Державний біотехнологічний університет

Навчально-науковий інститут «Кіберпорт»

Кафедра автоматизації та комп'ютерно-інтегрованих технологій

Методичні вказівки

до практичних занять

з дисципліни

"Теорія ймовірностей та математична статистика"

(для студентів, що навчаються за напрямком підготовки
123 - "Комп’ютерна інженерія")

2023

Укладач:

Абраменко Іван Григорович, к.т.н., доцент,

Рекомендовано до друку:

на засіданні кафедри автоматизації та комп’ютерно-інтегрованих технологій, протокол № від . .2023 р.

ЗМІСТ

ПЕРЕДМОВА

Ефективність діяльності і ступінь конкурентоспроможності студентів-випускників у наш час у першу чергу визначається їхнім практичним умінням самостійно знаходити розв'язки конкретних технічних завдань за мінімально можливий час, а значить від ступеня засвоєння ними сучасних комп'ютерних технологій. Високий рівень складності розрахункових завдань в системах керування обумовлює складність відповідних обчислювальних алгоритмів і їх реалізацій на конкретних мовах програмування.

Теорія ймовірностей - це математична наука, що вивчає закономірності в масових випадкових явищах.

Випадкові відхилення завжди супроводжують будь-яке явище. Елемент невизначеності, складності, багатопричинності, що притаманний випадковим явищам, обумовлює створення спеціальних методів для їх вивчення.

На практиці спостереження за масовою сукупністю однорідних випадкових явищ відкривають у них цілком певні, властиві саме їм закономірності, свого роду стійкості. Мета ймовірнісних (статистичних) методів полягає в тому, щоб, обминаючи надто складне і часто практично неможливе дослідження окремих випадкових явищ, звернутися безпосередньо до законів, що керують їх масовими проявами. Вивчення цих законів дозволяє здійснювати прогноз середнього масового результату, цілеспрямовано впливати на хід явиш, контролювати їх, обмежувати сферу дії випадковості, звужувати її вплив.

Теорія ймовірностей служить фундаментом, на якому грунтуються важливі для практичних застосувань її органічні доповнення - математична статистика та теорія випадкових процесів.

Математична статистика - це розділ математики, в якому вивчаються методи збору, обробки й аналізу великих масивів стохастичних дослідних даних з метою виявлення закономірностей. Користуючись апаратом теорії ймовірностей, математична статистика дозволяє оцінювати ступінь точності та надійності висновків, що одержуються при обробці дослідних даних. Вона застосовується при плануванні й організації виробництва, аналізі технологічних процесів, контролі якості продукції, вивченні закономірностей еволюції систем прикладного характеру, що розвиваються в умовах стохастичної невизначеності, та в багатьох інших сферах.

Практично немає галузі науки, техніки чи суспільного життя, де б не використовувалися статистичні методи. Ознайомлення з ними необхідне сьогодні кожному освіченому фахівцю, оскільки його не можна вважати професійно грамотним, якщо він не може дати кількісної оцінки правильності вибору дій в умовах стохастичної невизначеності, здатних привести до виграшу чи втрат, статистично обґрунтувати вибір прийнятого рішення з оперативного керування виробництвом.

Основні поняття, методи, теореми та формули теорії ймовірностей ефективно використовують у науці, техніці, економіці, зокрема теоріях надійності та масового обслуговування, в плануванні та організації виробництва, страховій та податковій справах, соціології та політології, демографії та охороні здоров’я.

Призначення дисципліни "Моделювання систем автоматичного керування" – ознайомити студентів з найбільш характерними типами завдань розрахунку систем автоматичного керування, розв'язуваних із застосуванням комп'ютерних технологій, привити навики використання відповідних числових методів, засвоїти особливості їх застосування. При цьому є дуже важливим правильний вибір програмних засобів для використання в процесі навчання. З одного боку, ці засоби повинні бути широко розповсюдженими, так, щоб, потім, після приходу на виробництво, в організацію, випускник легко міг їх знайти і використати. З іншого боку, така програма повинна надавати відповідну математичну базу. Нарешті, бажаним є дружній інтерфейс, легкий в оволодінні.

Таким засобом був вибраний програмний комплекс Matlab - одна з найстарших, ретельно пророблених і перевірених часом систем автоматизації математичних розрахунків, яка побудована на застосуванні матричних операцій. Це знайшло відбиття в назві системи - Matrix Laboratory - матрична лабораторія. До складу Matlab ходить програма Simulink, яка дає можливість імітувати різноманітні реальні системи автоматичного керування і окремі пристрої, задаючи їх моделями, складеними з функціональних блоків. Simulink має велику і розширювану користувачами бібліотеку блоків і прості засоби їх завдання і зміни.

Практичне заняття № 1. Випадкові події

Мета. Одержання практичних навичок визначення ймовірностей випадкових подій шляхом використання методів теорії ймовірностей.

Теоретична частина. Під *експериментом (випробуванням*) розуміють деяку сукупність умов, в яких спостерігається те або інше явище, фіксується той або інший результат.

Подією (або випадковою подією) називається усякий факт, який в результаті експерименту може відбутися або не відбутися.

Ймовірністю події називається чисельна міра ступеня об'єктивної можливості появи цієї події в результаті нового експерименту.

Ймовірність події A позначається як $P(A)$.

Достовірною називається подія U, яка в результаті експерименту неодмінно відбувається. Для достовірної події $P(U)=1$.

Неможливою називається подія, яка в результаті експерименту не може відбутися. Ймовірність появи неможливої події дорівнює нулю.

Ймовірність власне випадкової події A лежить у межах від нуля до одиниці - $0<P(A)<1$.

Повною групою подій називається декілька попарно несумісних подій таких, що в результаті експерименту одна з них неодмінно відбувається.

Декілька подій називаються *несумісними*, якщо ніякі дві з них не можуть відбутися одночасно в одному експерименті.

Декілька подій називаються *рівноможливими*, якщо вони володіють рівним ступенем об'єктивної можливості відбутися в результаті експерименту.

Якщо результати експерименту утворюють повну групу несумісних рівноможливих подій, то вони називаються *випадками*.

Випадок називається сприятливим до події A, якщо його поява тягне за собою появу події A.

Якщо результати експерименту зводяться до схеми випадків, то ймовірність події A обчислюється за формулою

$$P(A)=\frac{m}{n} \qquad (1)$$

де n – загальне число випадків; m – число випадків, що сприятливі до події A.

Часто для підрахунку величин n і m у формулі (1) використовують формули комбінаторики:

для числа сполучень з n елементів по m – формулу

$$C_n^m=\frac{n!}{(n-m)!m!}=C_n^{n-m};$$

для числа розміщень з n елементів по m – формулу

$$A_n^m=\frac{n!}{(n-m)!};$$

для числа перестановок з m елементів – формулу

$$P_n=n!=A_n^n.$$

При цьому $n!=1\cdot 2\cdot\ldots\cdot n$. Виключення $0!=1$.

Завдання. Провести розв’язок прикладів 1-5.

Приклад 1. Кидають одночасно дві гральні кістки. Знайти ймовірності подій: A – сума очок на обох кістках дорівнює 6; B – добуток очок на обох кістках дорівнює 8; C – сума та добуток очок на обох кістках дорівнює 8.

Розв’язання. Загальне число можливих елементарних результатів експерименту $n=36$, оскільки випадання очок на одній кістці має 6 варіантів, і кожен варіант однієї кістки може поєднуватися з 6 варіантами іншої кістки. Всі результати складають повну групу несумісних рівноможливих подій.

Сприятливими до події A є наступні результати кидання кісток: 2+6; 3+5; 4+4; 5+3; 6+2, тобто $m=5$. Шукана ймовірність за формулою (1)

$$P(A)=\frac{5}{36}.$$

Сприятливими до події B є два результати: $2\cdot 4$ і $4\cdot 2$, тобто $m=2$. За формулою (1)

$$P(B)=\frac{2}{36}=\frac{1}{18}.$$

Сприятливих до події C результатів немає, тобто $m=2$, і $P(C)=0$.

Приклад 2. В ящику 100 деталей, з них 10 – браковані. Навмання витягують 4 деталі. Знайти ймовірність події A – наявність рівно трьох стандартних деталей серед витягнутих.

Розв'язання. Загальне число можливих виходів експерименту $n = C_{100}^{4}$.

Всі вони утворюють повну групу несумісних рівноможливих подій. Підрахуємо число результатів, що сприяють події A. Три стандартні деталі з 90 наявних в ящику можна витягнути C_{90}^{3} способами. З кожною вибіркою з трьох стандартних деталей може поєднуватися одна нестандартна деталь з 10, тобто кожне поєднання трьох стандартних деталей з одною бракованою деталлю може здійснюватися C_{10}^{1} способами.

Кількість сприятливих випадків становить $m = C_{90}^{3} C_{10}^{1}$. Отже

$$P(A) = \frac{m}{n} = \frac{C_{90}^{3} C_{10}^{1}}{C_{100}^{4}} = 0{,}3.$$

Приклад 3. На десяти картках написані цифри 0, 1, …, 9. Три з них вибираються навмання і укладаються на стіл в порядку появи. Знайти ймовірність того, що: а) вийде число 245 (подія A); б) з вибраних цифр можна скласти число 245 (подія B).

Розв'язання. Загальне число всіх можливих виходів експерименту – це число розміщень з 10 елементів по 3. Отримані з'єднання елементів (карток) можуть відрізнятися один від одного або самими елементами, або порядком їх входження. Всі результати утворюють повну групу несумісних рівноможливих подій у кількості $n = A_{10}^{3} = 720$. Із загального числа результатів тільки один сприятливий отриманню числа 245, тобто число сприяючих результатів $m = 1$.

Тоді шукана ймовірність події A за формулою (1)

$$P(A) = \frac{m}{n} = \frac{1}{720}.$$

На відміну від події A для події B загальне число результатів експерименту обчислюється як число з'єднань з 10 по 3, оскільки порядок вибору елементів не грає ролі. Шукана ймовірність

$$P(B)=\frac{1}{C_{10}^{3}}=\frac{1}{120}.$$

Приклад 4. З п'яти букв розрізної азбуки складено слово «КНИГА». Дитина, що не уміє читати, розсипала букви, а потім зібрала їх в довільному порядку. Знайти ймовірність того, що у нього знову вийшло слово «КНИГА».

