Міністерство освіти і науки України
Національний університет водного господарства та природокористування
Навчально-науковий інститут автоматики, кібернетики та обчислювальної техніки
Кафедра вищої математики

04-02-55М

МЕТОДИЧНІ ВКАЗІВКИ

до самостійної роботи та підготовки до практичних занять з навчальної дисципліни «Статистичний аналіз в обробці експериментальних даних»
для здобувачів вищої освіти другого (магістерського) рівня
за освітньо-професійною програмою
«Промислове та цивільне будівництво»
спеціальності 192 «Будівництво та цивільна інженерія»
денної та заочної форм навчання

Рекомендовано науково-методичною
радою з якості ННІ БА
Протокол № 8 від 20.06.2023 р.

Рівне – 2023

Методичні вказівки до самостійної роботи та підготовки до практичних занять з навчальної дисципліни «Статистичний аналіз в обробці експериментальних даних» для здобувачів вищої освіти другого (магістерського) рівня за освітньо-професійною програмою «Промислове та цивільне будівництво» спеціальності 192 «Будівництво та цивільна інженерія» денної та заочної форм навчання [Електронне видання] / Цецик С. П., Іващук Я. Г. – Рівне : НУВГП, 2023. – 74 с.

Укладачі:

Цецик С. П., кандидат пед. наук, доцент кафедри вищої математики;

Іващук Я. Г., кандидат фіз.-мат.н., доцент кафедри вищої математики.

Відповідальний за випуск: Тадеєв П. О., к.фіз.-мат.н., д.п.н., професор, завідувач кафедри вищої математики.

Керівник групи забезпечення спеціальності:
Масюк Г. Х., к.т.н., професор кафедри промислового, цивільного будівництва та інженерних споруд.

ЗМІСТ

ВСТУП

Методичні вказівки до самостійної роботи та підготовки до практичних занять навчальної дисципліни «Статистичний аналіз в обробці експериментальних даних» складені відповідно до освітньо-професійної програми «Промислове та цивільне будівництво» спеціальності 192 «Будівництво та цивільна інженерія» за другим (магістерським) рівнем вищої освіти, затвердженої Вченою радою НУВГП, протокол № 6 від 29.06.2017 р.

Вивченню дисципліни передує отримання компетентностей з дисципліни «Вища математика». Знання та навички, отримані під час вивчення дисципліни, допоможуть здобувачам вищої освіти оволодіти компетентностями з дисциплін: «Залізобетонні конструкції інженерних споруд», «Сучасні технології зведення будівель і споруд» та «Варіаційні методи в розрахунках будівельних конструкцій», а також допоможуть успішно написати та захистити кваліфікаційну роботу.

1. Програма навчальної дисципліни «Статистичний аналіз в обробці експериментальних даних»

Ступінь вищої освіти	магістр
Освітня програма	Промислове і цивільне будівництво
Спеціальність	192 Будівництво та цивільна інженерія
Рік навчання, семестр	1-й рік, 1 семестр
Кількість кредитів	3,0
Лекції:	20 годин/2 години
Практичні заняття:	10 годин/8 годин
Самостійна робота:	60 годин/ 80 годин
Форма навчання	денна / заочна
Форма підсумкового контролю	залік
Мова викладання	українська

2. Мета і завдання навчальної дисципліни

Мета освітньої компоненти «Статистичний аналіз в обробці експериментальних даних»: формування у студентів теоретичних основ математично-статистичної обробки даних експериментальних досліджень та уміння планувати експеримент, обирати адекватні методи обробки експериментального матеріалу і коректно їх використовувати.

Основними завданнями вивчення дисципліни є формування теоретичних знань та практичних навичок у відповідності до поставленої мети.

3. Програмні результати навчання

Згідно з освітньо-професійною програмою «Промислове та цивільне будівництво» спеціальності 192 «Будівництво та цивільна інженерія» за другим (магістерським) рівнем вищої освіти навчальна компонента спрямована на те, щоб за її результатами студенти могли:

- РН 1. Вміння провести постановку і проведення експериментів, метрологічне забезпечення, збір, обробку та аналіз результатів, ідентифікацію теорії і експерименту.
- РН 2. Здатність та уміння сприймати та розуміти науково-технічну вітчизняну та іноземну літературу зі спеціальності, складати науково-технічну документацію та спілкуватися на професійні теми, у тому числі іноземною мовою.
- РН 3. Вміння провести розробку інноваційних матеріалів, технологій, конструкцій і систем, розрахункових методик, в тому числі з використанням наукових досягнень.
- РН 10. Вміння застосувати системний підхід до вирішення інженерних проблем на основі досліджень в процесі проектуванні, зведенні, експлуатації та утриманні об'єктів будівництва та цивільної інженерії.
- РН 11. Здатність застосовувати набуті теоретичні знання з фундаментальних і прикладних дисциплін в інженерній практиці відповідно до спеціалізації.

- ВРН 14. Здатність пояснювати процеси, що відбуваються на основних етапах дослідження, проектування, експлуатації, утримання, реконструкції об'єктів будівництва та цивільної інженерії.

4. Компетентності здобувача

Під час вивчення дисципліни «Вища математика» у здобувача вищої освіти магістерського рівня за освітньо-професійною програмою «Промислове та цивільне будівництво» спеціальності 192 «Будівництво та цивільна інженерія» формуються наступні компетентності:

загальні компетентності:

- ЗК1. Здатність до абстрактного мислення, аналізу та синтезу.

- ЗК 3. Навички використання інформаційних і комунікаційних технологій.
- ЗК 4. Здатність до пошуку, оброблення та аналізу інформації з різних джерел та проведення досліджень на відповідному рівні.

спеціальні (фахові) компетентності:

- СК 4. Здатність проводити збір, систематизацію та аналіз вихідних даних для моніторингу проектування забудови, реконструкції будівель і споруд.
- СК 6. Здатність обґрунтовувати варіанти проектних рішень, методів організації та впровадження робіт на різних стадіях проектування, зведення і експлуатації об’єктів будівництва та цивільної інженерії.
- СК 7. Здатність розуміти і враховувати потреби користувачів, соціальні, екологічні, етичні, економічні та комерційні міркування, у процесі проектування будівельних об’єктів та реалізації технічних рішень в будівництві.
- СК 9. Здатність застосовувати системний підхід до вирішення інженерних проблем на основі досліджень в рамках спеціалізації.
- СК 11. Здатність до самостійної науково-дослідної роботи у галузі методології архітектурно-конструктивного проектування житлових, громадських й промислових будівель і споруд.
- СК 15. Здатність враховувати сучасні тенденції проектування в галузі будівництва та цивільної інженерії та вміти вибирати і застосовувати на практиці методи дослідження, планування і проводити необхідні експерименти, інтерпретувати результати і робити висновки щодо оптимальності рішень, що приймаються.

- ВСК 16. Здатність проводити обстеження, діагоностику, розрахунки, проектування, випробування та експлуатацію об’єктів в будівництві та цивільній інженерії.

5. Теоретична частина

5.1. Основні завдання та методи математичної статистики. Емпіричні розподіли. Показники вибірки

Математична статистика – це сучасна галузь знань математичної науки, яка займається статистичним описом результатів експериментів і спостережень, а також побудовою математичних моделей, що містять поняття ймовірності. Теоретичною основою математичної статистики є теорія ймовірностей.

В структурі математичної статистики виділяють два основні розділи: описова статистика та статистичні висновки.

Описова статистика використовується для:

- узагальнення показників однієї змінної (статистика випадкової вибірки);
- виявлення взаємозв'язків між двома і більше змінними (кореляційно-регресійний аналіз).

Описова статистика дає можливість отримати нову інформацію, швидше зрозуміти і всебічно оцінити її, тобто виконує *функцію опису* об'єктів дослідження. Тобто методи описової статистики перетворюють сукупність окремих емпіричних даних на систему наочних для сприйняття форм і чисел: розподіли частот; показники тенденцій, варіативності, зв'язку. Цими методами розраховуються статистики випадкової вибірки, що є основою для здійснення статистичних висновків.

Статистичні висновки дають можливість:

- оцінити точність, надійність і ефективність вибіркових статистик, виявити похибки, які виникають у процесі статистичних досліджень (статистичне оцінювання);

- узагальнити параметри генеральної сукупності, отримані на підставі вибіркових статистик (перевірка статистичних гіпотез).

Головна мета наукових досліджень – отримання нових знань про великі класи явищ, осіб або подій, які називають *генеральною сукупністю*.

Генеральною називається вся сукупність однотипних об’єктів, яка підлягає вивченню.

Вибірковою сукупністю або просто *вибіркою* називається сукупність *випадково* відібраних об’єктів із генеральної сукупності.

Обсягом сукупності (генеральної або вибіркової) називається число об’єктів цієї сукупності. Обсяг генеральної сукупності позначають літерою N, а вибіркової — n.

Для правомірності висновків про досліджувану ознаку об’єктів генеральної сукупності на підставі опрацювання вибірки необхідно, щоб об’єкти вибірки правильно представляли генеральну сукупність, тобто вибірка повинна володіти властивістю репрезентативності (представницькості). Випадковість відбору об’єктів у вибіркову сукупність і використання закону великих чисел дозволяють вирішити питання про репрезентативність вибірки.

На практиці використовуються різні способи утворення вибірки, які принципово розподіляються на два види:

1) відбір, що не вимагає розчленування генеральної сукупності на частини (*простий (власне випадковий) відбір*);

2) відбір, при якому генеральна сукупність розбивається на частини (*типовий* відбір, *механічний* відбір, *серійний* відбір, *комбінований* відбір).

Простим випадковим називається такий відбір, при якому об’єкти відбираються по одному випадковим чином із усієї генеральної сукупності. Проста випадкова вибірка може бути *повторною* або *безповторною. Повторною* називається вибірка, при утворенні якої відібраний об’єкт (перед відбором наступного) повертається в генеральну сукупність. *Безповторною* називається вибірка, в процесі утворення якої відібраний об’єкт в генеральну сукупність не повертається.

Нехай досліджується кількісна ознака X об’єктів генеральної сукупності. Будемо вважати, що вона є одновимірною випадковою величиною. Після опрацювання n об’єктів вибіркової сукупності отримуються n чисел

$$x_1, x_2, x_3, \ldots, x_i, \ldots, x_n,$$

які називаються *варіантами* і утворюють *ряд варіант* або *простий статистичний ряд.*

Первинна обробка ряду варіант полягає у групуванні рівних варіант цього ряду. Ряд варіант розташуємо в порядку зростання і у відповідності з цим перенумеруємо їх. В результаті одержиться послідовність чисел

$$x_1 \le x_2 \le x_3 \le \ldots \le x_i \le \ldots \le x_n,$$

яка називається *варіаційним рядом*. Якщо серед цієї послідовності є однакові варіанти, тому їх ще раз перенумеруємо, залишаючи один і той самий номер однаковим варіантам. Нехай у варіаційному ряді варіанта x_1 повторюється n_1 разів, x_2 — n_2 разів, ..., x_i — n_i разів. Числа n_i називаються *частотами (абсолютними частотами),* а їх відношення до обсягу вибірки $\omega_i = n_i / n$ — *відносними частотами*. Із цих означень випливають рівності:

$$\sum_{i=1}^{k} n_i = n, \quad \sum_{i=1}^{k} w_i = 1. \tag{1.1}$$

Статистичним розподілом вибірки називається відповідність між варіантами та частотами або відносними частотами:

$$\begin{array}{c|cccc} x_i & x_1 & x_2 & \ldots & x_k \\ \hline n_i & n_1 & n_2 & \ldots & n_k \end{array}, \tag{1.2}$$

$$\begin{array}{c|cccc} x_i & x_1 & x_2 & \ldots & x_k \\ \hline w_i = {}^{n_i}\!/_{n} & w_1 & w_2 & \ldots & w_k \end{array}. \tag{1.3}$$

У більшості випадків статистичний розподіл вибірки у вигляді (1.2) або (1.3) використовується тоді, коли ряд варіант є реалізацією *дискретної* випадкової величини X. Якщо ж X - неперервна випадкова величина, то статистичний розподіл вибірки задається у вигляді відповідності між інтервалами і частотами або відносними частотами тих варіант, які потрапляють у ці інтервали, тобто у вигляді таблиць:

$$\begin{array}{ccccc} [x_i, x_{i+1}) & [x_1, x_2) & [x_2, x_3) & \dots & [x_k, x_{k+1}] \\ n_i & n_1 & n_2 & \dots & n_k \end{array}, \qquad (1.4)$$

$$\begin{array}{c|cccc} [x_i, x_{i+1}) & [x_1, x_2) & [x_2, x_3) & \dots & [x_k, x_{k+1}] \\ \hline w_i = {}^{n_i}\!/_{n} & w_1 & w_2 & \dots & w_k \end{array}. \qquad (1.5)$$

Ці таблиці називаються *інтервальним статистичним розподілом вибірки.*

Приклад 1. Під час дослідження кількісної ознаки X із генеральної сукупності було отримано вибірку:

1, 6, 4, 5, 1, 5, 9, 3, 5, 10, 2, 6, 8, 9, 2, 6, 5, 3, 7, 9.

Знайти обсяг вибірки, побудувати варіаційний ряд вибірки, її статистичний розподіл і розподіл відносних частот.

Розв'язання. Обсяг вибірки – 20.

x_i	1	2	3	4	5	6	7	8	9	10	$\sum$
n_i	2	2	2	1	4	3	1	1	3	1	20
ω_i	0,1	0,5	0,1	0,05	0,2	0,15	0,05	0,05	0,15	0,05	1

При побудові інтервального статистичного розподілу на основі ряду варіант розглядається k інтервалів однакової довжини. Кількість інтервалів можна визначати наближено за формулою:

$$k = 1+3,322 \lg n \qquad (1.6)$$

Оптимальну ширина інтервалу h – за формулою:

$$h = \frac{x_{\max} - x_{\min}}{1 + 3{,}322 \cdot \lg n}, \quad (1.7)$$

де $x_{\max}, x_{\min}$ - найбільше і найменше значення варіанти у вибірці; n – обсяг вибірки.

Одна із задач математичної статистики — оцінка (наближене знаходження) невідомої функції розподілу $F(x)$ імовірностей кількісної ознаки X об'єктів генеральної сукупності. За означенням $F^*(x) = P(X<x)$.

Емпіричною функцією розподілу вибірки називається функція $F^*(x)$, яка для будь-якого значення x визначає відносну частоту події, що задовольняє умові $X < x$, тобто випадкова величина прийме значення менше за x:

$$F^*(x) = \frac{n_x}{n}, \quad (1.8)$$

де n_x – сума частот варіант для значень аргументу, менших x, n – обсяг вибірки.

Тобто, емпірична функція розподілу визначається шляхом послідовного додавання відносних частот варіант, менших за x.

Приклад 2. Записати та побудувати емпіричну функцію розподілу за заданим статистичним розподілом вибірки

x_i	1	3	5	7	9
n_i	2	3	6	5	4

Розв'язання.

