; 25; 27; 19; 17; 18; 22; 23; 24; 19; 26; 21; 25; 25; 23; 27. Розподіліть отриману сукупність в ранжирований варіаційний ряд. Охарактеризуйте як варіює досліджувана ознака. Визначіть ліміти варіації, розмах варіації, величину класового інтервалу.

Завдання 2. Побудова варіаційного інтервального ряду розподілу та побудова кривих

Визначіть розмах розподілу, відносну частоту (%), і накопичення частот (%). На основі заповненої таблиці намалюйте гістограму. Зробіть висновки. Для виконання завдання необхідно згенерувати вхідні дані. Для цього виконати наступні дії: натиснути «Данные» → «Анализ нанных» → «Генерация случайных чисел» (рисунок 2.8).

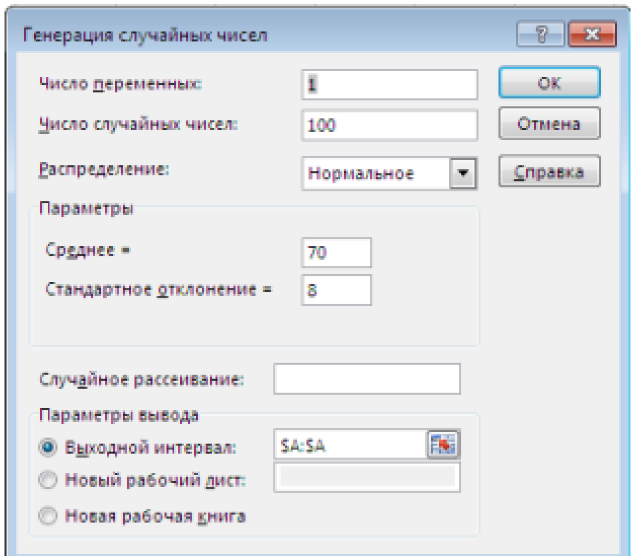


Рисунок 2.8 – Вікно «Генерация случайных чисел»

Контрольні запитання

1. Що вивчає біометрія, коли і де її застосовують?
2. Що таке ознака, які ознаки бувають? Що таке варіювання? Дайте визначення варіанти.
3. Поняття про вибіркову та генеральну сукупність. Які вибірки називають великими і малими?
4. Що таке варіаційний ряд і які його основні особливості?
5. Як розраховують величину класового проміжку?
6. Від чого залежить кількість класів у варіаційному ряді?
7. Який порядок визначення меж першого класу варіаційного ряду?
8. Який клас у варіаційному ряді називають модальним і чим він характеризується?
9. Що таке варіаційна крива і порядок її побудови.
10. В яких випадках будують варіаційну криву, в яких гістограму? Які бувають типи варіаційних кривих у залежності від характеру розподілу?

Лабораторна робота №3
ВИЗНАЧЕННЯ СЕРЕДНЬОЇ АРИФМЕТИЧНОЇ І БІОМЕТРИЧНИХ ПОКАЗНИКІВ РЯДУ РОЗПОДІЛУ. ГРАФІКИ РОЗПОДІЛУ

Мета роботи: навчитись визначати середні величини і біометричні показники ряду розподілу.

ТЕОРЕТИЧНА ЧАСТИНА

Однією із найважливіших характеристик варіюючих ознак експериментів є їхня середня величина. Середній показник, нівелюючи значення кожної окремо взятої ознаки, дозволяє визначити середню величину вибіркової сукупності, за якою можна порівнювати окремі піддослідні групи тварин, стада та популяції.

Аналіз вибіркової сукупності починається з групування даних, під яким розуміють процес систематизації, або розташування виражених кількостями даних з метою одержання вміщеної в них інформації, встановлення закономірності, якій підпорядковується ознака, що вивчають, або явище.

При роботі з кількісними ознаками і при великому числі варіант групування виконують за класами варіаційного ряду.

Варіаційний (ранжирований) ряд – це упорядковане розміщення варіант у сукупності відповідно до наростання або спадання їх чисельних значень.

Середні величини – це абстрактні, цілі або дробові кількості, які кількісно характеризують ознаки тієї чи іншої сукупності.

Основна мета обчислення середніх значень у селекційній роботі – це характеристика селекційних груп, стад, ліній, родин за рівнем продуктивності або іншим біологічним показником. Порівнюючи середні двох послідовних поколінь, селекціонер може визначити фактично одержаний генетичний прогрес. Важлива роль відводиться середнім величинам при оцінці плідників за якістю потомства.

Існує декілька видів середніх показників. До них належать: середня арифметична, середня зважена, середня геометрична, середня квадратична, середня гармонічна, середня кубічна, мода і медіана.

Середні величини в математичній статистиці належать до класу степеневих середніх, які описує загальна формула:

$$\overline{X} = \sqrt[k]{\frac{\sum x^k}{n}} , \qquad (3.1)$$

де $\overline{X}$ – середня величина варіант;

n – кількість варіант (спостережень);

k – показник ступеню середньої, це величина, по якій визначають вигляд середньої величини.

Якщо $k=1$ – середня арифметична;

$k=0$ – середня геометрична;

$k=-1$– середня гармонійна;

$k=2$– середня квадратична;

$k=3$– середня кубічна.

Із ступеневих середніх найчастіше використовують середню арифметичну, рідше – середню гармонійну та середню геометричну, а середню квадратичну використовують для розрахунків показників варіації.

Крім степеневих, в біології використовують структурні середні – медіану, моду і т.д. Середні величини можуть характеризувати тільки однорідну сукупність варіант. Для розрахунків степеневих середніх використовують наступні формули – середнє арифметичне:

Для обґрунтування вибору форми та виду середньої, насамперед, необхідно визначити характер вихідного матеріалу, тобто: що буде ознакою (усереднюванню величиною) x_i, а що (частотою) f_i При визначенні середньої величини за формулами арифметичної та гармонійної використовується тільки необхідна вихідна інформація.

Середня арифметична, позначається буквою M або X – це одна з основних характеристик вибіркової сукупності, яку найчастіше визначають у біологічних дослідженнях. Вона має кілька властивостей. Перша – займає середнє місце і знаходиться у центрі розподілу варіантів. Друга властивість полягає у тому, що середній показник може біти величиною абстрактною, тобто приймати значення, яке не існує у природі, наприклад, багатоплідність – 8,2 поросяти. Головна властивість стверджує, що сума відхилення всіх варіант від середньої арифметичної завжди дорівнює нулю.

Середня арифметична є величина поіменована і виражається у тих самих одиницях, що й варіанти.

Найбільш розповсюдженим видом середніх величин є ***середня арифметична,*** котра, як і всі середні, у залежності від характеру наявних даних, буває *двох видів*:

– ***проста:***

$$\bar{x} = \frac{x_1 + x_2 + ... + x_i + ... + x_n}{n} = \frac{\sum_{i=1}^{n} x_i}{n}, \qquad (3.2)$$

де x_i – значення ознаки (варіанти); n – кількість варіантів.

– ***зважена***

$$\bar{x} = \frac{x_1 f_1 + x_2 f_2 + ... + x_n f_n}{f_1 + f_2 + ... + f_n} = \frac{\sum_{x=1}^{n} x_i \cdot f_i}{\sum_{x=1}^{n} f_i} \qquad (3.3)$$

Якщо замість частот у процесі обчислення середньої за даними ряду розподілу використовуються частки $w = \frac{f_i}{\sum_{i=1}^{n} f_i}$, то формула середньої арифметичної зваженої набуває вигляду:

$$\bar{x} = \frac{\sum_{x=1}^{n} x_i \cdot f_i}{\sum_{x=1}^{n} f_i} = \sum_{x=1}^{n} x_i \cdot \frac{f_i}{\sum_{x=1}^{n} f_i} = \sum_{x=1}^{n} x_i \cdot w_i \qquad (3.4)$$

Досить часто розрахунок середньої здійснюється за даними не тільки інтервальних рядів розподілу, коли варіанти ознаки подаються у вигляді інтервалу (від... до). Тому для обчислення середньої спочатку потрібно перетворити інтервальний ряд на дискретний, для чого в кожній групі визначають середнє значення інтервалу.

Середнє значення інтервалу знаходять як напівсуму його верхньої та нижньої межі. Середнє значення відкритого інтервалу визначають з величини інтервалу наступної групи або попередньої. Після визначення середнього значення інтервалів, подальші розрахунки здійснюють так само, як і дискретному варіаційному ряду: варіанти перемножуються на частоти і суму добутків ділять на суму частот.

Середня гармонійна зважена застосовується в тих випадках, коли відомі індивідуальні значення ознаки та їх добутки на частоти, значення ж частот невідомі. Вид середньої гармонійної зваженої наступний:

Як і середня арифметична, середня гармонійна може бути проста і зважена.

Формула середньої гармонійної простої має такий вигляд:

$$\bar{x}_{гар.} = \frac{n}{\frac{1}{x_1} + \frac{1}{x_2} + ... + \frac{1}{x_n}} = \frac{n}{\sum \frac{1}{x}}, \tag{3.5}$$

де $\sum \frac{1}{x_i}$ – сума обернених значень варіант; n – число варіант.

Таким чином, ***середня гармонійна*** – це величина, обернена середній арифметичній із обернених значень ознаки. Вона використовується в тому випадку, коли обсяг явищ, тобто добутки за кожною ознакою рівні.

Середня гармонійна зважена має вигляд:

$$\bar{x}_{гар.} = \frac{\sum W}{\sum \frac{1}{x} \cdot W}, \tag{3.6}$$

оскільки, $W = x \cdot f$, *то* $f = \frac{W}{x} = \frac{1}{x} \cdot W$

де W – добуток варіант на частоти;

$\frac{1}{x}$ – обернені значення варіантів.

Вона використовується в тому випадку, коли обсяг явищ, тобто добутки ознаки на частоту за кожною ознакою нерівні.

Для аналізу рядів динаміки, зокрема, при визначенні середнього темпу росту, застосовують ***середню геометричну.*** При визначенні середнього темпу росту динамічного ряду з рівними проміжками часу використовують формулу середньої геометричної простої:

$$\bar{T} = \sqrt[n]{K_1 \cdot K_2 ... K_n} \cdot 100, \tag{3.7}$$

де T – змінні коефіцієнти росту;

n – число змінних темпів росту.

Для розрахунку можна використовувати абсолютні рівні ряду:

$$\overline{T} = \sqrt[n-1]{\frac{y_1}{y_0} \cdot \frac{y_2}{y_1} \cdot \frac{y_3}{y_2} \ldots \frac{y_n}{y_{n-1}}} \cdot 100 = \sqrt[n-1]{\frac{y_n}{y_0}} \cdot 100, \qquad (3.8)$$

де y_0 та y_n – відповідно початковий та останній рівні ряду.

Темп приросту $T_{пр.}$ показує на скільки % порівнюваний рівень більший або менший від рівня, взятого за базу порівняння, тобто:

$$T_{пр.} = \overline{T} - 100{,}0 \qquad (3.9)$$

Для розрахунку середнього рівня моментного ряду використовують середню хронологічну. ***Середньою хронологічною*** називають середню, обчислену із значень, які змінюються в часі. Розрізняють середню хронологічну просту і зважену.

Середня хронологічна проста застосовується для визначення середнього значення моментного ряду з рівними інтервалами часу і має вигляд:

$$\bar{x} = \frac{\frac{x_1}{2} + x_2 + \ldots + x_{n-1} + \frac{x_n}{2}}{n-1} \qquad (3.10)$$

Середня хронологічна зважена застосовується для визначення середнього значення моментного ряду з нерівними інтервалами часу й обчислюється за формулою:

$$\bar{x} = \frac{(x_1 + x_2)t_1 + (x_2 + x_3)t_2 + \ldots + (x_{n-1} + x_n)t_{n-1}}{2(t_1 + t_2 + \ldots + t_{n-1})}, \qquad (3.11)$$

де $x_1 \ldots\ldots x_n$ – рівні ряду динаміки;

t – тривалість терміну часу між рівнями.

В інтервальних рядах розподілу з рівними інтервалами частоти і частки дають уявлення про ступінь заповнення їх одиницями сукупності та можливість їх порівнювати. В інтервальних рядах з нерівними інтервалами такого порівняння здійснювати не можна, тому визначають показник, який називається *щільністю інтервалів або розподілу.*

Щільність розподілу (D) показує, скільки одиниць сукупності або який їх відсоток припадає в середньому на одиницю величини інтервалу групувальної ознаки.

Вона визначається діленням частоти або части (f_i або w_i) відповідного інтервалу на його величину (h_i). Розрізняють:

– абсолютну щільність:

$$D_a = \frac{f_i}{h_i} \tag{3.12}$$

– відносну щільність:

$$D_v = \frac{w_i}{h_i} \tag{3.13}$$

Для статистичного вивчення рядів розподілу використовують наступні характеристики центру розподілу:

- середня;
- мода;
- медіана.

На відміну від середньої для більш глибинного розкриття властивостей ряду розподілу обчислюють особливі показники, які у статистиці називають структурними середніми величинами. Це *мода й медіана*, які є допоміжними описовими характеристиками розподілу варіаційної ознаки.

Мода (Мо) – це величина ознаки (варіанта), яка найчастіше зустрічається у даній сукупності. У варіаційному ряду це – варіант, що має найбільшу частоту.