Розв'язання. З п'яти букв дитина може скласти різні буквосполучення, які відрізняються один від одного тільки порядком входження букв. Тому число всіх результатів експерименту обчислимо як число перестановок з 5 елементів - $n=P_5=5!=120$. Всі результати утворюють повну групу несумісних рівноможливих подій, з яких тільки одна сприяє появі події A – відновленню слова «КНИГА». Отже, шукана ймовірність

$$P(A)=\frac{m}{n}=\frac{1}{120}.$$

Приклад 5. З шести букв розрізної азбуки складено слово «*АНАНАС*». Знайти, як і в попередньому завданні, ймовірність відновлення слова.

Розв'язання. Загальне число можливих результатів експерименту $n=P_6=6!=720$. Число сприятливих результатів m більше, ніж в попередньому завданні. Слід врахувати, що перестановка місцями двох букв *Н*, значення слова не змінює. Відповідне число перестановок визначається як $P_2=2!=2$. Але з кожною перестановкою букв *Н* може поєднуватися перестановка з трьох букв *А*. Загальне число перестановок цих трьох букв визначається як $P_3=3!=6$. Таким чином, число сприяючих результатів $m=P_2\cdot P_3=12$. Шукана ймовірність

$$P(A)=\frac{m}{n}=\frac{12}{720}=\frac{1}{60}.$$

Практичне заняття № 2. Основні теореми теорії ймовірностей

Мета. Одержання практичних навичок використання основних теорем визначення теорії ймовірностей.

Теоретична частина. *Сумою двох подій A* і *B* називають подію *C*, що полягає в появі хоча би однієї з подій *A* і *B*.

Сумою декількох подій називають подію, що полягає в появі хоча би однієї з цих подій.

Добутком двох подій A і *B* називають подію *D*, що полягає в сумісній появі подій *A* і *B*.

Добутком декількох подій називають подію, що полягає в сумісній появі всіх цих подій.

Теорема про ймовірність суми подій. Ймовірність суми двох подій дорівнює сумі їх ймовірностей за відрахуванням імовірності добутку цих подій

$$P(A+B)=P(A)+P(B)-P(AB) \tag{2}$$

У разі *несумісних* подій імовірність їх суми визначається за формулами:

$$P(A+B)=P(A)+P(B) \text{ – для двох подій } A \text{ і } B; \tag{3}$$

$$P(A_1+A_2+\ldots+A_n)=P(A_1)+P(A_2)+\ldots+P(A_n) \text{ – для } n \text{ подій.} \tag{4}$$

Дві несумісні події A і $\bar{A}$ називаються *протилежними*, якщо вони складають повну групу. Сума ймовірності протилежних подій дорівнює одиниці

$$P(A)+P(\bar{A})=1. \tag{5}$$

Теорема про ймовірність добутку двох подій. Ймовірність добутку двох подій дорівнює добутку ймовірності одного з них на умовну ймовірність іншої за умови, що перша подія відбулася

$$P(AB)=P(A)\cdot P_A(B)=P(B)\cdot P_B(A) \tag{6}$$

де – $P_A(B)$ умовна ймовірність події *B* за умови, що відбулася подія *A*; $P_B(A)$ – умовна ймовірність події *A* за умови, що відбулося події *B*.

У разі *незалежних* подій формула (6) спрощується:

$$P(AB)=P(A)\cdot P(B) \text{ – для двох подій } A \text{ і } B; \tag{7}$$

$$P(A_1 \cdot A_2 \cdot \ldots \cdot A_n) = P(A_1) \cdot P(A_2) \cdot \ldots \cdot P(A_n) \text{ – для } n \text{ подій.} \quad (8)$$

Якщо в умовах експерименту подія *A* з'являється спільно з одним з повної групи несумісних подій (гіпотез) H_i $(i \in \{1, 2, \ldots, n\})$, то середня ймовірність події *A* визначається за *формулою повної* (середньою) *ймовірності*

$$P(A) = \sum_{i=1}^{n} P(H_i) \cdot P_{H_i}(A / H_i), \quad (9)$$

де $P(H_i)$ – ймовірність гіпотези H_i; $P(A / H_i)$ – умовна ймовірність події *A* при здійсненні гіпотези H_i.

Якщо відома апріорна ймовірність гіпотез $P(H_i)$ і відомо, що подія *A* відбулася, то апостеріорна ймовірність гіпотез обчислюється за *формулою Байєса*

$$P(H_j / A) = \frac{P(H_j) \cdot P(A / H_j)}{\sum_{i=1}^{n} P(H_i) \cdot P(A / H_j)}. \quad (10)$$

Завдання. Провести розв'язок прикладів 6-9.

Приклад 6. Ймовірність влучення в ціль при одному пострілі з першої гармати дорівнює 0,7; з другої – 0,8. Знайти ймовірність ураження цілі при одному залпі з двох гармат.

Розв'язання. Введемо позначення. Хай *A* – подія, яка полягає у влученні в ціль першою гарматою; *B* – другою. Ці події є сумісними і незалежними. Отже, подія *C* (ураження мішені при залпі) є сумою двох сумісних подій. Ймовірність події *C* можна визначити за формулою (2)

$$P(C) = P(A+B) = P(A) + P(B) - P(AB) = P(A) + P(B) - P(A) \cdot P(B) =$$
$$= 0{,}7 + 0{,}8 - 0{,}7 \cdot 0{,}8 = 0{,}94.$$

Цю задачу можна вирішити й іншим способом.

Ціль буде уражена, якщо відбудеться одна з трьох несумісних подій: $A_1 \cdot \bar{A}_2$ – в ціль влучила перша гармата та не влучила друга; $\bar{A}_1 \cdot A_2$ – в ціль не влучила перша гармата та влучила друга; $A_1 \cdot A_2$ – в ціль влучили обидві гармати. В цьому випадку, застосовуючи теорему про суму несумісних подій у формі (4), а потім теорему про добуток незалежних подій, отримаємо

$$P(C)=P(A_1\cdot\bar{A}_2)+P(\bar{A}_1\cdot A_2)+P(A_1\cdot A_2)=P(A_1)\cdot P(\bar{A}_2)+P(\bar{A}_1)\cdot P(A_2)+P(A_1)\cdot P(A_2)=$$
$$=0,7\cdot(1-0,8)+(1-0,7)\cdot 0,8+0,7\cdot 0,8=0,94.$$

Найпростіший спосіб вирішення цієї задачі полягає в подані ймовірності події *C* через імовірність протилежної події $\bar{C}$ – промах обох гармат

$$P(C)=1-P(\bar{C})=1-P(\bar{A}_1\cdot\bar{A}_2)=1-P(\bar{A}_1)\cdot P(\bar{A}_2)=1-(1-0,7)(1-0,8)=0,94.$$

Приклад 7. Студент прийшов на іспит, знаючи відповіді на 15 з 20 питань програми. Знайти ймовірність того, що студент відповість на три екзаменаційних питання.

Розв'язання. Подія *C* (студент знає відповіді на всі три питання) є добутком трьох залежних подій: A_1 (студент знає відповідь на перше питання); A_2 (студент знає відповідь на друге питання за умови, що він відповів на перше) і A_3 (студент знає відповідь на третє питання за умови, що він відповів на перше і друге). Обчислимо ймовірності цих подій

$$P(A_1)=\frac{15}{20};\ P_{A_1}(A_2)=\frac{14}{19};\ P_{A_1\cdot A_2}(A_3)=\frac{13}{18}.$$

За теоремою про ймовірність добутку залежних подій

$$P(C)=P(A_1)\cdot P_{A_1}(A_2)\cdot P_{A_1\cdot A_2}(A_3)=\frac{15}{20}\cdot\frac{14}{19}\cdot\frac{13}{18}=0,368.$$

Приклад 8. З п'яти букв розрізної азбуки складено слово «*КНИГА*». Дитина, що не уміє читати, розсипала букви, а потім зібрала в довільному порядку. Знайти ймовірність того, що у нього знову склалося слово «*КНИГА*».

Розв'язання. Ця задача вже розглядалася (див. Приклад 4). Наведемо другий варіант вирішення, використовуючи основні теореми теорії ймовірностей.

Щоб в порядку появи букв склалося слово «*КНИГА*», першою повинна з'явитися буква *К*. Ймовірність такої події становить $P(K)=1/5$, оскільки з п'яти можливих результатів тільки один сприяє появі букви *К*. Припустимо, що ця подія відбулася. Тоді ймовірність того, що з чотирьох букв, що залишилися, наступною з'явиться *Н*, визначається як $P_K(H)=1/4$. Аналогічно обчислюється ймовірність послідовної появи букв *И, Г і А* - $P_{K\cdot H}(И)=1/3$, $P_{K\cdot H\cdot И}(Г)=1/2$,

$P_{К\cdot Н\cdot И\cdot Г}(A)=1$. За теоремою про ймовірність добутку залежних подій знайдемо шукану ймовірність

$$P(К\cdot Н\cdot И\cdot Г\cdot A)=\frac{1}{5}\cdot\frac{1}{4}\cdot\frac{1}{3}\cdot\frac{1}{2}\cdot 1=\frac{1}{120}.$$

Приклад 9. На склад надходить продукція з трьох фабрик, при цьому частка продукції першої фабрики становить 20%, другої – 46%, третьої – 34%. Відомо, що середній відсоток нестандартних виробів для першої фабрики становить 3%, другої – 2%, третьої – 1%. Знайти ймовірність того, що: *а*) навмання узятий виріб виявиться нестандартним; *б*) виріб виготовлений на першій фабриці, якщо він виявився нестандартним; *в*) виріб виготовлений на другій фабриці, якщо він виявився стандартним; *г*) визначити, на якій фабриці найімовірніше було виготовлено виріб, якщо він виявився стандартним?