Обсяг вибірки – 20.

$$
F^*(x) = \begin{cases} \dfrac{0}{20} = 0, \; якщо \; x \le 1, \\ \dfrac{2}{20} = 0{,}1, \; якщо \; 1 < x \le 3, \\ \dfrac{2+3}{20} = 0{,}25, \; якщо \; 3 < x \le 5, \\ \dfrac{2+3+6}{20} = 0{,}55, \; якщо \; 5 < x \le 7, \\ \dfrac{2+3+6+5}{20} = 0{,}8, \; якщо \; 7 < x \le 9, \\ \dfrac{2+3+6+5+4}{20} = 1, \; якщо \; x > 1. \end{cases}
$$

У процесі аналізу статистичних даних важливу роль відіграє геометрична ілюстрація цих даних. Для наочності будують різні графіки статистичних розподілів, зокрема полігон і гістограму.

Полігоном частот називають ламану, відрізки якої з'єднують точки $(x_1;n_1), (x_2;n_2), \ldots, (x_k;n_k)$, де x_i – варіанти вибірки; n_i – відповідні частоти.

Для побудови полігона частот на осі абсцис відкладають варіанти, а на осі ординат — відповідні їм частоти (рис. 1).

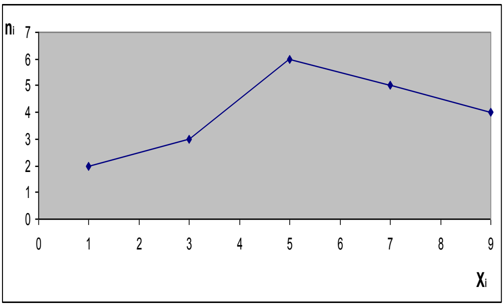


Рис. 1. Полігон частот

Полігоном відносних частот називають ламану, відрізком якої з'єднують точки $(k_1;\omega_1), \ldots, (x_k;\omega_k)$, де ω_i – відповідні емпіричні ймовірності.

Гістограмою частот називається сходинкова фігура, що складається із прямокутників, основами яких є частинні інтервали довжиною h, а висоти дорівнюють відношенню $\frac{n_i}{h}$ *(густина частоти)*. Для побудови гістограми частот на осі абсцис відкладаються частинні інтервали, а над ними проводяться прямолінійні відрізки, паралельні осі абсцис на віддалі $\frac{n_i}{h}$. Площа i-го частинного прямокутника дорівнює $h \cdot \frac{n_i}{h} = n_i$, тобто сумі частот тих варіант, що потрапляють в i-ий інтервал. Тому площа гістограми частот дорівнює сумі всіх частот вибірки n.

Приклад 3. Задано інтервальний статистичний розподіл вибірки:

$(x_i ; x_{i+1}]$	$(0;2]$	$(2;4]$	$(4;6]$	$(6;8]$	$(8;10]$
n_i	7	8	10	9	6
ω_i	0,18	0,20	0,25	0,22	0,15
$x_{i\,\text{сер}}$	1	3	5	7	9

Побудувати:

а) полігон відносних частот;

б) гістограму частот;

в) гістограму відносних частот.

Розв'язання.

а) Побудуємо полігон відносних частот, де $\omega_i = n_i / n$, $n = 40$ (рис. 2).

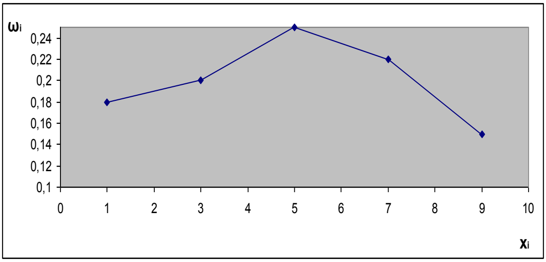


Рис. 2. Полігон відносних частот

б) Побудуємо гістограму частот (рис. 3). Визначимо:

$$h = x_{i+1} - x_i = 2; \quad n_{1h} = \frac{n_1}{h} = \frac{7}{2} = 3{,}5; \quad n_{2h} = \frac{n_2}{h} = \frac{8}{2} = 4;$$

$$n_{3h} = \frac{n_3}{h} = \frac{10}{2} = 5; \quad n_{4h} = \frac{9}{2} = 4{,}5; \quad n_{5h} = \frac{6}{2} = 3.$$

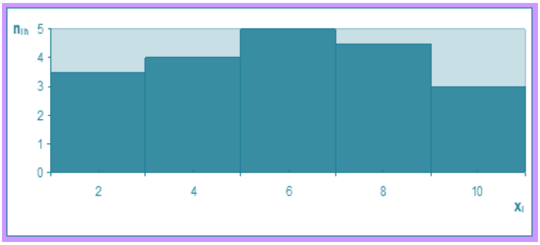


Рис. 3. Гістограма частот

в) Побудуємо гістограму відносних частот (рис. 4). Обчислимо щільності відносних частот ω_i / h:

$$\omega_1^h = \omega_1 / h = 0{,}18 / 2 = 0{,}09; \quad \omega_2^h = \omega_2 / h = 0{,}2 / 2 = 0{,}1;$$

$$\omega_3^h = \omega_3 / h = 0{,}25 / 2 = 0{,}125; \quad \omega_4^h = \omega_4 / h = 0{,}22 / 2 = 0{,}11;$$

$$\omega_5^h = \omega_5 / h = 0{,}15 / 2 = 0{,}075.$$

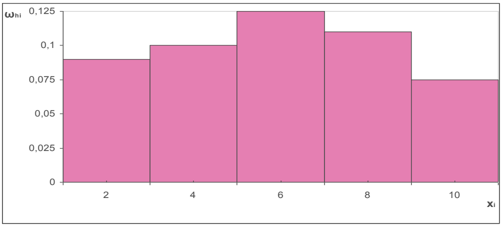


Рис. 4. Гістограма відносних частот

Визначення найбільш суттєвих особливостей статистичного розподілу вибірки передбачає знаходження її числових характеристик, які є оцінками (наближеними значеннями) невідомих параметрів розподілу кількісної ознаки генеральної сукупності.

Основними числовими характеристиками статистичного розподілу вибірки є: характеристики положення (вибіркова середня, мода, медіана); показники варіації (розмах варіації, вибіркова дисперсія, вибіркове середнє квадратичне відхилення, коефіцієнт варіації); показники асиметрії та ексцесу.

Середнє арифметичне називають *вибірковим середнім* і позначають $\bar{x}$.

Якщо кожна з варіант у точковому варіаційному ряді зустрічається тільки один раз, то середнє арифметичне називається *простим* і обчислюється за формулою:

$$\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i = \frac{x_1 + x_2 + ... + x_n}{n}. \qquad (1.9)$$

Нехай задано дискретний статистичний розподіл вибірки:

x	x_1	x_2	…	x_i	…	x_k
n	n_1	n_2	…	n_i	…	n_k

Варіанти у ньому повторюється певне число разів, то вибіркове середнє називають *зваженим* **(**середньозваженим) і обчислюють за формулою:

$$\overline{x} = \frac{1}{n}\sum_{i=1}^{k} x_i n_i = \frac{x_1 n_1 + x_2 n_2 + ... + x_i n_i + ... + x_k n_k}{n}. \qquad (1.10)$$

Означення. *Вибірковим середнім* $(\overline{x})$ статистичного розподілу вибірки називається середнє арифметичне значення її варіант x_i з урахуванням їхніх частот.

Вибіркове середнє характеризує середнє значення ознаки X .

При обчисленні вибіркової середньої інтервального варіаційного ряду, як варіант, беруть середини відповідних інтервалів.

Медіаною M_e дискретного статистичного розподілу вибірки називається таке число, яке ділить варіаційний ряд, що «породжує» цей розподіл, на дві рівні за кількістю варіант частини. Якщо число варіант непарне, тобто n=2m+1, тоді $Me^* = x_{m+1}$. Якщо ж обсяг вибірки є парним числом, тобто n=2m, тоді медіана дорівнює середньому арифметичному «середньої» (медіанної) пари варіант:

$$M_e = \frac{x_m + x_{m+1}}{2}. \qquad (1.11)$$

Формула для обчислення медіани інтервального статистичного розподілу :

$$M_e = x_{M_e} + h\frac{0,5 \cdot n - S_{M_e - 1}}{n_{M_e}}, \qquad (1.12)$$

де x_{M_e} – нижня (мінімальна) меже медіанного інтервалу (медіанним є інтервал на який припадає перша нагромаджена частота, що перевищує половину всього обсягу сукупності;

h – величина інтервалу;

S_{M_e-1} – сума нагромаджених частот інтервалу, який передує медіанному;

n_{M_e} – частота медіанного інтервалу.

Модою M_0 дискретного статистичного розподілу називається варіанта, якій відповідає найбільша частота.

В інтервальних варіаційних рядах мода обчислюється за формулою:

$$M_0 = x_{M_0} + h\frac{n_{M_0} - n_{M_{0-1}}}{2n_{M_0} - n_{M_0-1} - n_{M_0+1}} \qquad (1.13)$$

де, x_{M_0} - нижня (мінімальна) межа модального інтервалу;

h - величина інтервалу;

$n_{M_0}, n_{M_0-1}, n_{M_0+1}$ – частоти відповідно модального, домодального і післямодального інтервалів.

Приклад 4. Знайти моду, медіану, вибіркову середньозважену заданого інтервального ряду розподілу та побудувати кумулятивну криву цього розподілу.

x, *см*, студ.	150-154	154-158	158-162	162-166	166-170	170-174	174-178	178-182
$x_{сер}$	152	156	160	164	168	172	176	180
n_i, к-ть студ.	2	5	15	20	28	16	10	4
S_i	2	7	22	42	70	86	96	100

Розв'язання.

$$M_o = 166 + 4\frac{28-20}{2\cdot 28 - 20 - 16} = 166 + \frac{33}{20} = 167,6\,(см).$$

$$M_e = 166 + 4\frac{0,5\cdot 100 - 42}{20} = 167,6\,(см).$$

$$\bar{x} = \frac{152\cdot 2 + 156\cdot 5 + 160\cdot 15 + 164\cdot 20 + 168\cdot 28 + 172\cdot 16 + 176\cdot 10 + 180\cdot 4}{100} =$$

= 167 см.

Для побудови кумуляти (рис. 5) складемо таблицю:

x	150-154	154-158	158-162	162-166	166-170	170-174	174-178	178-182
$x_{кін}$	154	158	162	166	170	174	178	182
n_i	2	5	15	20	28	16	10	4
ω_i	0,02	0,05	0,15	0,2	0,28	0,16	0,1	0,04
ω_i^H	0,02	0,07	0,22	0,42	0,70	0,86	0,96	1,00

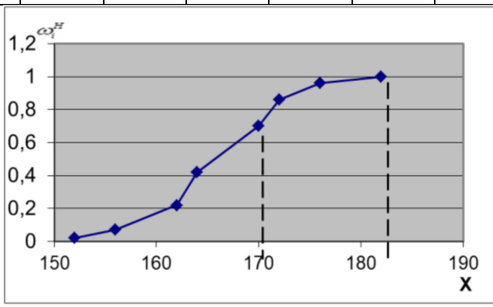


Рис. 5. Кумулята

Одним із недоліків середніх показників є те, що вони не дають уявлення про варіацію (мінливість) значень випадкової величини.

Варіацією ознаки називається наявність відмінності в числових значеннях ознак серед елементів сукупності.

Розмахом вибірки R називають різницю між найбільшим і найменшим значеннями її варіант, тобто:

$$R = x_{\max} - x_{\min}\,. \qquad (1.14)$$

Середнє лінійне відхилення – це середнє арифметичне з абсолютних значень відхилень окремих варіант від середньої арифметичної:

- просте $\bar{l} = \frac{1}{n}\sum_{i=1}^{n}\left|x_i - \bar{x}\right|$, (1.15)
- зважене $\bar{l} = \frac{1}{n}\sum_{i=1}^{k}\left|x_i - \bar{x}\right| n_i, \quad n = \sum_{i=1}^{k} n_i$, (1.16)

Основним недоліком середнього лінійного відхилення є те, що в ньому не враховуються знаки (напрямки) відхилень, а сума всіх відхилень значень випадкової величини від їх середнього значення дорівнює нулю: $\sum_{i=1}^{k}\left(x_i - \bar{x}\right) = 0$.

Вибірковою дисперсією $\overline{D}$ статистичного розподілу називається середнє арифметичне квадратів відхилень варіант від середньої вибіркової. При цьому вирізняють вибіркову дисперсію:

- просту $\overline{D} = \overline{\sigma^2} = \frac{1}{n}\sum_{i=1}^{n}\left(x_i - \bar{x}\right)^2$, (1.17)
- зважену $\overline{D} = \overline{\sigma^2} = \frac{1}{n}\sum_{i=1}^{k}\left(x_i - \bar{x}\right)^2 \cdot n_i$. (1.18)

На практиці зручніше користуватися так званою *розрахунковою формулою для обчислення дисперсії:*

$$D_в = \overline{x^2} - (\bar{x}_в)^2 = \frac{1}{n}\sum_{i=1}^{k} x_i^2 n_i - (\bar{x}_в)^2. \quad (1.19)$$

Недоліком $\overline{D}$ є її розмірність. $\overline{D}$ характеризує середню величину розкиду варіант навколо $\bar{x}_в$ в квадратних одиницях. Для виправлення цього недоліку використовується інша числова характеристика:

вибіркове середнє квадратичне відхилення $\overline{\sigma}$:

- просте $\overline{\sigma} = \sqrt{\frac{1}{n}\sum_{i=1}^{n}\left(x_i - \bar{x}\right)^2}$, (1.20)
- зважене $\overline{\sigma} = \sqrt{\frac{1}{n}\sum_{i=1}^{k}\left(x_i - \bar{x}\right)^2 \cdot n_i}$, (1.21)

або $$\overline{\sigma} = \sqrt{\overline{D}}\,. \qquad (1.22)$$

Якщо $\bar{x}_в$ відмінна від нуля, тоді для порівняння двох статистичних розподілів з точки зору їх розмірності відносно середньої вибіркової вводиться показник *коефіцієнт варіації,* який дорівнює відношенню середнього квадратичного відхилення до середньої вибіркової і виражений у відсотках:

$$V = \frac{\overline{\sigma}}{\overline{x}} \cdot 100\%. \qquad (1.23)$$

Показники асиметрії (несиметричності) та ексцесу (гостровершинності) використовують, щоб оцінити відхилення статистичного розподілу вибірки від нормального розподілу.

Асиметрія (несиметричність) є мірою несиметричності варіаційного ряду, що обчислюється за формулою:

$$A_s = \frac{1}{n\overline{\sigma}^3} \sum_{i=1}^{k} \left(x_i - \overline{x}\right)^3 n_i\,. \qquad (1.24)$$

Якщо $\overline{x} > M_o$, то асиметрія ($As > 0$) називається правосторонньою, при $\overline{x} < M_o$ - лівосторонньою ($As < 0$). Для симетричного варіаційного ряду $As = 0$.

Ексцес є характеристикою більшої чи меншої «вершинності» полігону чи гістограми порівняно з нормальною кривою, що обчислюється за формулою:

$$E_x = \frac{1}{n\overline{\sigma}^4} \sum_{i=1}^{k} \left(x_i - \overline{x}\right)^4 n_i - 3. \qquad (1.25)$$

Якщо $Ex > 0$, то розподіл значень випадкової величини X називається високо вершинним, при $Ex < 0$ - низьковершинним, при - нормальним.