Медіаною (Ме) в статистиці називають варіант, що знаходиться в середині упорядкованого (ранжованого) варіаційного ряду, тобто ділить його на дві рівні частини: *одна* частина має значення варіюючої ознаки менше між середня, а *друга* – більше. Медіана показує величину варіюючої ознаки, якої досягла половина одиниць сукупності.

На відміну від дискретних в інтервальних варіаційних рядах визначення моди і медіани має певні особливості. Мода визначається в наступній послідовності: спочатку необхідно визначити модальний інтервал, якому відповідає найбільша частота або частка, а значення моди в інтервальному ряду обчислюється за ***формулою Орженцького:***

$$Mo = x_{Mo} + h_{Mo} \cdot \frac{f_{Mo} - f_{Mo-1}}{(f_{Mo} - f_{Mo-1}) + (f_{Mo} - f_{Mo+1})}, \tag{3.14}$$

де x_{Mo} – нижня межа модального інтервалу;

h_{Mo} – величина модального інтервалу;

f_{Mo} – частота модального інтервалу;

f_{Mo-1} – частота інтервалу, що передує модальному;

f_{Mo+1} – частота інтервалу, що йде за модальним.

Для визначення медіани необхідно знайти медіанний інтервал, кумулятивна частота якого дорівнює або перевищує половину суми частот. Кумулятивна частота визначається шляхом поступового додавання частот, починаючи з інтервалу, який має найменше значення ознаки.

Значення медіани обчислюють за ***формулою Фехнера***:

$$Me = x_{Me} + h_{Me} \frac{\sum \frac{1}{2} f_i - S_{Me-1}}{f_{Me}}, \tag{3.15}$$

де x_{Me} – нижня межа медіанного інтервалу;

h_{Me} – величина медіанного інтервалу;

$\sum_{i=1}^{n} f_i$ – сума частот ряду;

S_{Me-1} – сума накопичених частот передмедіанного інтервалу;

f_{Me} – частота медіанного інтервалу.

Для будь-якої сукупності біологічних об'єктів характерною є наявність різноманітності між її членами за різними ознаками. Навіть при однакових середніх арифметичних двох або кількох вибірок у них можна встановити суттєві відмінності в характері варіювання. Найбільш поширеними показниками, які характеризують ступінь варіювання або мінливості ознак в сукупності є ліміт (lim), середнє квадратичне відхилення (σ) та коефіцієнт варіації (Cν).

Ліміт – це найбільш простий показник мінливості ознаки і характеризує мінімальне і максимальне значення досліджуваної ознаки у вибірковій сукупності.

$$lim = x_{max} - x_{min} \tag{3.16}$$

Середнє квадратичне відхилення (σ) – це основний найпоширеніший показник мінливості, який показує на скільки в середньому відхиляється кожна варіанта від середньої арифметичної даної вибірки. Середнє квадратичне відхилення – величина завжди позитивна, пойменована і виражається в тих же одиницях, в яких виміряна ознака.

За числовим значенням σ відносно $\overline{X}$ можна робити висновки про ступінь мінливості ознаки. Чим більша σ, тим більша мінливість і навпаки.

У багаточисельній вибірковій сукупності при розрахунку середнього квадратичного відхилення користуються даними з таблиці варіаційного ряду, і розрахунок здійснюється за формулою:

$$\sigma = k \cdot \sqrt{\frac{\sum f(a)^2}{n} - b^2}\ , \tag{3.17}$$

де k – величина класового проміжку;

f – частоти;

a – відхилення класів від умовної середньої;

n – об'єм вибірки;

b – поправка до умовної середньої ($b = \frac{\sum fa}{n}$)

$\sum$ – знак суми.

Коефіцієнт варіації (Cν), як і середнє квадратичне відхилення, також характеризує ступінь мінливості ознак. Його використовують для порівняння ступеня варіювання різнойменних ознак однієї вибіркової сукупності або однойменних ознак різних вибіркових сукупностей.

Коефіцієнт варіації – це відношення середнього квадратичного відхилення до середньої арифметичної, виражене у відсотках.

$$Cv = \frac{\sigma}{\overline{X}} \cdot 100\% \ , \qquad (3.18)$$

Розрахунок *Cv* проводять за цією формулою незалежно від того велика вибірка чи мала.

В залежності від числового значення коефіцієнта варіації виділяють мінливість сильну **(Cv≥ 15%), середню (Cv>5%< 15%) та слабку (Cv<5%).**

Помилки середніх величин розраховують для того, щоб середні величини, одержані у вибірці, перенести для характеристики всієї генеральної сукупності, а також для оцінки достовірності розрахованих статистичних показників вибірки.

Характеристика генеральної сукупності на основі вибірки, складеної за принципом випадковості, буде завжди неточною, тому що ціле характеризується на основі даних його частини. Помилки, які виникають при цьому, називають *помилками репрезентативності*.

Якщо дослідженням охоплено всю сукупність, то помилки репрезентативності не розраховують.

Помилки записують через знак "±" поряд з тим біометричним показником, для якого вони розраховані ($\overline{X} \pm \mathbf{S}_{\overline{X}}$; $Cv \pm S_{Cv}$; $\sigma \pm S_{\sigma}$).

Статистичні помилки для біометричних показників розраховують за формулами:

– для середньої арифметичної $S_{\bar{x}} = \frac{\sigma}{\sqrt{n}}$;

– для середнього квадратичного відхилення $S_{\sigma} = \frac{\sigma}{\sqrt{2n}}$;

– для коефіцієнту варіації $S_{Cv} = \frac{Cv}{\sqrt{2n}}$.

Приклад розв'язання типового завдання

Завдання 1. *Під час виконання досліджень якості атмосферного повітря на хімічному підприємстві були відібрані проби повітря. Отримали наступні значення концентрації оксидів азоту: 10, 15, 8, 4, 5, 9, 3, 18, 11, 6, 12, 2, 16, 7, 13 мкг. Визначіть середній вміст оксидів азоту в суміші, розмах варіації, середнє лінійне відхилення, дисперсію; середнє квадратичне відхилення; відносні показники* *варіації. Побудувати стовпчикову діаграму для половини розмаху варіації* $\frac{R}{2}$ *середнього лінійного відхилення L та середнього квадратичного відхилення* σ_x *.*

<u>*Розв'язання*</u>

У комірках A1:A15 і B1:B15 записати ряди $x_1, x_2, ..., x_n$ та $f_1, f_2, ..., f_n$ відповідно (при $M = 15$). У стовпці C сформувати величини $\frac{1}{x}$: для цього в комірку C1 записати формулу =1/A1 і продовжити її для всіх значень x. У стовпці D сформувати величини x^2: для цього в комірку D1 записати сформувати =A1^2 і продовжити її для всіх значень x. У стовпці E сформувати величини x^f : для цього в комірку E1 записати формулу =СТЕПЕНЬ(A1;B1) і продовжити її для всіх значень x.

Далі в 17-му рядку сформувати суми та добуток, які, згідно з робочими формулами, потрібні для обчислення зважених середніх величин. Для цього:

у комірку A17 записати формулу =СУММПРОИЗВ(A1:A15;B1:B15),

у комірку B17 – формулу =СУММ(B1:B15),

у комірку C17 – формулу =СУММПРОИЗВ(C1:C15;B1:B15),

у комірку D17 – формулу =СУММПРОИЗВ(D1:D15;B1:B15),

у комірку E17 – формулу =ПРОИЗВЕД(E1:E15).

У комірці B19 обчислити величину, обернену до суми частот, яка потрібна для добування складного кореня, за формулою =1/B17. Далі приступити до обчислення середніх величин. У комірці A19 обчислити середнє арифметичне за формулою =A17/B17. У комірці C19 обчислити середнє гармонійне за формулою =B17/C17. У комірці D19 обчислити середнє квадратичне за формулою =КОРЕНЬ(D17/B17). У комірці E19 обчислити середнє геометричне за формулою =СТЕПЕНЬ(E17;B19).

У комірках A20 та A21 знайти мінімальне та максимальне значення ряду $x_1, x_2, ..., x_n$ записавши туди формули =МИН(A1:A15) та =МАКС(A1:A15), відповідно.

Сума частот уже попередньо, за формулою =СУММ(B1:B15), обчислена в комірці B17. Також попередньо в комірці A17 обчислена сума попарних добутків ознаки на частоту за формулою =СУММПРОИЗВ(A1:A15;B1:B15).

У комірці A22 обчислити зважене середнє арифметичне за формулою =A17/B17, а в A23 – розмах варіації за формулою =A21–A20.

У комірці H1 записати формулу =ABS(A1–A22) і продовжити її до H15. У комірці H17 записати формулу =СУММПРОИЗВ(H1:H15;B1:B15), а в комірці H19 обчислити зважене середнє лінійне відхилення за формулою =H17/B17.

У комірці I1 записати формулу =H1^2 і продовжити її до I15. У комірці I17 записати формулу =СУММПРОИЗВ(I1:I15;B1:B15), у комірці I19 обчислити зважену дисперсію за формулою =I17/B17, а в I20 – середнє квадратичне відхилення за формулою =КОРЕНЬ(I19). Нарешті в комірці I21 обчислити коефіцієнт варіації за формулою =100*I20/A22.

Завдання 2. *При виконанні дослідження морфологічних особливостей видів зелених насаджень, зокрема площі листової пластинки клена звичайного було отримані наступні дані 2,8, 2,5, 3,6, 1,25, 2,65, 2,87, 2,89, 2,14, 2,22, 2,36, 2,54, 2,89, 3,15, 3,25, 3,14, 2,48, 3,66, 2,17, 2,05. Розрахуйте середні величини досліджуваних показників з використанням вкладки «Аналіз даних» MS Excel. Знайти числові характеристики та побудувати графіки статистичного розподілу за допомогою вбудованого аналізу даних.*

Розв'язання

1. Завантажите MS Excel. Внесіть показники досліджень до листа. Оберіть вкладку «Дані» → «Аналіз даних» → «Описова статистика» і натисніть кнопку «Ок» (Рисунок 3.1).

2. У діалоговому вікні «Описова статистика» (Рисунок 3.2) виконайте наступні дії:

- виділити вхідний інтервал (стовпець даних по довжині жолудя);
- згрупувати за стовпцями;
- мітки в першій стоці (ця опція відзначається, якщо в першому рядку внесений нецифровий запис, наприклад, у нас це «довжина», «ширина», «вага» тощо);
- вихідний інтервал;
- підсумкова статистика;
- рівень надійності 95%.

3. Після вибору необхідних опцій, натискаємо кнопку «Ок». З'являться статистичні показники вибірки (Рисунок 3.3).

4. Для побудови гістограми та кумуляти необхідно виконати наступні дії: натиснути «Данные» → «Анализ нанных» → «Описательная статистика» → «Гистограма» (рисунок 3.1).

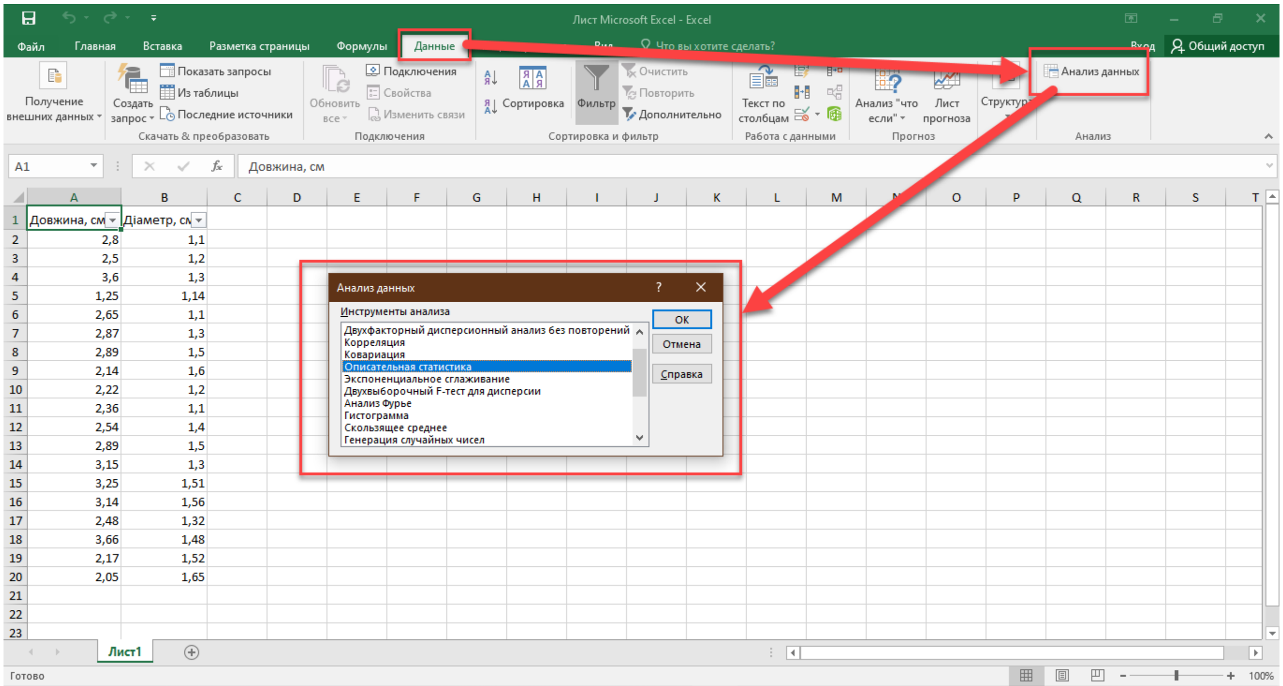


Рисунок 3.1 – Вікно опції «Описова статистика»