Розв'язання. *а*) Навмання узятий виріб може бути виготовлений або на першій фабриці (гіпотеза H_1), або на другій (гіпотеза H_2), або на третій (гіпотеза H_3). Всі гіпотези несумісні і складають повну групу. Ймовірність кожної гіпотези визначається приведенням процентної частки продукції відповідної фабрики в безрозмірну величину, тобто діленням заданої за умовами частки на 100%. Так, $P(H_1)=0,2$; $P(H_2)=0,46$; $P(H_3)=0,34$. Аналогічно визначається умовна ймовірність події A (виріб є нестандартним): $P(A/H_1)=0,03$; $P(A/H_2)=0,02$; $P(A/H_3)=0,01$. Тепер, використовуючи формулу (9), можна отримати шукану повну ймовірність події A

$$P(A)=\sum_{i=1}^{3}P(H_i)\cdot P(A/H_i)=P(H_1)\cdot P(A/H_1)+P(H_2)\cdot P(A/H_2)+P(H_3)\cdot P(A/H_3)=$$
$$=0,2\cdot 0,03+0,46\cdot 0,02+0,34\cdot 0,01=0,0186.$$

б) Відомо, що подія A вже відбулося. Потрібно визначити апостеріорну ймовірність гіпотези H_1. Шукану ймовірність знайдемо за формулою Байєса

$$P(H_1/A)=\frac{P(H_1)\cdot P(A/H_1)}{\sum_{i=1}^{n}P(H_i)\cdot P(A/H_j)}=\frac{P(H_1)\cdot P(A/H_1)}{P(A)}=\frac{0,2\cdot 0,03}{0,0186}=0,3226.$$

в) По умові завдання виріб виявився стандартним, тобто в прийнятих нами позначеннях відбулася подія A . Необхідно знайти апостеріорну ймовірність гіпотези H_2.

За формулою Байєса

$$P(H_2/\bar{A})=\frac{P(H_2)\cdot P(\bar{A}/H_2)}{P(\bar{A})}.$$

Події A і $\bar{A}$ є протилежними. З урахуванням (5) $P(\bar{A})=1-P(A)=1-0,0186=0,9814$. Аналогічного обчислюємо умовну ймовірність події за умови, що здійснилася гіпотеза H_2 - $P(\bar{A}/H_2)=1-P(A/H_2)=1-0,02=0,98$.

Підставляючи знайдену ймовірність у формулу Байєса, отримаємо шукану ймовірність

$$P(H_2/\bar{A})=\frac{P(H_2)\cdot P(\bar{A}/H_2)}{P(\bar{A})}=\frac{0,46\cdot 0,98}{0,9814}=0,4593.$$

г) Щоб визначити, на якій фабриці найімовірніше було виготовлено стандартний виріб, необхідно порівняти між собою апостеріорні ймовірності гіпотез $P(H_1/\bar{A})$, $P(H_2/\bar{A})$, $P(H_3/\bar{A})$. Найбільша з цих ймовірностей визначає шукану фабрику. Одна з вказаних ймовірність була тільки що визначена, а саме $P(H_2/\bar{A})=0,4593$. Аналогічно визначимо інші апостеріорні ймовірності гіпотез $P(H_1/\bar{A})=0,1977$, $P(H_3/\bar{A})=0,343$. Найбільша апостеріорна ймовірність відповідає другій гіпотезі. Отже, стандартний виріб найімовірніше був виготовлений на другій фабриці.

Практичні заняття № 3-4. Випадкові величини та їх закони розподілу

Мета. Одержання практичних навичок опису випадкових величин шляхом використання методів теорії ймовірностей.

Теоретична частина. *Випадковою величиною* називають величину, яка в результаті експерименту приймає заздалегідь невідоме значення.

Дискретною випадковою величиною називають випадкову величину, можливі значення якої належать доліченої множини (скінченої або нескінченої).

Неперервною випадковою величиною називають випадкову величину, можливі значення якої належать до Неперервної множини (обмеженої або необмеженої).

Закон розподілу – це вичерпна характеристика випадкової величини, що встановлює зв'язок між можливими значеннями випадкової величини і відповідними ймовірностями.

Закон розподілу дискретної випадкової величини може бути заданий у вигляді *ряду розподілу* або *інтегральної функції розподілу.*

Ряд розподілу – таблиця, що складається з двох рядків. У першому рядку перераховуються всі можливі значення випадкової величини в порядку їх зростання, а в другій – відповідна ймовірність.

x_i	x_1	x_2	…	x_n
p_i	p_1	p_2	…	p_n

Властивість ряду розподілу для будь-якої дискретної випадкової величини

$$\sum_{i=1}^{n} p_i = 1. \qquad (11)$$

Інтегральною функцією розподілу випадкової величини X називають функцію $F(x)$, що при кожному значенні свого аргументу x дорівнює ймовірності того, що випадкова величина X прийме значення, менше x

$$F(x) = P(X < x). \tag{12}$$

Закон розподілу неперервної випадкової величини може бути заданий *інтегральною функцією розподілу* (12) або *щільністю розподілу*

Щільність розподілу ймовірності є першою похідною від інтегральною функції розподілу

$$f(x) = \frac{dF}{dx}. \tag{17}$$

Властивості щільності розподілу

$$f(x) \geq 0, \tag{18}$$

$$\int_{-\infty}^{+\infty} f(x)dx = 1. \tag{19}$$

Інтегральна функція розподілу $F(x)$ може бути виражена через щільність розподілу (зворотне перетворення)

$$F(x) = \int_{-\infty}^{x} f(xt)dt. \tag{20}$$

Властивості інтегральної функції розподілу

$$F(-\infty) = 0; \tag{21}$$

$$F(+\infty) = 1; \tag{22}$$

$$\text{якщо } x_2 > x_1, \text{ то } F(x_2) \geq F(x_1). \tag{23}$$

Ймовірність влучення Неперервної випадкової величини на задану ділянку $[\alpha,\beta]$

$$P(\alpha \leq X < \beta) = F(\beta) - F(\alpha) = \int_{\alpha}^{\beta} f(x)dx. \tag{24}$$

Основні числові характеристики:

математичне сподівання дискретної випадкової величини

$$m_x = M[X] = \sum_{i}^{n} x_i p_i; \tag{25}$$

математичне сподівання неперервної випадкової величини

$$m_x = M[X] = \int_{-\infty}^{+\infty} x \cdot f(x)dx; \tag{26}$$

дисперсія дискретної випадкової величини

$$D_x = D[X] = \sum_i (x_i - m_x)^2 p_i; \tag{27}$$

дисперсія неперервної випадкової величини

$$D_x = D[X] = \int_{-\infty}^{+\infty} (x - m_x)^2 f(x)dx; \tag{28}$$

середнє квадратичне відхилення

$$\sigma_x = \sqrt{D_x}; \tag{29}$$

другий початковий момент дискретної випадкової величини

$$\alpha_2 = M[X^2] = \sum_i x_i^{\ 2} p_i; \tag{30}$$

другий початковий момент неперервної випадкової величини

$$\alpha_2 = M[X^2] = \int_{-\infty}^{+\infty} x^2 f(x)dx. \tag{31}$$

Формула зв'язку дисперсії з другим початковим моментом і математичним очікуванням

$$D_x = \alpha_2 - m_x^2. \tag{32}$$

Завдання. Провести розв'язок слідуючих п'яти прикладів.

Приклад 10. Схожість насіння оцінюється ймовірністю 0,8. Розглядається випадкова величина X – число насінин, що зійшли, серед п'яти посіяних. Визначити закон розподілу у вигляді ряду розподілу. Побудувати графік інтегральної функції розподілу $F(x)$. Знайти математичне сподівання, дисперсію і середнє квадратичне відхилення σ_x випадкової величини X. Знайти також імовірність влучення випадкової величини X в інтервал значень (-5; 3,5).

Розв'язання. Випадкова величина X в умовах цієї задачі може приймати одне з числових значень: 0, 1, 2, 3, 4, 5. Для побудови ряду розподілу залишається визначити тільки відповідні ймовірності. У різних задачах ці ймовірності визначаються за різними методиками, розглянутими в попередніх розділах курсу. В умовах даної задачі найбільш доцільним способом визначення шуканих ймовірностей є формула Бернуллі, яка використовується для n незалежних випробувань, у кожному з яких подія *A* з’являється з однаковою ймовірністю *p*. Формула Бернуллі визначає ймовірність того, що в цих випробуваннях подія *A* відбудеться рівно *k* разів (байдуже в якій послідовності)

$$P_n(k) = C_n^k \cdot p^k \cdot (1-p)^{n-k}.$$

В нашому прикладі *n*=5; *p*=0,8. Підставляючи замість *k* послідовно всі можливі значення випадкової величини, отримаємо відповідні ймовірності: для $x_1 = 0 - p_1 = 0{,}00032$; для $x_2 = 1 - p_2 = 0{,}0064$; для $x_3 = 2 - p_3 = 0{,}0512$; для $x_4 = 3 - p_4 = 0{,}2048$; для $x_5 = 4 - p_5 = 0{,}4096$; для $x_6 = 5 - p_6 = 0{,}32768$. Шуканий ряд розподілу має вигляд:

x_i	0	1	2	3	4	5
p_i	0,00032	0,0064	0,0512	0,2048	0,4086	0,32768

Ряд розподілу містить вичерпну інформацію про випадкову величину, тому його можна використовувати для знаходження відповідей на решту питань задачі. Зокрема – для побудови графіка інтегральної функції *F(x).*

При побудові графіка *F(x)* вісь абсцис розбивається можливими значеннями випадкової величини на $(n+1)$ діапазон: 1-й діапазон – $x^I \le 0$; 2-й – $0 < x^{II} \le 1$; 3-й – $1 < x^{III} \le 2$; 4-й – $2 < x^{IV} \le 3$; 5-й – $3 < x^{V} \le 4$; 6-й – $4 < x^{VI} \le 5$; 7-й – $3 < x^{VII}$ У кожному з таких діапазонів функція *F(x)* має постійне значення (рис.1).