Показники асиметрії $Ex = 0$ (несиметричності) та ексцесу (гостровершинності) використовують, щоб оцінити відхилення статистичного розподілу вибірки від нормального розподілу.

Якщо $Ex > 0$, то розподіл значень випадкової величини X називається високо вершинним, при $Ex < 0$ – низьковершинним, при – нормальним.

5.2. Статистичні оцінки параметрів генеральної сукупності

Наближені значення параметрів розподілу елементів генеральної сукупності, знайдені на основі вибірки, називають *статистичними оцінками цих параметрів*.

Статистичні оцінки поділяються на точкові та інтервальні.

Точковою називається оцінка, яка визначається одним числом.

Інтервальною називають оцінку, яка визначається двома числами - кінцями інтервалу, що покривають параметр, який оцінюється.

Як правило, основними оцінками параметрів генеральної сукупності є середнє вибіркове $\overline{x}$, яке є аналогом математичного сподівання $M(X)$, і вибіркова дисперсія $\overline{D}$ – аналог дисперсії $D(X)$.

Нехай маємо параметр θ, а θ^* – його вибіркова оцінка. Для того, щоб оцінка θ^* достатньо повно характеризувала параметр генеральної сукупності, необхідно, щоб вона мала наступні властивості: *незміщеність, спроможність, ефективність.*

Оцінка θ^* параметра θ називається *незміщеною*, якщо її математичне сподівання дорівнює заданому параметру, тобто

$$M(\theta^*) = \theta,$$

(наприклад, $\overline{x}_{\text{ген.}} \approx M(\overline{x})$).

Оцінка називається *спроможною (обгрунтованою)*, якщо вона при $n \to \infty$ прямує за ймовірністю до параметра, який оцінюють, тобто

$$\lim_{n\to\infty} P\left(\theta^* \to \theta\right) = 1.$$

Оцінка називається *ефективною*, якщо при заданому об'ємі досліджень, вона має найменшу дисперсію.

Вибіркове середнє $\overline{x}$ є незміщеною оцінкою середнього генерального $\overline{x}_{.ген} = a$, тобто

$$M\left(\overline{x}\right) = \overline{x}_{\text{ген.}} = a.$$

Вибіркова дисперсія є зміщеною оцінкою генеральної дисперсії $D_{\text{ген.}}$, тобто математичне сподівання вибіркової дисперсії не дорівнює генеральній дисперсії:

$$M\left(\overline{D}\right) = \frac{n-1}{n} D_{\text{ген.}}.$$

Величину

$$S_x^2 = S^2 = \frac{n}{n-1}\overline{D} \qquad (2.1)$$

називають *виправленою статистичною дисперсією вибірки*. Вона є незміщеною оцінкою генеральної дисперсії

$$M\left(S_x^2\right) = D_{\text{ген.}}.$$

Незміщеною оцінкою вибіркового середнього квадратичного відхилення є виправлене середнє квадратичне відхилення:

$$S = \sqrt{S^2}. \qquad (2.2)$$

Приклад 5. Задано дискретний статистичний розподіл вибірки

x_i	3	3,5	4	4,5	5
n_i	5	4	9	4	8

Знайти: вибіркове середнє та незміщену оцінку дисперсії (виправлену дисперсію), виправлене середнє квадратичне відхилення, коефіцієнт варіації. *Розв'язання.* Об'єм вибірки: $n = 5+4+9+4+8 = 30$.

Знаходимо вибіркову середню:

$$\bar{x} = \frac{1}{n}\sum_{i=1}^{k} x_i n_i = \frac{3\cdot 5 + 3{,}5\cdot 4 + 4\cdot 9 + 4{,}5\cdot 4 + 5\cdot 8}{30} = 4{,}1.$$

Вибіркова дисперсія:

$$\bar{D} = \frac{1}{n}\sum_{i=1}^{5} x_i^{\,2} n_i - \left(\bar{x}\right)^2 = \frac{3^2\cdot 5 + 3{,}5^2\cdot 4 + 4^2\cdot 9 + 4{,}5^2\cdot 4 + 5^2\cdot 8}{30} - (4{,}1)^2 = 0{,}49.$$

Виправлена дисперсія (незміщена оцінка дисперсії):

$$S^2 = \frac{n}{n-1}\bar{D} = \frac{30}{30-1}\cdot 0{,}49 \approx 0{,}505.$$

Середнє квадратичне відхилення: $\bar{\sigma} = \sqrt{\bar{D}} = \sqrt{0{,}49} = 0{,}7.$

Виправлене середнє квадратичне відхилення

$$S = \sqrt{S^2} = \sqrt{0{,}505} \approx 0{,}71.$$

Коефіцієнт варіації: $V = \dfrac{\bar{\sigma}}{\bar{x}}\cdot 100\% = \dfrac{0{,}7}{4{,}1}\cdot 100\% = 0{,}17\cdot 100\% = 17\%.$

Нехай за даними вибірки знайдена статистична оцінка θ^* невідомого параметра θ, яке будемо вважати постійним числом. Очевидно, що θ^* тим точніше визначає параметр θ, чим менша за абсолютною величиною різниця $\left|\theta^* - \theta\right|$.

Число δ, для якого виконується нерівність $\left|\theta^* - \theta\right| < \delta$, називають *точністю оцінки.*

Надійністю оцінки θ по θ^* називають ймовірність γ, з якою виконується нерівність $\left|\theta^* - \theta\right| < \delta$, тобто, $\gamma = P\left(\left|\theta^* - \theta\right| < \delta\right)$.

Надійним інтервалом називають інтервал $\left(\theta^* - \delta; \theta^* + \delta\right)$, який з заданою надійністю γ накриває невідомий параметр θ.

Для оцінки математичного сподівання a нормально розподіленої кількісної ознаки X за вибірковою середньою $\bar{x}$ при відомому середньому квадратичному відхиленні σ генеральної сукупності використовують надійний інтервал

$$\bar{x} - \sigma \frac{t}{\sqrt{n}} \le a \le \bar{x} + \sigma \frac{t}{\sqrt{n}}, \qquad (2.3)$$

де $\sigma \frac{t}{\sqrt{n}} = \delta$ - точність оцінки;

n - об'єм вибірки;

t - таке значення аргумента функції Лапласа $\Phi(t)$, при якому $\Phi(t) = \frac{\gamma}{2}$.

Надійність γ задають наперед, це наприклад, 0,95, 0,98, 0,99, 0,999.

Окрім надійності γ, використовують ще величину α, що визначає рівень значущості $\alpha = 1 - \gamma$. Ця величина набуває значення: 0,05 , 0,01, 0,001 й ін.

Якщо середнє квадратичну відхилення σ - не відоме (n>30), то для оцінки математичного сподівання a нормально розподіленої ознаки X використовують надійний інтервал

$$\bar{x} - t_\gamma \frac{S}{\sqrt{n}} \le a \le \bar{x} + t_\gamma \frac{S}{\sqrt{n}}, \qquad (2.4)$$

де S - виправлене середнє квадратичне відхилення,

t_γ - значення з таблиці (додаток на с. 72), що знаходять за рівнем значущості γ та об'ємом вибірки n, тобто $t_\gamma = t(\gamma, n)$.

Приклад 6. Знайти надійний інтервал для оцінки з надійністю 0,95 невідомого математичного сподівання a нормально розподіленої ознаки X генеральної сукупності, якщо задано генеральне середнє квадратичне відхилення $\sigma = 5$, вибіркове середнє $\overline{x} = 12{,}57$ і об'єм вибірки $n = 100$.

Розв'язання.

Надійний інтервал знайдемо за формулою (2.3). Знайдемо параметр *t*. Оскільки $\Phi(t) = \frac{\gamma}{2} = \frac{0{,}95}{2} = 0{,}475,$ то за таблицею значень функції $\Phi(t)$ (додаток на с. 70) отримаємо $t = 1{,}96$. Тоді

$$12{,}57 - \frac{5 \cdot 1{,}96}{\sqrt{100}} \le a \le 12{,}57 + \frac{5 \cdot 1{,}96}{\sqrt{100}},$$

$$11{,}59 \le a \le 13{,}56.$$

Отже, з надійністю 95% будь-яке число з цього інтервалу можна взяти за генеральне математичне сподівання.

Для оцінки середнього квадратичного відхилення σ нормально розподіленої кількісної ознаки X з надійністю γ при відомому виправленому вибірковому середньому квадратичному відхиленні *S* використовують надійні інтервали :

$$s(1-q) \le \sigma \le s(1+q), \ (q < 1), \tag{2.5}$$

$$\sigma \le s(1+q), \ q > 1, \tag{2.6}$$

де параметр $q = q(\gamma, n)$ визначають за таблицею (додаток на с.73) за відомими γ, n.

5.3. Перевірка статистичних гіпотез

Статистичною називають гіпотезу про властивості генеральної сукупності, що перевіряється на основі вибірки.

У математичній статистиці виділяють два основні типи статистичних гіпотез:

- гіпотези про закон розподілу ймовірностей випадкової величини (ознаки генеральної сукупності);
- гіпотези про значення параметрів розподілу випадкової величини (ознаки генеральної сукупності).

Статистичні гіпотези першого типу називають *непараметричними*, а другого типу – *параметричними.*

Основною (нульовою) називають висунуту гіпотезу і позначають H_0.

Альтернативною (конкуруючою) називають гіпотезу, яка повністю або частково логічно заперечує нульову гіпотезу, і позначають H_1.

Висунута статистична гіпотеза може бути правильною або хибною. Для перевірки її правильності використовують статистичні дані і статистичні методи, тому перевірку називають *статистичною*.

У результаті статистичної перевірки гіпотези може бути прийняте одне з двох правильних рішень:

1) гіпотеза приймається і вона істинна;

2) гіпотеза відхиляється і вона неістинна.

Поряд із тим у результаті статистичної перевірки статистичної гіпотези можуть бути допущені помилки (прийняті неправильні рішення) двох типів:

1) гіпотеза відхиляється, але вона істинна (помилка першого роду);

2) гіпотеза приймається, але вона неістинна (помилка другого роду).

Виявляється, що помилка першого роду має вагоміші наслідки, ніж помилка другого роду. Щоб застрахувати себе від помилки першого роду або принаймні звести до мінімуму ризик її допущення, вводиться спеціальне число α, яке виражає ймовірність відхилення правильної гіпотези.

Імовірність допущення помилки першого роду називають *рівнем значущості* і позначають через α . Число α задають наперед і найчастіше його вибирають рівним 0,1; 0,05; 0,01. Якщо α =0,05, то це означає, що ймовірність допустити помилку першого роду є мала, а саме – ми ризикуємо її допустити у 5-ти випадках зі 100.

Для перевірки нульової гіпотези вводять певну *числову характеристику,* яку обчислюють на основі вибірки і на підставі якої вирішують: прийняти основну гіпотезу чи альтернативну. Зрозуміло, що вибрана числова характеристика для різних вибірок матиме, загалом кажучи, різні значення, і тому вона є випадковою величиною.

Статистичним критерієм (або просто *критерієм*, чи *статистикою*) називається випадкова величина *К*, яка використовується для перевірки основної гіпотези і закон розподілу якої (точний або наближений) відомий. Для кожного конкретного випадку величина *К* спеціально підбирається і може позначатися різними літерами: *U* або *Z*, якщо вона нормально розподілена, *F* або v^2 — по закону Фішера-Снедокора, χ^2 — по закону “хі-квадрат” і т. д.

Значення випадкової величини *К*, обчислене на основі даних певної вибірки, називають *емпіричним значенням критерію* гіпотези і позначають K_{emp}.

Сукупність значень критерію *К,* за яких нульова гіпотеза H_0 відхиляється, називається *критичною областю*, а сукупність значень критерію *К*, за яких нульову гіпотезу H_0 приймають, називається *областю прийняття гіпотези*.

Точки, що відділяють критичну область від області прийняття, називають *критичними*.

Критерій узгодженості Пірсона (критерій χ^2)

Відповідно до даного критерію емпіричний розподіл вибіркової сукупності, що спостерігається, та який виражено емпіричними частотами n_i згрупованого ряду, порівнюється з припустимим теоретичним розподілом генеральної сукупності, який відображено теоретичними частотами n_i'. Якщо число спостережень дуже велике $(n \to \infty)$, то закон розподілу випадкової величини незалежно від того, якому закону розподілу підпорядкована генеральна сукупність, наближається до розподілу χ^2 з k ступенями свободи, а сам критерій називають критерієм згоди "хі – квадрат" або критерієм Пірсона.

Для перевірки нульової гіпотези H_0 обчислюють величину

$$\chi^2_{емп} = \sum_{i=1}^{s} \frac{(n_i - n_i')^2}{n_i'}, \tag{3.1}$$

де s – кількість інтервалів згрупованого ряду розподілу, n_i – емпіричні частоти, n_i' – теоретичні частоти.

Спостережень n_i в кожному інтервалі повинно бути не менше п'яти відсотків від загального числа спостережень: $n_i \geq 0{,}05n$. Якщо їх буде менше, то необхідно укрупнити інтервали.

Знайдена за формулою (3.1) величина порівнюється з критичними значеннями $\chi^2_\alpha(k)$, які знаходять у спеціальних довідкових таблицях. Число ступенів свободи k визначається за формулою: $k = s - 3$, де s – кількість укрупнених інтервалів.

Якщо $\chi^2{}_{емп} < \chi^2_{0,05}(k)$, то нульова гіпотеза H_0 приймається, тобто припустимий закон розподілу відповідає емпіричним даним, при цьому ми помиляємось у п'яти випадках із ста, приймаючи можливо хибну гіпотезу (похибка другого роду).

Якщо $\chi^2{}_{емп} > \chi^2_{0,01}(k)$, то нульову гіпотезу слід відкинути, тобто припустимий закон розподілу не відповідає емпіричним даним, при цьому ми помиляємось в одному випадку із ста, відкидаючи можливо правильну гіпотезу (похибка першого роду).

Якщо $\chi^2_{0,05}(k) < \chi^2{}_{емп} < \chi^2_{0,01}(k)$, то маємо невизначеність і можна використати інші критерії.

Величина α визначає рівень значущості. Для критерію Пірсона розглядатимемо два рівня значущості: $\alpha = 0{,}05$ і $\alpha = 0{,}01$.

Зауваження. Теоретичні частоти кожного ряду розподілу, тобто, частоти інтервалів (a_i, b_i), при умові, що ознака Х розподілена за нормальним законом обчислюється за формулою:

$$n_i' = n \cdot \left(\Phi(\beta_i) - \Phi(\alpha_i)\right), \tag{3.2}$$

де, $\alpha_i = \dfrac{a_i - \overline{x}}{\overline{\sigma}}$, $\beta_i = \dfrac{b_i - \overline{x}}{\overline{\sigma}}$.