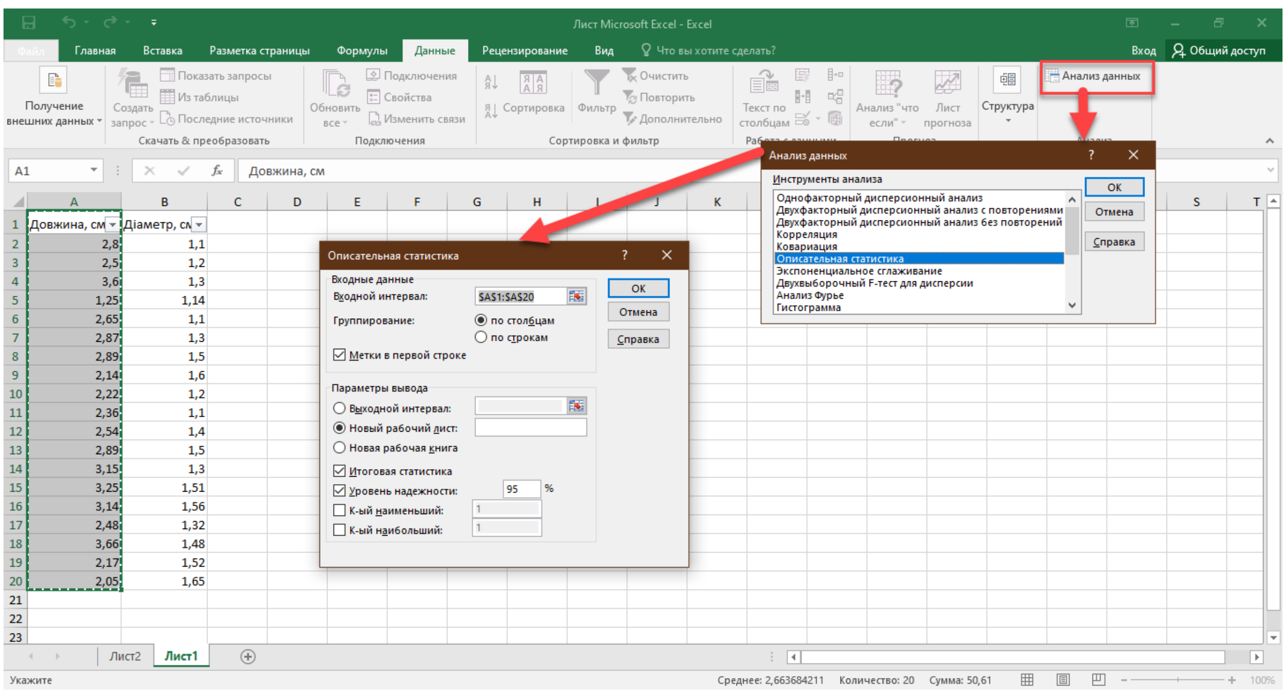


Рисунок 3.2 – Робота з діалоговим вікном «Описова статистика»

	A	B
1	Довжина, см	
2		
3	Среднее	2,663684211
4	Стандартная ошибка	0,133533207
5	Медиана	2,65
6	Мода	2,89
7	Стандартное отклонение	0,582057753
8	Дисперсия выборки	0,338791228
9	Эксцесс	0,691100293
10	Асимметричность	-0,371284244
11	Интервал	2,41
12	Минимум	1,25
13	Максимум	3,66
14	Сумма	50,61
15	Счет	19
16	Уровень надежности(95,0%)	0,280542857
17		

Рисунок 3.3 – Вигляд вікна результатів даних описової статистики

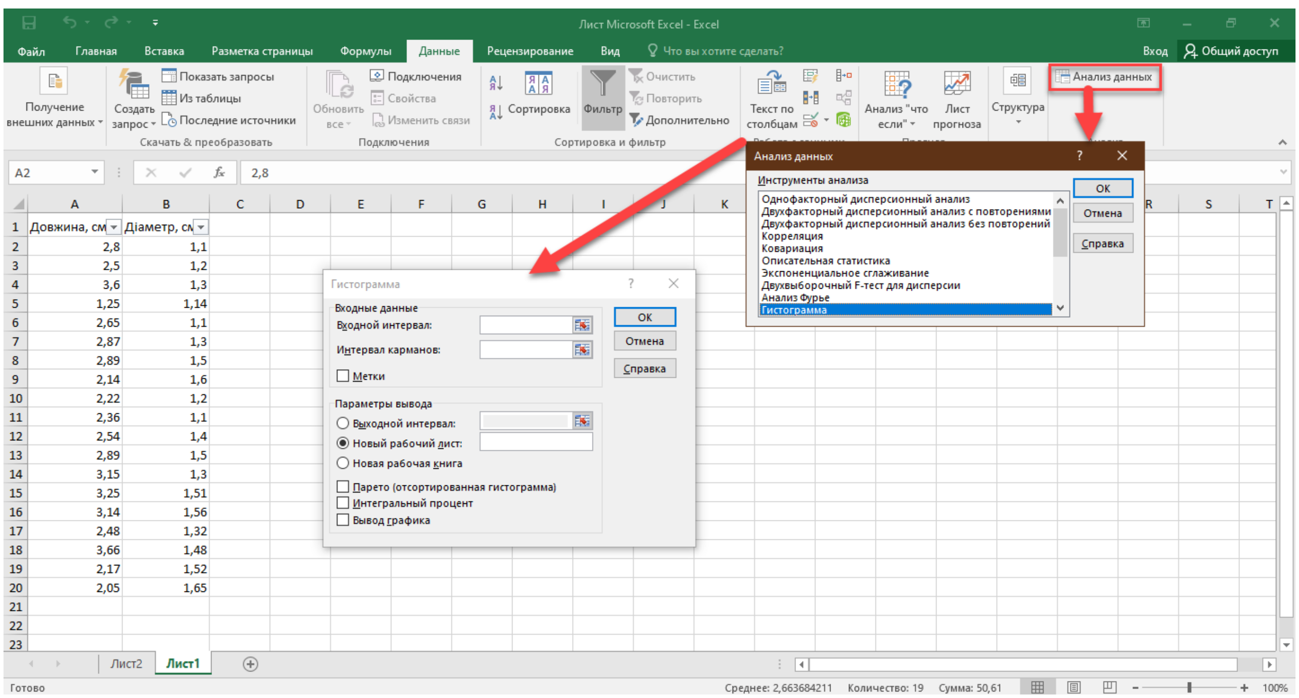


Рисунок 3.4 – Перехід до вікна «Гістограма»

У полі «Входной интервал» ввести вхідні дані. У полі «Интервал карманов» – внести дані «середина інтервалу». Виділити поле для формування даних. У вікні необхідно вказати: вхідний інтервал, вихідний інтервал, вивід графіка (рисунок 3.5).

Після натискання кнопки «Ок» будуть: автоматично розраховані інтервали значень, їх можна округлити до цілих значень; побудована таблиця частот;

виведена гістограма розподілу (Рисунок 3.6). Карман гістограми розподілу це стовпчик, який представляє ряд значень.

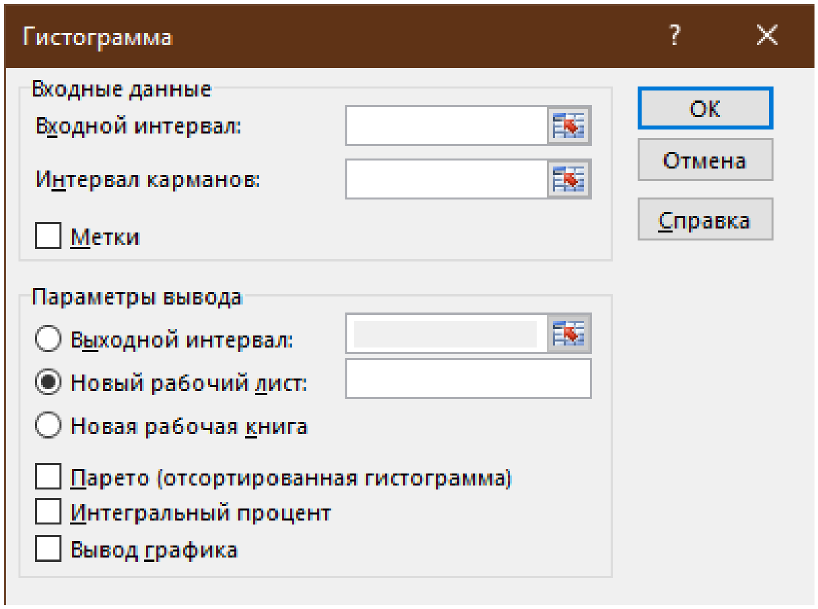


Рисунок 3.5 – Введення даних в діалоговому вікні для побудови діаграми

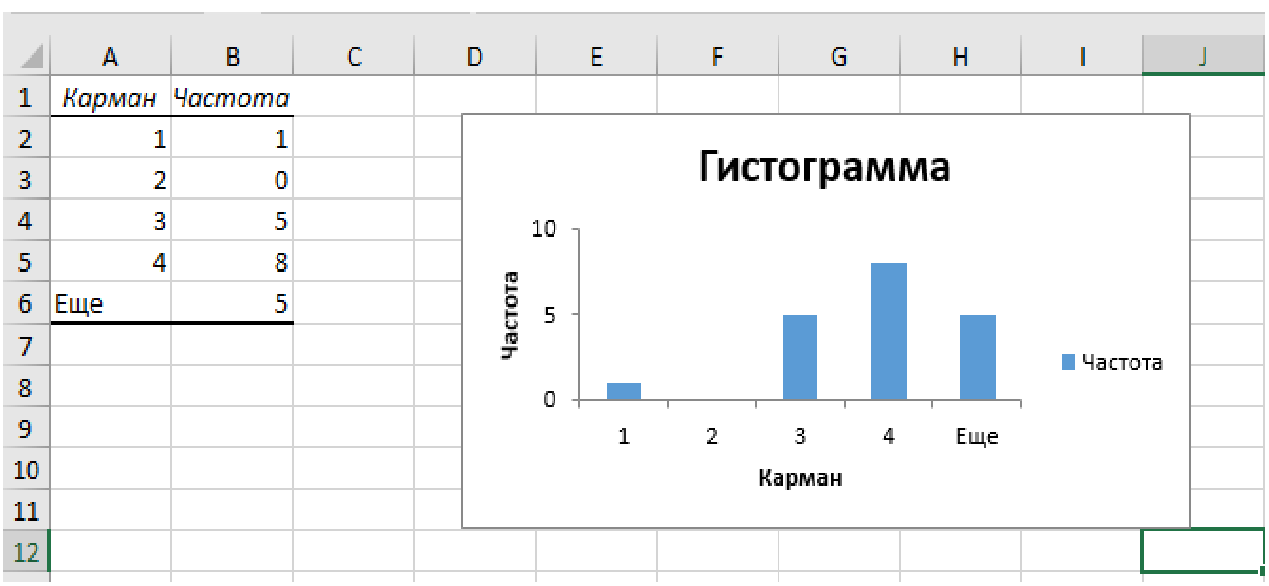


Рисунок 3.6 – Вигляд гістограми

ПРАКТИЧНА ЧАСТИНА

Ознайомитись з теоретичною частиною роботи і виконати наступні завдання:

Завдання 1. Під час виконання досліджень якості атмосферного повітря на хімічному підприємстві були відібрані проби повітря. Отримали наступні значення концентрації триокису сірки в суміші (у відсотках) – таблиця 3.1. Визначіть середній вміст оксидів азоту в суміші, розмах варіації, середнє лінійне відхилення, дисперсію; середнє квадратичне відхилення; відносні показники варіації. Побудувати стовпчикову діаграму для половини розмаху варіації $\frac{R}{2}$ середнього лінійного відхилення L та середнього квадратичного відхилення σ_x .

Таблиця 3.1 – Аналізів вмісту триокису сірки в суміші (у відсотках)

f	**Варіанти**									
	x									
	1	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**
2	10	14	17	16	22	6	32	38	30	11
3	15	16	10	18	14	7	22	29	14	15
4	8	10	9	17	29	8	21	22	19	17
5	4	7	4	13	19	9	19	31	31	19
2	5	15	16	23	16	11	15	17	27	23
6	9	4	5	15	15	19	13	25	9	24
5	3	10	11	14	17	10	14	30	29	26
1	18	20	19	25	30	20	35	40	33	35
2	11	6	6	9	20	12	16	32	32	27
3	6	8	3	13	18	13	17	33	23	29
2	12	5	12	24	24	14	24	23	16	28
4	2	18	3	21	26	15	20	27	15	31
2	16	11	8	20	27	16	34	18	28	32
5	7	9	10	19	25	17	28	35	25	33
2	13	13	18	11	28	18	30	37	17	34

Завдання 2. Обміряно 10 діаметрів дерев, які мають наступні значення: 8, 12, 18, 16, 8, 14, 12, 12, 14, 16 см. Визначити середній діаметр дерева за способом розрахунку простої середньої арифметичної і середньої квадратичної величини. Визначіть похибку середньої арифметичної та мінливість варіаційної ознаки: середнє квадратичне відхилення, коефіцієнт мінливості та репрезентативні помилки ($M, \sigma, C_v, m_M, m_\sigma, m_{c_v}, t_m$).