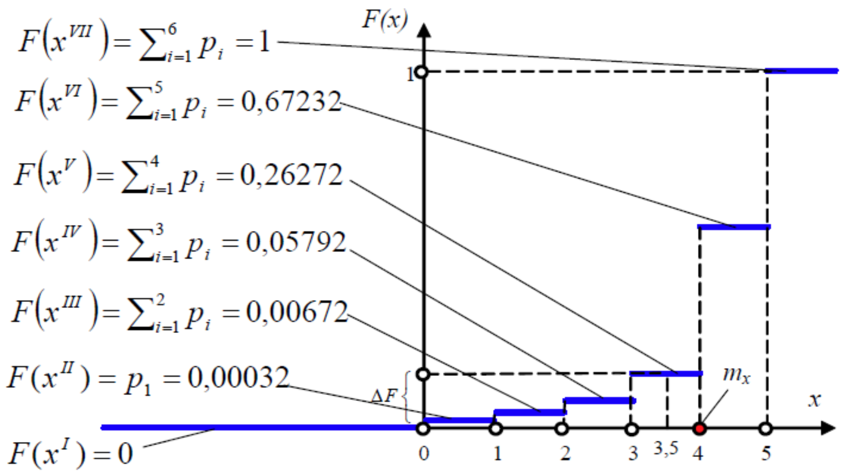


Рис. 1– Графік інтегральної функції розподілу *F(x)*

Математичне сподівання m_x випадкової величини X визначається за формулою (25) при n=6:

$$m_x = \sum_{i=1}^{6} x_i p_i = x_1 p_1 + x_2 p_2 + x_3 p_3 + x_4 p_4 + x_5 p_5 + x_6 p_6 =$$
$$= 0 \cdot 0,00032 + 1 \cdot 0,0064 + 2 \cdot 0,0512 + 3 \cdot 0,2048 + 4 \cdot 0,4096 + 5 \cdot 0,032768 = 4.$$

Математичне сподівання m_x є числовою характеристикою випадкової величини X, лежить в області визначення останньою і зображається крапкою на осі абсцис (див. рис.1).

Дисперсія D_x випадкової величини X визначається по формулі (27) при n=6

$$D_x = \sum_i (x_i - m_x)^2 p_i = (0-4)^2 0,00032 + (1-4)^2 0,0064 + (2-4)^2 0,0512 +$$
$$+ (3-4)^2 0,2048 + (4-4)^2 0,4096 + (5-4)^2 0,32768 = 0,8.$$

Середнє квадратичне відхилення випадкової величини X визначається по формулі (29)

$$\sigma_x = \sqrt{D_x} = \sqrt{0,8} = 0,8944.$$

Ймовірність попадання випадкової величини в заданий інтервал значень від $x_1 = -5$ до $x_2 = 2,7$ можна визначити двома способами:

а) за допомогою графіка інтегральної функції *F(x)*

$$P(-5 < X < 3,5) = F(3,5) - F(-5) = 0,26272 - 0 = 0,26272,$$

на рис.1 цій ймовірності відповідає відрізок ΔF;

б) за допомогою основних теорем теорії ймовірності як ймовірність складної події

$$P(-5 < X < 3,5) = P(X = 0) + P(X = 1) + P(X = 2) + P(X = 3) =$$
$$= 0,00032 + 0,0064 + 0,0512 + 0,2048 = 0,26272.$$

Приклад 14. Неперервна випадкова величина X задана щільністю розподілу

$$f(x) = \begin{cases} 0 \ \ при \ \ x \notin [0;1]; \\ cx \ \ при \ \ x \in [0;1]. \end{cases}$$

Знайти значення постійної величини c, математичний вираз інтегральної функції розподілу *F(x)*. Побудувати графік щільності розподілу *f(x)* і графік інтегральної функції розподілу *F(x)*. Визначити математичне сподівання m_x, дисперсію D_x, середнє квадратичне відхилення σ_x і ймовірність попадання випадкової величини в інтервал (0; 0,5).

Розв'язання. Постійна величина c визначається за допомогою 2-ої властивості щільності розподілу (19). Обчислюється визначний інтеграл в нескінчених межах від заданої функції щільності розподілу *f(x)* і прирівнюється одиниці.

Отримане рівняння вирішується відносно постійної c.

Оскільки задана *f(x)* – кусково-неперервна функція, то інтеграл від неї розпадається на три інтеграли

$$\int_{-\infty}^{+\infty} f(x)dx = \int_{-\infty}^{0} 0\cdot dx + \int_{0}^{1} cxdx + \int_{1}^{+\infty} 0\cdot dx = 0 + \left.\frac{cx^2}{2}\right|_0^1 + 0 = \frac{c}{2}.$$

Прирівнюючи отриманий вираз одиниці, отримуємо рівняння $c/2 = 1$, з якого $c = 2$.

З урахуванням знайденої константи щільність розподілу набуде вигляду

$$f(x) = \begin{cases} 0 \ \text{при} \ x \notin [0;1]; \\ 2x \ \text{при} \ x \in [0;1]. \end{cases}$$

Математичний вираз інтегральної функції розподілу *F(x)* знаходиться за допомогою зворотного перетворення (20), при цьому перетворення здійснюється для кожного «відрізку» функції *F(x)* окремо

при $x \le 0$ $F(x) = \int_{-\infty}^{x} 0\cdot dt = 0;$

при $0 < x \le 1$ $F(x) = \int_{-\infty}^{0} 0\cdot dt + \int_{0}^{x} 2t\cdot dt = 0 + t^2\Big|_0^x = x^2;$

при $1 < x$ $F(x) = \int_{-\infty}^{0} 0\cdot dt + \int_{0}^{1} 2t\cdot dt + \int_{1}^{x} 0\cdot dt = 0 + t^2\Big|_0^1 + 0 = 1.$

Остаточно отримуємо

$$F(x)=\begin{cases}0 \ \textit{якщо} \ x\le 0;\\ x^2 \textit{якщо} \ 0<x\le 1;\\ 1 \ \textit{якщо} \ x>0.\end{cases}$$

Графіки *f(x)* і *F(x)* мають вигляд, показаний відповідно на рис.2 і рис.3.

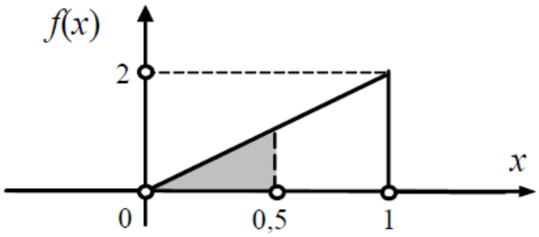


Рис. 2 – Графік щільності розподілу *f(x)*

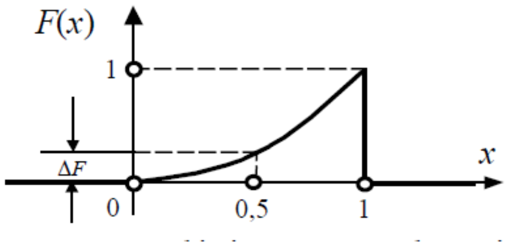


Рис. 3 – Графік інтегральної функції розподілу *F(x)*

Математичне сподівання m_x неперервної випадкової величини X визначається за формулою (26). При цьому інтеграл від кусково-неперервної функції *f(x)* в нескінченних межах розпадається на три інтеграли відповідно до числа «кусків» підінтегральної функції

$$m_x=\int_{-\infty}^{+\infty}x\cdot f(x)dx=\int_{-\infty}^{0}x\cdot 0\cdot dx+\int_{0}^{1}x\cdot 2xdx+\int_{1}^{+\infty}x\cdot 0\cdot dx=0+\left.\frac{2x^3}{3}\right|_0^1+0=\frac{2}{3}.$$

Дисперсію D_x неперервної випадкової величини X доцільно визначати за допомогою формули (32), для чого заздалегідь знаходять другий початковий момент за формулою (31)

$$\alpha_2=\int_{-\infty}^{+\infty}x^2 f(x)dx=\int_{-\infty}^{0}x^2\cdot 0\cdot dx+\int_{0}^{1}x^2\cdot 2xdx+\int_{1}^{+\infty}x^2\cdot 0\cdot dx=0+\left.\frac{2x^4}{4}\right|_0^1+0=\frac{1}{2};$$

$$D_x=\alpha_2-m_x^2=\frac{1}{2}-\left(\frac{2}{3}\right)^2=\frac{1}{18}.$$

Середньоквадратичне відхилення випадкової величини X визначається за формулою (29)

$$\sigma_x=\sqrt{D_x}=\sqrt{1/18}=0{,}2357\,.$$

Ймовірність влучення неперервної випадкової величини в заданий інтервал значень від $x_1 = 0$ до $x_2 = 0{,}5$ можна визначити двома способами: *а*) за допомогою інтегральної функції; *б*) за допомогою щільності розподілу.

а) $P(0 < X < 0{,}5) = F(0{,}5) - F(0) = \left(\frac{1}{2}\right)^2 - 0^2 = \frac{1}{4}.$

На рис.3 цій ймовірності відповідає відрізок вісі ΔF.

б) $P(-5 < X < 3{,}5) = \int_0^{1/2} f(x)dx = \int_0^{1/2} 2xdx = \frac{1}{4}.$

На рис.2 цій ймовірності відповідає площа, що виділена сірим фоном.

Приклад 15. Неперервна випадкова величина X задана інтегральною функцією розподілу

$$F(x) = \begin{cases} 0 \ \text{якщо} \ x \le 0; \\ c(1 - \cos x) \ \text{якщо} \ 0 < x \le \pi; \\ 1 \ \text{якщо} \ x > \pi. \end{cases}$$

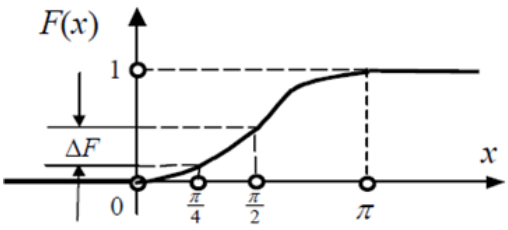
Знайти значення константи *с*, математичний вираз щільності розподілу *f(x)*. Побудувати графік інтегральної функції розподілу *F(x)* і графік щільності розподілу *f(x)*. Визначити математичне сподівання m_x, дисперсію D_x, середнє квадратичне відхилення σ_x і ймовірність попадання випадкової величини в інтервал $\left(\frac{\pi}{4}; \frac{\pi}{2}\right)$.

Розв'язання. Постійна величина *с* визначається за допомогою 2-ої властивості інтегральної функції розподілу (22) $F(+\infty) = 1$. В умовах задачі рівність (22) еквівалентна рівності $F(\pi) = 1$. Замінюючи в останній рівності $F(\pi)$ відповідним значенням, отримаємо рівняння $c(1 - \cos\pi) = 1$, з якого $c = 0{,}5$.