Приклад 7. Задано емпіричні n_i та теоретичні частоти n_i'. Перевірити за критерієм Пірсона при рівні значущості $\alpha = 0{,}05$ гіпотезу про нормальний розподіл випадкової величини, якщо його параметри були оцінені за вибіркою

n_i	8	16	40	72	36	18	10
n_i'	6	18	36	76	39	18	7

Розв'язання. Обчислимо емпіричне значення критерію Пірсона за формулою (3.1) при $s = 7$:

$$K_{емп} = \sum_{i=1}^{7} \frac{(n_i - n_i')^2}{n_i'} = \frac{(8-6)^2}{6} + \frac{(16-18)^2}{18} + \frac{(40-36)^2}{36_i} + \frac{(72-76)^2}{76} +$$

$$+ \frac{(36-39)^2}{39} + \frac{(18-18)^2}{18} + \frac{(10-7)^2}{7} = 3{,}06.$$

За таблицею критичних точок розподілу χ^2 для рівня значущості $\alpha = 0{,}05$ і числа ступенів свободи $k = s - 3 = 7 - 3 = 4$ знаходимо (додаток на с. 71) $k_{кр} = \chi^2_{кр}(0{,}05;4) = 9{,}5.$

Отже, $K_{емп} < k_{кр}$, тому розходження між емпіричними й теоретичними частотами – не значна і гіпотезу H_0 про нормальний розподіл випадкової величини приймаємо.

5.4. Елементи теорії кореляції

Між двома змінними x та y розрізняють два основних види зв'язків: *функціональну* та *стохастичну (статистичну)*.

Функціональна залежність між двома змінними x та y існує, якщо кожному значенню однієї змінної x відповідає певне значення другої змінної – y, і має місце рівняння $y = f(x)$, або навпаки. При функціональній залежності графіки рівнянь $y = f(x)$ і $x = \varphi(y)$ однакові, тобто однаково, яку змінну вважати незалежною змінною, а яку функцією. Тобто функціональна залежність не реагує на напрямленість причинно-наслідкових зв'язків.

Для випадкових величин X і Y не завжди можна встановити функціональну залежність. Між такими величинами існує зв'язок, при якому зі зміною однієї величини змінюється розподіл іншої величини. Такий зв'язок називається *стохастичним* (статистичним). При статистичних зв'язках розрізняють дві компоненти:

1) *стохастична* – пов'язана з безпосередньою залежністю між *функціональною ознакою* і фактором;

2) *випадкова* – пов'язана з впливом випадкових факторів на залежність між функціональною ознакою та фактором. Відсутність другої компоненти приводить до функціональної залежності. При статистичних зв'язках важливі причинно-наслідкові зв'язки, тобто яку із змінних вважати функціональною ознакою, а яку незалежним фактором.

При реальних статистичних даних ми ніколи не отримаємо просту геометричну лінію. Завжди будуть відхилення залежної змінної, які обумовлені помилками вимірів, впливом неврахованих факторів.

Кореляційні зв'язки є частковим випадком статистичних зв'язків. У залежності від причинно-наслідкових зв'язків лінії, які зв'язують змінні x та y, будуть різними.

Кореляційною залежністю називається така залежність між двома випадковими величинами, при якій із зміною однієї з них змінюється середнє значення іншої.

Умовним середнім $\overline{y}_{x_i}$ називають середнє арифметичне значень *Y*, які відповідають $x = x_i$.

Кореляційною залежністю *Y* від *X* називають функціональну залежність умовного середнього $\overline{y}_x$ від *x*. При кореляційній залежності *Y* від *X* будують рівняння умовного середнього $\overline{y}_x$ від x:

$$\overline{y}_x = f(x).$$

За результатами одного і того ж випробування можуть бути отримані дві залежності: $\overline{y}_x$ і $\overline{x}_y$, тобто необхідно вказати, яку із змінних вважати фактором, а яку функціональною ознакою (де причина, а де наслідок).

Рівняння $\overline{y}_x = f(x)$ називають *рівнянням регресії Y* на *X*, функцію $f(x)$ – *регресією* *Y* на *X*, а графік – *лінією регресії* *Y* на *X*. Можна ввести і іншу залежність: $\overline{x}_y = \varphi(y)$ – рівняння регресії *X на* Y.

Поняття регресії і кореляції безпосередньо пов'язані між собою. У той час, як у кореляційному аналізі оцінюється сила стохастичного зв'язку, в регресійному аналізі досліджується його форма. Обидва види аналізу призначені для встановлення причинних співвідношень між явищами та для означення наявності або відсутності зв'язку.

Розрізняють такі *види кореляції та регресії*:

1) *відносно характеру* кореляції та регресії маємо додатну і від'ємну кореляцію та регресію. *Додатна* – коли зі зростанням (зменшенням) аргументу x зростає (зменшується) функція *y*.

2) *відносно числа змінних*: *парна та множинна* кореляція та регресія. Наприклад, $\overline{y} = f(x)$ – парна регресія, $\overline{y} = f(x_1, x_2, x_3, x_4)$ – множинна регресія;

3) *відносно форми зв'язку*: *лінійна* та *нелінійна* кореляція та регресія.

Найпростішим випадком регресійної моделі є лінійна регресія, коли вибіркове рівняння регресії записують:

$$\overline{y}_x = ax + b. \tag{4.1}$$

У цьому випадку точкові оцінки для параметрів a та b задовольняють основні вимоги до точкових оцінок. Основним методом отримання точкових оцінок для параметрів a та b є метод найменших квадратів.

Нехай задано незгруповану вибірку $(x_1, y_1),\ (x_2, y_2), \ldots,\ (x_n, y_n)$ обсягу n.

Ідея методу найменших квадратів полягає в тому, що за точкові оцінки $\overline{a}$ та $\overline{b}$ параметрів a та b вибирають такі числа, для яких пряма $\overline{y}_x = \overline{a}x + \overline{b}$ є найближчою до точок $(x_1, y_1),\ (x_2, y_2), \ldots,$ (x_n, y_n). За міру відхилення шуканої прямої від точок (x_i, y_i) вибирають величину

$$S(a,b) = \sum_{i=1}^{n} \left(y_i - (ax_i + b)\right)^2, \tag{4.2}$$

тобто суму квадратів різниць між ординатами прямої та ординатами точок (x_i, y_i) для одних і тих же значень $x = x_i$. Числа a та b вибирають такими, щоб функція $S(a,b)$ набула найменшого значення, тобто пряма $\overline{y}_x = \overline{a}x + \overline{b}$ найменше відхилялася від точок (x_i, y_i).

Знайшовши частинні похідні функції (4.2) та скориставшись необхідною умовою екстремуму функції двох змінних отримаємо систему

$$\begin{cases} \dfrac{\partial S}{\partial a} = 2\sum\limits_{i=1}^{n}\left(ax_i + b - y_i\right)\cdot x_i = 0, \\ \dfrac{\partial S}{\partial b} = 2\sum\limits_{i=1}^{n}\left(ax_i + b - y_i\right) = 0, \end{cases}$$

або

$$\begin{cases} a\sum\limits_{i=1}^{n} x_i^2 + b\sum\limits_{i=1}^{n} x_i = \sum\limits_{i=1}^{n} x_i y_i \\ a\sum\limits_{i=1}^{n} x_i + \sum\limits_{i=1}^{n} b = \sum\limits_{i=1}^{n} y_i. \end{cases}, \tag{4.3}$$

Із системи (4.3) отримаємо:

$$a = \frac{\overline{xy} - \overline{x}\cdot\overline{y}}{\overline{x^2} - \left(\overline{x}\right)^2}\ , \tag{4.4}$$

$$b = \overline{y} - a\overline{x}. \tag{4.5}$$

Параметр a називають коефіцієнтом регресії Y на X і позначають $\rho_{y/x}$.

Лінійне рівняння регресії Y на *X:*

$$\overline{y}_x - \overline{y} = r_в \frac{\overline{\sigma}_y}{\overline{\sigma}_x}\left(x - \overline{x}\right). \tag{4.6}$$

Відповідно *лінійне рівняння регресії X на Y:*

$$\overline{x}_y - \overline{x} = r_в \frac{\overline{\sigma}_x}{\overline{\sigma}_y}\left(y - \overline{y}\right). \tag{4.7}$$

У рівняннях (4.6) та (4.6) введено величину

$$r_в = a\frac{\overline{\sigma}_x}{\overline{\sigma}_y} = \frac{\overline{xy} - \overline{x}\cdot\overline{y}}{\overline{\sigma}_x\cdot\overline{\sigma}_y}\ , \tag{4.8}$$

яка називається *вибірковим коефіцієнтом кореляції.*

Вибірковий коефіцієнт кореляції має такі *властивості:*

1. Величина $r_в$ є безрозмірною, тобто не залежить від вибору одиниць виміру випадкових величин X та Y.

2. Вибірковий коефіцієнт кореляції $r_в$ за модулем не перевищує одиницю, тобто $|r_в| \le 1.$
3. Вибірковий коефіцієнт кореляції $r_в = \pm 1$ тоді і тільки тоді, коли між випадковими величинами X та Y існує лінійний функціональний зв'язок.
4. Якщо $|r_в| < 0{,}5,$ то зв'язок між випадковими величинами X та Y – слабкий, якщо $|r_в| > 0{,}8,$ то зв'язок – сильний.
5. Якщо між випадковими величинами X та Y відсутній хоча б один із кореляційних зв'язків, то вибірковий коефіцієнт кореляції $r_в = 0.$

Точковою оцінкою коефіцієнта кореляції є вибірковий коефіцієнт кореляції $\hat{r} = r_в$.

Для інтервальної оцінки генерального коефіцієнта кореляції використовують надійний інтервал:

$$r_в - t\sigma_r < r_Г < r_в + t\sigma_r \,, \tag{4.9}$$

де для $n \ge 50$ $\sigma_r = \dfrac{1 - r_в^2}{\sqrt{n}}$, при цьому значення параметра знаходять з умови $\Phi(t) = \dfrac{\gamma}{2}.$

Приклад 8. За даними вибірки пари випадкових величин X та Y знайти вибірковий коефіцієнт кореляції, встановити щільність зв'язку між цими величинами та записати рівняння лінійної регресії Y на X.

x_i	1,00	1,50	3,00	4,50	5,00
y_i	1,25	1,40	1,50	1,75	2,25

Розв'язання. Складемо таблицю:

n	x_i	y_i	$x_i^{\ 2}$	$y_i^{\ 2}$	$x_i \cdot y_i$
1	1,00	1,25	1,00	1,5625	1,250
2	1,50	1,40	2,25	1,96	2,100
3	3,00	1,50	9,00	2,25	4,500
4	4,50	1,75	20,25	3,0625	7,875
5	5,00	2,25	25,00	5,0625	11,250
$\sum$	$\sum x_i = 15{,}00$	$\sum y_i = 8{,}15$	$\sum x^2{}_i = 57{,}50$	$\sum y^2{}_i = 13{,}8975$	$\sum x_i y_i = 26{,}975$

Задана вібірка – незгрупована. Використовуючи обчислення, наведені в таблиці, маємо:

$$\overline{x} = \frac{1}{n}\sum_{i=1}^{n} x_i = \frac{1}{5}\sum_{i=1}^{5} x_i = \frac{15}{5} = 3\,,$$

$$\overline{y} = \frac{1}{n}\sum_{i=1}^{n} y_i = \frac{1}{5}\sum_{i=1}^{5} y_i = \frac{8{,}15}{5} = 1{,}63.$$

$$\overline{xy} = \frac{1}{n}\sum_{i=1}^{n} x_i y_i = \frac{1}{5}\sum_{i=1}^{5} x_i \cdot y_i = \frac{26{,}975}{5} = 5{,}395.$$

$$\overline{D}_x = \frac{1}{n}\sum_{i=1}^{n} x_i^2 - \left(\overline{x}\right)^2 = \frac{1}{5}\sum_{i=1}^{5} x_i^2 - \left(\overline{x}\right)^2 = \frac{57{,}50}{5} - 3^2 = 2{,}5,$$

$$\overline{\sigma_x} = \sqrt{\overline{D_x}} = \sqrt{2{,}5} \approx 1{,}58.$$

$$\overline{D}_y = \frac{1}{n}\sum_{i=1}^{n} y_i^2 - \left(\overline{y}\right)^2 = \frac{1}{5}\sum_{i=1}^{5} y_i^2 - \left(\overline{y}\right)^2 = \frac{13{,}8975}{5} - (1{,}63)^2 \approx 0{,}123.$$

$$\overline{\sigma_y} = \sqrt{\overline{D_y}} = \sqrt{0{,}123} \approx 0{,}35.$$

Обчислимо вибірковий коефіцієнт кореляції:

$$r_в = \frac{\overline{xy} - \overline{x}\cdot\overline{y}}{\overline{\sigma}_x \cdot \overline{\sigma}_y} = \frac{5{,}395 - 3\cdot 1{,}63}{1{,}58\cdot 0{,}35} \approx 0{,}91.$$

Оскільки $r_в \approx 0{,}91 > 0,$ то зв'язок між випадковими величинами X та Y прямий і сильний.

Запишемо лінійне рівняння регресії Y на X:

$$\overline{y}_x - \overline{y} = r_в \frac{\overline{\sigma}_y}{\overline{\sigma}_x}\left(x - \overline{x}\right) \Rightarrow \qquad \overline{y}_x - 1{,}63 = 0{,}91\frac{0{,}35}{1{,}58}\left(x - 3\right)$$

або $\overline{y}_x = 0{,}202x + 1{,}024$.

Приклад 9. Задано кореляційну таблицю випадкових величин X та Y. Знайти вибірковий коефіцієнт кореляції, встановити щільність зв’язку між цими величинами та записати рівняння лінійної регресії Y на X.

y \ x	2	3	4	5	n_y
2	1	2			3
3	1	6	3		10
4		5	5		10
5			1	1	2
n_x	2	13	9	1	25

Розв’язання. Знаходимо числові характеристики двовимірної випадкової величини:

$$\overline{x} = \frac{1}{n}\sum_{i=1}^{4} x_i n_{x_i} = \frac{1}{25}\left(2\cdot 2 + 3\cdot 13 + 4\cdot 9 + 5\cdot 1\right) = 3{,}36,$$

$$\overline{y} = \frac{1}{n}\sum_{i=1}^{4} y_j n_{y_j} = \frac{1}{25}\left(2\cdot 3 + 3\cdot 10 + 4\cdot 10 + 5\cdot 2\right) = 3{,}44,$$

$$\overline{xy}=\frac{1}{n}\sum_{i=1}^{k}\sum_{j=1}^{m}n_{ij}x_iy_i=\frac{1}{25}\Big(\big(1\cdot(2\cdot 2)+2\cdot(3\cdot 2)\big)+\big(1\cdot(2\cdot 3)+6\cdot(3\cdot 3)+3\cdot(4\cdot 3)\big)+$$

$$+\big(5\cdot(3\cdot 4)+5\cdot(4\cdot 4)\big)+\big(1\cdot(4\cdot 5)+1\cdot(5\cdot 5)\big)\Big)=\frac{1}{25}(16+96+140+45)=11{,}88,$$

$$\overline{D}_x=\frac{1}{n}\sum_{i=1}^{k}x_i^2n_{x_i}-\left(\overline{x}\right)^2=\frac{1}{25}(4\cdot 2+9\cdot 13+16\cdot 9+25\cdot 1)-(3{,}36)^2=$$

$$=11{,}76-11{,}29=0{,}47;$$

$$\overline{D}_y=\frac{1}{n}\sum_{i=1}^{m}y_j^2n_{y_j}-\left(\overline{y}\right)^2=\frac{1}{25}(4\cdot 3+9\cdot 10+16\cdot 10+25\cdot 2)-(3{,}44)^2=$$

$$=12{,}48-11{,}83=0{,}65;$$

$$\overline{\sigma_x}=\sqrt{\overline{D_x}}=\sqrt{0{,}47}\approx 0{,}69;$$

$$\overline{\sigma_y}=\sqrt{\overline{D_y}}=\sqrt{0{,}65}\approx 0{,}81.$$

Обчислимо вибірковий коефіцієнт кореляції:

$$r_в=\frac{\overline{xy}-\overline{x}\cdot\overline{y}}{\overline{\sigma}_x\cdot\overline{\sigma}_y}=\frac{11{,}88-3{,}36\cdot 3{,}44}{0{,}69\cdot 0{,}81}\approx 0{,}57.$$

Оскільки $r_в\approx 0{,}57>0,$ то зв'язок між випадковими величинами X та Y прямий і середньої щільності.