Завдання 3. При виконанні дослідження морфологічних особливостей видів насаджень за чотирма ознаками, зокрема, площа листової пластинки, довжина жолудя, товщина листкової пластини, показника зміни зростання чисельності насаджень було отримані дані, наведені в таблиці 3.1 Розрахуйте середні величини досліджуваних показників з використанням вкладки «Аналіз даних» MS Excel.

Таблиця 3.1 – Дані досліджуваних біологічних ознак зелених насаджень

Площа листової пластинки, см^3	Довжина жолудя, мм	Товщина листової пластинки, мкм	Показники зміни зростання чисельності насаджень, шт. на 1 га
5,8	25	0,18	561
5,9	30	0,26	801
6,2	32	0,62	2015
6,3	31	0,20	1350
5,4	26	0,44	1670
5,7	27	0,12	2230
5,9	28	0,17	2450
5,4	20	0,15	2830
6,0	24	0,32	3120
5,8	32	0,52	3430
5,5	26	0,09	3710
5,2	20	0,18	3820
5,6	27	0,22	3950
5,7	30	0,32	4010
5,6	31	0,63	4185
5,8	22	0,25	4230
5,9	24	0,28	4310

Контрольні запитання

1. Що означає середня арифметична і в яких одиницях вона виражається?
2. Охарактеризуйте основні біометричні показники, які вказують на ступінь мінливості ознак у вибірковій сукупності?
3. В яких випадках характеристики ступеня мінливості ознак застосовують σ, а в яких C_v?
4. В яких одиницях визначається коефіцієнт мінливості, середнє квадратичне відхилення?
5. З якою метою розраховують помилки статистичних величин?

Лабораторна робота №4
РОЗРАХУНОК ПОМИЛОК РЕПРЕЗЕНТАТИВНОСТІ

Мета роботи: отримати навички визначення достовірності емпіричних показників

ТЕОРЕТИЧНА ЧАСТИНА

Вивчаючи певну ознаку, неможливо дослідити усі об’єкти генеральної сукупності тому, що вона, як правило, є дуже численною, можливо навіть складається з нескінченно великого числа членів.

Тому робиться вибірка об’єктів, які і досліджуються. При цьому постає таке питання: чи можливо за результатами, отриманими при вивченні вибірки, робити висновки про всю генеральну сукупність? Характеризуючи цілу сукупність лише за її частиною, неможливо уникнути помилок, які називаються *помилками репрезентативності.* Навіть за ідеальної організації дослідницької роботи з’являються помилки такого типу.

Розходження між величиною середньої арифметичної $\bar{x}$ вибірки і величиною середньої арифметичної генеральної сукупності M називають ***помилкою репрезентативності***, тобто помилкою, що допускається не в самому процесі виміру, а в результаті випадкового відбору варіант із генеральної сукупності при утворенні вибірки.

Помилка репрезентативності середньої арифметичної залежить від двох величин: від різноманітності ознаки у генеральній сукупності і від чисельності вибірки. Чим менша степінь різноманітності (на її величину вказує середнє квадратичне відхилення) і чим більша кількість вибраних для дослідження об’єктів, тим меншою є величина помилки репрезентативності вибіркового середнього арифметичного.

Величина статистичної помилки окремо взятої варіанти дорівнює квадратичному відхиленню, так як будь-який набутий емпіричний розподіл, що відповідає нормальному закону, практично укладається в межах трьох дисперсій, тобто $x \pm 3\sigma$. Тому, помилку репрезентативності називають середньою квадратичною помилкою, або просто середньою помилкою m.

Таким чином, середня квадратична помилка окремо взятої варіанти виражається у вигляді $mx = \pm 3\sigma$. Вибіркова середня x відхиляється від свого математичного очікування чи середньої арифметичної x генеральної (теоретично розрахованої) сукупності менше в $\sqrt{n}$ разів порівняно з окремими варіантами даного розподілу, та визначається за формулою:

$$m_x = \frac{\sigma_x}{\sqrt{n}} \quad (4.1)$$

Оскільки весь варіаційний ряд випадкової величини, що відповідає нормальному розподілу, практично укладається в межах між $x - 3\sigma_x$ та $x - \sigma_x$, то можна відзначити, що генеральна середня M таких розподілів не виходить за межі потроєного значення середньої помилки середньої арифметичної будь-якої вибірки, взятої із даної генеральної сукупності, тобто вона знаходиться в межах від $x - 3m_x$ до $x + 3m_x$ або в межах $x \pm 3m_x$. Тому потроєне значення середньої квадратичної помилки називається точною помилкою середньої арифметичної вибіркової сукупності. А вираз x±3mx містить в собі так звані «правила потроєної помилки».
При визначенні помилки середньої арифметичної на малих вибірках (до $n < 30$) використовують число «ступеня свободи» ($n-1$) формула приймає наступний вигляд:

$$m_x = \frac{S_x}{\sqrt{n-1}}, \quad (4.2)$$

де S_x – дисперсія масиву, розраховується за формулою: $S_x = \sqrt{\frac{n \cdot \sum x^2 - (\sum x)^2}{n \cdot (n-1)}}$.

Середня помилка середнього квадратичного відхилення розраховується за формулою:

$$m_{S_x} = \frac{S_x}{\sqrt{2 \cdot n}}, \quad (4.3)$$

Середня помилка коефіцієнту варіації C_V визначається за формулою:

$$m_C = \frac{C_v}{\sqrt{2 \cdot n}} \cdot \sqrt{1 + 2\left(\frac{C_v}{100}\right)^2}, \quad (4.4)$$

де C_v – коефіцієнт варіації масиву розраховується за формулою: $C_v = \frac{S_x}{\bar{x}} \cdot 100\%$

Середня помилка показнику асиметрії визначається за формулою:

$$m_{As} = \sqrt{\frac{6 \cdot n(n-1)}{(n-2) \cdot (n+1) \cdot (n+3)}} \tag{4.5}$$

Помилку коефіцієнта ексцесу можна вирахувати за наступною формулою:

$$m_{Ex} = \sqrt{\frac{24}{n}} \tag{4.6}$$

ПРАКТИЧНА ЧАСТИНА

Ознайомитись з теоретичною частиною роботи і виконати наступні завдання.

Завдання 1. Під час дослідження морфологічних особливостей насаджень сосни лапландської були отримані наступні дані (таблиця 4.1). Визначіть для досліджуваних ознак:

- помилку окремо взятої варіанти;
- помилку середньої арифметичної.
- помилку середнього квадратичного відхилення.
- помилку коефіцієнта варіації.
- помилки показників асиметрії і ексцесу.

Таблиця 4.1 – Висота сосни лапландської, дм

Варіанти									
1	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**
10	15	25	18	30	15	14	28	19	24
12	17	20	19	32	16	15	29	18	25
6	12	21	17	28	17	18	24	17	21
5	19	22	16	27	18	19	25	16	22
8	20	23	15	26	19	20	23	15	23
9	18	24	19	29	12	22	22	17	24
15	15	19	20	28	13	23	21	22	26

10	14	18	14	26	14	24	18	21	29
8	13	24	13	25	20	25	19	20	27
13	16	25	17	24	19	19	17	23	26
17	18	17	18	23	22	18	15	24	18
14	22	19	19	31	23	22	13	21	19
11	21	20	22	33	24	25	14	19	17
13	20	21	20	19	26	26	16	26	18
15	22	22	14	18	25	17	10	23	15

Контрольні запитання

1. Що називають помилкою репрезентативності?
2. Що таке середня квадратична помилка і як вона розраховується?
3. Помилка окремо взятої варіанти.
4. Помилка середньої арифметичної.
5. Помилка середнього квадратичного відхилення.
6. Помилка коефіцієнта варіації.
6. Помилки показників асиметрії та ексцесу.
7. Оцінка достовірності відмінностей між дисперсіями.

Лабораторна робота №5

ВИЗНАЧЕННЯ ЧИСЛОВИХ ПОКАЗНИКІВ КОРЕЛЯЦІЇ І ЇХ ДОСТОВІРНОСТІ

Мета роботи: навчитись визначати числові показники кореляції,засвоїти методику виконання розрахунків за допомогою стандартних функцій MS Excel для вирішення завдань аналізу зв'язків.

ТЕОРЕТИЧНА ЧАСТИНА

З метою підвищення ефективності селекції одночасно за кількома ознаками рекомендується насамперед враховувати їхню взаємну зумовленість, тобто кореляцію.

Для цього виникла необхідність математичного усвідомлення такого явища та вираження його числовим показником, коефіцієнтом, за величиною якого можна було б говорити про тісноту чи силу зв'язку між окремими ознаками.

Для цього існує ***коефіцієнт кореляції*** – математичний вираз, за допомогою якого визначають ступінь зв'язку між окремими ознаками. Позначається коефіцієнт кореляції латинською буквою ***r***.

Кореляція у біології – це зв'язок між середніми показниками окремих ознак організмів вибіркових сукупностей зумовлений генетичними факторами в поєднанні з умовами навколишнього середовища. Сам метод оцінки тісноти чи ступеня зв'язку носить назву *кореляційного аналізу*.

Кореляція між двома ознаками має назву *простої*, а коли до уваги береться більше оцінюваних ознак то говорять про множинну кореляцію та, відповідно, *множинний* коефіцієнт кореляції.

Існує також ***функціональний зв'язок***, як правило – в неживій природі. За законами функціональної залежності, зміні одного показника на певну величину відповідає тільки одна конкретна величина іншого показника (наприклад, зменшення стовпчика ртуті в термометрах при зниженні температури).

Функціональна залежність (альтернативна) існує і в живій природі (заплідненість – народження). При кореляційній залежності між двома ознаками вони по відношенню одна до одної можуть приймати різні значення. Так, надій п'яти корів при зменшенні рівня протеїнового живлення також знизиться, але для кожної корови це зменшення буде різним. Якщо функціональні зв'язки однаково просто виявити як на одиничних, так і на групових об'єктах, то кореляційні можна встановити лише при вивченні групових об'єктів методами математичної статистики, з усвідомленням того, що кореляційний зв'язок типовий для об'єктів і процесів, які відбуваються у живій природі. Зв'язки між мінливими селекціонованими ознаками мають різні тенденції. В одному випадку, коли із збільшенням першої ознаки збільшується друга, говорять про *прямий і позитивний* (додатний) зв'язок або про пряму і позитивну кореляцію, а коефіцієнт кореляції позначається із знаком плюс ***(+r)***. В іншому випадку із збільшенням першої ознаки друга зменшується. У цьому разі говорять про *зворотний* або від'ємний зв'язок і коефіцієнт кореляції пишеться із знаком мінус ***(-r).*** Отже, коефіцієнт кореляції вказує не лише на величину сполучної мінливості ознак, але й на напрямок зв'язку.

Таким чином, кореляційний аналіз є кількісним методом з'ясування тісноти сполучної мінливості між ознаками, напрямку зв'язку та його форми. Проте необхідно пам'ятати, що кореляційний зв'язок усе ж таки залишається методом статистичного аналізу, а не біологічного. Тому, не дивлячись на велике значення цього методу в біометрії, все ж не слід його переоцінювати і тим більше підміняти формально статистичним методом глибокий біологічний аналіз результатів у дослідах чи спостереженнях.

Більшість криволінійних зв'язків, які зустрічаються е медичних, соціальних, біологічних дослідженнях близькі до прямолінійних і їх зручніше аналізувати з використанням коефіцієнта кореляції.

Необхідно пам'ятати, що кореляційні зв'язки проявляються тільки в якісно однорідній статистичній сукупності.

Між морфологічними елементами живих організмів існує певний взаємозв'язок, який полягає в тому, що із збільшенням або зменшенням розміру одного елементу (x) відповідно збільшуються або зменшуються розміри іншого елементу (y).

Кореляційний зв'язок розрізняють:

- ✓ за формою: прямолінійний або криволінійний;
- ✓ за напрямом: прямий (додатній) або зворотній (від'ємний);
- ✓ за тіснотою: дуже тісний, тісний, значний, помірний, слабкий.

Коефіцієнт кореляції (r) – числовий показник простої лінійної кореляції,який описує напрям і тісноту зв'язку між досліджуваними величинами, вимірює зв'язок лише при лінійній формі залежності, а його абсолютне значення знаходиться в межах від –1 до +1. При значенні $r = 0$ – зв'язок відсутній; при $+1$ – пряма кореляційна залежність; а при $–1$ – зворотня.

Коефіцієнт кореляції визначається за формулою:

$$r = \frac{\sum (x - \bar{x}) \cdot (y - \bar{y})}{\sqrt{(x - \bar{x})^2 \cdot (y - \bar{y})^2}}. \qquad (5.1)$$

Якщо зв'язку дати оцінку 1, то віддалення від 1 (0,9; 0,8; 0,7;….0,1; 0) буде характеризувати тісноту відповідного кореляційного зв'язку. Показник $r > 0{,}9$ свідчить про дуже тісний зв'язок, $r = 0{,}7 - 0{,}9$ – тісний зв'язок; $r = 0{,}5 - 0{,}7 -$ значний зв'язок; $r = 0{,}3 - 0{,}5$характеризує помірний зв'язок, а $r < 0{,}3$– слабкий зв'язок.