Щільність розподілу ймовірності *f(x)* визначається як похідна від *F(x)*

$$f(x) = \begin{cases} 0 \ \text{якщо} \ x \le 0; \\ 0{,}5 \cdot \sin x \ \text{якщо} \ 0 < x \le \pi; \\ 0 \ \text{якщо} \ x > \pi. \end{cases}$$

Шукані графіки *F(x)* і *f(x)* мають вигляд, показаний відповідно на рис.4 і рис.5.



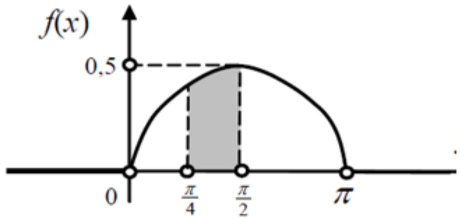


Рис. 4 – Графік інтегральної функції розподілу *F(x)* Рис. 5 – Графік щільності розподілу *f(x)*

Математичне сподівання m_x випадкової величини X (через симетричність закону розподілу) рівно $0{,}5\pi$. Цього ж значення можна набути за формулою (29). При цьому інтеграл від кусково-неперервної функції *f(x)* в нескінченних межах розпадається на три інтеграли відповідно до числа «кусків» підінтегральної функції:

$$m_x = \int_{-\infty}^{+\infty} x\cdot f(x)dx = \int_{-\infty}^{0} x\cdot 0\cdot dx + \int_{0}^{\pi} x\cdot 0{,}5\sin x dx + \int_{\pi}^{+\infty} x\cdot 0\cdot dx = \int_{0}^{\pi} x\cdot 0{,}5\sin x dx.$$

Інтегруючи по частинах (за формулою

$$\int_0^\pi UdV = UV\Big|_0^\pi - \int_0^\pi VdU$$

отримаємо

$$m_x = 0{,}5\int_0^\pi x\cdot\sin x dx = -0{,}5x\cos x\Big|_0^\pi - 0{,}5\int_0^\pi(\cos x)dx =$$
$$= -0{,}5\cos x\Big|_0^\pi + 0{,}5\sin x\Big|_0^\pi = 0{,}5\pi.$$

Дисперсію D_x випадкової величини X доцільно визначати за допомогою формули (32), для чого спочатку визначають другий початковий момент

$$\alpha_2 = 0{,}5\int_0^\pi x^2\sin x dx.$$

Інтегруючи двічі по частинах, знаходимо

$$\alpha_2 = 0{,}5\int_0^\pi x^2\sin x dx = 0{,}5(\pi^2-4).$$

Тоді

$$D_x = \alpha_2 - m_x^2 = 0{,}5(\pi^2-4) - (0{,}5\pi)^2 = 0{,}4649.$$

Середньоквадратичне відхилення випадкової величини X визначається по формулі (30)

$$\sigma_x = \sqrt{D_x} = \sqrt{0{,}4649} = 0{,}6818.$$

Ймовірність попадання випадкової величини в заданий інтервал значень від $x_1 = \pi/4$ до $x_2 = \pi/2$ можна визначити двома способами: *а*) за допомогою інтегральної функції; *б*) за допомогою щільності розподілу.

а) $P(\frac{\pi}{4} < X < \frac{\pi}{2}) = F\left(\frac{\pi}{2}\right) - F\left(\frac{\pi}{4}\right) = 0,5\left(1 - \cos\left(\frac{\pi}{2}\right)\right) - 0,5\left(1 - \cos\left(\frac{\pi}{4}\right)\right) = 0,3536.$

На рис.4 цій ймовірності відповідає відрізок вісі ΔF.

б) $P(\frac{\pi}{4} < X < \frac{\pi}{2}) = \int_{\pi/4}^{\pi/2} f(x)dx = \int_{\pi/4}^{\pi/2} 0,5\sin x dx = -0,5\cos x\Big|_{\pi/4}^{\pi/2} = 0,3536.$

На рис.5 цій ймовірності відповідає площа, що виділена сірим фоном.

Практичне заняття № 5-6. Статистичне оцінювання масових однорідних випадкових явищ

Мета. Одержання практичних навичок встановлення закономірностей масових випадкових явищ шляхом використання методів математичної статистики.

Теоретична частина. Математична статистика грунтується на *вибірковому методі*. *Вибіркової сукупністю* (або вибіркою) називається та частина об'єктів, яка відібрана для безпосереднього вивчення деякої якісної або кількісної ознаки з генеральної сукупності.

Генеральною сукупністю називають вся сукупність об'єктів, які підлягають вивченню.

Числа об'єктів (спостережень) в генеральній, або вибіркової сукупності називаються їх *обсягами*.

Вибірка повинна бути відібрана *випадково*. Випадковість досягається дотриманням принципу рівної можливості всім елементам генеральної сукупності бути відібраними у вибірку (жеребкування; використання випадкових чисел).

Повторною називають вибірку, при якій відібраний об'єкт (перед відбором наступного) повертається у генеральну сукупність.

Безповторною називають вибірку, при якій відібраний об'єкт у генеральну сукупність не повертається.

Репрезентативною (представницькою) називається вибірка яка правильно представляє пропорції генеральної сукупності.

Відбір об'єктів, що формують вибірку, може бути: *простим випадковим; типовим; механічним і серійним.*

Вид відбору визначає дослідник при рішенні певного завдання.

Варіантою називають спостережуване значення елемента вибірки.

Варіаційним рядом називають послідовність варіант, записаних у зростаючому порядку.

Частотами називають число спостереження варіанти.

Відносними частотами називають відношення частот до обсягу вибірки.

Якщо n − обсяг вибірки, $x_1, x_2, \ldots, x_k$ - варіани, $n_1, n_2, \ldots, n_k$ − частоти варіант, то $W_i = n_i/n$, де $i = \overline{1,k}$, є відносними частотами.

Статистичним рядом (розподілом вибірки) називають перелік варіант і відповідних їм частот або відносних частот, представлених у вигляді таблиці

i	1	2	…	k	
x_i	x_1	x_2	…	x_k	$\sum_{i=1}^{k}$
n_i	n_1	n_2	…	n_k	n
W_i	W_1	W_2	…	W_k	1

Для контролю правильності визначення відносних частот вибірки обсягу n використовують співвідношення $\sum_{i=1}^{k} W_i = 1$ і $\sum_{i=1}^{k} n_i = n$.

Статистичний розподіл можна задати також у вигляді послідовності інтервалів і відповідних частот.

Емпіричною функцією розподілу називають функцію $F^*(x)$, що визначає для кожного значення x *відносну частоту події* $X < x$.

Теоретичною функцією розподілу називають функцію $F(x)$, що визначає для кожного значення x ймовірність події $X < x$.

На підставі теореми Бернуллі при великих значеннях числа n $F^*(x)$ і $F(x)$ мало відрізняються, тобто доцільно використання емпіричної функції розподілу вибірки для наближеного подання теоретичної функції розподілу генеральної сукупності.

Полігоном частот називають ламану, відрізки якої з'єднують точки $(x_1;n_1)$, $(x_2;n_2)$,...,$(x_k;n_k)$.

Полігоном відносних частот називають ламану, відрізки якої з'єднують точки $(x_1;W_1)$, $(x_2;W_2)$,...,$(x_k;W_k)$.

Гістограмою частот називають східчасту фігуру, що складається із прямокутників з підставами на вісі абсцис (часткові інтервали) і висотами, які дорівнюють щільності частот. Площа гістограми частот дорівнює сумі всіх частот, тобто обсягу вибірки.

Гістограмою відносних частот називають східчасту фігуру, що складається із прямокутників, підставами яких служать часткові інтервали довжини h, а висоти дорівнюють відношенню W_i/h. Площа цієї гістограми дорівнює одиниці.

Статистичною оцінкою невідомого параметра теоретичного розподілу називають функцію від спостережуваних випадкових величин.

Зміщеною (зсунутою) називають статистичну оцінку, математичне очікування якої не дорівнює оцінюваному параметру.

Незміщеною (незсунутою) називають статистичну оцінку параметра, математичне очікування якої дорівнює оцінюваному параметру при будь-якому обсязі вибірки.

Ефективною називають статистичну оцінку, що має найменшу можливу дисперсію.

Переконливою називають статистичну оцінку, що при необмеженому зростанні об'єму вибірки прагне до оцінюваного параметра з ймовірністю, яка прямує до одиниці.

Генеральною середньою $\bar{x}_Г$ називають середнє арифметичне значень ознаки генеральної сукупності.

Вибірковою середньою $\bar{x}_B$ називають середнє арифметичне значення ознаки вибіркової сукупності.

Якщо всі x_i $(i=\overline{1,n})$ різні, то $\bar{x}_B = \frac{1}{n}\sum_{i=1}^{n} x_i$.

Якщо ж x_i мають частоті n_i $(i=\overline{1,k})$, то

$$\bar{x}_B = \frac{1}{n}\sum_{i=1}^{k} n_i x_i \text{. Тут } n = \sum_{i=1}^{k} n_i \text{.}$$

Близькість $\bar{x}_B$ до $\bar{x}_Г$ залежить від обсягу вибірки - чим обсяг вибірки більший, тим менше вибіркова середня відрізняється від генеральної.

Генеральною дисперсією $D_Г$ називають середнє арифметичне квадратів відхилень значень ознаки генеральної сукупності від їхнього середнього значення $\bar{x}_Г$.

Вибірковою дисперсією D_B називають середнє арифметичне квадратів відхилень значень ознаки вибіркової сукупності від їхнього середнього значення $\bar{x}_B$.

Якщо всі x_i $(i=\overline{1,n})$ різні, то $D_B = \frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x}_B)^2$.

Якщо ж x_i мають частоті n_i $(i=\overline{1,k})$, то

$D_B = \frac{1}{n}\sum_{i=1}^{k} n_i (x_i - \bar{x}_B)^2$. Тут $n = \sum_{i=1}^{k} n_i$.

Формулу (3.4) можна спростити

$$D_B = \overline{x^2} - (\bar{x})^2 .$$

Тобто дисперсія дорівнює середньому квадратів значень ознаки мінус квадрат його загальної середньої.