Запишемо лінійне рівняння регресії Y на X:

$$\overline{y}_x-\overline{y}=r_в\frac{\overline{\sigma}_y}{\overline{\sigma}_x}\left(x-\overline{x}\right)\Rightarrow\qquad \overline{y}_x-3{,}44=0{,}57\frac{0{,}81}{0{,}69}(x-3{,}36)$$

або $\overline{y}_x=0{,}67x+1{,}19.$

6. Практична частина

Розрахунок статистичних параметрів та кореляційної залежності двомірної вибіркової сукупності

Завдання 1. Розрахувати статистичні параметри і кореляційної залежності між урожайністю зернових культур і кількістю робітників, що працювали на 100 га ріллі в середньому за рік.

Вихідні дані: Об'єм вибірки $n = 50$, надійність $\gamma = 0,9$, рівень значущості $\alpha = 0,1$.

Постановка задачі.
1. Скласти вибірку за отриманими статистичними даними.
2. Скласти кореляційну таблицю та інтервальні ряди розподілів.
3. Знайти елементи рядів розподілу.
4. Зобразити графічно ряди розподілів, побудувавши полігони, гістограми і кумуляти.
5. Безпосередньо визначити числові характеристики рядів розподілу.
6. Знайти моду і медіану рядів розподілу. Показати їх на графіках.
7. Визначити числові характеристики методом моментів.
8. Обчислити вибірковий коефіцієнт кореляції шляхом використання умовних варіант.
9. Скласти вибіркове рівняння прямої регресії y на x і прямої регресії x на y.
10. Побудувати поле кореляції, емпіричні лінії регресії та обидві прямі регресії.
11. Знайти статистичні оцінки параметрів.
12. Знайти надійні інтервали для генеральних середніх та для генерального коефіцієнта кореляції.

13. Знайти надійні зони відхилень від прямих регресії. Побудувати їх на окремих графіках.
14. Перевірити гіпотези про нормальний розподіл обох статистичних величин X і Y.
15. Сформулювати висновки щодо статистичних параметрів і кореляційної залежності між величинами X і Y.
Позначимо ознаки X і Y :
X - кількість робітників, які працювали на 100 га ріллі в середньому за рік (в додатку А відповідає стовбець А5);
Y - урожайність зернових культур , ц/га (в додатку А відповідає стовбець А1).

Виконання роботи
1. Складаємо вибірку за отриманими статистичними даними.

Таблиця 1

Вибіркові дані спостережень ознак *X* і *Y*

№		Значення величин		№		Значення величин	
Вибір-ки	генерал. сукуп-ності	X, чол.	Y, ц/га	Виб ір-ки	генерал. сукуп-ності	X, чол.	Y, ц/га
1	78	18	21,7	**26**	09	20	29,7
2	25	13	16,1	**27**	42	20	31,1
3	93	17	18,2	**28**	06	18	23,2
4	73	23	38,8	**29**	94	18	28,3
5	56	21	29,7	**30**	95	22	33,6
6	34	17	19,3	**31**	91	14	20,1
7	97	20	26,2	**32**	69	22	31,8
8	36	21	34,3	**33**	20	17	20,6
9	43	21	28	**34**	35	20	26,5
10	07	14	16,2	**35**	02	14	20,3
11	21	24	40,6	**36**	13	22	30,3
12	74	20	24,4	**37**	08	23	38,3
13	58	23	37,6	**38**	26	16	22,8
14	64	25	39,9	**39**	90	22	29,5
15	82	22	38,8	**40**	67	19	25,2
16	03	16	21	**41**	27	15	19,8
17	18	20	30	**42**	60	21	28,8
18	17	20	25,5	**43**	51	21	31,4
19	92	18	26	**44**	55	20	31,9
20	32	22	35,8	**45**	63	20	25,1
21	29	19	25,3	**46**	83	22	36,5
22	54	16	24,6	**47**	39	16	15,6
23	66	20	24,2	**48**	99	16	28,2
24	01	16	28	**49**	72	21	37,9
25	79	20	28,9	**50**	59	21	32,6

2. Складаємо кореляційну таблицю та інтервальні ряди розподілів.

Кількість інтервалів наближено може бути визначена з нерівності: $k \geq 1+3,222 \cdot \lg n$, де $n=50$ - чисельність сукупності (об'єм вибірки).

k ≥ 1+ 3,22·lg50 = 6,47; k =7.

Мінімальне значення ознаки X: $x_{\min} = 13$.

Максимальне значення ознаки X: $x_{\max} = 25$.

Величина інтервалу $h_x = \dfrac{x_{\max} - x_{\min}}{k-1} = \dfrac{25-13}{7-1} = 2$.

Початок першого інтервалу:

$a_1 = X_{min} - \frac{1}{2} \cdot h_x = 13 - \frac{1}{2} \cdot 2 = 12;$

Кінець першого інтервалу: $b_1 = a_1 + h_x = 12 + 2 = 14$.

Отримаємо такі інтервали ознаки X:

12-14; 14-16; 16-18; 18-20; 20-22; 22-24; 24-26.

Для ознаки Y маємо:

$Y_{min} = 15{,}6;$ $Y_{max} = 40{,}6;$

$$h_y = \frac{y_{\max} - y_{\min}}{k-1} = \frac{40,6-15,6}{7-1} \approx 4{,}16 \approx 4$$

Початок першого інтервалу:

$a_1 = Y_{min} - \frac{1}{2} \cdot h_y = 15{,}6 - \frac{1}{2} \cdot 4 = 13{,}6;$

Кінець першого інтервалу: $= a_1 + h_y = 13{,}6 + 4 = 17{,}6;$

Отримаємо такі інтервали ознаки Y:

13,6-17,6 ; 17,6-21,6; 21,6-25,6 ; 25,6-29,6; 29,6-33,6; 33,6-37,6; 37,6-41,6.

Якщо при групуванні вибіркових даних ознак X і Y деякі значення співпадатимуть з межею двох інтервалів, то ці значення відносять до другого із цих інтервалів.

Таблиця 2

Групування вибіркових даних ознак X і Y

Y / X	13,6 – 17,6	17,6 – 21,6	21,6 – 25,6	25,6 – 29,6	29,6 – 33,6	33,6 – 37,6	37,6 – 41,6
12 – 14	│						
14 – 16	│	│ │ │					
16 – 18	│	│ │ │ │	│ │	│ │			
18 – 20			│ │ │ │	│ │			
20 – 22			│ │ │ │ │	│ │	│ │ │ │ │		
22 – 24				│ │	│ │ │	│ │ │	│ │ │ │ │
24 –26							│ │

Підрахувавши частоти кожної клітинки та просумувавши їх в рядках і стовпчиках, дістанемо кореляційну таблицю розподілу чоловіків (X, чол.) і розподілу урожайності зернових культур (Y, ц/га).

Таблиця 3

Кореляційна таблиця розподілу

Y / X	13,6 – 17,6	17,6 – 21,6	21,6 – 25,6	25,6 – 29,6	29,6 – 33,6	33,6 – 37,6	37,6 - 41,6	n_i
12-14	1							1
14-16	1	3						4
16-18	1	4	2	2				9
18-20			4	2				6
20-22			5	4	5			14
22-24				2	3	4	5	14
24-26							2	2
n_j	3	7	11	10	8	4	7	50

Для кожного з цих варіаційних рядів знаходимо варіанти x_i і y_j (як середини відповідних інтервалів) та нагромаджені частоти n_i^H і n_j^H .

3. Знаходимо елементи рядів розподілу

Таблиця 4

Елементи ряду розподілу ознаки Х

X	12 – 14	14 – 16	16 – 18	18 – 20	20 – 22	22 – 24	24-26
x_i	13	15	17	19	21	23	24
n_i	1	4	9	6	14	14	2
n_i^H	1	5	14	20	34	48	50

З табл. .4 видно, що медіанним інтервалом ознаки X є інтервал 20-22, бо на цей інтервал припадає перша нагромаджена частота 34, яка є більшою за половину всього обсягу сукупності, тобто більшою за 25. Модальним інтервалом ознаки X є інтервал 20-22, або 22-24, бо цей інтервал має найбільшу частоту.

Таблиця 5

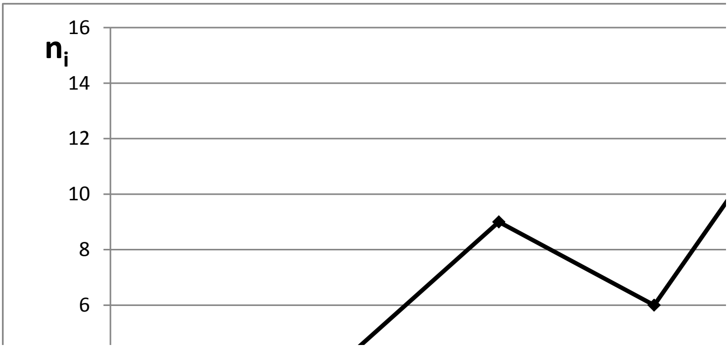
Елементи ряду розподілу ознаки Y

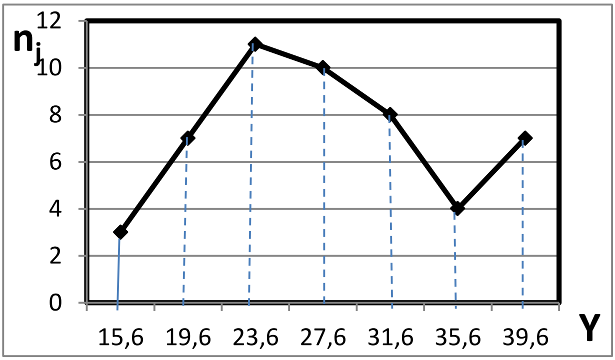
Y	13,6-17,6	17,6-21,6	21,6-25,6	25,6-29,6	29,6-33,6	33,6-37,6	37,6-41,6
y_j	15,6	19,6	23,6	27,6	31,6	35,6	39,6
n_j	3	7	11	10	8	4	7
n_j^H	3	10	21	31	39	43	50

З табл. 5 видно, що медіанним інтервалом ознаки Y є інтервал 25,6-29,6, а модальним інтервалом є інтервал 21,6-25,6.

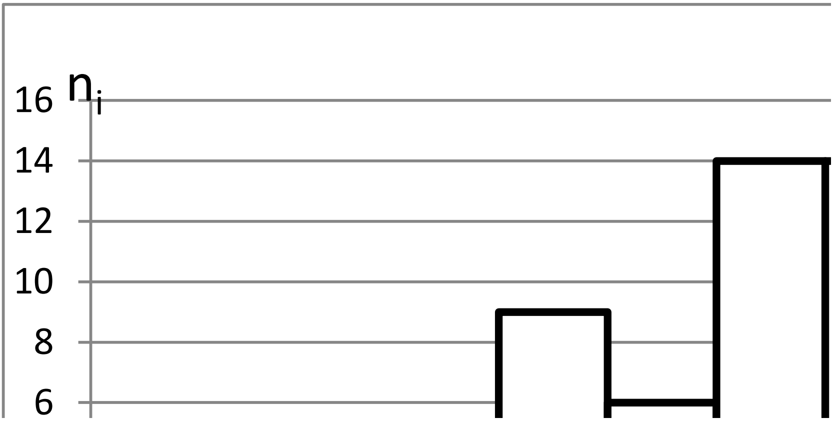
4. За даними таблиць 4 і 5 будуємо відповідні полігони, гістограми і кутуляти ознак X і Y

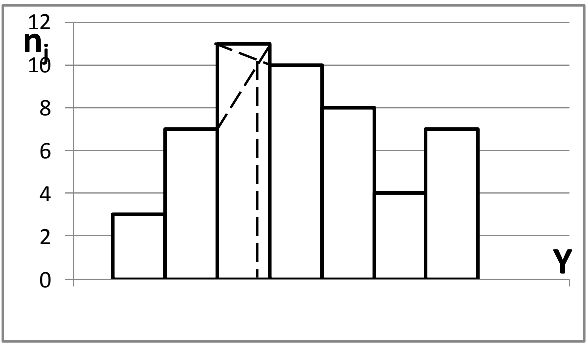
1) Полігони розподілу (з’єднуються відрізками прямих точки $(x_i;n_i)$ ознаки X і точки $(y_j;n_j)$ ознаки Y





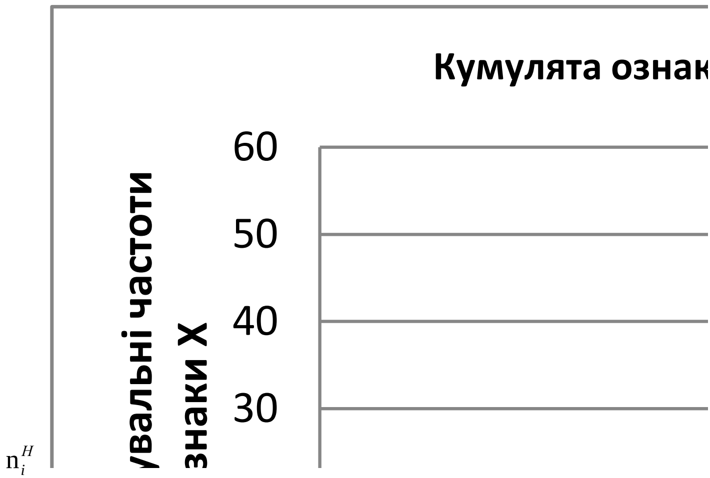
2) Гістограми (будуються прямокутники з основами $(a_i;b_i)$ і висотами n_i для ознаки X і прямокутники з основами $(a_j;b_j)$ і висотами n_j для ознаки Y):

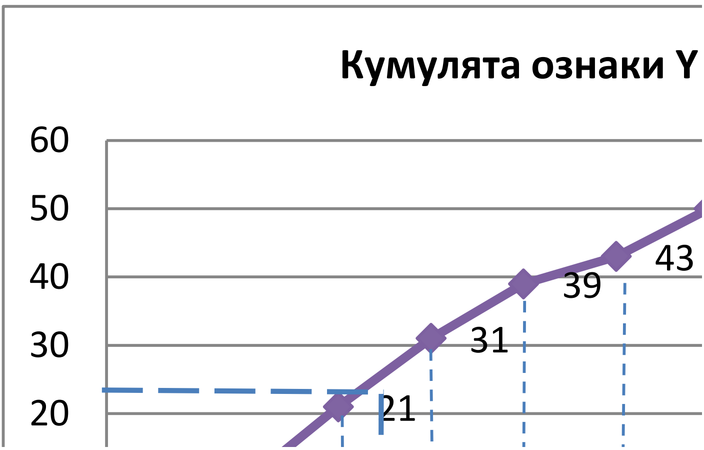




13,6 17,6 21,6 25,6 29,6 33,6 37,6 41,6

3) Кумуляти (з'єднуються відрізками прямих точки $\left(b_i ; \mathrm{n}_i^H\right)$ для ознаки X, і точки $\left(b_j ; \mathrm{n}_j^H\right)$ для ознаки Y):


Кумулята ознак
60
50
40
30
увальні частоти
знаки X
n_i^H


Кумулята ознаки Y
60
50
40
30
20
43
39
31
21

5. Обчислюємо вибіркові числові характеристики і кореляційні показники.