При прямому кореляційному зв'язку, зі зміною значення однієї ознаки змінюється значення іншої ознаки, яка була взаємопов'язана з нею в тому ж напрямку (наприклад, з підвищенням температури тіла у людини збільшується частота пульсу). Позначається прямий зв'язок знаком (+).

При зворотному кореляційному зв'язку зі зміною значення однієї ознаки змінюється значення іншої ознаки, яка була взаємопов'язана з нею в зворотному напрямку (наприклад, восени, чим нижче температура навколишнього середовища, тим вище захворюваність серед дитячого населення. Зворотній зв'язок позначається знаком (−).

Крім того, кореляційний зв'язок може бути прямолінійним та криволінійним.

Прямолінійний зв'язок характеризується відносно рівномірною зміною значень однієї ознаки, при рівній зміні іншої ознаки (наприклад, при вимірюванні артеріального тиску відповідність між максимальним і мінімальним рівнями АТ).

При криволінійному кореляційному зв'язку інше співвідношення: при рівномірній зміні значення однієї ознаки можуть спостерігатися зростаючі або спадаючі значення іншої ознаки.

Вимірювання і оцінка зв'язку між явищами при прямолінійному кореляційному зв'язку здійснюється за допомогою коефіцієнта кореляції, а при криволінійному зв'язку − кореляційним відношенням.

Найбільш поширеними способами розрахунку є:

– метод квадратів (Пірсона);

– метод рангів (Спірмена).

Метод квадратів (Пірсона). При обчислення коефіцієнта лінійної кореляції (Метод Пірсона) застосовують наступну формулу:

$$r_{xy} = \frac{\sum d_x \cdot d_y}{\sqrt{\sum d_x^2 \cdot \sum d_y^2}}, \quad (5.2)$$

де d_x, d_y – відхилення кожного значення x і y від відповідних середніх арифметичних (X_x, X_y).

x, y – значення варіант першої та другої ознак.

Достовірність коефіцієнта лінійної кореляції оцінюють за допомогою середньої помилки (помилки репрезентативності) за формулою:

$$m_r = \sqrt{\frac{1 - r_{xy}^2}{n - 2}}, \qquad \text{при } n < 30 \quad (5.3)$$

$$m_r = \frac{1 - r_{xy}^2}{\sqrt{n}}, \qquad \text{при } n > 30 \quad (5.4)$$

де m_r – середня помилка коефіцієнта лінійної кореляції;

r^2 – коефіцієнт лінійної кореляції в квадраті;

n – кількість досліджуваних пар.

Критерій достовірності (критерій Стьюдента – t) для даного методу розраховується за формулою:

$$t = \frac{r}{m_r}, \quad (5.5)$$

де t – критерій достовірності (критерій Стьюдента);

r – коефіцієнт лінійної кореляції;

m_r – помилка коефіцієнта лінійної кореляції.

Метод рангів (Спірмена). Метод рангів відноситься до непараметричних критеріїв оцінки взаємозв’язку між досліджуваними ознаками. Особливість цього коефіцієнта в простоті обчислення, хоча за рахунок цього й втрачається точність дослідження. Він найчастіше використовується для орієнтовної оцінки зв’язку між ознаками. Його застосовують при наявності якісних або напівякісних ознак. Він ґрунтується на ранжуванні досліджуваних ознак.

Формула розрахунку коефіцієнта рангової кореляції:

$$\rho_{xy} = 1 - \frac{6\Sigma d^2}{n(n^2 - 1)}, \qquad (5.6)$$

де x, y – явища, між якими визначається зв'язок;

ρ – коефіцієнт рангової кореляції;

d^2 – різниця рангів в квадраті;

n – кількість досліджуваних пар.

Для оцінки достовірності коефіцієнта рангової кореляції користуються формулою визначення середньої помилки:

$$m_\rho = \sqrt{\frac{1 - \rho_{xy}^2}{n - 2}}, \text{ при n<30}, \qquad (5.7)$$

$$m_\rho = \frac{1 - \rho_{xy}^2}{\sqrt{n}}, \text{ при n>30}, \qquad (5.8)$$

де m_ρ – середня помилка коефіцієнту рангової кореляції;

ρ – коефіцієнт рангової кореляції в квадраті;

n – кількість досліджуваних пар.

Критерій достовірності (критерій Стьюдента – t) для даного методу розраховується за формулою:

$$t = \frac{\rho_{xy}}{m_\rho}, \qquad (5.9)$$

де t – критерій достовірності (критерій Стьюдента);

ρ – коефіцієнт рангової кореляції;

m_ρ – помилка коефіцієнта лінійної кореляції.

Після проведення розрахунку проводять оцінення критерію достовірності по таблиці значень критерію Стьюдента.

Необхідно вміти в повному обсязі і послідовно оцінювати коефіцієнт кореляції, при цьому враховувати наявність зв’язку, його напрямок і силу залежності.

Наприклад, якщо отриманий коефіцієнт кореляції **(+**0,5), з рівнем достовірності t=2,1 при кількості спостережень більше 30 між такими ознаками, як успішність і відвідуваність студентів, то висновок буде наступний:

– кореляційний зв’язок між ознаками існує;

– кореляційний зв’язок – прямий (+);

– між ознаками встановлена середня залежність (0,5) однієї ознаки від іншої;

– кореляційний зв’язок достовірний, його достовірність становить понад 95,0% та помилка становить менш ніж 5%.

ПРАКТИЧНА ЧАСТИНА

Ознайомитись з теоретичною частиною роботи і виконати наступні завдання.

Завдання 1. Проаналізувати тісноту зв’язку між діаметрами і висотами у 10 дерев сосни звичайної. Зробити висновки.

Варіанти																			
1		**2**		**3**		**4**		**5**		**6**		**7**		**8**		**9**		**10**	
x	*y*	*x*	*y*	*x*	*y*	*x*	*y*	*x*	*y*	*x*	*y*	*x*	*y*	*x*	*y*	*x*	*y*	*x*	*y*
10	15	25	18	30	15	14	28	19	24	17	18	18	24	18	10	25	15	28	19
12	17	20	19	32	16	15	29	18	25	14	22	22	21	19	12	20	16	29	18
6	12	21	17	28	17	18	24	17	21	11	21	25	19	17	6	21	17	24	17
5	19	22	16	27	18	19	25	16	22	13	20	26	26	18	5	22	18	25	16
8	20	23	15	26	19	20	23	15	23	15	22	17	23	15	8	23	19	23	15
9	18	24	19	29	12	22	22	17	24	14	18	15	24	18	9	24	12	22	17
15	15	19	20	28	13	23	21	22	26	18	19	13	21	19	15	19	13	21	22
10	14	18	14	26	14	24	18	21	29	19	22	14	19	17	10	18	14	18	21
8	13	24	13	25	20	25	19	20	27	13	15	16	26	18	8	24	20	19	20
13	16	25	17	24	19	19	17	23	26	15	13	10	23	15	13	25	19	17	23

Завдання 2. Провести аналіз залежності між довжиною 20 окремих листків озимої пшениці та їх площами (таблиця 5.1), визначених на основі індивідуальних вимірів у MS EXCEL Зробити висновки.

Таблиця 5.1 – Результати дослідження показників схожості пшениці

Номери листків (пар)	Довжина листа, см (х)	Площа листа, см^2 (у)
1	16,1	7,4
2	17,3	8,7
3	18,6	10,3
4	20,0	11,2
5	21,3	12,9
6	21,6	13,2
7	21,8	13,7
8	22,0	14,1
9	22,4	14,3
10	22,8	14,8
11	23,1	15,2
12	23,3	16,2
13	23,3	16,7
14	23,7	17,0
15	24,0	17,4
16	24,1	19,2
17	25,2	19,3
18	26,0	20,3
19	26,5	21,4
20	26,4	22,3

Для визначення коефіцієнт кореляції у MS EXCEL розрахунок поводять за допомогою статистичної функції КОРЕЛЛ, в полях Массив1 і Массив2, куди вводять послідовно посилання на діапазон значення x і на діапазон значення y відповідно і натискають кнопку «ОК».

Контрольні запитання

1. Що таке кореляція і якою вона буває?
2. Що таке основний показник кореляційного аналізу та його використання?

Лабораторна робота №6
РЕГРЕСІЙНИЙ АНАЛІЗ. ПРОСТА ЛІНІЙНА РЕГРЕСІЯ. ВИЗНАЧЕННЯ КОЕФІЦІЄНТІВ РЕГРЕСІЇ

Мета роботи: закріпити теоретичні знання та отримати практичні навички щодо використання ПЕОМ з метою оцінки параметрів та підбору емпіричних формул за допомогою простої вибіркової лінійної регресії методом найменших квадратів.

ТЕОРЕТИЧНА ЧАСТИНА

При оцінці ступеня взаємозв'язку статистичних величин важливо провести математичне моделювання, тобто підібрати аналітичне рівняння, яке відповідало б природі досліджуваного явища з метою передбачення поведінки незалежної характеристики об'єкта при зміні залежного параметра. Динаміка взаємної залежності між змінними величинами отримала назву регресії, а методика дослідження регресії носить назву *регресійного аналізу*.

Метою регресивного аналізу є встановлення певного виду рівняння, графічне вираження якого (лінія регресії) добре апроксимує розподіл фактичних значень ознаки. Будь-яка залежність може бути описана рівнянням виду: $y = f(x)$, де y – залежна ознака (значення функції), x – незалежна факторна ознака (аргумент функції).

Регресія – це зміна значення функції в залежності віл зміни одного або декількох аргументів.

Емпіричний ряд регресії – це подвійний ряд значень ознак аргументу і значень відповідних ознак функції.

Якщо при зростанні чи спаданні аргумента функція також пропорційно зростає чи спадає, регресія є прямолінійною, в іншому випадку – криволінійною.

Найпростішим прикладом регресії є рівняння прямої: $y = a + bx$, де y – залежна ознака; x – незалежна ознака; a – вільний член рівняння (ордината точки перетину прямої з віссю ординат); b – коефіцієнт лінійної регресії (абсциса точки перетину прямої з віссю абсцис).

Коефіцієнт лінійної регресії (b) – це число, яке вказує напрям і середню величину зміни залежної ознаки при зміні факторної на одиницю виміру.

Коефіцієнт b має знак коефіцієнта кореляції. Коефіцієнт a приймає додатнє значення, якщо лінія регресії перетинає вісь OY над початком координат і від'ємне значення, якщо лінія регресії проходить нижче початку координат. Чим більший коефіцієнт b, тим більший кут нахилу прямої.

До завдання регресивного аналізу належить:

1) обчислення коефіцієнтів рівняння;

2) встановлення достовірності коефіцієнтів рівняння;

3) знаходження теоретичних (найбільш ймовірних) значень залежної ознаки;

4) обчислення середньо квадратичного відхилення від регресії (помилка рівняння);

5) вирівнювання емпіричних рядів;

6) оцінка точності вирівнювання;

7) визначення ефекту регресії при вимірюванні варіації залежної ознаки;

8) вибір рівняння, яке найбільш точно описує існуючу залежність;

9) визначення стандартної помилки обчисленого значення.

При обчисленні коефіцієнтів рівнянь можна використати наступні способи: графічний, спосіб вибраних точок, спосіб найменшої помилки, за центральними відхиленнями, за способом Маркова, за коефіцієнтом кореляції, спосіб найменших квадратів, за числовими коефіцієнтами (спосіб Труля).

У основі регресійного аналізу лежать дві гіпотези.

1. Передбачається, що досліджувана сукупність параметрів має внутрішній статистичний зв’язок, який може бути виявлений і формалізований у вигляді кореляційної (отже лінійної) залежності одного параметра від іншого або від інших. Тобто вважається, що існує внутрішній лінійний зв’язок середніх значень цих параметрів.

2. Передбачається, що випадковий розкид (дисперсія) значень (кожного) параметру має регулярну компоненту, яка залежить від деякого аргументу (“сигналу”), і випадкову компоненту (“шум”). Випадкова компонента (“шум”) розподілена за нормальним законом.

Початковою інформацією для побудови лінійної однофакторної регресійної моделі є сукупність із n двовимірних точок (x_i, y_i), де кожна координата точки, як правило, має свій фізичний сенс, наприклад x_i – зріст людини в сантиметрах, y_i – її вага в кілограмах.

Під час формалізації постановки задачі розглянемо двовимірну випадкову величину (X, Y), над якою проведено n незалежних випробувань і в результаті випробувань отримана вибірка – n пар чисел (координат точки): $(x_1, y_1), (x_2, y_2), \ldots, (x_n, y_n)$, де x_i – значення випадкової величини X у i-му випробуванні; y_i – значення випадкової величини Y у i-му випробуванні.

Необхідно знайти наближене зображення значень однієї з випадкових величин як функції значень другої випадкової величини.

Вибірковим *рівнянням регресії* Y на $X\,(y \to x)$ називається рівняння, яке встановлює залежність змінної y від змінної x, тобто коли змінна у **вважається** *функцією, а змінна* x *– аргументом*: $y = f(x)$, при цьому початковою інформацією є вибірка з n пар чисел.