Вибірковим середнім квадратичним відхиленням називають квадратний корінь із вибіркової дисперсії

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

Розмахом варіації R називають різницю між найбільшою і найменшою варіантами: наиб наим.

Варіанта, яка має найбільшу відносну частоту, називається *модою*, і позначається Mo_B.

Варіанта, яка ділить варіаційний ряд на дві частини, рівні між собою по числу варіант, називається *медіаною* Me_B.

Виправлена дисперсія – це вибіркова дисперсія, помножена на виправляючий коефіцієнт

$$S^2 = \frac{n}{n-1} D_B .$$

При досить великих n вибіркова і виправлена дисперсії практично не розрізняються.

Точечною називають оцінку, що визначається одним числом.

Інтервальною називають оцінку, що визначається двома числами – кінцями інтервалу.

Надійністю (довірчою ймовірністю) оцінки параметра розподілу називають імовірність: 0,95; 0,99 або 0,999, з якою статистичний параметр розподілу відповідає теоретичному.

Довірчим називають інтервал, що покриває невідомий параметр розподілу із заданою надійністю.

Завдання. Провести розв’язок слідуючих п’яти прикладів.

Приклад 13. Задано вибірку об’єму $n = 170$, наведену у таблиці.

187	193	199	197	196	184	200	193	198	191	193	188	193	195	197	199	202
193	190	197	195	182	201	202	184	197	205	178	191	200	223	188	192	188
194	183	207	183	195	184	175	195	212	197	194	184	175	198	189	194	185
213	192	200	194	173	206	163	204	174	183	199	203	185	199	196	196	188
169	196	190	205	189	189	190	175	190	193	209	190	183	191	193	191	190
192	191	185	202	173	184	176	199	182	186	189	193	185	168	192	193	205
171	193	191	206	187	193	192	189	191	190	182	194	194	197	207	198	180
175	193	191	188	187	191	191	192	192	214	171	208	185	195	190	214	193
183	193	193	187	198	203	181	173	189	195	180	180	205	194	179	191	201
195	195	189	185	199	194	187	173	211	190	165	182	182	194	168	176	192

Провести статистичну обробку даної сукупності значень і з цією метою виконати слідуючі етапи:

1. Представити вибірку у графічному вигляді;
2. Визначити і побудувати варіаційний ряд;
3. Визначити і побудувати статистичний ряд;
4. Визначити і побудувати інтервальний статистичний ряд;

5. Визначити і побудувати гістограму відносних частот;

6. Визначити і побудувати емпіричну функцію розподілу $F^*(x)$.

Розв'язання.

1. Для графічного представлення вибірки використати слідуючу програму

```
close all;clear all;clc
x=load('C:\data.scv');X=x';X=X(:)';
n=length(X);%Об'єм вибірки
figure(1);plot(X,'o','Color','r','LineWidth',2); hold all;grid on
title('Вибіркова сукупність');
xlabel('Номер виміру');ylabel('Результати вимірів');
```

Рис. – Графік вибіркової сукупності

2. Для виконання цього пункту слід у програму добавити слідуючі оператори

```
x_i=unique(X);
figure(2);plot(x_i,'o','Color','b','LineWidth',2); hold all;grid on
title('Варіаційний ряд');
xlabel('Номер варіанти');ylabel('Значення варіанти');
```

Одержаний варіаційний ряд наведено у слідуючій таблиці, а його графічний вигляд на рис.

i	1	2	3	4	5	6	7	8	9	10	11	12
Варіанта x_i	163	165	168	169	171	173	174	175	176	178	179	180

i	13	14	15	16	17	18	19	20	21	22	23	24
Варіанта x_i	181	182	183	184	185	186	187	188	189	190	191	192
i	25	26	27	28	29	30	31	32	33	34	35	36
Варіанта x_i	193	194	195	196	197	198	199	200	201	202	203	204
i	37	38	39	40	41	42	43	44	45	46		
Варіанта x_i	205	206	207	208	209	211	212	213	214	223		

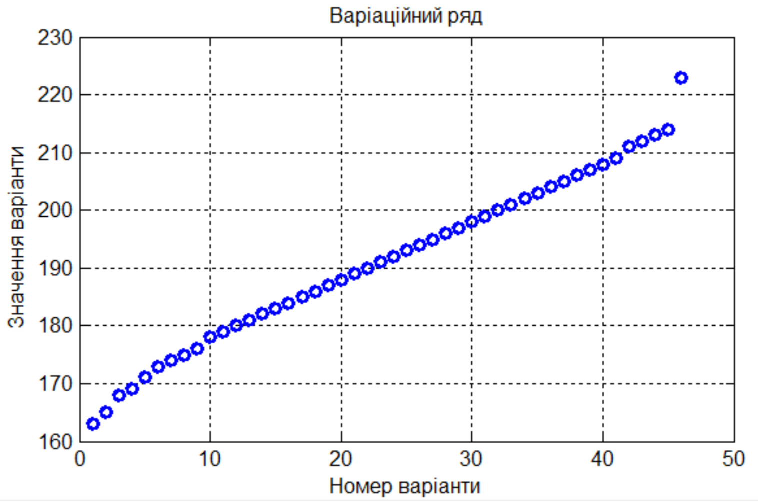


Рис. – Варіаційний ряд

3. Для визначення статистичного ряду треба спочатку розрахувати частоти n_i і відносні частоти $W_i = n_i / n$ варіант.

Для цього слід у програму добавити слідуючі оператори

```
k=length(x_i);%Кількість варіант
for i=1:k;n_i(i)=0;
for j=1:n;if X(j)==x_i(i);n_i(i)=n_i(i)+1;end;end
W_i(i)=n_i(i)/n;
end
figure(3);subplot(1,2,1);plot(x_i,n_i,'o','Color','m','LineWidth',2); hold all;grid on
title('Статистичний ряд');
xlabel('Значення варіанти');ylabel('Частота варіанти');
subplot(1,2,2);plot(x_i,W_i,'o','Color','m','LineWidth',2); hold all;grid on
```

```
title('Статистичний ряд');
xlabel('Значення варіанти');ylabel('Відносна частота варіанти');
for j=1:k;s1=sum(n_i);s2=sum(W_i);end%Перевірка розрахунків
```

Одержаний статистичний ряд наведено у слідуючій таблиці, а його графічний вигляд на рис.

i	1	2	3	4	5	6	7	8	9	10
x_i	163	165	168	169	171	173	174	175	176	178
n_i	1	1	2	1	2	4	1	4	2	1
W_i	0,005882	0,005882	0,011765	0,005882	0,011765	0,023529	0,005882	0,023529	0,011765	0,005882
i	11	12	13	14	15	16	17	18	19	20
x_i	179	180	181	182	183	184	185	186	187	188
n_i	1	3	1	5	5	5	6	1	5	5
W_i	0,005882	0,017647	0,005882	0,029412	0,029412	0,029412	0,035294	0,005882	0,029412	0,029412
i	21	22	23	24	25	26	27	28	29	30
x_i	189	190	191	192	193	194	195	196	197	198
n_i	7	9	11	8	15	9	8	4	6	4
W_i	0,041176	0,052941	0,064706	0,047059	0,088235	0,052941	0,047059	0,023529	0,035294	0,023529
i	31	32	33	34	35	36	37	38	39	40
x_i	199	200	201	202	203	204	205	206	207	208

n_i	6	3	2	3	2	1	4	2	2	1
W_i	0,035294	0,017647	0,011765	0,017647	0,011765	0,005882	0,023529	0,011765	0,011765	0,005882
i	41	42	43	44	45	46				
x_i	209	211	212	213	214	223	$\sum_{i=1}^{k}$			
n_i	1	1	1	1	2	1	170			
W_i	0,005882	0,005882	0,005882	0,005882	0,011765	0,005882	1			

4. Для визначення інтервального статистичного ряду необхідно:

- увесь діапазон $[x_{\min}, x_{\max}] = [x_1, x_k]$ значень варіант розділити на m елементарних проміжків $[a_0, a_1]$, $[a_0, a_1], [a_1, a_2], \ldots, [a_{m-2}, a_{m-1}], [a_{m-1}, a_m]$, з довжинами $h_i = a_i - a_{i-1}$, $i = \overline{1, m}$ так, що кінець попереднього служить початком наступного, причому $a_0 = x_1$ і $a_m = x_k$;

- знайти частоти n_i і відносні частоти $W_i = n_i / n$ попадання значень випадкової величини X в i-й частинний інтервал з кінцями a_{i-1} і a_i, $i = \overline{1, m}$.

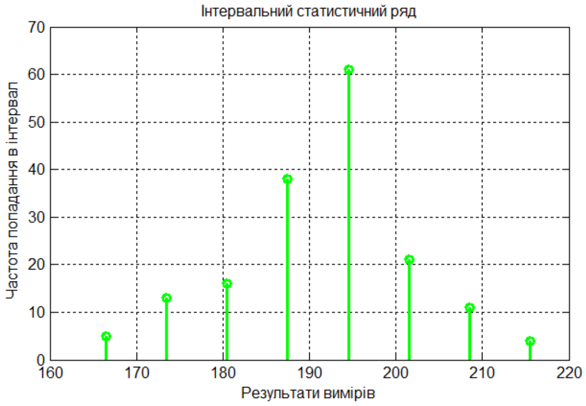
При групуванні даних з метою побудови інтервального статистичного ряду виникає питання раціонального вибору числа m елементарних інтервалів. Використаємо наближене співвідношення $m = \mathrm{int}(1 + 3{,}322 \ln n)$, де $\mathrm{int}(x)$ – ціла частина дійсного числа $1 + 3{,}322 \ln n$. Довжину елементарного проміжка h визначимо за формулою $h = (x_{\max} - x_{\min}) / m$.