Таблиця 6

Проміжні дані ознаки X

№ з/п	x_i	n_i	$x_i\, n_i$	$x_i - \overline{x}$	$\left(x_i - \overline{x}\right)^2$	$\left(x_i - \overline{x}\right)^2 n_i$
1	13	1	13	- 7	49	49
2	15	4	60	- 5	25	100
3	17	9	153	- 3	9	81
4	19	6	114	- 1	1	6
5	21	14	294	1	1	14
6	23	14	322	3	9	126
7	25	2	50	5	25	50
$\sum$	133	50	1006	-7	119	426

Для ознаки X маємо: $\overline{x} = \frac{1}{n}\sum_{i=1}^{7} x_i\, n_i$ =1006/50=20 (чол),

$\sigma_x^2 = \frac{1}{n}\sum_{i=1}^{7}\left(x_i - \overline{x}\right)^2 n_i$ =426/50=8,52, $\sigma_x = \sqrt{\sigma_x^2} = \sqrt{8,52} \approx 3$,

Вибірковий коефіцієнт варіації

$V_x = \frac{\sigma_x}{\overline{x}} \cdot 100\% = \frac{3}{20} 100\% = 15\ \%$.

Аналогічно робимо для розподілу урожайності зернових культур Y :

Таблиця 7

Проміжні дані ознаки Y

№ з/п	y_j	n_j	$y_j n_j$	$y_j - \bar{y}$	$\left(y_j - \bar{y}\right)^2$	$\left(y_j - \bar{y}\right)^2 n_j$
1	15,6	3	46,8	- 12,24	149,82	449,46
2	19,6	7	137,2	- 8,24	67,9	475,46
3	23,6	11	259,6	- 4,24	17,98	197,78
4	27,6	10	276	-0,24	0,06	0,6
5	31,6	8	252,6	3,76	14,14	113,12
6	35,6	4	142,4	7,76	60,22	240,88
7	39,6	7	277,2	11,76	138,29	968,03
$\sum$	193,2	50	1391,8	-1,68	448,41	2445,17

$\bar{y} = \frac{1391{,}8}{50} = 27{,}84$(ц/га); $\sigma_y^2 = \frac{2445{,}17}{50} = 48{,}9$;

$\sigma_y = 7$ (ц/га); $V_y = \frac{7}{27{,}84} \cdot 100\% \approx 25\%$;

6. Знаходимо моду та медіану рядів розподілу. Показуємо їх на графіках.

Для ознаки X: $Mo\left(X\right) = x_{Mo} + h\frac{n_{Mo} - n_{Mo-1}}{2 \cdot n_{Mo} - n_{Mo-1} - n_{Mo+1}}$,

де $x_{Mo} = 20$ - нижня межа модального інтервалу; $h = 2$ - величина інтервалу; $n_{Mo-1} = 6$ - частота інтервалу, що передує модальному інтервалу; $n_{Mo} = 14$ - частота модального інтервалу; $n_{Mo+1} = 14$ - частота інтервалу, наступного за модальним.

$Mo\left(X\right) = 20 + 2\frac{14 - 6}{2 \cdot 14 - 6 - 14} = 22$ (чол.), або

$$Mo(X) = 22 + 2\frac{14-14}{2\cdot 14 - 14 - 2} = 22 \text{ (чол.)}$$

Для інтервального варіаційного ряду медіану Me визначають за формулою:

$$Me = x_{Me} + h\cdot\frac{0,5\cdot n - n^{H}_{M_e-1}}{n_{Me}},$$

де $x_{Me} = 20$ - нижня (мінімальна) межа медіанного інтервалу (медіанним є інтервал на який припадає перша нагромаджена частота, що перевищує половину всього обсягу сукупності); $h = 2$ - величина інтервалу; $n^{H}_{M_e-1} = 20$ - сума нагромаджених частот інтервалу, який передує медіанному; $n_{Me} = 14$ - частота медіанного інтервалу.

$$Me(X) = 20 + 2\cdot\frac{0,5\cdot 50 - 20}{14} = 20,71 \approx 21 \text{(чол.)}$$

Отже, для ознаки X маємо: $\overline{x} = 20$ чол., $Me(X)$=21 чол., $Mo(X)$=22 чол., тобто виконується нерівність $\overline{x} < Me(X) < Mo(X)$, з якої випливає висновок, що розподіл ознаки X має лівосторонню асиметрію.

Аналогічно знаходимо моду $Mo(Y)$ і медіану $Me(Y)$ для ознаки Y: $Mo(Y) = y_{Mo} + h\dfrac{n_{Mo} - n_{Mo-1}}{2\cdot n_{Mo} - n_{Mo-1} - n_{Mo+1}}$,

$$Mo(Y) = 21,6 + 4\cdot\frac{11-7}{2\cdot 11 - 7 - 10} = 24,8 \text{ (ц/га)};$$

$$Me(Y) = y_{Me} + h_y\cdot\frac{0,5\cdot n - n^{H}_{M_e-1}}{n_{Me}},$$

$$Me(Y) = 25,6 + 4 \cdot \frac{0,5 \cdot 50 - 21}{10} = 27,2 \text{ (ц/га).}$$

Отже, для ознаки Y маємо: $\bar{y} = 27,84$ ц/га, $Me(Y) = 27,2$ ц/га, $Mo(Y) = 24,8$ ц/га, тобто виконується нерівність $Mo(Y) < Me(Y) < \bar{y}$ з якої випливає висновок, що розподіл ознаки Y має правосторонню асиметрію.

7. Визначаємо числові характеристики методом моментів.

Таблиця 8

Статистичний розподіл

X \ Y	15,6	19,6	23,6	27,6	31,6	35,6	39,6	n_{x_i}	$\bar{y}_{x_i}$
13	1							1	15,6
15	1	3						4	18,6
17	1	4	2	2				9	21,8
19			4	2				6	24,9
21			5	4	5			14	27,6
23				2	3	4	5	14	35
25							2	2	39,6
n_{y_j}	3	7	11	10	8	4	7	**50**	
$\bar{x}_{y_j}$	15	16,1	19,5	20,2	21,8	23	23,6		

Визначимо середні арифметичні, дисперсії і середні квадратичні відхилення величин X **і** Y методом моментів. Для цього у варіаційних рядах перейдемо до умовних варіантів, обравши за «несправжні нулі» середини модальних інтервалів. Для X за нуль беремо значення, $C_x = 21,$ а для Y беремо $C_y = 23,6$.

Таблиця 9

Варіаційний ряд для варіанти u

X	13	15	17	19	21	23	25
u_i	-4	-3	-2	-1	0	1	2
$n_{x_i} = n_{u_i} = n_i$	1	4	9	6	14	14	2
$u_i\, n_i$	-4	-12	-18	-6	0	14	4
$u_i^{\,2} n_i$	16	36	36	6	0	14	8

Початковий момент 1-го порядку (зважене середнє арифметичне):

$$M_1 = \bar{u} = \frac{1}{n}\sum_{i=1}^{7} u_i\, n_i = -\frac{22}{50} = -0,44.$$

Початковий момент 2-го порядку:

$$M_2 = \frac{1}{n}\sum_{i=1}^{7} u_i^{\,2} n_i = \frac{116}{50} = 2,32$$

Центральний момент 2-го порядку (дисперсія варіанти u):

$\mu_2 = \sigma_u^2 = M_2 - M_1^2 = 2{,}32 - (-0{,}44)^2 = 2{,}13.$

Середнє квадратичне відхилення в умовних варіантах:

$$\sigma_u = \sqrt{\mu_2} = \sqrt{2{,}13} \approx 1{,}45.$$

Тоді вибіркова середня ознаки X **:**

$$\bar{x} = M_1 \cdot h + C_x = -0,44 \cdot 2 + 21 \approx 20 \text{ (чол.)}$$

Вибіркова дисперсія : $\sigma_x^2 = h_x^2 \cdot \mu_2 = 2^2 \cdot 2{,}13 = 8{,}4.$

Вибіркове середнє квадратичне відхилення :

$$\sigma_x = h_x \cdot \sigma_u = 2 \cdot 1{,}45 = 2{,}9 \approx 3 \text{ (чол)}.$$

Таблиця 10

Варіаційний ряд для варіанти v

Y	15,6	19,6	23,6	27,6	31,6	35,6	39,6
n_j	3	7	11	10	8	4	7
v_j	-2	-1	0	1	2	3	4
$v_j n_j$	-6	-7	0	10	16	12	28
$v_j^2 n_j$	12	7	0	10	32	36	112

Початковий момент 1-го порядку (зважене середнє арифметичне):

$$M_1 = \bar{v} = \frac{1}{n}\sum_{j=1}^{7} v_j n_j = \frac{53}{50} = 1,06 .$$

Початковий момент 2-го порядку:

$$M_2 = \overline{v^2} = \frac{1}{n}\sum_{j=1}^{7} v_j^2 n_j = \frac{116}{50} = 4,18 .$$

Центральний момент 2-го порядку (дисперсія варіанти **v**):

$\mu_2 = \sigma_v^2 = M_2 - M_1^2 = 4,18 - 1,06^2 = 3,06.$

Середнє квадратичне відхилення в умовних варіантах:

$\sigma_v = \sqrt{\mu_2} = \sqrt{3,06} \approx 1,75;$

Тоді вибіркова середня ознаки **Y:**

$$\bar{y} = h \cdot M_1 + C_y = 4 \cdot 1,06 + 23,6 = 27,84 \text{ (ц/га)}.$$

Вибіркова дисперсія: $\sigma_y^2 = h_y^2 \cdot \mu_2 = 4^2 \cdot 3,06 = 48,96.$

Вибіркове середнє квадратичне відхилення :

$$\sigma_y = h_y \cdot \sigma_v = 4 \cdot 1,75 \approx 7 \text{ (ц/га)}.$$

8. Обчислюємо вибірковий коефіцієнт кореляції шляхом використання умовних варіант.

При вивченні кореляційного зв'язку виникає необхідність визначити ступінь тісноти зв'язку між ознаками. Тіснота зв'язку характеризується за допомогою спеціального відносного показника, який називається коефіцієнтом кореляції.

Коефіцієнт кореляції може набувати значень від 0 до ± 1. Якщо коефіцієнт кореляції дорівнює 0, то зв'язок відсутній, якщо 1, то зв'язок функціональний. Знак при коефіцієнті кореляції вказує на напрямок зв'язку ("+" – прямий, "-" – зворотний). Чим ближче коефіцієнт кореляції до 1, тим зв'язок між ознаками тісніший.

Вибірковий коефіцієнт кореляції обчислимо за формулою:

$$r_в = \frac{\sum_{i=1}^{7}\sum_{j=1}^{7} u_i\, v_j\, n_{ij} - n\bar{u}\,\bar{v}}{n\sigma_u\,\sigma_v} \;.$$

Позначимо: $\bar{v}_j = \sum_j v_j\, n_{ij}$, $\bar{u}_i = \sum_i u_i\, n_{ij}$,

$$\left(\sum_j v_j\, n_{ij}\right)\cdot u_i = \bar{v}_j\, u_i\,, \quad \left(\sum_i u_i\, n_{ij}\right)\cdot v_j = \bar{u}_i\, v_j\,.$$

Для обчислення суми $\sum_{i=1}^{7}\sum_{j=1}^{7} u_i\, v_j\, n_{ij}$ складаємо таблицю :

Таблиця 11

Обчислення вибіркового коефіцієнта кореляції

v \ u	-2	-1	0	1	2	3	4	$\overline{v}_j$	$\overline{v}_j u_i$
-4	-2 1 -4							**-2**	8
-3	-2 1 -3	-3 3 -9						**-5**	15
-2	-2 1 -2	-4 4 -8	0 2 -4	2 2 -2				**-4**	8
-1			0 4 -4	2 2 -2				**2**	-2
0			0 5 0	4 4 0	10 5 0			**14**	0
1				2 2 2	6 3 3	12 4 4	20 5 5	**40**	40
2							8 2 4	**8**	16
$\overline{u}_i$	-9	-17	-8	-2	3	4	9		**87**
$\overline{u}_i v_j$	18	17	0	-2	6	12	36	**87**	

Отже, вибірковий коефіцієнт кореляції

$$r = \frac{87 - 50(-0{,}44)(1{,}06)}{50 \cdot 1{,}45 \cdot 1{,}75} = 0{,}87$$

9. Складаємо вибіркове рівняння прямої регресії y на x і прямої регресії x на y.

Рівняння $\overline{y_x} - \bar{y} = r \cdot \frac{\sigma_y}{\sigma_x}(x - \bar{x})$ регресії Y на X буде таким: $(y\_x)^{-} - 27{,}84 = 0{,}87 \cdot 7/3\ (x - 20)$.

Спростивши отримаємо: $\boxed{\overline{y_x} = 2{,}03 \cdot x - 12{,}76}$.

Рівняння $\overline{x_y} - \bar{x} = r \cdot \frac{\sigma_x}{\sigma_y}(y - \overline{y})$ регресії X на Y матиме вигляд: $\overline{x_y} - 20 = 0{,}87 \cdot \frac{3}{7}(y - 27{,}84)$.

Спростивши , отримаємо: $\overline{\mathrm{x_y}} = 0{,}37\mathrm{y} + 9{,}62$.

10. Знаходимо надійну зону відхилень від прямої регресії.

Попередньо знайдемо середню квадратичну похибку рівняння прямої регресії Y на X за формулою:

$$S(y/x) = \sqrt{\frac{\sum_{i=1}^{k}\left(\bar{y}_i - \bar{y}_{x_i}\right)^2 n_i}{n}}$$

де: $\bar{y}_i$ - умовні середні ознаки Y, що відповідають значенням x_i ознаки X; $\bar{y}_{x_i}$ - значення ознаки Y, обчисленні для тих самих значень x_i за рівнянням прямої регресії $\overline{y_x} = 2{,}03 \cdot x - 12{,}76$.