Вибірковим рівнянням регресії X на $Y(x \to y)$ називається рівняння $x = \varphi(y)$, у якому при тій же початковій інформації вже змінна x уважається функцією, а змінна y – її аргументом.

Лінійною називається регресія у випадку, коли залежності $f(x)$ і $\varphi(y)$ є лінійними функціями. Тоді рівняння регресії мають вигляд: $y = a \cdot x + b,\ x = c \cdot y + d$.

Порядок розрахунку параметрів рівняння регресії розглянемо без виведення і для двох варіантів умов: при відсутності і при наявності збіжних точок у вибірці.

1. Нехай серед точок (x_1, y_1) вибірки збіжних точок немає. Для того щоб скласти вибіркове рівняння прямої лінії регресії, виконуються наступні розрахунки:

а) обчислюються середні значення величин: $\bar{x}, \bar{y}, \overline{x \cdot y}, \overline{x^2}, \overline{y^2}$ і знаходяться середні квадратичні відхилення $S_x = \sigma_x, S_y = \sigma_y$ із використанням формул, перерахованих з обліком доцільної послідовності їх застосування:

$$\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i\ ; \quad \bar{y} = \frac{1}{n}\sum_{i=1}^{n} y_i\ ; \quad \overline{x^2} = \frac{1}{n}\sum_{i=1}^{n} x_i^2\ ; \quad \overline{y^2} = \frac{1}{n}\sum_{i=1}^{n} y_i^2\ ;$$

$$\overline{x \cdot y} = \frac{1}{n}\sum_{i=1}^{n} x_i \cdot y_i\ ; \quad S_x^2 = \overline{x^2} - \left(\bar{x}\right)^2\ ; \quad S_x = \sqrt{S_x^2}\ ; \quad S_y^2 = \overline{y^2} - \left(\bar{y}\right)^2\ ; \quad S_y = \sqrt{S_y^2}\ ;$$

б) обчислюється значення вибіркового коефіцієнту кореляції r:

$$r = \frac{\overline{x \cdot y} - \bar{x} \cdot \bar{y}}{S_x \cdot S_y}, \qquad (6.1)$$

Вибірковий коефіцієнт кореляції характеризує рівень лінійного кореляційного зв’язку двох випадкових величин. Чим ближче $|r|$ до одиниці, тим більш сильним є зв’язок двох величин, чим ближче $|r|$ до нуля, тим зв’язок слабше;

в) для одержання рівняння регресії Y на X: $y = a \cdot x + b$ обчисляємо коефіцієнти a і b даного рівняння за формулами:

$$a = r \cdot \frac{S_y}{S_x};\ b = \overline{y} - a \cdot \overline{x}, \qquad (6.2)$$

Для одержання рівняння регресії X на Y: $x = c \cdot y + d$ обчисляємо коефіцієнти c і d даного рівняння за формулами:

$$c = r \cdot \frac{S_x}{S_y};\ d = \overline{x} - a \cdot \overline{y}. \qquad (6.3)$$

Обидві прямі лінії $y = a \cdot x + b$ і $x = c \cdot y + d$ проходять через точку $(\overline{x}, \overline{y})$. Для зображення обох прямих ліній на одному графіку друге рівняння варто подати у вигляді: $y = x/c - d/c$.

2. При великій кількості точок n у вибірці значення x_i може зустрітися m_i разів, значення y_j може зустрітися n_j разів. Та сама пара чисел (x_i, y_j) може зустрітися n_{ij} разів. У цьому випадку вибірку зручно подати у вигляді кореляційної таблиці (див. табл. 1) так, що кількість повторень m_i значень координат x_i, кількість повторень n_j значень координат y_j і обсяг вибірки n дорівнюватимуть:

$$m_i = \sum_{j=1}^{\ell} n_{ij}, \quad n_j = \sum_{i=1}^{k} n_{ij}; \qquad n = \sum_{i=1}^{k}\sum_{j=1}^{\ell} n_{ij} = \sum_{j=1}^{\ell} n_j = \sum_{i=1}^{k} m_i = n.$$

Таблиця 6.1 – Кореляційна таблиця

x_i	Y_j				
	y_1	y_2	…	y_ℓ	$\boldsymbol{m_i}$
x_1	$\boldsymbol{n_{11}}$	$\boldsymbol{n_{12}}$	…	$\boldsymbol{n_{1\ell}}$	$\boldsymbol{m_1}$
x_2	$\boldsymbol{n_{21}}$	$\boldsymbol{n_{22}}$	…	$\boldsymbol{n_{2\ell}}$	$\boldsymbol{m_2}$
…	…	…	…	…	…
x_k	$\boldsymbol{n_{k1}}$	$\boldsymbol{n_{k2}}$	…	$\boldsymbol{n_{k\ell}}$	$\boldsymbol{m_k}$
$\boldsymbol{n_j}$	$\boldsymbol{n_1}$	$\boldsymbol{n_2}$	…	$\boldsymbol{n_\ell}$	$\boldsymbol{n}$

Вибіркове рівняння прямої лінії регресії знаходимо аналогічно першому випадку, розрахунки величин $\overline{x}, \overline{y}, \overline{x \cdot y}, \overline{x^2}, \overline{y^2}$ виконуємо з урахуванням наявності повторюваних значень змінних за формулами:

$$\overline{x}=\frac{1}{n}\sum_{i=1}^{k}x_i\cdot m_i\,;\quad \overline{y}=\frac{1}{n}\sum_{i=1}^{\ell}y_i\cdot n_j\,;\quad \overline{x^2}=\frac{1}{n}\sum_{i=1}^{k}x_i^2\cdot m_i\,;\quad \overline{y^2}=\frac{1}{n}\sum_{i=1}^{\ell}y_i^2\cdot n_j\,;\quad \overline{x\cdot y}=\frac{1}{n}\sum_{i=1}^{k}\sum_{j=1}^{\ell}x_i\cdot y_i\cdot n_{ij}\,.$$

Якщо розглядається вибірка з генеральної сукупності безперервних випадкових величин X і Y, то кореляційна таблиця буде містити інтервали $[a_{i-1},a_i)\ i\,[b_{j-1},b_j)$.

У цьому випадку для обчислення величин $\overline{x},\overline{y},\overline{x\cdot y},\overline{x^2},\overline{y^2}$ спочатку потрібно перейти до дискретних рядів, а потім виконати обчислення за розглянутими вище формулами.

У ряді практичних задач обробки результатів вимірів, у тому числі значень випадкових величин, виникає необхідність згладженого подання значень однієї величини (наприклад, величини y), як функції іншої величини (наприклад, величини x). Одним із найбільш поширених способів такого наближеного зображення є ***метод найменших квадратів***.

За допомогою методу найменших квадратів розв'язується задача добору такої аналітичної залежності $y(x)=\Psi_m(x,a_0,a_1,...,a_m)$, графік якої *не обов'язково* проходив би через *усі* задані точки, але максимально "згладжував"
би випадкові похибки вимірюваних ординат функції $y_i = f(x_i)\,(i=0,1,2,...,n)$ тобто щоб сума квадратів відхилень значень аналітичної залежності $\Psi_m(x,a_0,a_1,...,a_m)$ від значень вимірюваних ординат $y_i = f(x_i)$ у цих точках була мінімальною:

$$S=\sum_{i=1}^{n}[y_i-\Psi_m(x_i)]^2 \Rightarrow \min\Psi_m(x_i)\,, \tag{6.4}$$

Апроксимація за методом найменших квадратів виконується у два етапи:

– на першому етапі вибирають вигляд $\Psi_m(x,a_0,a_1,...,a_m)$ шуканої формули;
– на другому етапі для формули обраного вигляду "підбирають" значення параметрів $a_0,a_1,...,a_m$, виходячи з вимоги.

У методі найменших квадратів дані параметри визначаються з умови мінімуму наступного критерію:

$$R=\sum_{i=1}^{n}(y_i-\overline{y_i})^2=\sum_{i=1}^{n}(y_i-a_0-a_1\cdot x_i)^2\,. \tag{6.5}$$

ПРАКТИЧНА ЧАСТИНА

Ознайомитись з теоретичною частиною роботи і виконати наступні завдання.

Завдання 1. Обчисліть коефіцієнти регресії прямої для біологічних ознак сосни звичайної наведених в таблиці 6.1. У таблиці представлені результати вимірювання діаметра сосни *X* в сантиметрах і її висоти *Y* в метрах (N=20):

Таблиця 6.1 – Результати досліджень біологічних ознак сосни звичайної

n	*1*	*2*	*3*	*4*	*5*	*6*	*7*	*8*	*9*	*10*
X	14,5	20,8	19,6	23,4	21,5	25,3	19,2	19,4	33,7	28,6
Y	13,4	19,5	20,4	19,2	18,3	21,3	17,4	18,2	25,1	21,5
n	*11*	*12*	*13*	*14*	*15*	*16*	*17*	*18*	*19*	*20*
X	34,2	31,1	16,6	14,9	18,4	28,5	20,5	22,7	20,3	21,2
Y	24,4	23,4	14,5	13,2	17,2	22,2	19,4	20,1	19,5	20,3

Контрольні запитання

1. Що таке регресія?
2. Побудова емпіричних рядів регресії.
3. Що таке рівняння регресії?
4. Що таке коефіцієнт регресії?
5. Які є способи визначення коефіцієнтів регресії?

Лабораторна робота №7
ДИСПЕРСІЙНИЙ АНАЛІЗ ОДНОФАКТОРНИХ РІВНОМІРНИХ І НЕРІВНОМІРНИХ КОМПЛЕКСІВ МАЛИХ ГРУП

Мета роботи: отримати навички вивчення статистичного впливу одного або декількох факторів на результативну ознаку (навчитись проводити дисперсійний аналіз).

ТЕОРЕТИЧНА ЧАСТИНА

У практиці нерідко виникає необхідність в оцінці цілих комплексів кількісних показників, необхідність порівнювати між собою одночасно не дві, а кілька вибірок, об'єднаних в єдиний комплекс.

У будь-якому експерименті середні значення досліджуваних величин змінюються у зв’язку зі зміною основних факторів (кількісних та якісних), що визначають умови досліду, а також і випадкових факторів. Дослідження впливу тих чи інших факторів на мінливість середніх є задачею дисперсійного аналізу, який був розроблений Р.А. Фішером (1925).

Дисперсійний аналіз використовує властивість адитивності дисперсії випадкової величини, що обумовлено дією незалежних факторів. У залежності від числа джерел дисперсії розрізняють *однофакторний та багатофакторний дисперсійний аналіз.*

Дисперсійний аналіз полягає у виділенні і оцінці окремих факторів, що викликають зміну досліджуваної випадкової величини. При цьому проводиться розклад сумарної вибіркової дисперсії на складові, обумовлені незалежними факторами. Кожна з цих складових є оцінкою дисперсії генеральної сукупності. Щоб вирішити, чи дієвий вплив даного фактору, необхідно оцінити значимість відповідної вибіркової дисперсії у порівнянні з дисперсією відтворення, обумовленою випадковими факторами.

Перевірка значимості оцінок дисперсії проводять по критерію Фішера. Коли розрахункове значення критерію Фішера виявиться меншим табличного, то вплив досліджуваного фактору немає підстав вважати значимим. Коли ж розрахункове значення критерію Фішера виявиться більшим табличного, то цей фактор впливає на зміни середніх.

Факторами звуться причини зміни характеристик біологічних об'єктів, а ті характеристики біологічних об'єктів, які змінюються під їх впливом, звуться *результативними ознаками.*

Методика дисперсійного аналізу зводиться до деякої загальної схеми – алгоритму (рисунок 7.1).

№	Зміст організаційної або математичної дії
1	Групування вибіркових матеріалів у комбінаційну таблицю дисперсійного комплексу
2	Визначення значень: середнього арифметичного всього комплексу ($\bar{x}$) і групових середніх за градаціями організованого фактора (x_i)
3	Визначення загальної суми квадратів відхилень (D_y), тобто суми квадратів відхилень варіант від загальної середньої: $D_y = \Sigma x^2 - \frac{(\Sigma x)^2}{N}$
4	Визначення міжгрупової суми квадратів відхилень, яка дорівнює сумі квадратів відхилень групових середніх від загальної середньої з урахуванням статистичної ваги (n_i) групових середніх: • у випадку рівних чисел варіант в градаціях комплексу – $D_x = n_i \Sigma \bar{x}_i^2 - \frac{(\Sigma x)^2}{N}$; • у випадку різної кількості варіант в градаціях комплексу – $D_x = \Sigma[n_i(x_i - x)^2]$;
5	Визначення внутрішньогрупової суми квадратів, тобто суми квадратів відхилень групових варіант від групових середніх: $D_z = \Sigma x^2 - n\Sigma \bar{x}_i^2$
6	Визначення дисперсій (середніх квадратів відхилень): • загальна: $\sigma_{заг}^2 = D_y / (N-1)$; • факторна: $\sigma_{факт}^2 = D_x / (a-1)$, де a – кількість груп; • остаточна: $\sigma_{ост}^2 = D_z / (N-a)$
7	Визначення фактичного значення критерію $F_{факт.} = \frac{\sigma^2_{факт}}{\sigma^2_{ост}}$;
8	Порівняння фактичного значення критерію F з його табличним (стандартним) значенням для відповідного рівня значимості (p) і даних чисел

Рисунок 7.1 – Алгоритм дисперсійного аналізу однофакторних комплексів

У MS EXCEL однофакторний дисперсійний аналіз проводиться наступним чином: Сервіс / Аналіз даних / Однофакторний дисперсійний аналіз / вкажіть діапазон вхідних значень, групування за стовпцями, прапорець Мітки зніміть, вкажіть клітинку вихідного діапазону комірок. Проаналізуйте отримані результати: порівняйте дисперсію всередині груп і між групами. Якщо вони значимо відрізняються (рівень значимості P = 0,05), то фактор вважається надає статистично значимий вплив на досліджувану змінну. Порівняйте розрахункове F і критичне значення статистики Фішера. Відмінність вважається значимою, якщо розрахункове значення більше критичного.