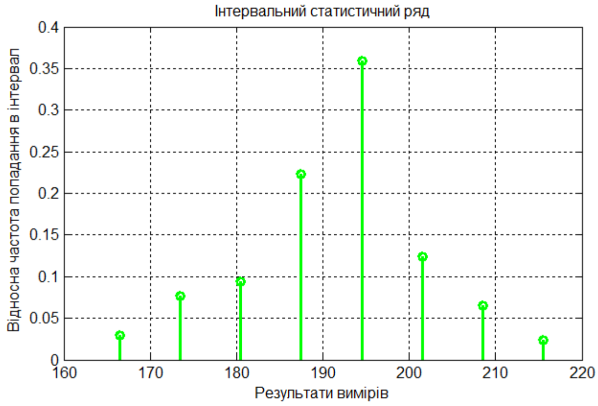
Для визначення інтервального статистичного ряду добавимо у програму слідуючі оператори

```
m=ceil(1+3.322*log10(n));h=ceil((x_i(k)-x_i(1))/m);
x_lev(1)=x_i(1);x_pr(1)=x_lev(1)+h;
for j=2:m;x_lev(j)=x_lev(j-1)+h;x_pr(j)=x_pr(j-1)+h;end
for i=1:m;ni_i(i)=0;
for j=1:n;if X(j)>=x_lev(i)&X(j)<x_pr(i);ni_i(i)=ni_i(i)+1;end
end
Wi_i(i)=ni_i(i)/n;
end
for j=1:m;s3=sum(ni_i);s4=sum(Wi_i);end%Перевірка розрахунків
figure(4)
for i=1:m-1
subplot(1,2,1);stem((x_lev(i)+x_lev(i+1))/2,ni_i(i),'o','Color','g','LineWidth',2);
hold all;grid on
title('Інтервальний статистичний ряд');
xlabel('Результати вимірів');ylabel('Частота попадання в інтервал');
sub-
plot(1,2,2);stem((x_lev(i)+x_lev(i+1))/2,Wi_i(i),'o','Color','g','LineWidth',2); hold
all;grid on
end
title('Інтервальний статистичний ряд');
xlabel('Результати вимірів');ylabel('Відносна частота попадання в інтер-
вал');
```

Одержаний інтервальний статистичний ряд наведено у слідуючій таблиці,

Інтервал	1	2	3	4	5	6	7	8	9
Ліва межа	163	170	177	184	191	198	205	212	219
Права межа	170	177	184	191	198	205	212	219	226
n_i	5	13	16	38	61	21	11	4	1
W_i	0,029412	0,076471	0,094118	0,223529	0,358824	0,123529	0,064706	0,023529	0,005882





5. Гістограми відносних частот є статистичним емпіричним аналогом графіка щільності розподілу. Гістограма представляє собою східчасту фігуру, складену з m елементарних прямокутників, у кожного з яких основою служить відповідний частинний інтервал з кінцями a_{i-1} і a_i, а висота дорівнює відповідному значенню $f^*(x) = W_i / h_i$, $x \in [a_{i-1}, a_i]$, $i = \overline{1,m}$ щільності відносної частоти W_i.

Для її побудови використаємо інтервальний статистичний ряд. Для цього у програму добавимо слідуючі оператори

```
for i=1:m;f_i(i)=Wi_i(i)/h;end
figure(5)
for i=1:m
Ox=[x_lev(i),x_lev(i),x_pr(i),x_pr(i)];Oy=[0,f_i(i),f_i(i),0];
plot(Ox,Oy,'Color','k','LineWidth',2);hold all
end
grid on
title('Гістограма відносних частот');
xlabel('Результати вимірів');ylabel('Щільність відносної частоти');
```

Одержаний дані для побудови гістограми наведені у слідуючій таблиці, а сама гістограма на рис.

Інтервал	1	2	3	4	5	6	7	8	9
Ліва межа	163	170	177	184	191	198	205	212	219
Права межа	170	177	184	191	198	205	212	219	226
$f^*(x)$	0,004202	0,010924	0,013445	0,031933	0,051261	0,017647	0,009244	0,003361	0,00084

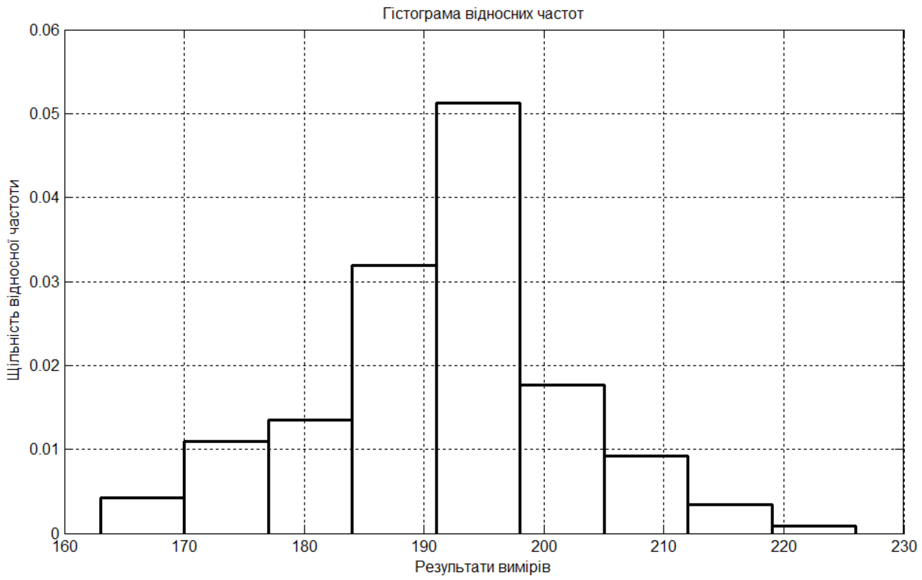


Рис. - Гістограма відносних частот

6. Емпірична функція розподілу $F^*(x)$ визначає за вибіркою для кожного значення x відносну частоту події $X < x$. Щоб знайти $F^*(x)$, треба підрахувати число варіант $n(x)$, у яких випадкова величина X прийняла значення, менші x. Тоді $F^*(x) = n(x)/n$.

Графік емпіричної функції розподілу $F^*(x)$ називають кумулятивною кривою (кумулятою).

Для будь-якої випадкової величини X – дискретної чи неперервної – кумулята є розривною кусково-сталою східчастою лінією, скінченні стрибки якої відповідають значенням варіант і за величиною дорівнюють їх відносним частотам.

Для визначення інтервального статистичного ряду добавимо у програму слідуючі оператори

```
H=x_pr(1);
for i=1:m;F_i(i)=0;
for j=1:k;if x_i(j)<H;F_i(i)=F_i(i)+1;end;end
F_i(i)=F_i(i)/k;H=H+h;
end
figure(6)
for i=1:m
Ox=[x_lev(i),x_pr(i)];Oy=[F_i(i),F_i(i)];
plot(Ox,Oy,'Color','k','LineWidth',2);hold all
Ox=[x_lev(i),x_lev(i)];Oy=[0,F_i(i)];
plot(Ox,Oy,'--','Color','k','LineWidth',1)
end
Ox=[x_pr(m),x_pr(m)];Oy=[0,F_i(m)];
plot(Ox,Oy,'--','Color','k','LineWidth',1)
title('Емпірична функція розподілу');
xlabel('Значення варіант');ylabel('Кумулята');
grid on
```

Одержаний дані для побудови емпіричної функції розподілу $F^*(x)$ наведені у слідуючій таблиці, а сама функція на рис.

x	170	177	184	191	198	205	212	219	226
$n(x)$	177	184	191	198	205	212	219	226	219
$F^*(x)$	0,086957	0,195652	0,326087	0,478261	0,630435	0,782609	0,913043	0,978261	1

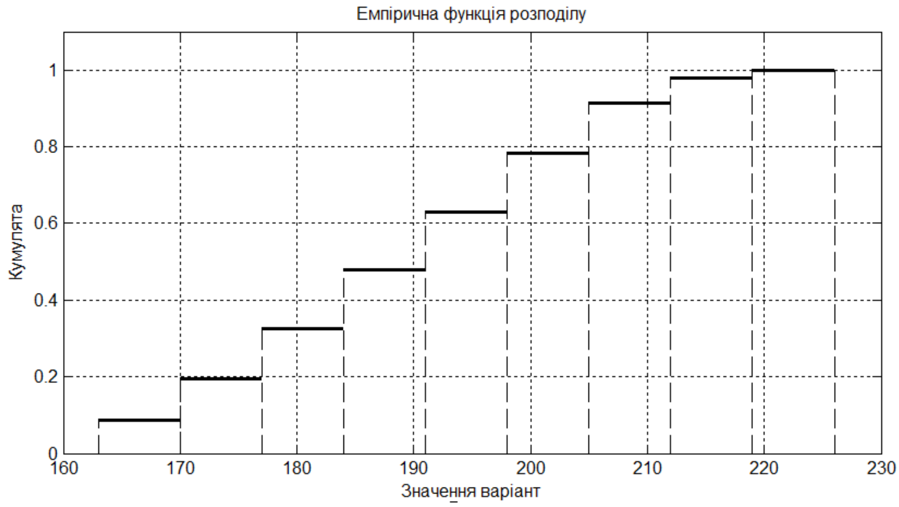


Рис. - Емпірична функція розподілу $F^*(x)$

Приклад 14. Задано вибірку об'єму $n = 50$, наведену у таблиці.

0,1	0,4	0,23	0,12	0,35	0,46	0,11	0,04	0,51	0,27
0,31	0,34	0,094	0,18	0,49	0,33	0,3	0,22	0,14	0,5
0,41	0,25	0,48	0,32	0,29	0,31	0,31	0,46	0,44	0,38
0,39	0,13	0,47	0,4	0,53	0,374	0,16	0,44	0,39	0,27
0,25	0,46	0,2	0,11	0,32	0,41	0,48	0,224	0,35	0,52

Провести самостійно статистичну обробку даної сукупності значень і з цією метою виконати слідуючі етапи:

1. Представити вибірку у графічному вигляді;
2. Визначити і побудувати варіаційний ряд;
3. Визначити і побудувати статистичний ряд;
4. Визначити і побудувати інтервальний статистичний ряд;
5. Визначити і побудувати гістограму відносних частот;
6. Визначити і побудувати емпіричну функцію розподілу $F^*(x)$.

Практичне заняття № 7. Числові характеристики вибірки

Мета.

Теоретична частина. До основних числових характеристик вибірки належать: вибіркове середнє (середнє арифметичне), дисперсія, середньоквадратичне відхилення, мода, медіана, коефіцієнт варіації.