Таблиця 12

Знаходження середньої квадратичної похибки рівняння прямої регресії Y на X

x_i	n_i	$\overline{y}_i$	$\overline{y}_{x_i}$	$\overline{y}_i-\overline{y}_{x_i}$	$\left(\overline{y}_i-\overline{y}_{x_i}\right)^2$	$\left(\overline{y}_i-\overline{y}_{x_i}\right)^2 n_i$
13	1	15,6	13,63	1,97	3,88	3,88
15	4	18,6	17,69	0,91	0,83	3,31
17	9	21,8	21,75	0,05	0,0025	0,0225
19	6	24,9	25,81	-0,91	0,83	4,97
21	14	27,6	29,87	-2,27	5,15	72,14
23	14	35	33,93	1,07	1,14	16,03
25	2	39,6	38	1,6	2,56	5,12
$\sum$						105,5

$S\left(y/x\right)=\sqrt{\dfrac{105,5}{50}}=\sqrt{2,11}\approx 1,45$ ц/га

При заданій надійності $\gamma=0,90$ із співвідношення $2\Phi\left(t\right)=\gamma$ спочатку знаходимо значення t: $\Phi\left(t\right)=\dfrac{\gamma}{2}=0,45$, $\Rightarrow t=1,64$ (див. таблицю значень інтегральної функції $\Phi\left(x\right)$) , а потім знаходимо граничну похибку рівняння прямої регресії Y на X: $\Delta=t\cdot S\left(y/x\right)=1,64\cdot 1,45=2,38$ ц/га

11. Знаходимо статистичні оцінки параметрів.

Точкові оцінки генеральних середніх:
вибіркова середня ознаки Х, $a_x=\overline{x}=20$ (чол.);
вибіркова середня ознаки У, $a_y=\overline{y}=27,84$ (ц/га).

Точкові оцінки генеральних дисперсій:

$$\tilde{\sigma}_x^2 = \frac{n}{n-1}\sigma_x^2 = \frac{50}{49}8,52 = 8,7\,; \tilde{\sigma}_y^2 = \frac{n}{n-1}\sigma_y^2 = \frac{50}{49}48,96 = 49,96$$

Точкові оцінки генеральних середніх квадратичних відхилень:

$\tilde{\sigma}_x = \sqrt{8,7} = 2,95 \approx 3$ (чол.); $\tilde{\sigma}_y = \sqrt{49,96} = 7,07$ (ц/га).

Точкові оцінки генеральних коефіцієнтів варіації:

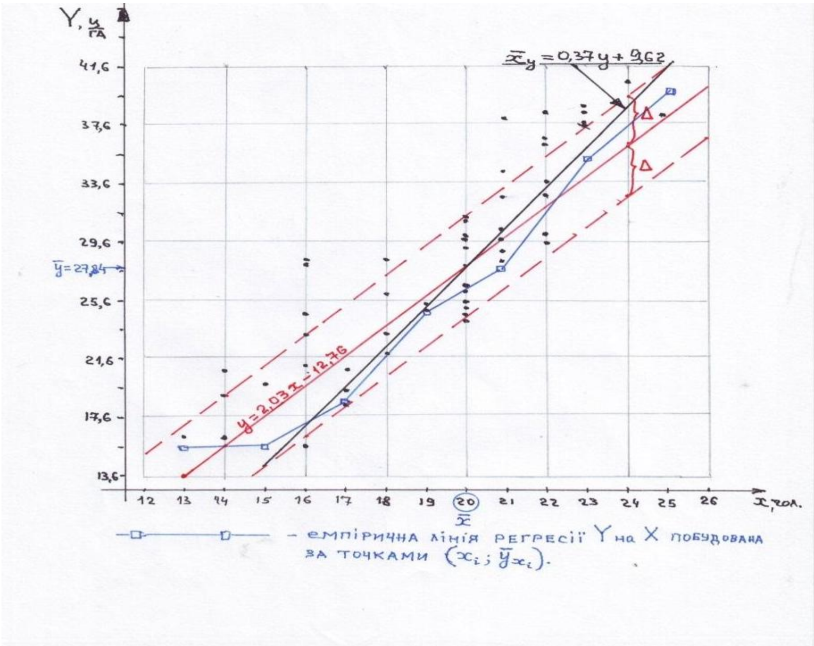
$$\tilde{V}_x = \frac{\tilde{\sigma}_x}{a_x}100\,\% = \frac{2,95}{20}\cdot 100\,\% = 14,75\,\%\,;$$

$$\tilde{V}_y = \frac{\tilde{\sigma}_y}{a_y}100\,\% = \frac{7,07}{27,84}\cdot 100\,\% = 25,4\,\%\,.$$

Точкова оцінка генерального коефіцієнта кореляції:

$r_г = r_b = 0,87$.

12. Будуємо поле кореляції, емпіричні лінії регресії та обидві прямі регресії.



13. Знаходимо інтервали надійності для генеральних середніх та для генерального коефіцієнта кореляції.

Надійний інтервал для генерального середнього $a_x = \overline{x} = 20$ чол. з заданою надійністю $\gamma = 0,9$ знаходимо за формулою

$$\mathrm{I}_{\gamma}\left(a_x\right) = \left(\overline{x} - \frac{t \cdot \sigma_x}{\sqrt{n}}; \overline{x} + \frac{t \cdot \sigma_x}{\sqrt{n}}\right).$$

При заданій надійності $\gamma = 0,90$ із співвідношення $2\Phi(t) = \gamma$ спочатку знаходимо значення t: $\Phi(t) = \frac{\gamma}{2} = 0,45$, $\Rightarrow t = 1,64$.

$$\mathrm{I}_{0,9}(a_x) = \left(20 - \frac{1,64 \cdot 2,93}{\sqrt{50}}; 20 + \frac{1,64 \cdot 2,93}{\sqrt{50}}\right) =$$

$$= (19,422; 20,578).$$

Аналогічно знаходимо надійний інтервал для генерального середнього $a_y = \bar{y} = 27,84$ (ц/га):

$$\mathrm{I}_{\gamma}(a_y) = \left(\bar{y} - \frac{t \cdot \sigma_y}{\sqrt{n}}; \bar{y} + \frac{t \cdot \sigma_y}{\sqrt{n}}\right) =$$

$$= \left(27,84 - \frac{1,64 \cdot 7,07}{\sqrt{50}}; 27,84 + \frac{1,64 \cdot 7,07}{\sqrt{50}}\right) = (26,2; 29,48)$$

Перед знаходженням інтервалу довір'я для генерального коефіцієнта кореляції обчислюємо середню квадратичну похибку вибіркового коефіцієнта кореляції за формулою:

$$\sigma_r = \frac{1 - r_в^2}{\sqrt{n}}.$$

Отже, похибка $\sigma_r = \frac{1 - 0,87^2}{\sqrt{50}} = 0,034$.

Тоді $I_{0,9}(r_b) = (r_b - t \cdot \sigma_r; r_b + t \cdot \sigma_r) =$

$$= (0,87 - 1,64 \cdot 0,034; 0,87 + 1,64 \cdot 0,034) = (0,814; 0,926).$$

14. Перевіряємо гіпотезу про нормальний розподіл обох статистичних величин X і Y.

Попередньо обчислюємо теоретичні частоти кожного ряду розподілу, тобто, частоти інтервалів, при умові, що кожна з ознак X і Y розподілена за нормальним законом за формулами:

$$n_i' = n \cdot \left(\Phi\left(\beta_i\right) - \Phi\left(\alpha_i\right)\right), \text{ де } \alpha_i = \frac{a_i - \overline{x}}{\sigma_x}, \; \beta_i = \frac{b_i - \overline{x}}{\sigma_x},$$

$$n_j' = n \cdot \left(\Phi\left(\beta_j\right) - \Phi\left(\alpha_j\right)\right), \text{ де } \alpha_j = \frac{c_j - \overline{y}}{\sigma_y}, \; \beta_j = \frac{d_j - \overline{y}}{\sigma_y}.$$

Враховуючи значення: $\overline{x} = 20$ чол., $\overline{y} = 27{,}84$ ц/га, $\sigma_x = 3$ чол., $\sigma_y = 7$ ц/га, складаємо дві таблиці.

Таблиця 13

Обчислення теоретичних частот ознаки X

№	a_i	b_i	α_i	β_i	$\Phi(\alpha_i)$	$\Phi(\beta_i)$	$\Phi(\beta_i)-\Phi(\alpha_i)$	n_i'
1	12	14	-2,67	-2	-0,496	-0,477	0,019	1
2	14	16	-2	-1,33	-0,477	-0,408	0,069	4
3	16	18	-1,33	-0,67	-0,408	-0,249	0,159	8
4	18	20	-0,67	0	-0,249	0	0,249	13
5	20	22	0	0,67	0	0,249	0,249	12
6	22	24	0,67	1,33	0,249	0,408	0,159	8
7	24	26	1,33	2	0,408	0,477	0,069	4
Σ								50

Таблиця 14

Обчислення теоретичних частот ознаки Y

№ з/п	c_j	d_j	α_j	β_j	$\Phi(\alpha_j)$	$\Phi(\beta_j)$	$\Phi(\beta_j)$- $\Phi(\alpha_j)$	n_j'
1	13,6	17,6	-2,03	-1,46	-0,479	-0,428	0,051	3
2	17,6	21,6	-1,46	-0,89	-0,428	-0,313	0,115	6
3	21,6	25,6	-0,89	-0,32	-0,313	-0,126	0,187	10
4	25,6	29,6	-0,32	0,25	-0,126	0,099	0,225	12
5	29,6	33,6	0,25	0,82	0,099	0,294	0,195	10
6	33,6	37,6	0,82	1,39	0,294	0,418	0,124	6
7	37,6	41,6	1,39	1,97	0,418	0,476	0,058	3
Σ								50

Знаючи емпіричні (n_i) і теоретичні (n_i') частоти кожного ряду розподілу, визначаємо фактичне значення критерію згоди Пірсона за формулою: $\chi^2 = \sum_{i=1}^{k} \frac{\left(n_i - n_i'\right)^2}{n_i'}$, де k - число інтервалів (класів, груп) на які розбито вибірковий розподіл. Для цього складемо додатково дві таблиці:

Таблиця 15

Обчислення фактичного значення критерію згоди Пірсона ознаки X

№ з/п	n_i	n_i'	$n_i - n_i'$	$\left(n_i - n_i'\right)^2$	$\frac{\left(n_i - n_i'\right)^2}{n_i'}$
1	1	1	0	0	0
2	4	4	0	0	0
3	9	8	1	1	0,125
4	6	13	-7	49	3,77
5	14	12	2	4	0,33
6	14	8	6	36	4,5
7	2	4	-2	4	1
Σ	50	50			9,725

Фактичне значення критерію згоди Пірсона ознаки Х $\chi^2_{ф.х} = 9,725$.

З таблиці «Критичні точки χ^2 - розподілу» знаходимо критичні значення критерію для ознаки Х при рівні значущості $\alpha = 0,1$ і числу ступенів свободи $\nu = k' - 3 = 6 - 3 = 3$: $\chi^2_{кр.х} = 6,3$ (тут $k' = 6$, бо перші два інтервали об'єднуються в один через їх малочисельність). Оскільки $\chi^2_{ф.х} > \chi^2_{кр.х}$, то для ознаки Х нульова гіпотеза відкидається, тобто вибірковий розподіл ознаки Х не узгоджується із її теоретичним розподілом.

Таблиця 16

Обчислення фактичного значення критерію згоди Пірсона ознаки Y

№ з/п	n_j	n_j'	$n_j - n_j'$	$\left(n_j - n_j'\right)^2$	$\dfrac{\left(n_j - n_j'\right)^2}{n_j'}$
1	3	3	0	0	0
2	7	6	1	1	0,17
3	11	10	1	1	0,1
4	10	12	-2	4	0,33
5	8	10	-2	4	0,4
6	4	6	-2	4	0,67
7	7	3	4	16	5,33
Σ	50	50			7,0

Фактичне значення критерію згоди Пірсона ознаки Y $\chi^2_{ф.у} = 7{,}0$. З таблиці «Критичні точки χ^2- розподілу» знаходимо критичні значення критерію для ознаки Y при рівні значущості $\alpha = 0{,}1$ і числу ступенів свободи $\nu = k' - 3 = 7 - 3 = 4$: $\chi^2_{кр.у} = 7{,}8$. Оскільки $\chi^2_{ф.у} < \chi^2_{кр.у}$, то для ознаки Y нульова гіпотеза приймається, тобто ознака Y фактично розподілена за нормальним законом.

Загальні висновки

1. Середнє число робітників, які працюють на 100 га ріллі в середньому за рік складає 20 чоловік, причому, з надійністю 90 % можна стверджувати, що їхня кількість коливається в межах від 19 до 21 чоловіка.

2. Середнє значення урожайності зернових культур складає 27,84 ц/га, при цьому з надійністю 90 % можна стверджувати, що урожайність коливається в межах від 26,2 ц/га до 29,48 ц/га.

3. Оскільки $\overline{x} = 20$ чол., $Me(X) = 21$ чол., $Mo(X) = 22$ чол., тобто, виконується нерівність $\overline{x} < Me(X) < Mo(X)$, то вибірка статистичних даних кількості працюючих робітників характеризується незначною лівосторонньою асиметрією. Урожайність зернових культур для якої $\overline{y} = 27,84$ ц/га, $Me(Y) = 27,2$ ц/га, $Mo(Y) = 24,8$ ц/га характеризується правосторонньою асиметрією $\overline{y} > \mathrm{Me} > \mathrm{Mo}$.

3. Оскільки коефіцієнт варіації для ознаки X менший від коефіцієнта варіації для ознаки Y ($V_x = 15\,\%$, $V_y = 25\,\%$), то урожайність зернових культур характеризується більшим ступенем розсіювання , ніж статистичні дані кількості працюючих робітників.

4. З надійністю 90 % при рівні значущості $\alpha = 0,1$ приймається гіпотеза про нормальний розподіл , як статистичних даних кількості працюючих робітників, так і урожайності зернових культур.

5. Оскільки вибірковий коефіцієнт кореляції додатній ($r_b > 0$, $r_b = 0,87$) то існує сильний зв'язок між кількістю працюючих робітників та урожайністю зернових культур. Цей зв'язок описується рівнянням регресії $\overline{y_x} = 2,03 \cdot x - 12,76$ з можливим відхиленням на 2,38 ц/га(це стверджуємо з надійністю 90%).

6. Оскільки вибірковий коефіцієнт кореляції $r_b = 0,87$, а генеральний коефіцієнт кореляції лежить у межах від 0,814 до 0,962 то з надійністю 90% можна стверджувати , що зв'язок між кількістю працюючих робітників та урожайністю зернових культур достатньо сильний.

6. Література

1. Бахрушин В. Є. Методи аналізу даних : навчальний посібник для студентів. Запоріжжя : КПУ, 2011. 268 с.
2. Бобик О. І. Берегова І. Г., Копитко Б. І. Теорія ймовірностей і математична статистика : навч. підручник. К., 2006. 440 с.
3. Бугір М. К. Посібник з теорії ймовірності та математичної статистики / МОН України. Тернопіль : Підручники і посібники. 1998. 176 с.
4. Валєєв К. Г., Джалладова І. А. Збірник задач з теорії ймовірностей та математичної статистики : навч. посіб. / 2-е вид. без змін. К. : КНЕУ, 2008. 352 с.
5. Єріна А. М. Статистичне моделювання та прогнозування. Київ : КНТЕУ, 2001. 196 с.
6. Жлуктенко В. І., Наконечний С. І. Теорія ймовірностей і математична статистика : навч.-метод. посібник / У 2 ч. Ч. 2. Математична статистика. К. : КНЕУ, 2000. 304 с.
7. Мармоза А. Т. Практикум з математичної статистики : навч. посіб.. К. : Кондор, 2004. 264 с.
8. Пушак Я. С., Лозовий Б. Л. Теорія ймовірностей і елементи математичної статистики : навчальний посібник. Львів : „Магнолія 2006”. 2007. 276 с.
9. Руденко В. М. Математична статистика : навчальний посібник. К. : Центр учбової літератури, 2012. 304 с.
10. Суліма І. М., Яковенко В. М. Вища математика. Теорія ймовірностей. Математична статистика. К. Видавничий центр НАУ, 2004. 238 с.
11. Турчин В М. Математична статистика : посібник. К. : Видавничий центр «Академія», 1999. 238 с.