ПРАКТИЧНА ЧАСТИНА

Завдання 1. Користуючись таблицею 6.1 провести дисперсійний аналіз експериментальних даних наведених в прикладі. Досліджувався вплив кількості доглядів (розпушень ґрунту) на ріст 2-х річних сіянців сосни звичайної після висаджування їх в ґрунт. Дослід організовано на однорідній території з виділенням дослідних площ розміром 0,5 га. Сукупність рослин на кожній дослідній площі є відповідною генеральною сукупністю. Вибіркові сукупності формувались шляхом відбору по 200 сіянців (за принципом імовірності) для кожної дослідної площадки. Дослід закладався в 3-х повторностях з градаціями регульованого фактору (розпушування) – відсутність (0), два, чотири, шість. Всього закладено 12 пробних площ, на яких у дослідних рослин заміряні восени річні прирости за висотою. Результати замірів згруповані в таблиці 7.1.

Таблиця 7.1 – Заміри приросту сіянців сосни звичайної

Варіанти досліду	Середній приріст рослин за повторностями (варіантами), см				Середній приріст по всіх повторностях
	1	2	3	n_j	
Контроль	6	4	8	3	
Огляди проводилися: два рази	8	8	12	3	
Чотири рази	12	14	16	3	
Шість разів	12	12	12	3	

Однакова кількість повторностей по всіх варіантах досліду і участь лише одного регульованого фактора підтверджує класифікацію комплексу як однофакторіального і рівномірного.

Завдання 2. Проведіть дисперсійний аналіз вимірів приростів лісових культур у висоту (см/міс) при внесенні різних доз мінеральних добрив. Вихідні дані наведені в таблиці 7.2.

Таблиця 7.2 – Результати лабораторних досліджень схожості лісових культур

Варіанти факторів	Повторюваність за варіантами, x_j			
20%	2	4	4	6
40%	3	5	5	6
60%	6	7	7	7

Контрольні питання

1. Що таке дисперсійний аналіз?
2. Передумови та постановка задачі однофакторного дисперсійного аналізу.
3. Загальна, факторна та залишкова суми квадратів відхилень та зв'язок між ними.
4. Алгоритм однофакторного дисперсійного аналізу за Фішером.

ПОРЯДОК ВИКОНАННЯ ТА ОЦІНЮВАННЯ ЛАБОРАТОРНИХ РОБІТ

Перед початком лабораторних робіт викладач проводить інструктаж з техніки безпеки проведення лабораторних робіт і пожежної безпеки з оформленням у відповідних журналах і з підписами студентів. Студенти, які не пройшли інструктаж з техніки безпеки, до виконання робіт не допускаються.

Перед лабораторними заняттями викладач перевіряє опанування студентами теоретичної частини та методики виконання лабораторної роботи. Опитування студентів перед початком виконання лабораторної роботи відбувається у формі співбесіди, а зарахування після її виконання – у формі тестового контролю.

З усіма теоретичними питаннями, що виникають у процесі виконання лабораторної роботи, студенти звертаються до викладача.

Кожна робота повинна бути захищена. Захист лабораторної роботи складається з уміння здобувача вищої освіти викласти основні теоретичні положення теми, методики дослідження, проаналізувати отримані результати.

Для перевірки роботи студент подає рукописний матеріал/машинний набір та електронний варіант документу (файл MS Excel) який буде мати наступну назву: *ЛБ1_Прізвище студента_№варіанту.*

Результати лабораторних робіт заносять в лабораторний журнал (Додаток 1, 2).Оформлену роботу студент захищає і здає викладачу.

ТЕСТОВІ КОНТРОЛЬНІ ЗАВДАННЯ ДЛЯ ПЕРЕВІРКИ ЯКОСТІ ЗНАНЬ ЗДОБУВАЧІВ ВИЩОЇ ОСВІТИ

1) *Предметом біометрії є:*
    A. розміри і кількісні співвідношення між масовими явищами в природі, закономірності їх формування, розвитку, взаємозв'язку;
    B. розташування об'єкта відносно один одного, колір, спадковість ознаки, взаємовплив;
    C. багатогранність об'єкта, масовість ознаки, середня величина, медіана, мода.
2) *Біометрія це :*
    A. наука про взаємовплив навколишнього середовища на живі організми;
    B. наука про рух групової інформації в популяціях;
    C. наука про статистичний аналіз групових властивостей біологічних об'єктів.
3) *Генеральна сукупність* :
    A. це весь масив об'єктів однієї категорії, подібних за однаковими ознаками і що різняться за іншими;
    B. частина типових представників генеральної сукупності, яка її відображає, тобто, це частина від цілого.
    C. відбір даних до неї проводиться за принципом рендемічності, випадковості.
4) *Частина типових представників генеральної сукупності, яка її відображає, тобто, частина від цілого називається:*
    A. медіаною;
    B. вибіркою;
    C. генеральною сукупністю.
5) *Коливання значень однієї і тієї ж ознаки в певних межах, які спостерігаються у загальній кількості числових значень це:*
    A. вибірка;
    B. варіація;
    C. ознака.
6) *Термін "біометрія" був введений:*
    A. Гальтоном у 1889р.;

B. Гамільтоном у 1883р.;

C. Ж. Ламарком у 1785 р.

7) *В даний час при біометричних дослідженнях і аналізі даних широко застосовуються методи:*

A. варіаційної статистики (основи побудови варіаційних рядів, властивості генеральної і часткової сукупності величин, закони розподілу варіант);

B. кореляційного та регресивного аналізу даних емпіричних спостережень;

C. теоретичні методи аналізу та самоаналізу.

8) *Впорядковане розміщення варіант у сукупності відповідно до наростання або спадання їх чисельних значень:*

A. варіаційний ряд;

B. ранжирований ряд;

C. продуктивний ряд.

9) *Частка від ділення суми всіх варіант (х) сукупності на їх загальну кількість (n)* називається:

A. середня арифметична*;*

B. середня валідна величина;

C. середня імовірність прояву ознаки.

10) *Середні значення, що застосовуються в біології, діляться на:*

A. цілі та відсоткові значення;

B. параметричні та непараметричні;

C. степеневі та порядкові.

11) *Параметричні середні:……*

12) *Непараметричні* середні:…………

13) *Загальна формула параметричних (або степеневих) середніх така:*

14) *Середня арифметична може бути : ……….*

15) *При аналізі середніх арифметичних вирішуються такі завдання:*

16) *Для того,щоб середні величини мали об'єктивний характер, необхідно дотримуватись наступних умов*:…………………………………………………

17) *Найбільш розповсюдженими помилками є*: ……………………………………..

18) *Точність виміру це…………………………………………………………………..*

19) *Абсолютна помилка* це……………………………………………………

20) *Поясніть термін «середнє квадратичне відхилення», як воно обчислюється?*

21) *Наявність відмінностей в чисельних значеннях ознак у одиниць сукупності називають:*

A. абсолютною помилкою;
B. модою;
C. варіацією ознаки;
D. медіаною.

22) *Найпростішим показником варіації є розмах варіації, який представляє собою:*

A. відношення абсолютних показників варіації до середньої арифметичної (або медіани).
B. різницю між максимальним і мінімальним значеннями ознаки;
C. $R = x_{\max} - x_{\min}$
D. вимірювання ступеню коливання варіантів ознаки від рівня їх середньої величини;
E. розкладання загальної варіації ознаки на варіацію, що породжується систематичними та випадковими причинами.

23) *Зв'язок між окремими факторами, коли одному значенню данного елемента (x) відповідає декілька значень умовно залежного елементу (y), одержав назву:*

A. середнього квадратичного зв'язку;
B. дисперсійного розсіювання;
C. мофологфчно залежним зв'язком;
D. кореляційного зв'язку (залежності).

24) *Ступінь сполучення зв'язку між аргументом (x) і функцією (y) позначається:*

A. коефіцієнтом дисперсії (x) і характеризує щільність дисперсійного зв'язку;
B. коефіцієнтом регресії (R) і характеризує розсіювання регресійного зв'язку;
C. коефіцієнтом кореляції (r) і характеризує тісноту кореляційного зв'язку;
D. коефіцієнтом абсолютної похибки (a) і характеризує абсолютне відхилення.

25) *Виявлення і визначення кореляційних зв'язків між окремими елементами біологічних об'єктів дозволяє:*

A. досить легко проводити математичні обчислення;
B. активно залучати фахівців споріднених спеціальностей до спільних наукових розробок;

C. вдосконалювати методику досліджень живої матерії;

D. порівнювати значення медіани із середнім значенням.

26) *При визначенні критерію достовірності середнього значення нульова гіпотеза полягає в припущенні, що:*

A. між параметрами вибіркової і генеральної сукупності різниця відсутня, тобто $П_в - П_г = 0$;

B. опрацювання матеріалів експериментальних досліджень біологічного або екологічного об'єкту відповідає методами кореляційного аналізу і визначає номінальні ознаки;

C. одна з аналізованих ознак приймається як незалежна (аргумент) (x), друга − як сполучена ознака (у) (функція);

D. характеризується коефіцієнтом кореляції (r), який має значення в межах від 0 до +1 і від 0 до −1.

27) *Множинний коефіцієнт кореляції визначається тоді:*

A. коли знайдені межі довірчого інтервалу свідчать про те, що величина коефіцієнту кореляції в генеральній сукупності знаходяться в межах $0{,}18 < r < 0{,}74$;

B. коли коефіцієнт кореляції, який визначається для відповідної виборки варіант, так саме, як і окремі варіанти, що досліджуються, є величина випадкова;

C. коли на дану ознаку одночасно комплексно впливають дві інших ознаки (фактора) X і Z;

D. коли проводиться визначення номінальних ознак, що обрані для порівняння, шляхом вимірювання, зважування, встановлення кількості, визначення кольору та інше.

28) *Зміна значення функції в залежності від зміни значення аргументу називається:*

A. нульовою гіпотезою;

B. регресією;

C. прогресією;

D. критерієм t-Стьюдента.

29) *Міра відхилення показників варіаційного ряду від середнього арифметичного значення цього ряду (x), називається:*

A. регресією;

B. кореляцією;

C. критерієм Фішера;
D. дисперсією.

30)Нульова гіпотеза полягає:

A. в порівнянні з табличним значенням $F_{табл}$, які представлені за відповідними ступенями вільності;
B. в значеннях результативних ознак, що рівномірно групуються за градаціями комплексів;
C. в проведенні дисперсійного аналізу, в якому одночасно обробляються дані декількох вибіркових сукупностей, які складають дисперсійний комплекс;
D. в припущенні відсутності впливу організованого фактора на результативну ознаку.

СПИСОК РЕКОМЕНДОВАНОЇ ЛІТЕРАТУРИ

1. Атраментова Л.О., Утєвська О.М. Біометрія: підруч. для студ. вищ. навч. закл. – Х.: Ранок, 2007. – 176 с.
2. Багинский В.Ф. Биометрия в лесном хозяйстве: учебное пособие / В.Ф. Багинский, О.В. Лапицкая. – Гомель: ГГУ им. Ф. Скорины, 2010. – 374 с.
3. Близнюченко О. Г. Біометрія. – Полтава : РВВ "TERRA", 2003. – 346 с.
4. Горбунов Л.В. Биометрия: учеб. Пособие/ Л.В.Горбунов, Н.Ф. Клещев. Х.: НТУ «ХПИ», 2014. – 160 с.
5. Калінін М.І. Біометрія: Підручник для студентів вузів біологічних і екологічних напрямків./ Калінін М.І., Єлісєєв В.В. – Миколаїв: Вид-во МФ НаУКМА, 2000. – 204 с.
6. Чепур С.С. Біометрія: Методичний посібник. – Ужгород: Видавництво УжНУ «Говерла», 2015. – 40 с.
7. Горошко М.П. Біометрія / М.П. Горошко, С.І. Миклуш, П.Г. Хомюк. – Львів, Камула, 2004. – 285 с.
8. Горкавий В.К. Математична статистика: навч. посібн. / Горкавий В.К., Ярова В.В. – К.: ВД «Професіонал», 2004. – 384 с.
9. Вуколов Э. Л. Основы статистического анализа. Практикум по статистическим методам и исследованию операции с использованием пакетов STATISTICA и EXCEL: учебное пособие. – 2-е изд., исправ. и доп. — М.: ФОРУМ, 2008. – 464 с. – (Высшее образование).