Вибіркове середнє $\bar{x}$ визначається за формулою

$$\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i , \qquad (1)$$

або формулою

$$\bar{x} = \frac{1}{n}\sum_{i=1}^{k} x_i n_i ,$$

де x_i – варіанти; n_i – відповідні частоти; k – кількість груп варіант; n – об'єм вибірки.

Дисперсія вибірки $D_B(X)$ обчислюється за формулою

$$D_B(X) = \frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})^2 ,$$

або

$$D_B(X) = \frac{1}{n}\sum_{i=1}^{k}(x_i - \bar{x})^2 n_i ,$$

або

$$D_B(X) = \frac{1}{n}\sum_{i=1}^{k} x_i^2 n_i - (\bar{x})^2 .$$

Вибіркова дисперсія $D_B(X)$ є зміщеною оцінкою для $D(X)$. Незміщеною спроможною оцінкою дисперсії $D(X)$ служить виправлена дисперсія $\bar{D}_B(X)$, яка обчислюється за формулою

$$\bar{D}_B(X) = \frac{n}{n-1} D_B(X) .$$

При достатньо великих n виправлена і вибіркова дисперсії мало відрізняються.

Середньоквадратичне відхилення вибірки $\sigma_B(X)$ визначається як квадратний корінь із дисперсії

$$\sigma_B(X)=\sqrt{D_B(X)}\,.$$

Незміщеною спроможною оцінкою середнього квадратичного відхилення $\sigma(X)$ служить виправлене середнє квадратичне відхилення $S(X)$

$$\bar{\sigma}_B(X)=\sqrt{\bar{D}_B(X)}\,.$$

Мода вибірки Mo_B для статистичного ряду – це значення варіанти з максимальною частотою. Якщо задано інтервальний статистичний ряд, то мода Mo_B обчислюється за такою наближеною формулою

$$Mo_B=x_0+h\frac{n_s-n_{s-1}}{(n_s-n_{s-1})+(n_s-n_{s+1})},$$

де x_0 – початок інтервалу з максимальною частотою (модальний інтервал); h – його довжина; n_s – відповідна йому частота; n_{s-1} і n_{s+1} – частоти, що відповідають інтервалам попередньому і наступному за модальним.

Медіана Me_B – значення випадкової величини X, що відповідає середині ряду. При цьому: $Me_B=x_{(k+1)/2}$, якщо k – непарне число; б) $Me_B=(x_{k/2}+x_{k/2+1})/2$, якщо k – парне число

Медіана вибірки Me_B – це значення середнього елемента статистичного ряду

Если все элементы выборки различны, то медиана — это такое число выборки, что ровно половина из элементов выборки больше него, а другая половина меньше него.

Для інтервального варіаційного ряду медіана Me_B підраховується за формулою

$$Me_B=x_0+h\frac{n/2-T_{s-1}}{n_s},$$

де x_0 – початок медіанного інтервалу, тобто інтервалу, в якому утримується середній елемент; h – довжина медіанного інтервалу; n – об'єм вибірки; T_{s-1} –

сума частот інтервалів, які передують медіанному; n_s – частота медіанного інтервалу.

Коефіцієнтом варіації вибірки V називають виражене у процентах відношення середньоквадратичного відхилення до вибіркового середнього

$$V = \frac{\sigma_B(X)}{\bar{x}} \cdot 100\% .$$

Завдання. Провести розв'язок слідуючих п'яти прикладів.

Приклад 13. Задано вибірку об'єму $n = 170$, наведену у попередньому прикладі.

Визначити наступні числові характеристики вибірки для статистичного і інтервального статистичного рядів та вихідної вибірки.

1. Вибіркове середнє $\bar{x}$;
2. Дисперсію $D_B(X)$;
3. Середньоквадратичне відхилення $\sigma_B(X)$;
4. Моду Mo_B;
5. Медіану Me_B;
6. Коефіцієнт варіації V.

Розв'язання.

А. Використання статистичного ряду

Скористаємось формулами:

$$\bar{x} = \frac{1}{n}\sum_{i=1}^{46} x_i n_i ;$$

$$D_B(X) = \frac{1}{n}\sum_{i=1}^{46}(x_i - \bar{x})^2 n_i ;$$

$$\bar{D}_B(X) = \frac{n}{n-1} D_B(X) ;$$

$$\sigma_B(X) = \sqrt{D_B(X)} ;$$

$$\bar{\sigma}_B(X) = \sqrt{\bar{D}_B(X)} ;$$

$$Mo_B = x_i(n_{i\max}) ;$$

$$Me_B = (x_{k/2} + x_{k/2+1}) / 2 .$$

$$V = \frac{\bar{\sigma}_B(X)}{\bar{x}} \cdot 100\%;$$

Для обчислень по формулам добавимо у програму слідуючі оператори

```
Mx=sum(x_i.*n_i)/n
Dx=sum((x_i-Mx).^2.*n_i)/n
Dxns=Dx*n/(n-1)
SKOv=sqrt(Dx)
SKOvns=sqrt(Dxns)
N_imax=max(n_i);
for i=1:k;if n_i(i)==N_imax;j=i;end
end
Mo=x_i(j)
Me=(x_i(k/2)+x_i(k/2+1))/2
Vsr=100*SKOvns/Mx
```

В результаті обчислень одержуємо: Mx=191.0471; Dx=101.2448; Dxns=101.8439; SKOv =10.0620; SKOvns=10.0918; Mo=193; Me=191.5000; Vsr=5.2824.

Б. Використання інтервального статистичного ряду

Використаємо формули:

$$\bar{x} = \frac{1}{n} \sum_{i=1}^{9} x_i n_i \;;$$

$$D_B(X) = \frac{1}{n} \sum_{i=1}^{9} (x_i - \bar{x})^2 n_i \,;$$

в якій за варіанти x_i беруть середини інтервалів;

$$\bar{D}_B(X) = \frac{n}{n-1} D_B(X);$$

$$\sigma_B(X) = \sqrt{D_B(X)}\,;$$

$$\bar{\sigma}_B(X) = \sqrt{\bar{D}_B(X)}\,;$$

$$Mo_B = x_0 + h \frac{n_s - n_{s-1}}{(n_s - n_{s-1}) + (n_s - n_{s+1})};$$

$$Me_B = x_0 + h\frac{n/2 - T_{s-1}}{n_s};$$

$$V = \frac{\bar{\sigma}_B(X)}{\bar{x}} \cdot 100\%.$$

Для обчислень по формулам добавимо у програму слідуючі оператори

```
Mxi=sum((x_lev+x_pr)./2.*ni_i)/n
Dxi=sum(((x_lev+x_pr)./2-Mxi).^2.*ni_i)/n
Dins=Dxi*n/(n-1)
SKOi=sqrt(Dxi)
SKOins=sqrt(Dins)
Ni_imax=max(ni_i);
for i=1:m;if ni_i(i)==Ni_imax;j=i;end
end
Moi=x_lev(j)+h*(ni_i(j)-ni_i(j-1))/((ni_i(j)-ni_i(j-1))+(ni_i(j)-ni_i(j+1)))
Mei=x_lev((m+1)/2)+h*(n/2-sum(ni_i(1:(m+1)/2-1)))/ni_i((m+1)/2)
Visr=100*SKOins/Mxi
```

В результаті обчислень одержуємо: Mxi=191.6176; Dxi=111.5979; Dins=112.2583; SKOi=10.5640; SKOins=10.193.55565952; Moi=193.5556; Mei=192.4918; Visr=5.5293.

В. Використання вихідної вибірки

Скористаємось формулами:

$$\bar{x} = \frac{1}{n}\sum_{i=1}^{170} x_i$$

$$D_B(X) = \frac{1}{n}\sum_{i=1}^{170}(x_i - \bar{x})^2$$

$$\bar{D}_B(X) = \frac{n}{n-1}D_B(X)$$

$$\sigma_B(X) = \sqrt{D_B(X)}$$

$$\bar{\sigma}_B(X) = \sqrt{\bar{D}_B(X)}$$

$$V = \frac{\bar{\sigma}_B(X)}{\bar{x}} \cdot 100\%$$

Для обчислень по формулам добавимо у програму слідуючі оператори

MX=sum(X)/n
DX=sum((X-MX).^2)/n
*DXns=DX*n/(n-1)*
SKO=sqrt(DX)
SKOns=sqrt(DXns)
*V=100*SKOns/MX*

В результаті обчислень одержуємо

MX = 191.0471
DX = 101.2448
DXns = 101.8439
SKO = 10.0620
SKOns = 10.0918
V = 5.2824

	$\bar{x}$	$D_B(X)$	$\bar{D}_B(X)$	$\sigma_B(X)$	$\bar{\sigma}_B(X)$s	Mo_B	Me_B	V
Статист. ряд	191.0471	101.2448	101.8439	10.0620	10.0918	193	191.5000	5.2824
Інтеграл. Сстатист. ряд	191.6176	111.5979	112.2583	10.5640	10.5952	193	193.5556	5.5293
Вихідна вибірка	191.0471	101.2448	101.8439	10.0620	10.0918	-	-	5.2824

СПИСОК ДЖЕРЕЛ

1. Вентцель Е.С. Теория вероятностей и математическая статистика. – М.: Высш. школа, 1999. – 576 с.

2. Гмурман В.Е. Теория вероятностей и математическая статистика. – М.: Высш. шк., 2004. – 479 с.

3. Гмурман В.Е. Руководство к решению задач по теории вероятностей и математической статистике. – М.: Высш. шк., 2005. – 404 с.

4. Жалдак М.І., Кузьміна Н.М., Берлінська С.Ю. Теорія ймовірностей і математична статистика з елементами інформаційної технології. – К.: Вища шк., 1995. – 351 с.

5. Жевняк Р.М., Карпук А.А., Унукович В.Т. Теория вероятностей и математическая статистика. – Мн.: Харвест, 2000. – 384 с.

6. Копич І.М., Сороківський В.М. Елементи теорії ймовірностей і математичної статистики: теорія та практикум. – Львів: Вид-во ЛКА, 2001. – 336 с.

7. Кремер И.Ш. Теория вероятностей и математическая статистика. – М.: ЮНИТИ, 2006. – 573 с.

8. Теорія імовірностей і математична статистика / А.І. Колосов , Ю.Є. Печеніжський та ін. – Х.: ХНАМГ, 2011. – 129 с.