Додатки

Додаток А

Дані для статистичного аналізу урожайності зернових культур

№ з/п	А1	А2	А3	А4	А5	№ з/п	А1	А2	А3	А4	А5
1	28,0	79	133	36,0	16	**2**	20,3	69	78	21,8	14
3	21,0	70	107	20,2	16	**4**	37,4	90	222	42,0	24
5	27,6	80	150	30,0	17	**6**	23,2	75	130	26,9	18
7	16,2	71	85	23,3	14	**8**	38,3	86	168	28,2	23
9	29,7	77	160	26,8	20	**10**	28,0	85	169	34,8	25
11	26,6	77	141	29,6	21	**12**	29,0	77	148	26,7	19
13	30,3	84	160	37,1	22	**14**	37,5	92	228	36,6	20
15	15,7	66	71	18,1	14	**16**	34,7	87	206	37,0	21
17	25,5	74	118	35,1	20	**18**	30,0	75	183	29,0	20
19	15,8	67	79	20,8	12	**20**	20,6	69	101	25,0	17
21	40,6	89	212	41,9	24	**22**	41,8	90	226	44,0	24
23	27,2	69	150	31,1	21	**24**	28,6	80	153	32,8	20
25	16,1	60	87	27,1	13	**26**	22,8	67	93	19,1	16
27	19,8	75	81	21,8	15	**28**	28,8	84	158	31,8	20
29	25,3	80	135	28,6	19	**30**	29,2	79	173	30,9	20
31	16,9	63	79	24,5	17	**32**	35,8	84	212	34,5	22
33	31,4	80	179	37,2	18	**34**	19,3	70	97	25,7	17
35	26,5	73	137	26,5	20	**36**	34,3	89	197	39,0	21
37	20,4	75	103	22,2	16	**38**	33,7	82	199	36,5	21
39	15,6	79	91	27,1	16	**40**	20,5	67	20,5	24,4	17
41	19,0	69	93	21,2	17	**42**	31,1	84	175	32,2	20
43	28,0	81	147	33,7	21	**44**	28,6	80	158	33,7	22
45	34,4	86	209	37,5	22	**46**	19,0	68	104	25,2	14
47	33,8	83	205	35,8	22	**48**	24,4	72	103	26,8	18
49	19,4	70	88	24,3	19	**50**	22,4	77	88	23,1	22
51	31,4	83	197	37,6	21	**52**	14,0	59	58	15,0	12
53	24,8	75	136	32,0	18	**54**	24,6	76	142	21,1	16
55	31,9	82	190	33,9	20	**56**	29,7	82	172	30,9	21
57	36,3	90	220	39,6	23	**58**	37,6	81	210	43,0	23.

№ з/п	А1	А2	А3	А4	А5	№ з/п	А1	А2	А3	А4	А5
59	32,6	84	170	36,1	21	**60**	28,8	78	157	32,8	21
61	18,6	70	96	21,0	14	**62**	32,8	76	172	33,8	21
63	25,1	73	144	28,5	20	**64**	39,9	85	241	38,0	25
65	18,3	72	93	22,1	16	**66**	24,2	71	132	28,3	20
67	25,2	74	145	28.3	19	**68**	24,3	75	99	26,8	19
69	31,8	82	183	34,2	22	**70**	40,7	90	237	44,0	24
71	26,9	77	138	28,9	19	**72**	37,9	85	190	38,0	21
73	38,8	87	219	39,3	23	**74**	24,4	76	103	27,0	20
75	30,6	82	142	33,1	17	**76**	27,3	78	171	34,2	21
77	34,0	88	212	38,0	24	**78**	21,7	72	118	24,9	18
79	28,9	75	163	33,8	20	**80**	36,9	93	196	42,0	24
81	14,7	62	75	25,1	16	**82**	38,8	83	190	35,0	22
83	36,5	85	192	41,6	22	**84**	32,0	79	193	34,9	21
85	29,1	84	156	34,7	21	**86**	37,2	90	251	40,0	21
87	33,9	87	210	37,7	23	**88**	30,2	83	185	34,7	21
89	31,0	80	192	35,6	19	**90**	29,5	80	180	34,2	22
91	20,1	65	110	21,8	14	**92**	26,0	77	121	24,4	18
93	18,2	71	92	22,0	17	**94**	28,3	77	157	33,0	18
95	33,6	90	204	35,1	22	**96**	15,0	61	65	19,0	18
97	26,2	80	166	29,9	20	**98**	24,4	77	139	26,5	18
99	28,2	81	172	32,8	16	**100**	26,6	75	141	31,1	21
101	24,9	85	143	26,0	20	**102**	23,2	79	110	24,0	17
103	36,5	92	228	38,7	20	**104**	21,5	66	120	25,0	15
105	28,2	85	149	34,0	18	**106**	26,1	75	131	27,4	19
107	38,1	91	215	38,5	21	**108**	23,8	71	128	27,1	19
109	29,6	78	161	25,0	19	**110**	33,5	89	203	37,0	23
111	26,2	80	150	33,0	18	**112**	26,5	74	149	30,8	20
113	29,1	75	167	32,1	16	**114**	38,2	66	103	23,5	17
115	28,1	84	148	33,0	17	**116**	30,5	87	186	32,7	21

Таблиця значень функції Лапласа $\Phi(x)=\frac{1}{\sqrt{2\pi}}\int_0^x e^{-\frac{t^2}{2}}dt$

x	0	1	2	3	4	5	6	7	8	9
0,0	0,0000	0040	0080	0120	0160	0199	0239	0279	0319	0359
0,1	0398	0438	0478	0517	0557	0596	0636	0675	0714	0753
0,2	0793	0832	0871	0910	0948	0987	1026	1064	1103	1141
0,3	1179	1217	1255	1293	1331	1368	1406	1443	1480	1517
0,4	1554	1591	1628	1664	1700	1736	1772	1808	1844	1879
0,5	1915	1950	1985	2019	2054	2088	2123	2157	2190	2224
0,6	2257	2291	2324	2357	2389	2422	2454	2486	2517	2549
0,7	2580	2611	2642	2673	2703	2734	2764	2794	2823	2852
0,8	2881	2910	2939	2967	2995	3023	3051	3078	3106	3133
0,9	3159	3186	3212	3238	3264	3289	3315	3340	3365	3389
1,0	0,3413	3438	3461	3485	3508	3531	3554	3577	3599	3621
1,1	3643	3665	3686	3708	3729	3749	3770	3790	3810	3830
1,2	3849	3869	3883	3907	3925	3944	3962	3980	3997	4015
1,3	4032	4049	4066	4082	4099	4115	4131	4147	4162	4177
1,4	4192	4207	4222	4236	4251	4265	4279	4292	4306	4319
1,5	4332	4345	4357	4370	4382	4394	4406	4418	4429	4441
1,6	4452	4463	4474	4484	4495	4505	4515	4525	4535	4545
1,7	4554	4564	4573	4582	4591	4599	4608	4616	4625	4633
1,8	4641	4649	4656	4664	4671	4678	4686	4693	4699	4706
1,9	4713	4719	4726	4732	4738	4744	4750	4756	4761	4767
2,0	0,4772	4778	4783	4788	4793	4798	4803	4808	4812	4817
2,1	4821	4826	4830	4834	4838	4842	4846	4850	4854	4857
2,2	4861	4865	4868	4871	4875	4878	4881	4884	4887	4890
2,3	4893	4896	4898	4901	4904	4906	4909	4911	4913	4916
2,4	4918	4920	4922	4924	4927	4929	4931	4932	4934	4936
2,5	4938	4940	4941	4943	4945	4946	4948	4949	4951	4952
2,6	4953	4955	4956	4957	4959	4960	4961	4962	4963	4964
2,7	4965	4966	4967	4968	4969	4970	4971	4972	4973	4974
2,8	4974	4975	4976	4977	4977	4978	4979	4979	4980	4981
2,9	0,4981	4982	4983	4983	4984	4984	4985	4985	4986	4986

Для значень $x \geq 3{,}0$ значення функції $\Phi(x)$ такі:

x	3,0	3,1	3,2	3,3	3,4	3,5	3,6
Φ(x)	0,4986	0,499	0,4993	0,4995	0,4996	0,4997	0,4998

x	3,7	3,8	3,9	4,0	4,5	$x \geq 5{,}0$
Φ(x)	0,49989	0,49993	0,49995	0,499968	0,499997	0,5

Критичні точки χ^2 - розподілу

Число ступенів свободи, k	Рівень значущості α					
	0,99	0,95	0,10	0,05	0,01	0,001
1	0,000016	0,0039	2,7	3,8	6,6	10,8
2	0,020	0,103	4,6	6,0	9,2	13,8
3	0,115	0,352	6,3	7,8	11,3	16,3
4	0,30	0,711	7,8	9,5	13,3	18,5
5	0,55	1,15	9,2	11,1	15,1	20,5
6	0,87	1,64	10,6	12,6	16,8	22,5
7	1,24	2,17	12,0	14,1	18,5	24,3
8	1,65	2,73	13,4	15,5	20,1	21,6
9	2,09	3,33	14,7	16,9	21,7	27,9
10	2,56	3,94	16,0	18,3	23,2	29,6
11	3,05	4,57	17,3	19,7	24,7	31,3
12	3,57	5,23	18,5	21,0	26,2	32,9
13	4,11	5,89	19,8	22,4	27,7	34,5
14	4,66	6,57	21,1	23,7	29,1	36,1
15	5,23	7,26	22,3	25,0	30,6	37,7
16	5,81	7,96	23,5	26,3	32,0	39,3
17	6,41	8,67	24,8	27,6	33,4	40,8
18	7,01	9,39	26,0	28,9	34,8	42,3
19	7,63	10,1	27,2	30,1	36,2	43,8
20	8,26	10,9	28,4	31,4	37,6	45,3
21	8,90	11,6	29,6	32,7	38,9	46,8
22	9,64	12,3	30,8	33,9	40,3	48,3
23	10,2	13,1	32,0	35,2	41,6	49,7
24	10,9	13,8	33,2	36,4	43,0	51,2
25	11,5	14,6	34,4	37,7	44,3	52,6
26	12,2	15,4	35,6	38,9	45,6	54,1
27	12,9	16,2	36,7	40,1	47,0	55,5
28	13,6	16,9	37,9	41,3	48,3	56,9
29	14,3	17,7	39,1	42,6	49,6	58,3
30	15,0	18,5	40,3	43,8	50,9	59,7

Таблиця значень $t_\gamma = t(\gamma, n)$

γ n	0,95	0,99	0,999
5	2,78	4,60	8,61
10	2,26	3,25	4,78
15	2,15	2,98	4,14
20	2,093	2,861	3,883
25	2,064	2,797	3,745
30	2,045	2,756	3,659
35	2,032	2,720	3,600
40	2,023	2,708	3,558
45	2,016	2,692	3,527
50	2,009	2,679	3,502
60	2,001	2,662	3,464
70	1,996	2,649	3,439
80	1,991	2,640	3,418
90	1,987	2,633	3,403
100	1,984	2,627	3,392
120	1,980	2,617	3,374
∞	1,960	2,576	3,291

Таблиця значень $q = q(\gamma, n)$

γ / n	0,95	0,99	0,999
5	1,37	2,67	5,64
10	0,65	1,08	1,80
15	0,46	0,73	1,15
20	0,37	0,58	0,88
25	0,32	0,49	0,73
30	0,28	0,43	0,63
35	0,26	0,38	0,56
40	0,24	0,35	0,50
45	0,22	0,32	0,46
50	0,21	0,30	0,43
60	0,188	0,269	0,38
70	0,174	0,245	0,34
80	0,161	0,226	0,31
90	0,151	0,211	0,29
100	0,143	0,198	0,27
150	0,115	0,160	0,211
200	0,099	0,136	0,185
250	0,089	0,120	0,162

Таблиця випадкових чисел

1	2	3	4	5	6	7	8	9	10
5880	1257	6163	4439	7276	6353	6912	0731	9033	5294
9083	4260	5277	4998	4298	5204	3965	4028	8936	5148
1762	8713	1189	1090	8989	7273	3213	1935	9321	4820
2023	2509	1740	0424	8924	0005	1969	1636	7237	1227
7965	3855	4765	0703	1678	0841	7543	0308	9732	1289
7690	0480	8098	9629	4819	7219	7214	5128	3853	1921
9292	0426	9573	4903	5916	6517	8368	3270	6641	0038
0867	1651	7016	4220	2533	6345	8227	1904	5138	2537
0505	2127	8255	5276	2233	3956	4118	8199	6380	6340
6290	9795	1112	5765	2575	6837	3336	9222	7403	8345
6323	2615	3410	3365	1117	2417	3176	2434	5240	5455
8672	8536	2966	5773	5412	8114	0903	4697	6919	4569
1422	5507	7596	0670	3013	1351	3886	3268	9469	2534
2553	1473	5113	5735	1469	9545	9331	5303	9914	6394
0438	4376	3328	8649	8327	0110	4549	7955	5275	2890
2851	2157	0047	7085	1129	0460	6821	8323	2572	8962
7962	2753	3077	8718	7418	8004	1425	3706	8822	1494
3837	4098	0220	1217	4732	0150	1637	1097	1040	7372
8542	4126	9274	2251	0607	4301	8730	7690	6235	3477
0139	0765	8039	9484	3577	7859	1976	0623	1418	6685
6687	1943	4307	0579	8171	8224	8641	7034	3595	3875
6242	5582	5872	3197	4919	2792	5991	4058	9769	1918
6859	9606	0522	4993	0345	8958	1289	5825	6941	7685
6590	1932	6043	3623	1978	4112	1795	8465	2110	8045
3482	0478	0221	6738	7323	5643	4767	0106	2272	9862
5489	5583	3156	0835	1988	3912	0938	7460	0869	4420
3522	0935	7877	5665	7020	9555	7379	7124	7878	5544
7555	7579	2550	2487	9477	0864	2349	1012	8250	2633
5759	3554	5080	9074	7001	6249	3234	6368	9102	2672
6303	6895	3371	3196	7231	2918	7380	0438	7547	2644
7351	5634	5323	2623	7803	8374	2191	0464	0696	9529
7068	7803	8832	5119	6350	0120	5026	3684	5667	0304
3613	1428	1796	8447	0503	5654	3254	7336	9536	1944
5143	4534	2105	0368	7890	2473	4240	8562	9435	1422
9815	5144	7649	8638	6137	8070	5345	4865	2456	5708
5780	1277	6316	1013	2867	9938	3930	3203	5696	1769
1187	0951	5991	5245	5700	5564	7352	0891	6249	6568
4184	2179	4554	9083	2254	2435	2965	5154	1209	7069
2916	2972	9885	0275	0144	8043	8122	3213	7666	0230
5524	1341	9860	6565	6981	9842	0171	2284	2707	3008