ДОДАТКИ

Додаток 1

Стандартні значення критерію *t*-Стьюдента
(критерій достовірності- *p*)

Число ступенів вільності	**Рівень імовірності (значущості)**		
	0,95 (0,05)	0,99 (0,01)	0,999 (0,001)
1	12,7	63,7	637,0
2	4,3	9,9	31,6
3	3,2	5,8	12,9
4	2,8	4,6	8,6
5	2,6	4,0	6,9
6	2,4	3,7	6,0
7	2,4	3,5	5,3
8	2,3	3,4	5,0
9	2,3	3,3	4,8
10	2,2	3,2	4,6
11	2,2	3,1	4,4
12	2,2	3,1	4,3
13	2,2	3,0	4,1
14-15	2,1	3,0	4,1
16-17	2,1	2,9	4,0
18-20	2,1	2,9	3,9
21-24	2,1	2,8	3,8
25-28	2,1	2,8	3,7
29-30	2,0	2,8	3,7
31-34	2,0	2,7	3,7
35-42	2,0	2,7	3,6
43-62	2,0	2,7	3,5
62-175	2,0	2,6	3,4
176 і більше	2,0	2,6	3,3

Додаток 2

Стандартні значення критерію F для рівня імовірності 0,95 (значущості 0,05)

Число ступенів вільності (дисперсії знаменнику)	Число ступенів вільності для дисперсії в чисельнику												
	1	2	3	4	5	6	7	8	9	10	12	24	∞
1	164	200	216	225	230	234	237	239	241	242	245	249	254
2	18,5	19, .2	19,2	19,.3	19,4	19,4	19,4	19,4	19,4	19,4	19,4	19,5	19,5
3	10,1	9,6	9,3	9,1	9,0	8,9	8,8	8,8	8,8	8,8	8,7	8,6	8,5
4	7,7	6,9	6,6	6,4	6,3	6,2	6,1	6,0	6,0	6,0	5,9	5,8	5,6
5	6,6	5,8	5,4	5,2	5,1	5,0	4,9	4,8	4,8	4,7	4,7	4,5	4,4
6	6,0	5,1	4,8	4,5	4,4	4,3	4,2	4,2	4,1	4,1	4,0	3,8	3,7
7	5,6	4,7	4,4	4,1	4,0	3,9	3,8	3,7	3,7	3,6	3,6	3,4	3,2
8	5,3	4,5	4,1	3,8	3,7	3,6	3,5	3,4	3,4	3,3	3,3	3,11	2,9
9	5,1	4,3	3,9	3,6	3,5	3,4	3,3	3,2	3,2	3,1	3,1	2,9	2,7
10	5,0	4,1	3,7	3,5	3,3	3,2	3,1	3,2	3,0	3,0	2.9	2,7	2,5
11	4,8	4,0	3,6	3,4	3,2	3,1	3,0	3,0	2,9	2,9	2,8	2,6	2,4
12	4,8	3,9	3,5	3,3	3,1	3,0	2,9	2,8	2,8	2,8	2,8	2,7	2,3
13	4,7	3,8	3,4	3,2	3,0	2,9	2,8	2,8	2,7	2,7	2,6	2,4	2,2
14	4,6	3,7	3,3	3,1	3,0	2,9	2,8	2,7	2,6	2,6	2,5	2,3	2,1
15	4,5	3,7	3,3	3,1	2,9	2,8	2,7	2,6	2,6	2,6	2,5	2,3	2,1
16	4,5	3,6	3,2	3,0	2,9	2,7	2,7	2,6	2,6	2,5	2,4	2,2	2,0
17	4,5	3,6	3,2	3,0	2,8	2,7	2,6	2,6	2,5	2,4	2,4	2,2	2,0
18	4,4	3,6	3,2	2,9	2,8	2,7	2,6	2,5	2,5	2,4	2,3	2,1	1,9
19	4,4	3,5	3,1	2,9	2,7	2,6	2,6	2,5	2,4	2,4	2,3	2,1	1,9
20	4,4	3,5	3,1	2,9	2,7	2,6	2,5	2,4	2,4	2,4	2,3	2,1	1,8
22	4,3	3,4	3,1	2,8	2,7	2,6	2,5	2,4	2,4	2,3	2,2	2,0	1,8
24	4,3	3,4	3,0	2,8	2,6	2,5	2,4	2,4	2,3	2,3	2,2	2,0	1,7
26	4,2	3,4	3,0	2,7	2,6	2,5	2,4	2,3	2,3	2,2	2,2	2,0	1,7
28	4,2	3,3	3,0	2,7	2,6	2,4	2,4	2,3	2,2	2,2	2,1	1,9	1,7
30	4,2	3,3	2,9	2,7	2,5	2,4	2,3	2,3	2,2	2,2	2,.1	2,1	1,6
40	4,1	3,2	2,9	2,6	2,5	2,3	2,2	2,2	2,1	2,1	2,0	1,8	1,5
60	4,0	3,2	2,8	2,5	2,4	2,3	2,2	2,1	2,0	2,0	1,9	1,7	1,4
120	3,9	3,1	2,7	2,5	2,3	2,2	2,1	2,0	2,0	1,9	1,8	1,.6	1,3

Додаток 3

Стандартні значення критерію F для рівня імовірності 0,99 (значущості 0,01)

Число ступенів вільності	Число ступенів вільності для дисперсії в чисельнику												
(дисперсії знаменнику)	1	2	3	4	5	6	7	8	9	10	12	24	∞
1	змінюється від 4052 до 6366												
2	98,5	99,0	99,2	99,3	99,3	99,4	99,3	99,3	99,4	99,4	99,4	99,5	99,5
3	34,1	30,8	29,5	28,7	28,2	27,9	27,7	27,5	27,3	27,2	27,1	26,6	26,4
4	21,2	18,0	16,7	16,0	15,5	15,2	15,0	14,8	14,7	14,5	14,4	13,9	13,5
5	16,3	13,3	12,1	11,4	11,0	10,7	10,4	10,3	10,2	10,0	9,9	9,5	9,0
6	13,7	10,9	9,8	9,2	8,8	8,5	8,3	8,1	8,0	7,9	7,7	7,7	6,9
7	12,3	9,6	8,5	7,9	7,5	7,2	7,0	6,8	6,7	6,6	6,5	6,1	5,7
8	11,3	8,7	7,6	7,0	6,6	6,4	6,2	6,0	5,9	5,8	5,7	5,3	4,9
9	10,6	8,0	7,0	6,4	6,1	5,8	5,6	5,5	5,4	5,3	5,1	4,7	4,3
10	10,0	7,6	6,6	6,0	5,6	5,4	5,2	5,1	5,0	4,8	4,7	4,3	3,9
11	9,7	7,2	6,2	5,7	5,3	5,1	4,9	4,7	4,6	4,5	4,4	4,0	3,6
12	9,3	6,9	6,0	5,4	5,1	4,8	4,6	4,5	4,4	4,3	4,2	3,8	3,4
13	9,1	6,7	5,7	5,2	4,9	4,6	4,4	4,3	4,2	4,1	4,0	3,6	3,2
14	8,9	6,5	5,6	5,0	4,7	4,5	4,3	4,1	4,0	3,9	3,8	3,4	3,0
15	8,7	6,4	5,4	4,9	4,6	4,3	4,1	4,0	3,9	3,8	3,7	3,3	2,9
16	8,5	6, ,2	5,3	4,8	4,4	4,2	4,0	3,9	3,8	3,7	3,6	3,2	2,8
17	8,4	6,1	5,2	4,7	4,3	4,1	3,9	3,8	3,7	3,6	3,5	3,1	2,7
18	8,3	6,0	5,1	4,6	4,3	4,0	3,8	3,7	3,6	3,5	3,4	3,0	2,6
19	8,2	5,9	5,0	4,5	4,2	3,9	3,8	3,6	3,5	3,4	3,3	2,9	2,5
20	8,1	5,9	4,9	4,4	4,1	3,9	3,7	3,6	3,5	3,4	3,2	2,9	2,4
22	7,9	5,7	4,8	4,3	4,0	3,8	3,6	3,5	3,4	3,3	3,1	2,8	2,3
24	7,8	5,6	4,7	4,2	3,9	3,7	3,5	3,3	3,2	3,1	3,0	2,7	2,2
26	7,7	5,5	4,6	4,1	3,8	3,6	3,4	3,3	3,2	3,1	3,0	2,6	2,1
28	7,6	5,5	4,6	4,1	3,8	3,5	3,4	3,2	3,1	3,0	2,9	2,5	2,1
30	7,6	5,4	4,5	4,0	3,7	3,5	3,3	3,2	3,1	3,0	2,8	2,5	2,0
40	7,3	5,2	4,3	3,8	3,5	3,3	3,1	3,0	2,9	2,8	2,7	2,3	1,8
60	7,1	5,0	4,1	3,7	3,3	3,1	3,0	2,8	2,7	2,6	2,5	2,1	1,6
120	6,9	4,8	4,0	3,5	3,2	3,0	2,8	2,6	2,5	2,4	2,3	2,0	1,4
∞	6,6	4,6	3,8	3,3	3,0	2,8	2,6	2,5	2,4	2,3	2,2	1,8	1,0

Додаток 4

Значення Z для відповідних коефіцієнтів кореляції

r	Соті долі коефіцієнта кореляції *r*									
	0	1	2	3	4	5	6	7	8	9
0,0	0,000	0,010	0,020	0,030	0,040	0,050	0,060	0,070	0,080	0,090
0,1	0,100	0,111	0,121	0,131	0,141	0,151	0,161	0,172	0,182	0,192
0,2	0,203	0,213	0,224	0,234	0,245	0,255	0,266	0,277	0,288	0,299
0,3	0,310	0,321	0,332	0,343	0,354	0,365	0,373	0,388	0,400	0,412
0,4	0,424	0,436	0,448	0,460	0,472	0,485	0,498	0,510	0,523	0,536
0,5	0,549	0,563	0,576	0,590	0,604	0,618	0,633	0,648	0,663	0,678
0,6	0,693	0,709	0,725	0,741	0,758	0,776	0,793	0,811	0,829	0.848
0,7	0,867	0,887	0,908	0,929	0,951	0,973	0,996	1,020	1,045	1,071
0,8	1,099	1,127	1,157	1,188	1,221	1,256	1,293	2,092	1,376	1,422
0,9	1,472	1,528	1,589	1,658	1,738	1,832	1,946	2,092	2,298	2,647
0,99	2,647	2,700	2,759	2,826	2,903	2,995	3,106	3,250	3,453	3,800

Додаток 5

Значення функції Пуассона $P_n(\mathrm{m}) = \dfrac{a^m}{m!} e^{-a}$

m					Значення $a \cong pn^*$					
	0,5	0,7	1,0	1,5	2,0	3,0	4,0	5,0	7,0	10,0
0	6065	4966	3679	2231	1353	0498	0183	0067	0009	-
1	3033	34476	3679	3347	2707	1494	0733	0337	0064	0005
2	0758	1217	1839	2510	2707	2240	1465	0842	0223	0023
3	0126	0284	0613	1255	1805	2240	1687	1404	0521	0076
4	0016	0050	0253	0471	0902	1680	1898	1755	0912	0189
5	0002	0007	0031	0141	0361	1008	1708	1755	1277	0378
6		0001	0005	0035	0120	0504	1281	1462	1400	0631
7			0001	0008	0034	0216	0824	1044	1400	0901
8				0001	0009	0081	0463	0653	1304	1126
9					0002	0027	0232	0363	1014	1251
10						0008	0104	0181	0710	1251
11							0043	0082	04542	1137
12							0016	0034	0264	0948
13							0006	0013	0142	0729
14							0002	0005	0071	0521

* В разі необхідності значення будь-яких градацій 0,1 і більших значень *m* можна знайти у відповідних таблицях довідників по варіаційній статистиці.

Додаток 6

МІНІСТЕРСТВО ОСВІТИ І НАУКИ УКРАЇНИ

ЧЕРКАСЬКИЙ ДЕРЖАВНИЙ ТЕХНОЛОГІЧНИЙ УНІВЕРСИТЕТ

БУДІВЕЛЬНИЙ ФАКУЛЬТЕТ

КАФЕДРА ЕКОЛОГІЇ

ЖУРНАЛ ЗВІТІВ

лабораторних робіт
з освітньої компоненти «Біометрія»

Виконав:
здобувач вищої освіти __курсу,
групи _____
освітньо-професійної програми «Екологія та охорона навколишнього середовища»
спеціальності 101 «Екологія»
____
(прізвище та ініціали здобувача)
Керівник _____________
(прізвище та ініціали керівника)

Черкаси 20____ року

Додаток 7

МІНІСТЕРСТВО ОСВІТИ І НАУКИ УКРАЇНИ

ЧЕРКАСЬКИЙ ДЕРЖАВНИЙ ТЕХНОЛОГІЧНИЙ УНІВЕРСИТЕТ

БУДІВЕЛЬНИЙ ФАКУЛЬТЕТ

КАФЕДРА ЕКОЛОГІЇ

ЛАБОРАТОРНА РОБОТА №__

з освітньої компоненти «**Біометрія**»

Тема:____

Лабораторну роботу прийняв:

____

“___“__________20___ р.

Лабораторну роботу виконав:
студент групи ____________

____

“___“____________20___ р

Черкаси 20____ р